Structural Contributions of Substrates to Their Binding to P-Glycoprotein. A TOPS-MODE Approach

Ernesto Estrada[1,*], Enrique Molina[2], Delvin Nodarse[2] and Eugenio Uriarte[3]

[1]*Department of Mathematics and Statistics, Department of Physics, Institute of Complexity Systems, University of Strathclyde Glasgow, G1 1XQ, UK,* [2]*Department of Chemistry, Faculty of Chemistry, University of Camagüey, Camagüey 74650, Cuba,* [3]*Department of Organic Chemistry, Faculty of Pharmacy, University of Santiago de Compostela, Santiago de Compostela 15782, Spain*

Abstract: A topological substructural molecular design approach (TOPS-MODE) has been used to formulate structural rules for binding of substrates of P-glycoprotein (P-gp). We first review some of the models developed in the recent literature for predicting binding to P-gp. Then, we develop a model using TOPS-MODE, which is able to identify 88.4% of substrates and 84.2% of non-substrates. When the model is presented to an external prediction set of 100 substrates and 77 nonsubstrates it identifies correctly 81.8% of all cases. Using TOPS-MODE strategy we found structural contributions for binding to P-gp, which identifies 24 structural fragments responsible for such binding. We then carried out a chemico-biological analysis of some of the structural fragments found as contributing to P-gp binding of substrates. We show that in general the model developed so far can be used as a virtual screening method for identifying substrates of P-gp from large libraries of compounds.

Keywords: TOPS-MODE, knowledge generation, P-glycoprotein, QSAR, molecular modeling, Spectral Moments, graph-theory descriptors.

INTRODUCTION

In QSAR analysis a quantitative model is used to predict the biological response of a chemical based on a series of molecular descriptors or physicochemical properties [1]. However, the structural information contained in such descriptors or properties is many times encrypted [2] in a way that does not allow the extraction of structural rules to form a *knowledge base* similar to those provided by human expertise [3]. In the case of toxicological assessment of chemicals these knowledge bases are the heart of expert systems, such as DEREK [4] and TOPKAT [5], used to evaluate the toxicological profile of chemicals [6].

An important step in the development of any toxicological or metabolic activity of a chemical is the transport to the organs where the final effect takes place. One of the most important proteins in the transport of various molecules across extra- and intra-cellular membranes is P-glycoprotein (P comes from permeability). P-glycoprotein (P-gp), is a 170-kDa glycoprotein, which is a member of the ATP binding cassette (ABC) super-family of transport proteins and couples hydrolysis of ATP to the transport of compounds out of the cell [7-11]. P-gp behaves quite differently from conventional transporters in two respects [12]:

i. it does not transport a specific substrate but a wide variety of chemically most diverse compounds, and

ii. it seems to extract its substrates directly out of the membrane [13].

In MDR cell lines, drugs entering the cells through passive diffusion bind to P-gp and are actively pumped outward from the cells. There is some evidence that P-gp can also decrease the influx of cytotoxic drugs into the cell [14, 15].

In conjunction with drug-metabolizing enzymes, P-gp provides a protective physiological barrier capable of altering the rate and extent of xenobiotic entry into the systemic circulation [16].

1. REVIEW OF CONCEPTS AND MODELS

1.1. P-gp Efflux Substrates

P-gp is expressed in many normal tissues such as intestine, liver, kidney, lung, and endothelia of brain, testis, and placenta, consistent with its role as a natural detoxification system. Because of this activity it can have a high impact on many drugs' pharmacokinetics and pharmacodynamics [10, 17]. Particularly, P-gp has been shown to limit oral absorption, modulate hepatic, renal, or intestinal elimination, and restrict central nervous system entry of certain drugs [18-20]. In addition, because of its broad substrate specificity, P-gp mediated drug-drug interactions may occur when substrates and inhibitors are coadministered [21]. For example, P-gp inhibitors lead to an increase in the systemic exposure and tissue distribution of coadministered P-gp substrate drugs that could cause serious adverse effects [21, 22].

In comparison to most of other transport proteins that recognize specific chemical substrates, P-glycoprotein is unusual since it transports structurally and mechanistically unrelated agents out of the cell [9]. It is evident from the literature that compounds that interact with the P-gp efflux pump represent a wide spectrum of chemical structures as well as different classes of drugs [23]. Overexpression of this protein may result in multidrug resistance (MDR) and it is a major cause of the failure of cancer chemotherapy, in addition to decreasing the efficacy of antibiotics and antiviral agents [24, 25]. Other therapeutic agents are also affected by P-gp, such as HIV-protease inhibitors, detergents, antibiotics, immunosuppressives, antihypertensives and many others [10, 26]. Some of these agents are listed in Table **1**. As can be seen these compounds are chemically diverse; some of them are positively charged at physiological pH and, since most of them are relatively hydrophobic, they permeate the cell membrane by passive diffusion.

Several *in vitro* screening assays such as the monolayer efflux, ATPase activity, and calcein-AM fluorescence assays have been suggested to classify compounds as P-gp substrates [24, 25, 27-29]. Each of these assays provides different information and has advantages and disadvantages [23]. However, the assays are not designed to distinguish P-gp substrates from inhibitors [30-32] and do not directly measure transport. The limitations of each assay makes selection difficult [27]. Nevertheless, methods that increase the identification of P-gp substrate are therefore useful during early drug development.

Prediction of P-gp substrate specificity ($S_{P\text{-}gp}$) can be viewed as a constituent part of a compound's pharmaceutical profiling in drug design. This task is difficult to achieve due to several factors that raised many contradictory opinions [33]:

*Address correspondence to this author at the Department of Mathematics and Statistics, Department of Physics, Institute of Complexity Systems, University of Strathclyde Glasgow, G1 1XQ, UK; Tel: ??????????; Fax: ??????????; E-mail: ernesto.estrada@strath.ac.uk

1381-6128/10 $55.00+.00

i. the disparity between the $S_{P\text{-}gp}$ values obtained in different assays;

ii. the confusion between P-gp substrate and inhibitors;

iii. the confusion between lipophilicity and amphiphilicity of P-gp substrate;

iv. the dilemma of describing class-specific relationships when P-gp has no binding sites of high ligand specificity.

1.2. Outcome and Mechanisms for P-gp Interaction

P-glycoprotein is encoded by the *MDR1* gene in humans and the *mdr1* (also called *mdr1b*) and *mdr3* (also called *mdr1a*) genes in rodents [22, 34]. High levels of P-gp expression have been observed in the endothelial cells of the blood-brain barrier [9, 35, 36], in certain cells of the adrenal gland, liver, pancreas, kidney proximal tubules, colon, jejunum [9, 37], digestive tract [9, 38], cells of the luminal surface of the secretory epithelium of the gravid uterus [9, 39] and placental trophoblast. This demonstrates its protective roles in limiting drug absorption, contributing to pharmacokinetics, and probably impacting pharmacodynamics and toxicity [20, 40, 41]. High expression levels of P-gp are also observed in many cancer cells [9, 42].

The mechanism for P-glycoprotein modulation may be different for different classes of compounds [10, 23]. The three main mechanisms of modulators are:

i. direction interaction with one or more of the binding sites on P-gp thus blocking transport by acting as competitive or non-competitive inhibitors;

ii. inhibition of ATP binding, ATP hydrolysis or coupling of ATP hydrolysis to the translocation of the substrate;

iii. interaction with the lipid membrane of the cell thus perturbing the membrane environment or modifying the drug-membrane interaction.

A better understanding of this interplay would help to understand not only the mechanism of action of P-gp, but also possible unexpected physiological roles. On the one hand, lipids can modulate P-gp catalytic activity as well as drug binding. Therefore the physiological role(s) of P-gp may be highly tissue-dependent, as plasma membrane lipid composition depends on cell type. On the other hand, a possible role for P-gp in lipid transport or metabolism would involve it with a variety of cellular processes (see Fig. **1**) because increasing evidence shows that lipids have many different functions, such as signal transduction or modulation of peripheral or integral proteins [43, 44].

1.3. In Silico Studies

In vitro assays and *in silico* models for predicting P-gp substrates or inhibitors have been recognized to be valuable tools during early phases of drug development [22]. Using different sets of molecules and various types of *in vitro* assays to measure P-gp activities, a number of structure-property relationships (SPR) have been developed to elucidate the physicochemical properties characterizing the P-gp substrates [9, 12, 23, 32, 45-47]. More detailed 3D pharmacophore hypotheses for P-gp substrates and inhibitors have also been proposed [48-51].

In one of these works, Wang *et al*. [52] recently published a model based on an unsupervised machine learning approach for classifying potential P-gp substrates and inhibitors. It must be recognized that there are several difficulties in developing computational models due to broad substrate specificity, multiple binding P-gp sites and different modulator mechanisms (i.e. competitive, noncompetitive, alteration of cell membrane lipids, etc.) [10]. However, the availability of virtual screening tools for discriminating substrates and nonsubstrates, as well as inhibitors would be helpful in the design of new drugs. In particular, these models can be used for assessing the potential for drug-enzyme interactions of new candidates.

A data set of 609 diverse compounds tested for MDR reversing (MDRR) activity against P388/ADR-resistant cell lines was submitted to the MULTICASE computer program for structure-activity analysis in order to design more effective MDRR agents [15]. In such a way some substructural features related to MDRR activity were identified. Based on quantitative structure-activity relationship study of MDRR agents (modulators, chemosensitizers, reverters), some compounds with desired substructural features and activity were identified from the MACCS-II and National Cancer Institute DIS databases and tested experimentally.

On the other hand, Bakken and Jurs [53] develop classification models for MDRR activity based on structural descriptors. The same set of compounds employed by Klopman et. al. [15] is used this study. Structure-based descriptors are used to develop classification models using linear discriminant analysis (LDA). Predictive ability of all models developed is examined using external prediction sets. Models developed could be used to screen large libraries of compounds aiming the identification of those likely to display activity as MDRR agents. These models, validated using external data sets, can be implemented for virtual screenings in early phases of drug discovery. This work set the basis of our present study.

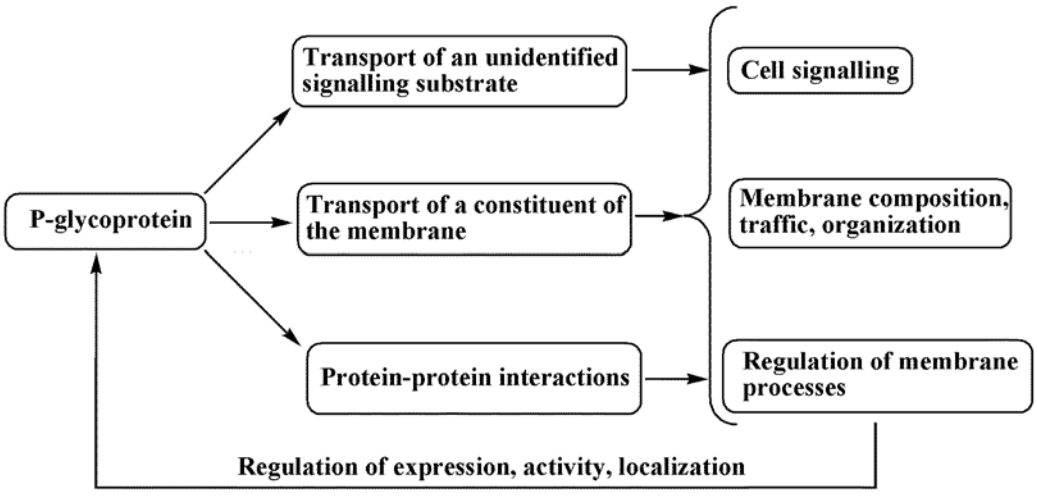


Fig. (1). Possible physiological role for P-glycoprotein beyond its protective functions toward xenobiotics.

1.4. Necessity for Automated Rule-Extraction

Classical QSARs permits the classification of chemicals as substrates/non-substrates, but their information cannot be easily incorporated on the existing expert systems due to the cryptic nature of the variables included in such models [2]. On the other hand, the traditional methods for extracting knowledge from human expertise requires a great amount of information about a set of chemicals permitting the experts their generalization. In addition, the rate of producing new chemical entities overtakes the rate of their toxicological profile evaluation. Thus a method that permits to extract knowledge from the minimum information available about a series of chemicals is necessary to keep expert systems updated. In this sense, an expert system can be considered as *knowledge archive* where a collection of knowledge is expressed using some formal representation language. An *automatic knowledge generator* is a methodology that will provide new structural alerts to the knowledge archive in a cyclic way keeping it updated. In previous works [54-56] we have shown that the so-called topological sub-structural molecular design (TOPS-MODE) approach [57-62] represents a useful platform for the automatic generation of toxicological structural alerts. In these works a general strategy for knowledge flow concerning skin sensitization and chromosome aberrations based on the combined use of TOPS-MODE and DEREK expert system was proposed [54-56].

Taking into consideration the above mentioned, this study describes a quantitative structure-activity relationship (QSAR) analysis on a dataset of compounds assayed for P-gp activity. Then, the main purpose of the current work is to generate structural alert rules that permit the identification of substrates of P-gp in chemicals database using information coded in their molecular structure. Thus we develop a classification model using the TOPS-MODE approach, which allows the calculation of the contribution of each part of a molecule to the activity under study. Using this information we identify structural regions responsible for the P-gp activity of chemicals and transform this information into structural alert rules which are ready to be implemented in expert systems such as DEREK.

2. DATA SET

All experimental procedures were carried out by Ramu and Ramu, and experimental details can be found in the literature [63-65]. Briefly, ED_{50} values were collected for a line of P388 murine leukemia cells that were resistant to adriamycin, ADR (ED_{50} values denote the drug concentration effective in inhibiting cell growth by 50%). The ED_{50} values were collected for cells in the presence of the drug and for cells in the presence of the drug and 200 nM ADR. The drug's ability to reverse MDR (R-Fold, RF) was measured as: RF = (ED_{50} with no ADR) / (ED_{50} with 200 nM ADR). An RF value of 1.0 indicates no ability to reverse MDR (non-substrate), while large RF values indicate excellent ability to reverse MDR (substrate). The compound structures and their associated RF values were taken from the literature [53]. Compounds were classified as P-gp substrate if there were reported to be bound and transported out of the cell by P-gp, and compounds were designed P-gp non-substrates if they were not bound neither transported by P-gp.

A data set of compounds was carefully assembled from literature [15, 53]. We left out 12 of the 609 compounds available due to software limitations (because isomer relations), leaving 597 compounds for analysis [53]. Compounds reported as moderate by Bakken-Jurs [53], 75 in total, were also removed by taking into account the following considerations:

i. we are looking for a discrimination "exclusively" between actives substrates and non-substrates, so we discard compounds having a moderate activity (marginal compounds according to Klopman et. al. [15]). Nevertheless, we keep in mind the same discrimination considered by Bakken & Jurs, taken as the first analysis approach in their work (see Table **6** in [53]);

ii. the substrate (moderates and actives) composition is 62.7 % and only 37.3 % for non-substrates [15, 53]. This could falsify the QSAR results. In the present work we based our study in one of two classifications given by both precedent works [15, 53], specifically the one referred to data set 1 (see Table **1** in [53]), which will be described later on in this paper;

iii. in fact the "moderates (marginal) compounds" are classified as active or inactives substrates by some authors [9. 25, 47, 49, 66], which gives an idea of the "marginality" of them, not showing a clear cut-off criterion for the classification of these compounds (actives, moderates, inactives). Some of these chemicals are shown in Table **1**.

Table 1. List of Moderate [53] or Marginal [15] Compound Classified as "Active" Substrate or as non-Substrate by Others Authors

Compounds	RF	S/NS	Reference
Procyclidine	2.2	NS	[66]
Chloropromazine	2.5	S	[47]
Triflupromazine	2.5	S	[25, 47, 49]
Mequitazine	2.7	S	[25]
Promazine	3.0	S	[47]
Desipramine	3.0	NS	[66]
Imipramine	3.0	NS	[25]
Clozapine	3.3	S	[9, 25]
Fluphenazine	4.0	S	[25, 47, 49]
Bepridil	4.2	S	[25, 47, 49]

S - Substrate; NS - Non Substrate

Consequently, our data is formed by 522 compounds, from which 299 are substrates and 223 are non-substrates (57.3 % and 42.7 % respectively).

In order to generate classification models, compounds were first grouped according to MDRR activity. Bakken-Jurs [53] suggests labeling compounds with RF ≤ 2.0 as inactive and compounds with RF > 4.2 as active. Such a division was used in the present study and will be referred to as data set 1 (see Table **1** in [53]). The division of compounds into classes based on activity is somewhat arbitrary [53]. Therefore, classification models were generated for the two-class problem (inactive/active).

This data set was subdivided according to [53] into two subsets one containing 345 compounds (199 actives and 146 inactives) used as a training set for developing the classification model. The other subset formed by 177 compounds (100 actives and 77 inactives) was used as a prediction set.

3. METHODOLOGY

3.1. The TOPS-MODE Approach

In the last 12 years we have developed an approach to QSAR/QSPR and molecular design. It is known as TOPS-MODE approach, which is the acronym for topological substructural molecular descriptors/design [57-62]. TOPS-MODE approach is based on the calculation of spectral moments of molecular bond matrices appropriately weighted to account for hydrophobic, electronic and steric molecular features. Spectral moments are the trace of the *k*th power of a matrix, i.e., the sum of all entries in the main diagonal of such matrices. The reader is referred to [57-59] to obtain full details of this method.

A bond matrix is a square symmetric matrix in which non-diagonal entries are ones or zeroes if the corresponding bonds have a common atom or not, respectively [67]. These matrices represent the molecular skeleton without taking into account hydrogen atoms. Bonds weights are placed as diagonal entries of such matrices and represent quantitative contributions to different physicochemical properties. Among bond weights currently in use in our approach we have standard bond distance (SD), standard bond dipole moments (DM), hydrophobicity (H) [68], polar surface area (PS) [69], polarizability (Pol) [70], molar refractivity (MR) [70], van der Waals radii (vdW) [71], and Gasteiger-Marsilli charges (Ch) [72].

The starting point for our approach is to calculate TOPS-MODE descriptors of the different types, e.g., H, PS, Pol, MR, vdW, and Ch, for the series of molecules under study. Then, we develop a quantitative model describing the property under study in terms of the spectral moments. In general this model can be of the following form:

$$P = b_0 + \sum_{j=1}^{L} b_j \mu_j \quad (1)$$

where P is the property under study, b_j are the coefficients of the quantitative model (linear regression or discriminant analysis) and b_0 is the error.

The *j*th spectral moment of the bond matrix can be expressed as a sum of bond moments, which are simply the corresponding entries of the *j*th power of the bond matrix:

$$\mu_j = \sum_{i=1}^{m} \mu_j(i), \quad (2)$$

where $\mu_j(i)$ is the bond moment of the *i*th bond in a molecule with *m* bonds. Then, model [2] can be written as:

$$P = b_0 + \sum_{j=1}^{L} b_j \sum_{i=1}^{m} \mu_j(i) = b_0 + \sum_{i=1}^{m} \sum_{j=1}^{L} b_j \mu_j(i), \quad (3)$$

where the right-hand side in (3) represents the contribution of bond *i* to the property *P* and is called the ''bond contribution'' and represented by $P(i)$:

$$P(i) = \sum_{j=1}^{L} b_j \mu_j(i), \quad (4)$$

and the property *P* can be expressed as an additive function of bond contributions:

$$P = \sum_{i=1}^{m} P(i) \quad (5)$$

Bond contributions are numeric characterization of bonds which permit to identify some groups or regions of a molecular framework which can be responsible for a property/activity [73]. By carefully analyzing similar regions in different molecules we can obtain general rules about the contributions of molecular fragments to a particular property/activity. They are based on the substructural nature of TOPS-MODE. This procedure consists in transforming a QSPR or QSAR model into a bond additive scheme in which a property can be calculated as the sum of bond contributions for a molecule.

3.2. Orthogonalization of TOPS-MODE Descriptors

One of the inherent characteristic of the TOPS-MODE approach is that spectral moments are collinear among them. This means that there is redundancy in the information contained in any pair of collinear descriptors. Two descriptors are called collinear if they have a significant linear correlation between them as measured by the linear correlation coefficient. The main drawback of collinearity from the point of view of a QSAR model is that of the stability of the coefficients in the linear regression model. This introduces a difficulty in interpreting the models obtained with collinear variables because the sign and magnitude of the coefficients in the regression model can be affected by the removal or introduction of a new variable in the model. In the case of the TOPS-MODE approach this can be traduced into false interpretation of bond contributions because the magnitude and sign of them can be falsified by the effect produced by the existence of collinear variables in the model. Consequently, we have implemented the Randić's method of orthogonalization [74-76] to eliminate the collinearities between the TOPS-MODE variables. In doing so, we have developed a new approach to extract the information contained in these variables after orthogonalization.

The Randić's method of orthogonalization has been described in details in several publications [74-78]. Thus, we will give only a general overview here. The first step in orthogonalizing the molecular descriptors is to select the appropriate order of orthogonalization, which in this case is the order in which the variables were selected in the forward stepwise search procedure of the linear discriminant analysis. The first variable (v_1) is taken as the first orthogonal descriptors ${}^1\Omega(v_1)$ and the second one is orthogonalized respect to it by taking the residual of its correlation with ${}^1\Omega(v_1)$. The process is repeated until all variables are completely orthogonalized and the orthogonal variables are then used to obtain the new model. Let consider the following QSAR/QSPR model: $P = a_0 + a_1X_1 + a_2X_2 + a_3X_3$, then the orthogonalization of the independent variables is carried out as follows:

i. Orthogonalizes $X_1 : \Omega(X_1) = X_1$,

ii. Orthogonalizes X_1 respect to X_2: $\Omega(X_2) = X_2 - \hat{X}_2$, where $\hat{X}_2 = b_0 + b_1X$,

iii. Orthogonalizes X_3 respect to $\Omega(X_2)$: $\Omega(X_3) = X_3 - \hat{X}_3$, where $\hat{X}_3 = b'_0 + b'_1 \Omega(X_2)$,

iv. Orthogonalizes $\Omega(X_3)$ respect to X_1: ${}^2\Omega(X_3) = \Omega(X_3) - \hat{\Omega}(X_3)$, where $\hat{\Omega}(X_3) = b''_0 + b''_1 X_1$

The orthogonalized variables are then: X_1, $\Omega(X_2)$ and ${}^2\Omega(X_3)$ and the coefficients in steps (ii) - (iv) are obtained by linear regression analysis.

In order to extract the information contained in the orthogonalized descriptors, i.e., bond contributions, we implemented the following iterative procedure:

i. Calculate bond contributions to $\Omega(X_1): C[\Omega(X_1)] = C(X_1)$,

ii. Calculate bond contributions to $\Omega(X_2): C[\Omega(X_2)] = C(X_2) - b_1 C(X_1)$, ,

iii. Calculate bond contributions to $\Omega(X_2): C[\Omega(X_3)] = C(X_3) - b'_1 C[\Omega(X_2)]$, ,

iv. Calculate bond contributions to ${}^2\Omega(X_3): C[{}^2\Omega(X_3)] = C[\Omega(X_3)] - b''_1 C(X_1)$.

The final bond contributions are $C(X_1)$, $C[\Omega(X_2)]$ and $C[{}^2\Omega(X_3)]$. This procedure represents the extraction of the information contained into a bond contribution of a variable which is duplicated by the other variables in the model.

4. CLASSIFICATION MODEL

A linear discriminant model was developed using our training data set (TSET) of 345 compounds. The model contains six TOPS-MODE descriptors accounting for hydrophobic, electronic and steric features of molecules. The model classifies correctly 86.7% of the total number of compounds in the TSET (88.4% for substrate and 84.3% of good classification for non-substrate compounds). The percentage of false actives in the training set was only 11.6% (23 inactive compounds were classified as actives from 345 cases), and 15.7% were false inactives (23 substrate compounds were classified as inactive).

The classification function to discriminate P-gp substrate from P-gp nonsubstrates, in the compounds belonging to the training set, is given below together with the statistical parameters of the LDA (λ is the *Wilks*' statistics, D^2 is the squared Mahalanobis distance and F is the Fisher ratio):

$$\begin{aligned} Class(1) = {} & 1.2221 \cdot 10^{-2} \mu_3^{Pol} - 7.1923 \cdot 10^{-12} \mu_{14}^{Pol} - 3.7016 \cdot 10^{-6} \mu_{11}^{Ch} \\ & + 1.4125 \cdot 10^{-10} \mu_{15}^{SD} + 3.4212 \cdot 10^{-3} \mu_4^{SD} - 5.2410 \cdot 10^{-3} \mu_5^{Ch} \\ & - 8.2998 \end{aligned} \quad (7)$$

$$N = 345 \quad Wilks - \lambda = 0.585 \quad F(6,\ 338) = 39.900 \quad D^2 = 2.891$$

This model shows the best performance in predicting P-glycoprotein activity in both the TSET and PSET among all the models generated using TOPS-MODE descriptors and linear discriminant analysis. In Table **2** we give the classification of all compounds used in the training and test sets by using this model.

As can be seen in the above equation, six variables are present in the model and some of them are spectral moments of high order. Taking into consideration that these variables could be mathematically collinear and over-fitting results can be produced, the Randic's orthogonalization procedure [74-76], in order to avoid the collinearity among different variables was carried out.

Then, we proceed to orthogonalize the variables in this model in order to eliminate any collinearity present among the variables included in the model. Following the Randic´s orthogonalization procedure previously described we generate the following orthogonal classification model:

$$\begin{aligned} Class(2) = {} & 1.7983 \cdot 10^{-2} \left[\Omega\left(\mu_3^{Pol}\right)\right] + 1.6431 \cdot 10^{-10} \left[\Omega\left(\mu_{15}^{SD}\right)\right] \\ & - 1.3710 \cdot 10^{-11} \left[\Omega\left(\mu_{14}^{Pol}\right)\right] - 6.3904 \cdot 10^{-7} \left[\Omega\left(\mu_{11}^{Ch}\right)\right] \\ & + 1.6223 \cdot 10^{-3} \left[\Omega\left(\mu_4^{SD}\right)\right] - 5.2410 \cdot 10^{-3} \left[\Omega\left(\mu_5^{Ch}\right)\right] \\ & - 7.1772 \end{aligned} \quad (8)$$

where ${}^m\Omega\left(\mu_k^w\right)$. Where, the symbol Ω means orthogonal, m is the degree of importance of the descriptor to explain the property determined by the order in which it is selected by forward stepwise analysis, μ_k is the kth spectral moment and w is the bond weight used. In this analysis the least squares method selected all orthogonal analogs of collinear variables. It ensured us that, in spite of collinear variables, each one has an amount of information not encoded in the others and the relative importance of each variable can be determined [74-76].

The variables in the model (Eq. 8) encoded specific structural information, being the most influential descriptors those weighted with standard bond distance, polarizability and the atomic charge. On the other hand, the inclusion of polarizability is a complement of the bond distance since both are influenced by the electronegativity of the atoms that form the bond. Another computational study of 22 diverse drugs revealed that molecular descriptors associated with strong hydrogen bonding strength and high polarizability promote increased P-gp ATPase activity [45]. In general, these three properties are taken several times into account in the scientific literature, emphasizing the bond distance and the atomic charges [12, 15, 22, 23, 25, 33, 41, 44, 45, 47, 49, 50, 52, 53].

We remark here that the classification of compounds using the orthogonalized model is exactly the same that those using the non-orthogonalized one and that the main differences are in the interpretation of the results. As can be seen there are not changes in the sign of coefficients. However, the relative contribution of the variables in the orthogonalized model is significantly different compared to those in the non-orthogonalized one. For instance, the variables μ_4^{SD} and μ_5^{Ch} have practically similar contributions (in absolute terms) in the non-orthogonalized model. However, in the orthogonalized model the contribution of μ_5^{Ch} is five times larger than that of μ_4^{SD}. These differences in the relative importance of the variables in both models can influence the contributions of the different bonds to the P-glycoprotein activity of the compounds under study, which are the main purpose of our research.

The most important criterion for the quality of the discriminant model is based on the statistics for the external prediction set. In the PSET the percentage of good classification is 81.8% (82.0% for substrate and 80.2% for non-substrate). For this prediction set the false active rate is 18.0% and 19.8% the false inactives one. It is desirable that the number of false active compounds be as low as possible because this number represents inactive compounds that will be sent to the biological assays with the consequent loss of time and resources. Finally, the predictive capacity of the computational model was assessed by three data sets. The first set of compounds was composed by those listed in Table **1**. These compounds are moderate according to [53], but classified as substrates and non-substrates by other authors (see Table **1**). In Table **3**, we present the results obtained by using Eq. (7) in predicting the activity of compounds in Table **1**.

The percentage of good classification is 80.0%, 71.4% for substrate (5/7) and 100% for non-substrate (3/3). Our classification

Table 2. Predictions Made by Using TOPS-MODE Classification Model for Substrate (1) and Non-Substrate (-1) Compounds in the TSET and PSET (Compounds Ordered from Highest to Lowest Based on the Value of RF [53])

No.	No. in [55]	CAS	Class.	Pred.	Prob.
1	545	a	1	1	99.8
2	292	16662-47-8	1	1	83.1
3	144	144236-78-2	1	1	99.7
4	287	52-53-9	1	1	80.3
5	301	58033-02-6	1	1	96.8
6	306	3625-06-7	1	1	75.8
7	541	a	1	1	95.3
8	542	a	1	1	96.1
9	543	a	1	1	98.0
10	226	56079-61-9	1	1	97.2
11	300	67018-83-1	1	1	74.9
12	120	298-55-5	1	1	98.9
13	320	286841-82-5	1	1	76.1
14	232	144236-87-3	1	1	99.5
15	231	144236-86-2	1	1	98.6
16	272	161161-56-4	1	1	74.3
17	294	16740-29-7	1	1	95.4
18	343	35898-87-4	1	1	90.0
19	344	7077-33-0	1	1	91.2
20	345	161161-57-5	1	1	96.4
21	540	a	1	1	97.2
22	544	a	1	1	99.1
23	43	a	1	1	81.8
24	208	53179-11-6	1	1	95.6
25	227	56079-81-3	1	1	97.6
26	228	54742-90-4	1	1	97.1
27	296	16359-24-3	1	1	95.4
28	66	79781-95-6	1	1	95.9
29	109	2062-78-4	1	1	99.4
30	291	78370-15-7	1	1	98.3
31	302	85247-76-3	1	1	99.4
32	536	a	1	1	97.0
33	590	25000-95-7	1	1	93.6
34	248	27076-46-6	1	1	87.7
35	575	16908-55-7	1	1	92.9
36	110	53179-12-7	1	1	99.8

(Table 2) Contd….

No.	No. in [55]	CAS	Class.	Pred.	Prob.
37	18	62030-88-0	1	1	99.8
38	535	a	1	1	81.4
39	290	78370-14-6	1	1	80.0
40	170	522-18-9	1	1	99.0
41	230	144236-85-1	1	1	98.7
42	295	78370-11-3	1	1	84.2
43	143	144236-77-1	1	1	99.5
44	243	67914-69-6	1	1	99.1
45	250	68576-86-3	1	1	56.6
46	19	62030-90-4	1	1	99.0
47	34	388-51-2	1	1	99.6
48	220	982-43-4	1	1	68.1
49	242	67915-35-9	1	1	96.0
50	256	52618-67-4	1	1	84.6
51	270	32487-03-9	1	1	60.0
52	329	22609-73-0	1	-1	39.6
53	339	510-74-7	1	1	70.4
54	539	a	1	1	98.9
55	595	16888-10-1	1	1	93.6
56	114	5522-39-4	1	1	99.3
57	298	41035-34-1	1	1	95.5
58	550	a	1	1	88.3
59	592	14343-19-2	1	1	89.4
60	27	316-81-4	1	1	74.5
61	241	65277-42-1	1	1	98.4
62	251	75706-37-5	1	1	68.8
63	252	103997-59-7	1	1	81.1
64	341	58581-89-8	1	1	90.7
65	280	55985-32-5	1	1	67.7
66	80	53772-82-0	1	1	96.6
67	148	82137-00-6	1	1	95.8
68	254	85673-87-6	1	1	97.7
69	459	53230-10-7	1	1	74.3
70	526	a	1	1	76.6
71	594	14464-97-2	1	1	98.6
72	13	5588-33-0	1	1	58.4
73	57	65509-66-2	1	1	59.9

(Table 2) Contd….

No.	No. in [55]	CAS	Class.	Pred.	Prob.
74	58	2058-52-8	1	1	75.1
75	61	14028-44-5	1	1	58.9
76	74	16378-21-5	1	1	90.9
77	106	144236-73-7	1	1	99.9
78	177	a	1	-1	U
79	215	5633-20-5	1	-1	31.1
80	258	83366-66-9	1	1	98.0
81	337	24678-13-5	1	1	77.6
82	374	70312-00-4	1	1	84.3
83	440	85-79-0	1	-1	34.8
84	444	51481-62-0	1	1	78.1
85	521	a	1	1	81.9
86	532	a	1	1	94.4
87	569	76203-97-9	1	1	93.8
88	274	39133-31-8	1	-1	36.3
89	516	5061-22-3	1	1	99.8
90	40	1096-72-6	1	1	74.9
91	113	79467-23-5	1	1	99.7
92	257	14728-33-7	1	1	99.6
93	118	569-65-3	1	1	98.9
94	335	72803-02-2	1	-1	7.1
95	14	14759-06-9	1	1	67.7
96	15	13093-88-4	1	1	75.6
97	24	117-89-5	1	1	94.5
98	29	2751-68-0	1	1	71.5
99	36	2470-73-7	1	1	83.1
100	37	49864-70-2	1	1	92.1
101	67	42239-60-1	1	1	U
102	97	390-64-7	1	1	85.3
103	98	57653-27-7	1	1	87.6
104	129	60607-34-3	1	1	97.6
105	134	80273-79-6	1	1	95.9
106	140	14176-10-4	1	-1	23.4
107	149	144236-80-6	1	1	93.6
108	151	a	1	1	95.6
109	167	90729-43-4	1	1	99.5
110	168	86-13-5	1	1	76.2

(Table 2) Contd….

No.	No. in [55]	CAS	Class.	Pred.	Prob.
111	169	524-83-4	1	1	86.6
112	191	117023-61-7	1	1	99.6
113	239	54063-40-0	1	1	95.7
114	246	153-87-7	1	1	84.5
115	285	78370-13-5	1	1	67.5
116	304	57558-44-8	1	1	U
117	307	15687-16-8	1	1	90.9
118	311	90-54-0	1	-1	U
119	331	88150-42-9	1	-1	21.1
120	445	56208-01-6	1	1	88.5
121	461	1301-42-4	1	1	91.7
122	522	a	1	1	95.3
123	523	a	1	1	93.5
124	527	a	1	1	97.6
125	528	a	1	1	79.9
126	570	58-32-2	1	1	58.0
127	573	16908-50-2	1	-1	36.7
128	578	77749-49-6	1	-1	35.5
129	378	57149-07-2	1	1	89.0
130	145	82117-51-9	1	1	92.7
131	51	31721-17-2	1	1	90.4
132	141	144236-75-9	1	1	94.2
133	534	a	1	-1	21.2
134	584	16908-52-4	1	1	98.3
135	587	54093-30-0	1	1	64.9
136	7	60-87-7	1	1	84.0
137	47	83166-17-0	1	-1	34.5
138	123	68-88-2	1	1	71.9
139	225	43016-37-1	1	-1	45.1
140	253	1480-19-9	1	1	74.5
141	336	77590-96-6	1	1	96.3
142	353	28820-28-2	1	1	92.4
143	361	442-52-4	1	1	79.2
144	397	53076-26-9	1	1	88.1
145	450	3692-16-8	1	1	96.5
146	474	77400-65-8	1	1	95.4
147	568	59831-63-9	1	1	99.2

(Table 2) Contd….

No.	No. in [55]	CAS	Class.	Pred.	Prob.
148	600	120593-58-0	1	1	53.2
149	121	298-57-7	1	1	96.9
150	23	58-38-8	1	1	80.5
151	152	a	1	1	88.2
152	173	144236-83-9	1	1	63.0
153	326	749-13-3	1	1	90.6
154	377	a	1	1	70.5
155	324	90-69-7	1	-1	30.4
156	396	37640-71-4	1	1	88.0
157	111	75558-90-6	1	1	88.8
158	12	50-52-2	1	1	76.8
159	49	739-71-9	1	1	69.7
160	119	82-95-1	1	1	99.7
161	199	64294-95-7	1	1	88.1
162	318	10540-29-1	1	1	94.6
163	562	72479-26-6	1	1	97.0
164	62	22013-23-6	1	1	59.5
165	68	314-03-4	1	1	U
166	83	15574-96-6	1	-1	34.0
167	122	52468-60-7	1	1	99.1
168	150	a	1	1	96.3
169	477	68844-77-9	1	1	98.9
170	483	1951-25-3	1	1	99.4
171	508	59831-64-0	1	1	89.9
172	552	60628-96-8	1	1	84.6
173	263	366-93-8	1	1	91.5
174	9	84-08-2	1	1	55.4
175	28	58-39-9	1	1	84.3
176	64	1977-11-3	1	1	66.3
177	65	67121-76-0	1	1	79.7
178	71	129-03-3	1	1	78.3
179	102	58473-73-7	1	-1	24.7
180	131	3601-19-2	1	1	89.0
181	166	147-20-6	1	-1	U
182	179	23239-78-3	1	1	56.8
183	190	972-02-1	1	1	58.7
184	255	67254-81-3	1	1	U

(Table 2) Contd….

No.	No. in [55]	CAS	Class.	Pred.	Prob.
185	281	5585-64-8	1	1	90.5
186	308	32665-36-4	1	1	58.1
187	424	a	1	-1	42.6
188	437	59831-65-1	1	1	78.2
189	513	1893-33-0	1	-1	U
190	548	a	1	1	90.0
191	101	3426-08-2	1	-1	32.3
192	4	61-01-8	1	1	73.8
193	22	84-97-9	1	1	70.2
194	127	17692-34-1	1	1	76.5
195	206	901-02-0	1	-1	39.6
196	376	a	1	1	84.0
197	393	74191-85-8	1	1	81.0
198	553	60628-98-0	1	1	93.4
199	561	22833-02-9	1	1	86.3
200	546	a	1 [b]	1	99.7
201	342	54-03-5	1 [b]	1	74.2
202	267	108704-90-1	1 [b]	1	64.8
203	293	67018-81-9	1 [b]	1	91.0
204	266	475-81-0	1 [b]	1	65.8
205	268	161273-29-6	1 [b]	1	54.7
206	130	68741-18-4	1 [b]	1	99.8
207	100	23891-60-3	1 [b]	1	85.7
208	284	59170-23-9	1 [b]	-1	13.6
209	338	357-66-4	1 [b]	1	87.0
210	531	a	1 [b]	1	83.3
211	538	a	1 [b]	1	96.7
212	589	16982-41-5	1 [b]	1	94.8
213	286	67018-79-5	1 [b]	1	70.5
214	209	53179-10-5	1 [b]	1	99.5
215	265	2688-77-9	1 [b]	1	U
216	537	a	1 [b]	1	98.6
217	50	58503-82-5	1 [b]	1	97.3
218	297	92302-55-1	1 [b]	1	77.3
219	571	16908-39-7	1 [b]	1	58.0
220	75	54341-02-5	1 [b]	1	98.1
221	76	3313-26-6	1 [b]	1	90.1

(Table 2) Contd….

No.	No. in [55]	CAS	Class.	Pred.	Prob.
222	269	30001-98-0	1 [b]	1	55.9
223	328	66085-59-4	1 [b]	-1	18.6
224	85	82140-22-5	1 [b]	-1	27.6
225	271	114697-88-0	1 [b]	1	67.2
226	78	4177-58-6	1 [b]	1	91.0
227	264	4747-98-2	1 [b]	-1	31.6
228	330	39562-70-4	1 [b]	-1	8.7
229	316	1178-99-0	1 [b]	1	99.2
230	327	51493-19-7	1 [b]	1	98.5
231	25	653-03-2	1 [b]	1	87.5
232	240	67915-31-5	1 [b]	1	99.1
233	33	17692-26-1	1 [b]	1	98.8
234	520	57818-92-5	1 [b]	1	99.2
235	30	2622-30-2	1 [b]	1	78.2
236	77	982-24-1	1 [b]	1	91.9
237	321	10448-84-7	1 [b]	1	97.4
238	325	52-86-8	1 [b]	1	79.8
239	577	17312-08-2	1 [b]	1	78.8
240	59	5800-19-1	1 [b]	1	66.8
241	60	1977-10-2	1 [b]	1	78.1
242	108	1841-19-6	1 [b]	1	99.3
243	146	82117-73-5	1 [b]	1	96.1
244	249	4448-96-8	1 [b]	1	83.1
245	401	85-10-9	1 [b]	1	63.5
246	460	130-95-0	1 [b]	1	61.7
247	322	1845-11-0	1 [b]	-1	38.1
248	79	a	1 [b]	1	94.7
249	313	6732-77-0	1 [b]	1	97.0
250	574	16982-40-4	1 [b]	1	94.1
251	26	1420-55-9	1 [b]	1	81.6
252	81	53772-85-3	1 [b]	1	96.6
253	96	13042-18-7	1 [b]	1	82.2
254	125	83881-37-2	1 [b]	1	65.3
255	142	144236-76-0	1 [b]	1	95.9
256	163	3703-76-2	1 [b]	1	80.4
257	174	82227-49-4	1 [b]	1	53.7
258	175	55837-21-3	1 [b]	1	55.9

(Table 2) Contd….

No.	No. in [55]	CAS	Class.	Pred.	Prob.
259	247	24360-55-2	1 [b]	1	86.4
260	273	3735-45-3	1 [b]	-1	37.5
261	305	4378-36-3	1 [b]	1	74.5
262	403	27848-84-6	1 [b]	1	90.7
263	512	63959-21-7	1 [b]	1	83.1
264	560	27523-40-6	1 [b]	1	93.9
265	585	16908-56-8	1 [b]	-1	36.0
266	598	96801-86-4	1 [b]	1	68.2
267	41	19410-02-7	1 [b]	-1	33.4
268	505	5636-83-9	1 [b]	1	80.3
269	593	14343-31-8	1 [b]	1	80.9
270	510	30271-90-0	1 [b]	1	60.9
271	214	4354-45-4	1 [b]	-1	28.7
272	283	22232-57-1	1 [b]	-1	24.8
273	362	60662-19-3	1 [b]	1	78.4
274	112	3416-26-0	1 [b]	1	99.5
275	107	26864-56-2	1 [b]	1	99.9
276	439	32421-46-8	1 [b]	1	71.8
277	530	a	1 [b]	1	77.8
278	315	53775-12-5	1 [b]	1	99.6
279	8	13461-01-3	1 [b]	1	58.4
280	95	15793-40-5	1 [b]	1	71.5
281	379	54063-58-0	1 [b]	1	53.6
282	501	525-66-6	1 [b]	-1	8.4
283	5	61-00-7	1 [b]	1	79.0
284	69	60085-78-1	1 [b]	-1	35.1
285	192	50679-08-8	1 [b]	1	99.3
286	38	800-22-6	1 [b]	1	68.5
287	73	50-48-6	1 [b]	1	57.0
288	164	525-01-9	1 [b]	-1	36.7
289	205	302-33-0	1 [b]	1	61.1
290	207	25021-35-6	1 [b]	-1	19.4
291	309	34758-83-3	1 [b]	1	57.6
292	310	469-62-5	1 [b]	1	54.9
293	493	69118-25-8	1 [b]	1	79.5
294	312	78-41-1	1 [b]	1	98.8
295	178	93076-89-2	1 [b]	1	96.6

(Table 2) Contd….

No.	No. in [55]	CAS	Class.	Pred.	Prob.
296	20	3819-00-9	1 [b]	1	81.6
297	183	21826-41-5	1 [b]	1	96.9
298	472	89365-50-4	1 [b]	-1	26.6
299	525	a	1 [b]	1	96.7
300	6	84-96-8	-1	-1	54.5
301	53	5585-93-3	-1	-1	55.3
302	147	144236-79-3	-1	1	1.4
303	188	511-45-5	-1	-1	U
304	212	514-65-8	-1	1	35.0
305	299	86656-06-6	-1	1	14.2
306	314	911-45-5	-1	1	1.3
307	332	72509-76-3	-1	-1	59.7
308	355	91-84-9	-1	-1	84.4
309	422	58-74-2	-1	-1	65.5
310	447	41717-30-0	-1	1	40.8
311	558	61318-90-9	-1	1	17.2
312	586	16908-48-8	-1	-1	63.1
313	88	a	-1	-1	64.5
314	92	980-71-2	-1	-1	76.6
315	153	15301-93-6	-1	-1	87.2
316	154	58-73-1	-1	-1	83.6
317	156	4024-34-4	-1	-1	72.4
318	186	467-60-7	-1	-1	79.4
319	189	968-63-8	-1	-1	60.1
320	234	961-71-7	-1	-1	74.5
321	354	91-81-6	-1	-1	86.4
322	365	50335-55-2	-1	-1	97.7
323	372	34661-75-1	-1	-1	54.1
324	373	83928-76-1	-1	-1	74.9
325	381	1649-18-9	-1	1	48.5
326	418	118-08-1	-1	1	34.4
327	432	24526-64-5	-1	-1	73.8
328	491	23887-41-4	-1	-1	89.1
329	496	15301-80-1	-1	1	36.7
330	529	a	-1	1	21.7
331	237	493-78-7	-1	-1	94.8
332	478	3215-70-1	-1	-1	96.3

(Table 2) Contd….

No.	No. in [55]	CAS	Class.	Pred.	Prob.
333	219	23744-24-3	-1	1	37.9
334	402	86-42-0	-1	1	34.4
335	414	63-12-7	-1	-1	U
336	515	7009-79-2	-1	1	37.4
337	46	303-69-5	-1	-1	85.2
338	72	3964-81-6	-1	-1	U
339	261	86-12-4	-1	-1	81.7
340	276	620-40-6	-1	1	29.6
341	303	57010-31-8	-1	-1	90.4
342	413	58-46-8	-1	-1	59.6
343	446	a	-1	1	21.8
344	466	447-41-6	-1	-1	80.7
345	471	77862-92-1	-1	1	39.5
346	475	510-53-2	-1	-1	62.6
347	554	23593-75-1	-1	1	26.5
348	567	84697-22-3	-1	1	19.0
349	400	54-05-7	-1	-1	U
350	44	10423-37-7	-1	-1	81.1
351	52	28797-61-7	-1	-1	80.6
352	70	31232-26-5	-1	1	45.6
353	86	3963-62-0	-1	-1	89.6
354	87	957-51-7	-1	-1	89.6
355	91	132-22-9	-1	-1	79.8
356	104	64-95-9	-1	-1	53.8
357	133	2759-28-6	-1	-1	98.4
358	138	1679-76-1	-1	-1	54.1
359	158	524-99-2	-1	-1	81.3
360	159	486-16-8	-1	-1	83.4
361	165	6703-39-5	-1	-1	89.7
362	180	47128-12-1	-1	-1	54.8
363	187	791-35-5	-1	-1	74.5
364	194	302-40-9	-1	-1	72.2
365	197	469-21-6	-1	-1	92.1
366	202	77-01-0	-1	-1	56.8
367	221	57-41-0	-1	-1	97.6
368	222	741-28-6	-1	-1	96.7
369	223	19283-01-3	-1	-1	96.0

(Table 2) Contd….

No.	No. in [55]	CAS	Class.	Pred.	Prob.
370	224	19283-04-6	-1	-1	94.3
371	259	91-75-8	-1	-1	84.4
372	262	a	-1	-1	73.9
373	275	103-49-1	-1	-1	94.9
374	340	528-52-9	-1	1	U
375	346	a	-1	-1	97.8
376	348	31793-07-4	-1	-1	93.4
377	349	50892-23-4	-1	-1	98.2
378	350	90-30-2	-1	-1	73.8
379	352	30544-61-7	-1	-1	76.1
380	357	91-85-0	-1	-1	95.1
381	359	91-79-2	-1	-1	97.0
382	360	91-80-5	-1	-1	97.6
383	363	92-54-6	-1	-1	98.3
384	367	299-48-9	-1	-1	90.8
385	369	60207-31-0	-1	-1	92.2
386	370	4004-94-8	-1	-1	96.9
387	384	70458-96-7	-1	-1	70.7
388	385	74011-58-8	-1	-1	87.7
389	387	70458-92-3	-1	-1	59.5
390	389	79660-72-3	-1	-1	72.6
391	390	82419-36-1	-1	-1	53.1
392	392	66969-81-1	-1	-1	80.4
393	399	4298-15-1	-1	-1	94.0
394	404	103-83-3	-1	-1	99.4
395	405	a	-1	-1	99.2
396	406	a	-1	-1	99.1
397	409	a	-1	-1	99.2
398	410	a	-1	-1	98.9
399	411	a	-1	-1	99.2
400	416	4747-99-3	-1	-1	94.6
401	417	485-33-6	-1	-1	88.2
402	420	476-70-0	-1	-1	59.6
403	423	18429-69-1	-1	-1	85.3
404	428	1485-70-7	-1	-1	96.2
405	431	54870-28-9	-1	-1	88.8
406	435	364-62-5	-1	-1	93.4

(Table 2) Contd….

No.	No. in [55]	CAS	Class.	Pred.	Prob.
407	436	71320-77-9	-1	-1	98.5
408	438	318-23-0	-1	-1	81.3
409	441	54824-20-3	-1	-1	56.9
410	451	61714-27-0	-1	-1	81.8
411	452	64-04-0	-1	-1	99.4
412	453	120-20-7	-1	-1	99.2
413	457	24558-01-8	-1	-1	96.9
414	458	101-82-6	-1	-1	97.7
415	463	7541-30-2	-1	-1	94.3
416	464	13392-18-2	-1	-1	96.3
417	465	136-70-9	-1	-1	90.1
418	468	395-28-8	-1	-1	88.1
419	476	27737-38-8	-1	-1	94.4
420	479	5588-21-6	-1	-1	99.4
421	482	6376-26-7	-1	-1	75.1
422	490	54063-23-9	-1	-1	96.5
423	492	23887-47-0	-1	-1	87.2
424	495	32527-55-2	-1	-1	94.1
425	498	26481-51-6	-1	-1	99.0
426	502	14860-49-2	-1	-1	87.5
427	503	3572-52-9	-1	-1	63.6
428	507	55837-25-7	-1	-1	93.3
429	511	553-65-1	-1	-1	84.0
430	514	14007-64-8	-1	-1	90.0
431	518	5696-09-3	-1	-1	81.2
432	519	3415-54-1	-1	1	44.5
433	547	a	-1	1	34.1
434	549	a	-1	1	4.4
435	551	4238-71-5	-1	-1	99.0
436	581	13144-82-6	-1	-1	79.7
437	582	16887-98-2	-1	-1	81.0
438	588	13120-27-9	-1	-1	99.6
439	601	a	-1	-1	96.5
440	602	a	-1	1	0.6
441	604	a	-1	1	15.9
442	605	a	-1	1	28.5
443	606	a	-1	-1	83.9

(Table 2) Contd….

No.	No. in [55]	CAS	Class.	Pred.	Prob.
444	608	a	-1	-1	68.9
445	609	a	-1	-1	68.1
446	116	82-92-8	-1 [b]	-1	53.3
447	235	493-80-1	-1 [b]	-1	60.7
448	260	4945-47-5	-1 [b]	-1	58.4
449	419	128-62-1	-1 [b]	1	31.1
450	449	55905-53-8	-1 [b]	1	42.1
451	421	58-00-4	-1 [b]	-1	67.1
452	35	7224-08-0	-1 [b]	1	2.4
453	124	55837-13-3	-1 [b]	-1	55.5
454	161	a	-1 [b]	-1	53.6
455	193	80387-96-8	-1 [b]	-1	71.2
456	368	55485-20-6	-1 [b]	-1	70.0
457	371	72822-12-9	-1 [b]	-1	71.6
458	382	19794-93-5	-1 [b]	1	21.3
459	395	21560-59-8	-1 [b]	-1	53.0
460	481	a	-1 [b]	1	14.0
461	564	64211-45-6	-1 [b]	1	6.9
462	487	5322-53-2	-1 [b]	-1	U
463	494	7762-32-5	-1 [b]	-1	64.6
464	334	21829-25-4	-1 [b]	-1	95.7
465	103	69365-65-7	-1 [b]	1	3.3
466	155	83-98-7	-1 [b]	-1	74.0
467	375	52867-74-0	-1 [b]	1	18.0
468	398	90-34-6	-1 [b]	-1	95.0
469	448	2545-39-3	-1 [b]	-1	80.2
470	480	92-12-6	-1 [b]	-1	81.9
471	82	a	-1 [b]	1	4.5
472	16	2622-26-6	-1 [b]	1	38.2
473	1 7	14008-44-7	-1 [b]	1	42.4
474	90	86-21-5	-1 [b]	-1	92.1
475	126	83881-51-0	-1 [b]	-1	60.9
476	137	7273-99-6	-1 [b]	-1	76.5
477	160	5560-77-0	-1 [b]	-1	84.9
478	176	82227-38-1	-1 [b]	-1	56.5
479	182	87848-99-5	-1 [b]	-1	U
480	201	60-46-8	-1 [b]	-1	83.2

(Table 2) Contd….

No.	No. in [55]	CAS	Class.	Pred.	Prob.
481	204	3737-09-5	-1 [b]	-1	74.2
482	229	144236-84-0	-1 [b]	-1	92.4
483	233	92-59-1	-1 [b]	-1	77.9
484	279	a	-1 [b]	-1	95.0
485	282	15687-41-9	-1 [b]	-1	90.4
486	347	91-66-7	-1 [b]	-1	96.7
487	351	135-88-6	-1 [b]	-1	72.8
488	358	553-13-9	-1 [b]	-1	97.0
489	364	98224-03-4	-1 [b]	-1	95.7
490	366	17692-31-8	-1 [b]	-1	97.8
491	380	98206-10-1	-1 [b]	1	31.5
492	383	4774-24-7	-1 [b]	-1	91.8
493	386	51940-44-4	-1 [b]	-1	97.4
494	391	15793-38-1	-1 [b]	-1	74.0
495	394	35795-16-5	-1 [b]	-1	60.3
496	407	6913-92-4	-1 [b]	-1	98.1
497	408	635-41-6	-1 [b]	-1	97.7
498	412	550-10-7	-1 [b]	-1	96.0
499	415	30418-38-3	-1 [b]	-1	83.7
500	425	483-18-1	-1 [b]	1	1.5
501	426	4914-30-1	-1 [b]	1	3.5
502	427	22407-74-5	-1 [b]	1	19.0
503	442	3436-11-1	-1 [b]	-1	96.8
504	443	94746-78-8	-1 [b]	-1	89.7
505	455	87-52-5	-1 [b]	-1	98.7
506	456	332-14-9	-1 [b]	-1	93.3
507	462	26652-09-5	-1 [b]	-1	94.5
508	467	34368-04-2	-1 [b]	-1	90.7
509	470	88578-07-8	-1 [b]	-1	79.4
510	473	148-07-2	-1 [b]	-1	U
511	489	23887-46-9	-1 [b]	-1	71.3
512	497	27591-97-5	-1 [b]	-1	52.8
513	500	67227-55-8	-1 [b]	-1	98.0
514	509	57808-66-9	-1 [b]	1	13.6
515	517	2179-37-5	-1 [b]	-1	75.3
516	533	a	-1 [b]	-1	64.0
517	5 72	13665-88-8	-1 [b]	-1	95.4

(Table 2) Contd….

No.	No. in [55]	CAS	Class.	Pred.	Prob.
518	580	75751-70-1	-1 [b]	-1	96.0
519	596	16888-13-4	-1 [b]	-1	63.6
520	599	96801-76-2	-1 [b]	-1	84.3
521	603	a	-1 [b]	-1	99.0
522	607	a	-1 [b]	1	18.3

U - **U**nclassified (compounds for which the percentage of classification as active and inactive does not differ by more than 5%); [a] see smile in supporting information [53]; [b] PSET

Table 3. Prediction Results for Compounds Related in Table 1 Using Eq. 7

Compounds	Reference	S/NS	RF	P-gp Substrate	Class.	Prob.
Chloropromazine	[47]	S	2.5	+	+	64.30
Triflupromazine	[25, 47, 49]	S	2.5	+	-	22.83
Mequitazine	[25]	S	2.7	+	-	43.51
Promazine	[47]	S	3.2	+	+	81.29
Clozapine	[9, 25]	S	3.3	+	+	62.18
Fluphenazine	[25, 47, 49]	S	4.0	+	+	87.90
Bepridil	[25, 47, 49]	S	4.2	+	+	73.47
Procyclidine	[66]	NS	2.2	-	-	71.18
Desipramine	[66]	NS	3.0	-	-	79.35
Imipramine	[25]	NS	3.0	-	-	70.28

S - substrate; NS - Non Substrate; + Positive values are for P-gp substrate compounds; - Negative values are for P-gp non-substrates compounds.

of Mequitazine (substrate) and Imipramine (non-substrate) coincides with predictions made by Cabrera et. al. [25], when using these compounds in an external validation set related to commercial drugs. The second set of compounds was composed by 35 commercial drugs with reported values of P-gp activity [25] (see Table **4**).

The predictivity of the model for this external validation set was 85.71% and the specificity was 71.43% for a general accuracy of 74.29%. From these compounds three drugs were erroneously predicted with a classification probability higher than 0.9. These drugs, Naloxone, Naltrexone and Indomethacin, have been described in the literature as potent inhibitors of the drug transporter, P-gp [80-82].

The third set of compounds was composed by 40 commercial drugs with reported values of P-gp activity, all of them as substrates [83] (see Table **5**). As provided in Table **5** IC_{50} values have been obtained for several commercially available compounds. In the case of those compounds having $IC_{50} > 90$ we have decided to evaluate only those for which their activity as substrate is confirmed by at least other data source as provided in the references given in Table **5**.

The percentage of good classification for this data set using our model is 83.3 %. We have to remark that from all compounds in this dataset we have excluded 18 due to the following reasons: one chemical (Propantheline) is electronically charged and current version of TOPS-MODE does not account for charged compounds; another compound is a Quinine isomer (Quinidine), not differentiated by current TOPS-MODE approach and the remaining 16 compounds have $IC_{50} > 90$ and are not confirmed by other sources to be substrate. Their numbers are: 6, 8, 10, 13, 17-19, 21, 25-29, 31, 34 and 37. Among those compounds which are not correctly classified by our model *chlorpromazine* has been considered to display only moderate activity [15, 53] (see Table **4**). On the other hand, itraconazole has been considered as moderate by some authors [15, 53] and as inactive by others [25], which agree with our classification of this compound as uncertain (U). Finally, amitryptiline and rofecoxib are wrongly classified and they display $IC_{50} > 50$.

5. DISCUSSION

Although several computational models have been carried out to classified the P-gp substrate only two studies had used reasonably large data sets (see Table **6**). A quantitative comparison among previous studies is not possible basically due to the differences in the composition of the data sets used. Nevertheless, a qualitative comparison may provide some results of the use of our model with respect to those reported in other studies. In Table **6** we show the results obtaioned by some methods previously reported in the literature and the present approach.

From all the data sets previously analyzed, the training set of used for developing our model was the larger one, having a ratio of number of compounds per number of variables included in the model equal to 57.5. Also, if substrates (S), non-substrates (NS), and accuracy (A) values for the test set are considered, all other models, with the exception of the one reported by Penzotti *et al*. [49], showed similar predictive quality, to each other. As can be seen in the Eq. 8, descriptors accounting for bond distance have positive contributions to the P-gp activity. We can think about

Table 4. Results of the Classification of the External Validation set of Commercial Drugs

No	CAS	Effluxed (Y/N)	Calcein-AM (Y/N)	P-gp Substrate	Class.	Prob.
1	514-65-8	N	N	-	+	35.02
2	298-46-4	N	N	-	-	60.97
3	78755-81-4	N	N	-	-	78.85
4	50-49-7	N	N	-	-	70.28
5	6740-88-1	N	N	-	-	88.45
6	57-53-4	N	N	-	-	98.01
7	465-65-6	N	N	-	+	1.56
8	16590-41-3	N	N	-	+	2.01
9	146-22-5	N	N	-	+	24.58
10	1088-11-5	N	N	-	+	27.73
11	51-34-3	N	N	-	-	90.30
12	14611-51-9	N	N	-	-	92.88
13	321-64-2	N	N	-	-	71.46
14	56775-88-3	N	N	-	-	56.41
15	60-80-0	N	N	-	-	89.23
16	29122-68-7	N	N	-	-	98.75
17	4205-90-7	N	N	-	-	56.81
18	469-21-6	N	N	-	-	92.05
19	5051-62-7	N	N	-	+	40.11
20	29110-47-2	N	N	-	-	73.65
21	53-86-1	N	N	-	+	1.01
22	100-92-5	N	N	-	-	95.33
23	37350-58-6	N	N	-	-	92.09
24	31828-71-4	N	N	-	-	93.45
25	6452-71-7	N	N	-	-	90.96
26	86-21-5	N	N	-	-	92.07
27	599-79-1	N	N	-	-	52.76
28	81-81-2	N	N	-	+	6.09
29	60-99-1	Y	Y	+	+	90.35
30	438-60-8	Y	Y	+	+	54.32
31	106266-06-2	Y	Y	+	-	46.11
32	68844-77-9	Y	Y	+	+	98.92
33	53179-11-6	Y	Y	+	+	95.59
34	29216-28-2	Y	Y	+	+	81.29
35	52-53-9	Y	Y	+	+	80.31

Table 5. ***In Vitro*** **IC_{50} Values for a Range of Commercial Compounds Using the Standard Assay Conditions for** ***In Vitro*** **Investigation of Inhibitors of P-gp.**

No	Compound	CAS	R-Fold [15]	IC_{50} (μM)	Class.	Prob.	Ref.
1	Amiodarone	1951-25-3	5.6	>10	+	99.35	[9,15,47,49,53,83]
2	Amitryptiline	50-48-6	---	59.1	–	42.09	[9,47,49,83]
3	Amprenavir	161814-49-9	---	>100	+	64.44	[22,25,27,46,83]
4	Donepezil	120014-06-4	---	28.6	+	87.32	[83]
5	Astemizole	68844-77-9	5.6	5.2	+	98.92	[15,22,25,2953,83]
6	Atorvastatin	134523-00-5	---	>100	---	---	[83]
7	Bepridil	64706-54-3	4.2	13.7	+	92.84	[9,15,47,49,53,83]
8	Captopril	62571-86-2	---	>100	---	---	[83]
9	Chlorpromazine	50-53-3	2.5	28.9	–	46.40	[15,22,53,83]
10	Clarithromycin	81103-11-9	---	>100	---	---	[83]
11	Cyclosporine	59865-13-3	---	1.6	+	98.63	[83]
12	Dexamethasone	50-02-2	---	>100	–	45.02	[22,25,27,46,47,49,83]
13	Diazepam	439-14-5	---	>100	---	---	[83]
14	Diltiazem	42399-41-7	---	7.0	+	70.19	[22.25,27,47,49,83]
15	Erythromycin	114-07-8	---	>100	+	80.61	[9,25,27,46,47,49,83]
16	Fentanyl Citrate	990-73-8	---	18.7	+	91.66	[83]
17	Fexofenadine	83799-24-0	---	>90	---	---	[83]
18	Flecainide	54143-55-4	---	>100	---	---	[83]
19	Galanthamine	357-70-0	---	>100	---	---	[83]
20	Haloperidol	52-86-8	15	40.0	+	79.75	[15,22,53,83]
21	Ibuprofen	15687-27-1	---	>100	---	---	[83]
22	Itraconazole	84625-61-6	4	1.3	+	**51.95 U**	[15,25,27,4653,83]
23	Ketoconazole	65277-42-1	16.7	3.7	+	98.38	[15,22,53,83]
24	Loperamide	53179-11-6	44.4	5.5	+	95.59	[9,15,25,27,46,47,49,53,83]
25	Memantine	19982-08-2	---	>100	---	---	[83]
26	Nifedipine	21829-25-4	1.4	>100	---	---	[15,53,83]
27	Propranolol	525-66-6	---	>100	---	---	[83]
28	Propantheline[a]	298-50-0	---	>100	---	---	[83]
29	Quinidine[b]	56-54-2	---	14.1	---	---	[9,25,27,46,47,49,83]
30	Quinine[b]	130-95-0	13.3	18.6	+	61.69	[9,15,25,27,46,47,49,53,83]
31	Ranitidine	66357-35-5	---	>100	---	---	[83]
32	Reserpine	50-55-5	---	1.2	+	92.82	[9,25,27,46,47,49,83]
33	Ritonavir	155213-67-5	---	28.2	+	90.99	[25,27,46,47,83]
34	Scopolamine	51-34-3	---	>100	---	---	[83]
35	Simvastatin	79902-63-9	---	16.2	+	86.97	[83]
36	Spironolactone	52-01-7	---	8.8	+	70.51	[23,25,83]

(Table 5) Contd….

No	Compound	CAS	R-Fold [15]	IC_{50} (μM)	Class.	Prob.	Ref.
37	Triamterene	396-01-0	---	>90	---	---	[83]
38	Verapamil	52-53-9	100	5.9	+	80.31	[15,22,25,29,53,83]
39	Vinblastine	865-21-4	---	17.8	+	91.99	[9,25,27,46,47,49,83]
40	Rofecoxib	162011-90-7	---	>50	–	40.60	[83]

Compounds which have not achieved a level of inhibition great enough at the highest concentration tested to allow calculation of an IC_{50} value are reported as >*x* μM, where *x* is the highest concentration tested. [a] Structures charged, [c] Isomer structures.

Table 6. Statistical Parameters Reported for Different Models for the P-gp Substrates Classification

	TSET				PSET			
Reference	N	V	R(N/V)	A %	N	S (%)	NS (%)	A %
Cabrera, et al. [25]	163	**6**	27.2	81.0	40	82.0	72.0	78.0
Gombar, et al. [46]	95	27	3.5	-	58	91.0	78.0	**86.2**
Xue, et al. [47]	142	22	6.4	79.4	59	84.2	66.7	**80.0**
Penzotti, et al. [49]	144	-	-	80.0	51	53.0	79.0	63.0
Bakken-Jurs, [53]	**345**	7	49.3	82.0	**177**	-	-	**81.9**
This Work	**345**	**6**	**57.5**	**86.7**	**177**	82.0	80.2	**81.8**

N - Number of compounds in the training (T) and prediction (P) sets; V - number of variables in the model; R - compound-variables ratio; S - substrate; NS - non-substrate: A - accuracy.

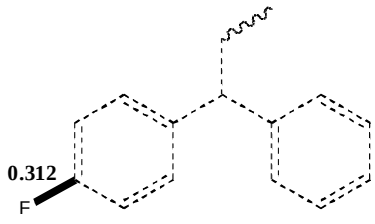
standard distance as in relation with the molecular size. Consequrntly, keeping the rest of parameters unaltered increasing molecular size tends to increase the chances for a compound to becaome a P-gp substrate. These results are in agreement with other reports [25, 46]. Another property which globally displays a positive contribution to P-gp activity is polarizability. In full agreement with results reported in other studies [45, 84, 85]. Finally, the effect of the atomic charge on the molecular bond is also related to the previous properties, showing that less electronegativity difference between the atoms that form a bond (C-N > C-O > C-F) is translated into larger contributions to the P-gp activity. This effect is explained by the negative coefficients in Eq. 8 of variables μ_5^{Ch} and μ_{11}^{Ch}. On the other hand, several relationships between simple molecular descriptors and P-gp activity class were studied. One striking relationship was observed between the first spectral moment weighted with the average bond distance (μ_1^D) and the P-gp classification (-1 for non-substrate and 1 for substrate) as shown in Fig. (**2**). Among the 345 TSET molecules, Fig (**2**) illustrates a high area under the ROC curve equal to 0.85, which is notably different from the value from a random classifier (diagonal line) [86, 87]. A subsequent principal components analysis of 345 heterogeneous compounds demonstrated that μ_1^D codifies information different to other descriptors used in substrate/non-substrate P-gp studies. In a general sense, bond distance could be considered as a more general property, providing new information for distinguishing P-gp substrates from nonsubstrates.

Despite this reductionist analysis of variable of different types we have to remark that a holistic analysis os needed in order to get a better understanding of the process. In this sense TOPS-MODE approach not only permits the correct classification of heterogeneous data set according to the P-gp activity, but it also allows the determination of fragment contributions, which are important to interpret this relation in structural terms.

5.1. Generation of Structural Rules

In order to generate structural rules we start by calculating bond contributions for all bonds in every molecule studied. Using a similar approach Cabrera et. al. [25] have found two fragments with opposite contributions to the P-gp activity. Here we report by first time four new fragments with positive contribution to the P-gp. The contribution of these fragments, hereafter designated as F1, F2, F3 and F4, are 0.15, 0.09, 0.07 and 0.03 respectively. These fragments are the nucleus of the rules for the structurtal alerts later on defined. These fragments and their contributions are shown in Fig. (**3**).

At this point we group molecules into chemical classes following a criterion of functional groups or similar molecular structural regions. For instance, we group together all compounds having Fluorine atom directly bonded to a benzene ring and we check that most of them have the same (qualitative) contribution for such similar regions, as displayed below:



When a pattern of substrate or P-gp is extended beyond a functional group to include a molecular region which is repeated in several compounds with the same fragment we select this region as the structural alert instead of the specific functional groups contained in such region, e.g.:

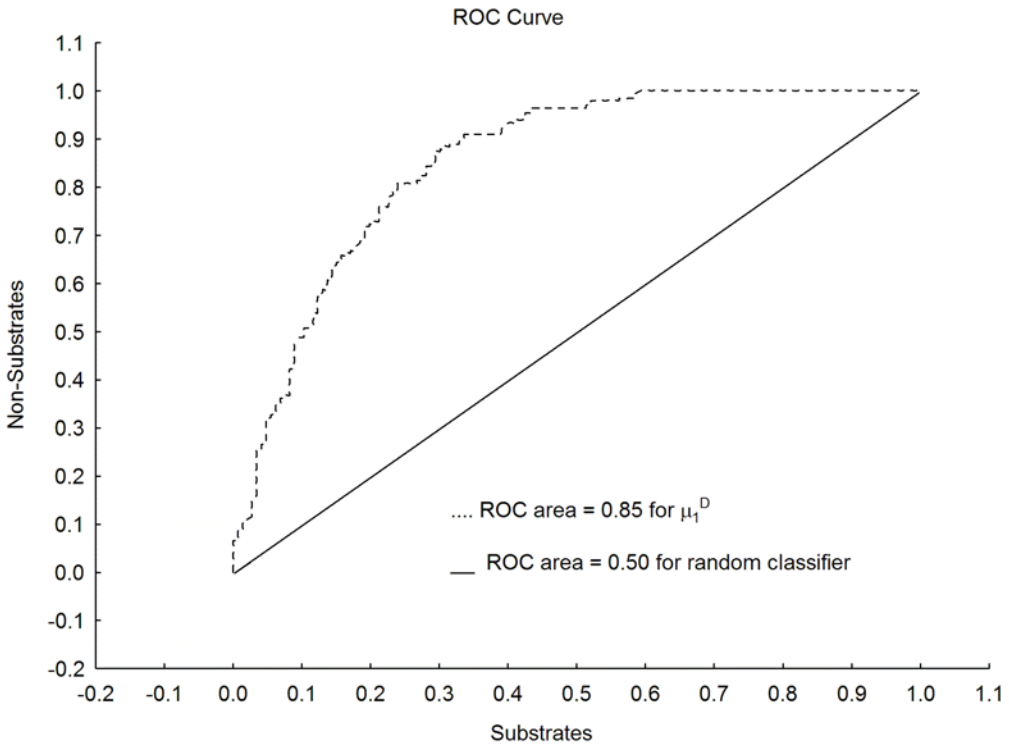


Fig. (2). Rule to classify P-gp substrates and nonsubstrates showing the ROC classifier for the first spectral moment of distance weighted bond matrices (dotted line) and that obtained for a random classifier (continuous line).

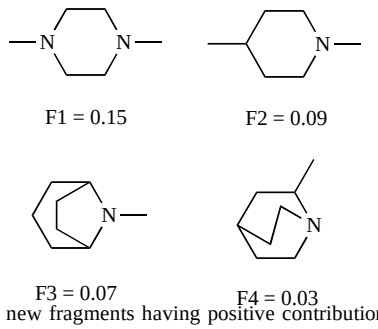
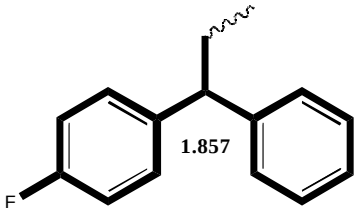


Fig. (3). List of new fragments having positive contributions to P-pg binding and their contributions determined by the model.



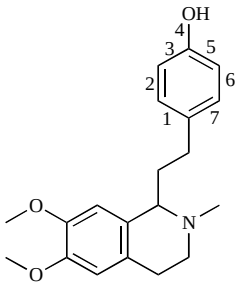
As a matter of example we give in Table **7** the values of the bond contributions for Tetrahydroisoquinoline. Phenolic group has been shown to have negative contribution to P-pg binding [15], which is also the case here. Note that after orthogonalizing the variables in the model the contributions of the bonds nearby -OH group also have negative contributions indicating the existence of a 'negative' region which extends beyond the functional group. More interestingly, if we sum together all bond contributions for the seven bonds of the phenolic group, the contribution before orthogonalization is positive, i.e., 0.299. However, as soon as the variables are orthogonalized this contribution is negative: -0.298, in agreement with previous findings [15].

In order to extract functional rules as the previously exemplified we use fragments/molecular-regions which are represented in a large pool of compounds or which sounded chemico-biological information about its possible role in P-glycoprotein has been reported. When such structural rules are created we test them for robustness using our prediction sets of compounds. In Table **8** we give details for all the structural rules generated in this work, which include the prototype structure defining the structural alert, compounds in the dataset that display this fragment/region and some examples of them, identifying the bond contributions of the region responsible for the P-glycoprotein activity. We have to remark that the way for using such structural alerts or rules is in the 'classical' IF-THEN way of expert systems. That is, IF 'structural alert x' is present in a molecule, THEN 'this molecule is a P-pg substrate'.

Using the previously mentioned criteria we have selected 24 structural alert rules accounting for the following structural classes of compounds: Benzimidazolone and Benzoindolinone; Benzimidazole; Indole; Benzodioxole and Benzodioxine; Pyrimido pyrimidine; Pteridine, Tetrahydroisoquinoline; Quinoline; Isoquinoline; Phenothiazine; Thioxanthene and Xanthene; Dibenzepine saturated and unsaturated; Benzoic acid derivatives; Benceneacetonitrile; Ethylphenylacetate and analogs; Diphenylmethane and Diphenylalkyl derivatives; 1H-Indene; 1,3-Dioxolane with Imidazole or Triazole; Imidazole; Azepane; compounds with piperidinyl ring; compounds with piperazinyl ring; compounds with nitrogen in an heterocyclic ring and compounds with nitrogen and oxygen in an heterocyclic ring. Note that obviously one compound can have more than one structural alert.

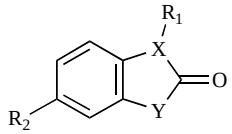
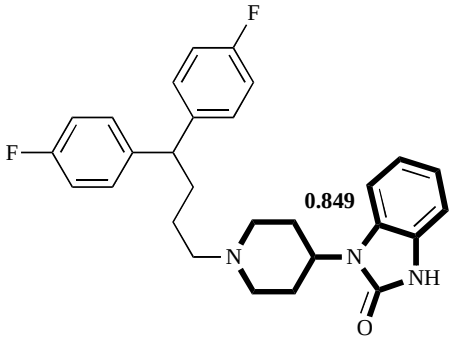
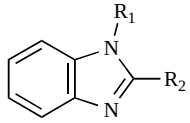
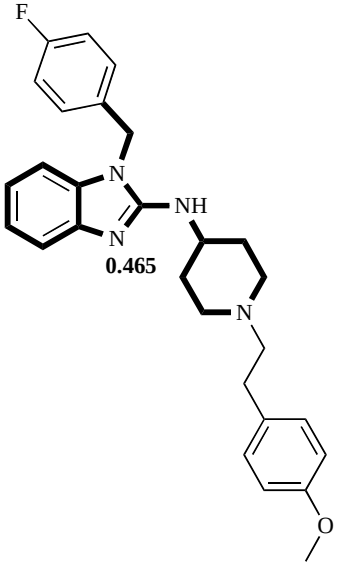
As can be seen in all the structural alerts rules defined here the secundary and tertiary forms of N are presents (see example in

Table 7. Bond Contributions for Tetrahydroisoquinoline Before (NO) and After (O) Orthogonalization of Variables in the TOPS-MODE Classification Model

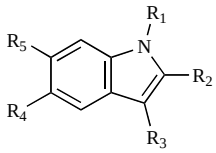
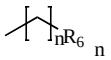
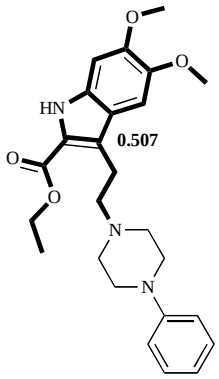
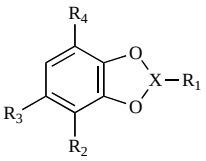
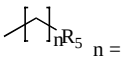
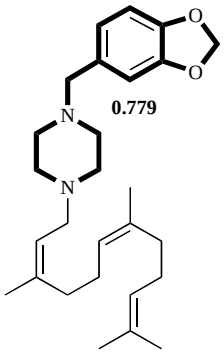
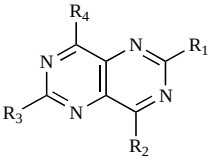
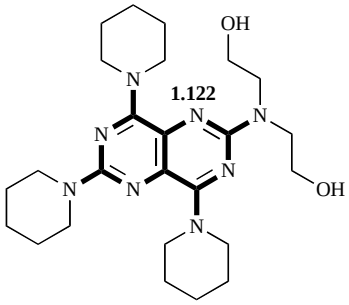
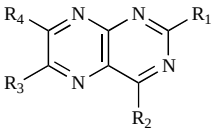
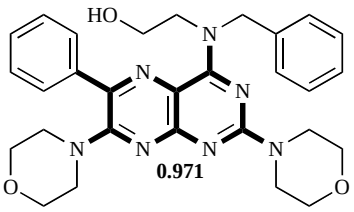
Structure Numbering (Phenolic Group)	Bond	C(NO)[a]	C(O)[b]
 **Compound 187**	1	0.077	0.114
	2	0.211	0.258
	3	0.025	-0.273
	4	-0.327	-0.496
	5	0.025	-0.273
	6	0.211	0.258
	7	0.077	0.114

[a] C(NO) - contributions before orthogonalization; [b] C(O) - contributions after orthogonalization

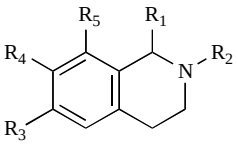
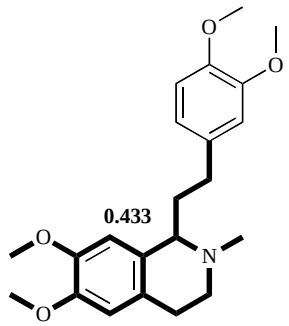
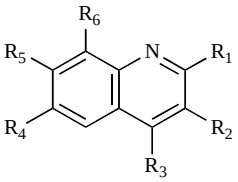
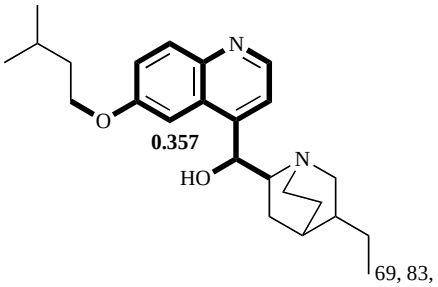
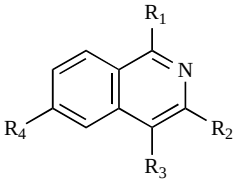
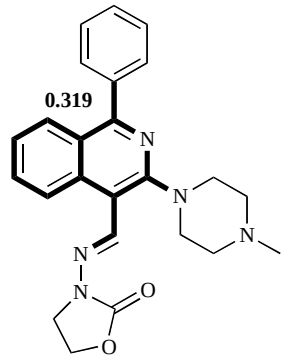
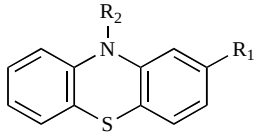
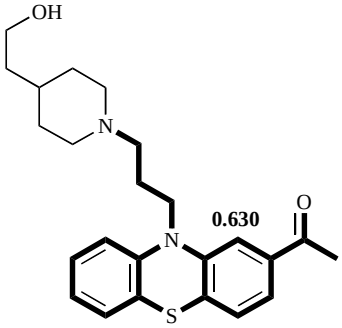
Table 8. Structural Alerts Selected by Means of the Bond Contributions Obtained from the TOPS-MODE Classification Model

Name	Description	Examples [References]
Benzimidazolone and Benzoindolinone	 X = N; Y = CH or NH R_1 = CH_3; C in aliphatic heterocycle;  n = 3 saturated C; R_2 = H; O; Cl R_3 = N in aliphatic heterocycle	 [**29**, 36, 104, 154, 171, 188, 196, 263]
Benzimidazole	 R_1 =  n = 1 or 2 saturated C; C_2H_5; R_2 = CH_2-R_4; NH-R_5; R_3 = CH_3; saturated or aromatic C; R_4 = C_2H_5; N in aliphatic heterocycle; R_5 = C in aliphatic heterocycle	 143, **169**, 206, 230, 273

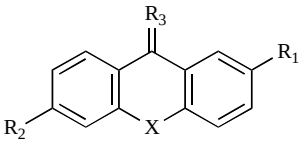
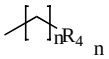
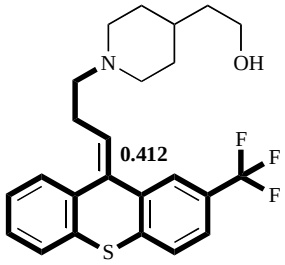
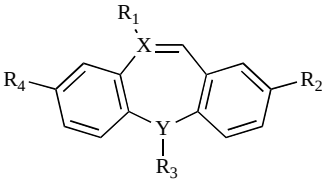
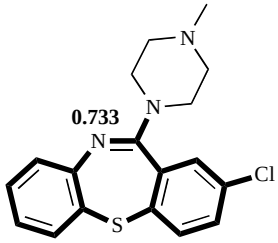
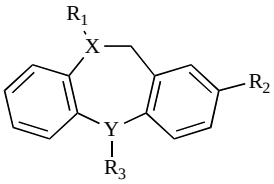
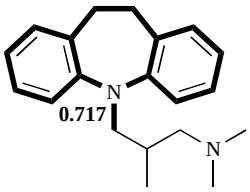
(Table 8) Contd….

Name	Description	Examples [References]
Indole	 R_1 = H; CH_3; R_2 = CH_3; unsaturated C (C=O); R_3 =  n = 2 saturated C; R_4 = O-R_7; R_5 = O-R_7, O in aliphatic heterocycle; R_6 = N in aliphatic heterocycle R_7 = CH_3	 **0.507** **34**, 114, 244, 254
Benzodioxole and Benzodioxine	 X = 1 or 2 saturated C; R_1 = H, unsaturated C (C=O); R_2 = H; O; R_3 = H;  n = 1 or 2 saturated C; R_4 = H; unsaturated C (C=O); R_5 = C and N in aliphatic heterocycle	 **0.779** **120**, 197, 225, 244, 246
Pyrimido pyrimidine	 R_1 = N in aliphatic heterocycle; N-R_5R_6; N-R_5R_7; R_2 = N in aliphatic heterocycle; N-R_5R_6; N-R_8R_9; R_3 = H; N in aliphatic heterocycle; N-R_5R_6; N-R_5R_7; R_4 = N in aliphatic heterocycle; N-R_5R_6; N-R_8R_9; S-R_9; R_5 = R_6 = CH_2CH_2OH; R_7 = $CH_2CH_2OCH_3$; R_8 = CH_3; R_9 = OCH_3; $CH_2C_6H_5$	 **1.122** 35, 126, 127, 128, 134, 135, 219, 239, **250**, 265
Pteridine	 R_1 = N in aliphatic heterocycle; R_2 = N in aliphatic heterocycle; N-R_5R_6; N-R_5R_7; N-R_7R_8; N-R_7R_9; N-R_7R_{10}; R_3 = aromatic C; Cl; S-R_8; R_4 = N in aliphatic heterocycle; N-R_7R_8; NH-R_8; R_5 = R_6 = CH_3; R_7 = CH_2CH_2OH; R_8 = $CH_2C_6H_5$; R_9 = C_2H_5; R_{10} = $CH_2CH(OH)CH_3$	 **0.971** 33, 55, 59, **71**, 148, 212, 266, 269

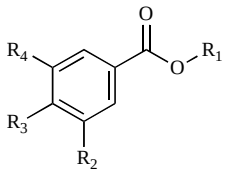
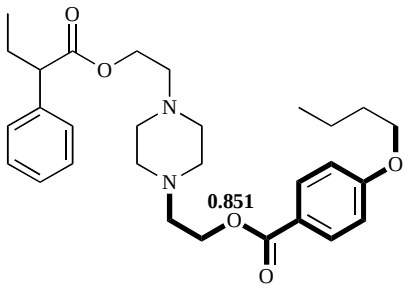
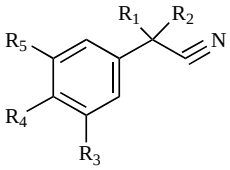
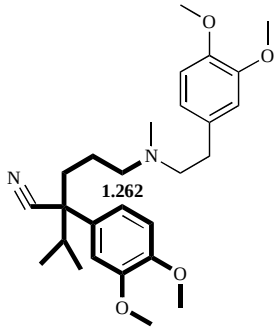
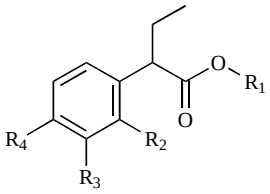
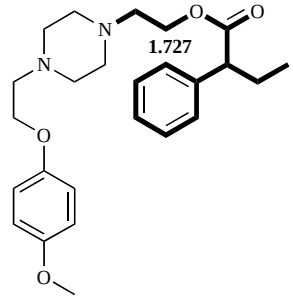
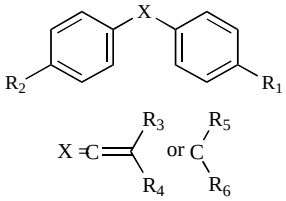
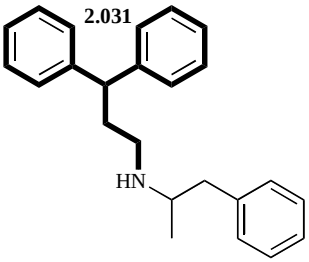
(Table 8) Contd….

Name	Description	Examples [References]
Tetrahydroisoquinoline	 R_1 =  n = 1 or 2 saturated C; C in aliphatic ring; R_2 = H; CH_3; R_3 = R_4 = O-R_7; R_5 = H; C in aliphatic ring; R_6 = aromatic C; R_7 = CH_3	 16, **51**, 187, 202, 204, 205, 215, 222, 225
Quinoline	 R_1 = H; CF_3; O-R_7; R_2 = H; CH_3; R_3 = saturated C; unsaturated C (C=O); N-R_7; R_4 = H; O-R_7; R_5 = H; Cl; R_6 = H; CF_3; R_7 = saturated C	 69, 83, **121**, 245, 246
Isoquinoline	 R_1 = H; CH_3; aromatic C; R_2 = H; CH_3; N in aliphatic heterocycle; Cl; R_3 = H; unsaturated C (C=C) R_4 = H; N in aliphatic heterocycle	 3, 43, 67, 107, 108, 130, **132**, 151, 168, 243, 255
Phenothiazine	 R_1 = H; O; Cl; CO-Alkyl; CF3; S-Alkyl; SO-CH_3; SO_2-CH_3; SO_2-NR_3R_4; R_2 =  n = 2 or 3 saturated C; unsaturated C (C=O); R_3 = R_4 = CH_3; R_5 = C and N in aliphatic heterocycle; N-R_3R_4	 37, 46, 47, 60, 72, 95, 96, 97, 98, 99, 100, 136, 150, 158, 174, 175, 192, 193, 231, 233, 235, 251, 279, 283, 286, 296

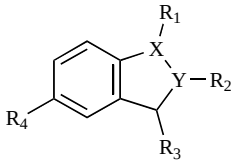
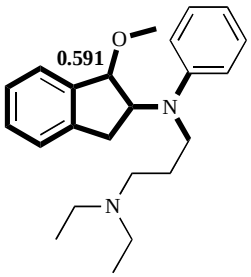
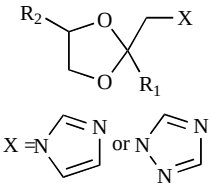
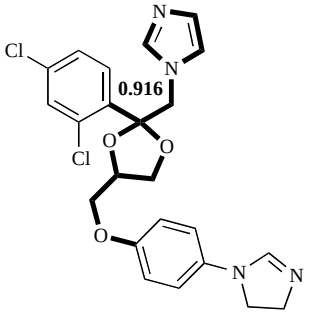
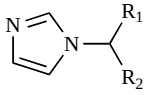
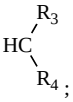
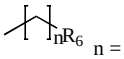
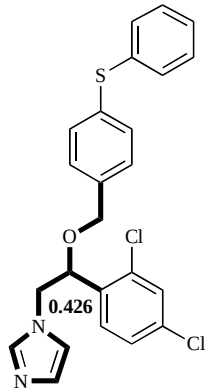
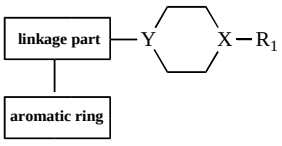
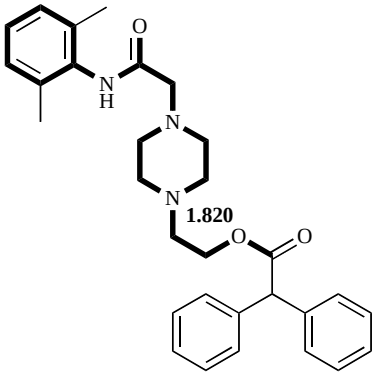
(Table 8) Contd….

Name	Description	Examples [References]
Thioxanthene and Xanthene	 X = S or O R_1 = H; Cl; CF_3; S-Alkyl; SO_2-N$(CH_3)_2$; SO_2-NR_3R_4; R_2 = H; F; R_3 = unsaturated C in aliphatic heterocycle;  n = 2 saturated C; R_4 = N in aliphatic heterocycle	 66, 165, 220, 221, 226, 236, **252**, 284
Dibenz[b,e]azepine Dibenz[b,f]oxazepine Dibenz[b,f]thiazepine Dibenz[b,f]oxepine Dibenzo[2,3;6,7]thiepine Dibenzo[a,d]cyclo heptenylidene	 X = N; Y = Csp3 X = N; Y = O X = N; Y = S X = Csp2; Y = O X = Csp2; Y = S X = Y = Csp2 R_1 = H; N in heterocycle ring; R_2 = H; CH_3; O; Cl; R_3 = H; saturated C; saturated C in aliphatic heterocycle; unsaturated C (C=C); unsaturated C in aliphatic heterocycle R_4 = H; F	 28, 73, **74**, 75, 164, 176, 177, 178, 240, 241
Dibenz[b,f]azepine 10,11-Dihydrodibenzo[a,d] cycloheptenylidene	 X = Csp3; Y = N X = O; Y = Csp2 X = Csp3; Y = Csp2 X = Y = Csp3 R_1 = H; N in aliphatic heterocycle; R_2 = H; Cl; R_3 = unsaturated C (C=C); unsaturated C (C=C) in aliphatic heterocycle; O-R_4 R_4 = saturated C in aliphatic heterocycle	 23, 76, 131, **159**, 248, 287

(Table 8) Contd….

Name	Description	Examples [References]
Benzoic acid derivatives	 R_1 =  n = 2, 3, 4 or 8 saturated C; $R_2 = R_4$ = H; O; R_3 = CH_3; O-Alkyl R_5 = N in aliphatic heterocycle; tertiary N	 6, 18, **86**, 88, 201, 207, 234
Benceneacetonitrile	 R_1 = saturated C; R_2 =  n = 2 or 3 saturated C; R_3 = H; CH_3; CF_3; O-Alkyl; Cl R_4 = H; O-Alkyl; F; Cl; R_5 = H; O-Alkyl; R_6 = tertiary N	 verapamil 2, **4**, 5, 11, 17, 27, 30, 31, 39, 42, 57, 115, 203, 213, 218
Etylphenylacetate and analogs	 R_1 =  n = 2 saturated C; $R_2 = R_4$ = H; CH_3; F; Cl; R_3 = H; CH_3; CF_3; O-Alkyl; Cl R_5 = N in aliphatic heterocycle; tertiary N	 1, 7, 8, 9, 19, 21, 22, 32, 54, 85, 86, 122, 123, 124, **125**, 200, 210, 211, 216, 261, 277, 299
Diphenylmethane and Dipheny-lalkyl derivatives	 R_1 = H; O-Alkyl; F; Cl; R_2 = H; CH_3; aromatic C; O; F R_3 = H; saturated C; C in aliphatic heterocycle; NO_2; R_4 = aromatic C; C in aliphatic heterocycle; R_5 =  n = 2 or 3 saturated C; saturated C; unsaturated C (C=O); OH; O-R_8 R_6 = H; saturated C; OH; OC_2H_5; R_7 = secondary N; R_8 = C in aliphatic heterocycle	 **102**, 103, 110, 111, 117, 152, 162, 179, 185, 195, 207, 237, 242, 248, 249, 253, 280, 289, 290, 294

(Table 8) Contd….

Name	Description	Examples [References]
1H-Indene	 X = Y = Csp3 or Csp2 R_1 = H; OCH_3; saturated C; aromatic C; R_2 = H; saturated C; tertiary N; R_3 = H; N in aliphatic heterocycle; R_4 = H; CF_3	 105, **144**, 156, 268
1,3-Dioxolane with Imidazole or Triazole	 R_1 = aromatic C; R_2 = CH_2-R_3; R_3 = O-Aryl	 44, 49, 61, **87**, 147, 232
Others Imidazole without 1,3-Dioxolane	 R_1 = H; aromatic C; R_2 = aromatic C;  R_3 = O-R_5; R_4 = aromatic C; R_5 =  n = 1 saturated C; R_6 = aromatic C	 **163**, 172, 198, 199, 264
Compounds with piperazinyl ring	 X = Y = N R_1 =  n = 2 saturated C; aromatic C; unsaturated C (C=O); NH; R_2 = O-R_3; R_3 = saturated and unsaturated C(C=O); **linkage part** $CH(OH)CH_2O$-N=CH; $(CH_2)_2CH(OH)$; CH_2CH=CH; $(CH_2)_3$; CH_2; $(CH_2)_3C$=O; $(CH_2)_2O(C$=O)NH; $CH_2(C$=O)NH; $CH_2CH(OH)CH(OCH_3)$; $(CH_2)_6O$; $(CH_2)_2O(C$=O)$CH(CH_3)$; $(CH_2)_2NH$; $(CH_2)_2O(C$=O)$CH(C_2H_5)O$; $CH_2CH(OH)CH_2O$ **aromatic ring** benzene and naphthalene	 12, **20**, 40, 45, 56, 58, 89, 91, 92, 93, 133, 140, 149, 160, 167, 184, 186, 274, 291

(Table 8) Contd….

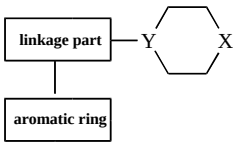
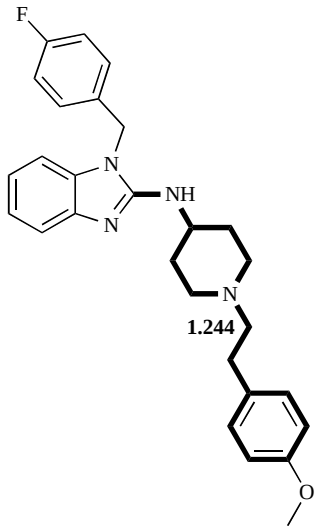
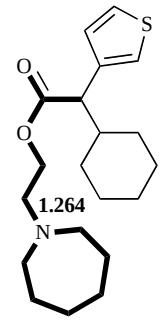
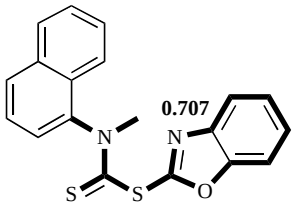
Name	Description	Examples [References]
Compounds with piperidinyl ring	 X = CH_2; CH-R_1; CH-R_2; CH-O-R_1; CR_3R_4; NH-R_5; Y = N R_1 = saturated C; R_2 = unsaturated C (C=O); R_3 = unsaturated C (C=O); OH; R_4 = aromatic C; N in aliphatic heterocycle; R_5 = C in aromatic ring **linkage part** CH_2; $(CH_2)_2$; $(CH_2)_3O$; $(CH_2)_3C=O$; $(CH_2)_3CH(OH)$; $(CH_2)_3CH(C_6H_5)$; $(CH_2)_3C(OH)(C_6H_5)$; $(CH_2)_2OCH(C_6H_5)$ **aromatic ring** benzene and naphthalene	 24, 77, 81, 109, 112, 153, **169**, 183, 189, 214, 238, 256, 275, 285
Compounds with N in heterocycle ring	Miscellaneous compounds (**Azabicycle**, Azepane, Triazolone, Pyrazolone, Phthalazinone, Thiazolopyrimidinone, Benzotriazinone)	 64, 68, 80, **106**, 110, 111, 121, 131, 161, 191, 267, 270, 295
Compounds with N and O in heterocycle ring	Miscellaneous compounds (**Benzoxazole**, Benzodiazole, Oxadiazole, Oxazolidinone, Compounds with morpholinyl ring)	 33, 48, 55, 59, 71, 128, 132, 139, 141, **142**, 255, 261, 269, 270, 288

Table **7**). It is well known [15, 23, 41, 50] that nitrogen atom (positively charged or not) has an important role in the binding of substrates to P-gp. According to Stouch [23], the most striking result of the study by Klopman [15] was the importance of a dialkyl-substituted amine. Some progresses have been made in uncovering pharmacophores responsible for activity. Even for reversing agents, Klopman *et al*. [15] have confirmed the importance of specific structural features and have proposed a nascent pharmacophore about the studied secondary amine. It should be noted that 260 of the 609 had a secondary amine.

A biophore referred to by Klopman [15] (in his Table **2**) in the generic form of C-C-*X*-C-C, where *X* = N, NH, or O (preferably a tertiary nitrogen), linked to two unsubstituted alkyl groups, is consistent with our results. In fact, this fragment is present in all alerts proposed here in two alternative geometric shapes *cis* and *trans*. This orientation is selected in agreement with the presentation about P-gp as a H-bond donor/acceptor and the neutral drug forms used [88]. In particular, stereoisomers of Thioxanthenes (rule number 11) can have two different shapes depending on the direction of the aliphatic chain toward the second position substituent in the ring system: in the same direction as (*cis* forms) or opposite to (*trans* form) the substituent [89].

The other biophores that appear in the above-mentioned Table **2** of [15] coincide with some of the alerts defined by us. For example, biophore # 2 with rule 8 (Quinoline); biophore # 4 with rule 17 (Diphenylmethane and Diphenylalkyl derivatives); biophore # 5 with particular cases of rules 12 and 13 (Dibenzazepine); biophore # 6 with rules 15 and 16 (Benceneacetonitrile; Ethylphenylacetate

and analogs) and biophores 10 and 11 with rules 14, 15, 16 and 17 (Benzoic acid derivatives, Benceneacetonitrile, Ethylphenylacetate and analogs, Diphenylmethane and Diphenylalkyl derivatives).

Most models developed up to now identify the minimal requirements of involving one or two hydrophobic centers including one or two aromatic rings, one to three H-bond acceptors (>N-, -OH, =O), and/or one H-bond donor (>N-, -OH, =NH, -NH-) [41]. The most active ligands should have more pharmacophore contact points simultaneously occupied [41]. Examples of P-glycoprotein substrates with pharmacophore contact points identified according to Raub [41] coincide with our structural alerts. Phenotiazines are one of the most abundant compounds in our datasets, having a total of 26 chemicals (18 in training serie and 8 in prediction one). They are characterized by our structural rule # 10. An outstanding aspect for this family selection is related to the fact that the model used here classifies and predict 100 % of these 26 structures. Two of them, *Promethazine* (compound 136) and *Prochlorperazine* (compound 150) are classified with 83.98 and 80.48 %, respectively. These compounds have RF values of 7.5 and 6.7, respectively and log (MDR) values of 0.279 and 0.415, respectively [89].

6. CONCLUSION

The prediction of the P-gp activity has been a goal of pharmaceutical research. For this reason, several *in silico* methods have been applied in order to predict this property in the early stage of drug development and some of them have become important tools to select new drug candidates. In this study, the TOPS-MODE approach has been a successful methodology for classifying P-gp substrate/non-substrate compounds. We have shown the relevance of bond distance, polarizability and the atomic charge as bond weights in the bond matrix for describing P-gp binding. The most relevant aspect of the current approach is that it has permitted the identification and quantification of fragment contributions that are responsible of the P-gp activity for any molecular structure. These fragment contributions have been expressed here as structural alerts/rules, which are easily implementable in expert systems and computer-aided tools. We hope that this work contributes to the efforts of finding predictive and interpretable models for pharmaceutical properties of molecules.

ACKNOWLEDGEMENT

EE thanks the Principal, University of Strathclyde for partial financial support under the New Professor's Fund. EM thanks Unilever and the University of Santiago de Compostela for a postdoctoral fellowship and the Xunta of Galicia (20304PR) by the partial financial supporting of this work.

REFERENCES

[1] Veith GD. On the nature, evolution and future of quantitative structure-activity relationships (QSAR) in toxicology. SAR QSAR Environ Res 2005; 5/6: 323-30.

[2] Trinajstić N, Randić M, Klein DJ. On the quantitative structure-activity relationships in drug research, Acta Pharm Jugosl 1986; 36: 267-79.

[3] Benfenati E, Gini G. Computational predictive programs (expert systems) in toxicology. Toxicology 1997; 119: 213-25.

[4] Sanderson DM, Earnshaw G. Computer prediction of possible toxic action from chemical structure; the DEREK system. Human Exp Toxicol 1991**;** 10: 261-73.

[5] Enslein K, Gombar VK, Blake BW. Use of SAR in computer-assisted prediction of carcinogenicity and mutagenicity of chemicals by the TOPKAT program Mutat Res 1994; 305: 47-61.

[6] Tunkel T, Mayo K, Austin C, *et al*. Practical considerations on the use of predictive models for regulatory purposes, Environ Sci Technol 2005; 39: 2188-99.

[7] Higgins CF. ABC transporters, from microorganisms to man, Annu Rev Cell Biol 1992; 8: 67-113.

[8] Doige CA, Ames GFL. ATP-dependent transport systems in bacteria and humans, relevance to cystic fibrosis and multidrug resistance, Annu Rev Microbiol 1993; 47: 291-319.

[9] Seelig A. A general pattern for substrate recognition by P-glycoprotein. Eur J Biochem (FEBS) 1998; 251: 252-261.

[10] Ambudkar SV, Dey S, Hrycyna CA, *et al*. Biochemical, cellular, and pharmacological aspects of the multidrug transporter. Annu Rev Pharmacol Toxicol 1999; 39: 361-98.

[11] Borst P, Evers R, Kool M, *et al*. The multidrug resistance protein family. Biochim Biophys Acta 1999; 1461: 347-57.

[12] Seelig A, Landwojtowicz E. Structure-activity relationship of P-glycoprotein substrates and modifiers. Eur J Pharm Sci 2000; 12: 31-40.

[13] Stein WD. Kinetics of the multidrug transporter (P-glycoprotein) and its reversal. Physiol Rev 1997; 77: 545-90.

[14] Gottesman MM, Pastan I. Biochemistry of multidrug resistance mediated by the multidrug transporter. Annu Rev Biochem 1993; 62: 385-427.

[15] Klopman G, Shi LM, Ramu A. Quantitative Structure-Activity Relationship of Multidrug Resistance Reversal Agents. Mol Pharmacol 1997; 52: 323-34.

[16] Taub ME, Podila L, Ely D, *et al*. Functional assessment of multiple P-gp probe substrates: influence of cell line and modulator concentration on P-gp Activity. Drug Metab Disp 2005; 33: 1679-87.

[17] Borst P, Evers R, Kool M, *et al*. A family of drug transporters: the multidrug resistance-associated proteins. J Natl Cancer Inst 2000; 92: 1295-1302.

[18] Benet LZ, Izumi T, Zhang Y, *et al*. Intestinal MDR transport proteins and P-450 enzymes as barriers to oral drug delivery. J Controlled Release 1999; 62: 25-31.

[19] Varma MVS, Sateesh K, Panchagnula R. Functional role of P-glycoprotein in limiting intestinal absorption of drugs: contribution of passive permeability to P-glycoprotein meditated efflux transport. Mol Pharm 2005; *2*: 12-21.

[20] Schinkel AH, Jonker JW. Mammalian drug efflux transporters of the ATP binding cassette (ABC) family: an overview. Adv Drug Delivery Rev 2003; 55: 3-29.

[21] Lin JH. Drug-drug interaction mediated by inhibition and induction of P-glycoprotein. Adv Drug Delivery Rev 2003; 55: 53-81.

[22] Crivori P. Computational Models for Identifying Potential P-gp Substrates and Inhibitors. Mol Pharm 2005; 3: 33-44.

[23] Stouch TR, Gudmundsson O. Progress in understanding the structure-activity relationships of P-glycoprotein. Adv Drug Delivery Rev 2002; 54: 315-28**.**

[24] Kim RB, Fromm MF, Wandel C, *et al*. The drug transporter P-glycoprotein limits oral absorption and brain entry of HIV-1 protease inhibitors. J Clin Invest 1998; 101: 289-94.

[25] Cabrera MA, Gonzalez I, Fernandez C, *et al*. A Topological Substructural Approach for the Prediction of P-Glycoprotein Substrates. J. Pharm. Sci 2006; 95: 589-606.

[26] Wacher VJ, Wu CY, Benet LZ. Overlapping substrate specificities and tissue distribution of cytochrome P450 3A and P-glycoprotein: implications for drug delivery and activity in cancer chemotherapy, Mol Carcinogen 1995; 13: 129-34.

[27] Polli JW, Wring SA, Humphreys JE, *et al*. Rational use of *in vitro* P-glycoprotein assays in drug discovery. J Pharmacol Exp Ther 2001; 299: 620-28.

[28] Schwab D, Fischer H, Tabatabaei A, *et al*. Comparison of *in vitro* Pglycoprotein screening assays: recommendations for their use in drug discovery. J Med Chem 2003; 46: 1716-25.

[29] Mahar-Doan KM, Humphreys JE, Webster LO, *et al*. Passive permeability and P-glycoprotein-mediated efflux differentiate central nervous system (CNS) and non-CNS marketed drugs. J Pharmacol Exp Ther 2002; 303: 1029-37.

[30] Scarborough GA. Drug-stimulated ATPase activity of the human Pglycoprotein. J Bioenerg. Biomembr 1995; 27: 37-41.

[31] Tiberghien F, Loor F. Ranking of P-glycoprotein substrates and inhibitors by calcein-AM fluorometry screening assay. Anticancer Drugs 1996; 7: 568-78.

[32] Litman T, Zeuthen T, Skovsgaard T, *et al*. Structure-activity relationships of P-glycoprotein interacting drugs: kinetic characterization of their effects on ATPase activity. Biochim Biophys Acta 1997; 1361: 159-68.

[33] Didziapetris R, Japertas P, Avdeef A, *et al*. Classification Analysis of P-Glycoprotein Substrate Specificity, J Drug Target 2003; 11: 391-406.

[34] Dahl SG, Sylte I, Ravna AW. Structures and models of transporter proteins. J Pharmacol Exp Ther 2004; 309: 853-60.

[35] Thiebaut F, Tsuruo T, Hamada H, *et al*. Immunohistochemical localization in normal tissues for different epitopes in the multidrug transport protein P-170, evidence for localization in brain capillaries and cross-reactivity of one antibody with a muscle protein, J Histochem Cytochem 1989; 37: 159-64.

[36] Cordon-Cardo C, O'Brien JP, Casals D, *et al*. Multidrug-resistance gene (P-gp) is expressed by endothelial cells at blood-brain barrier sites, Proc Natl Acad Sci 1990; 86: 695-98.

[37] Thiebaut F, Tsuruo T, Hamada H, *et al*. Cellular localization of the multidrug-re sistance gene product P-glycoprotein in normal human tissues, Proc Natl Acad Sci USA 1987; 84: 7735-38.

[38] Mukhopadhyay T, Batsakis JG, Kuo MT. Expression of the mdr (P-glycoprotein) gene in Chinese hamster digestive tracts, J Natl Cancer Inst 1988; 80: 269-75.

[39] Piekarz R, Cohen D, Horwitz SB. Progesterone regulates the murine multidrug resistance *mdr*1b gene, J Biol Chem 1993; 268: 7613-16.

[40] Ambudkar SV, Kimchi-sarfaty C, Sauna ZE, *et al*. P-glyco-protein: from genomics to mechanism. Oncogene 2003; 22: 7468-85.

[41] Raub TJ. P-Glycoprotein Recognition of Substrates and Circumvention through Rational Drug Design (reviews). Mol Pharm 2005; 3: 3-25.

[42] Cordon-Cardo C, O'Brian JP, Boccia J, *et al*. Expression of the multidrug resistance gene product (P-gp) in human and tumor tissues, J Histochem Cytochem 1990; 38: 1277-87.

[43] Divecha N, Irvine RF. Phospholipid signaling. Cell 1995; 80: 269-78.

[44] Ferte J. Analysis of the tangled relationships between P-glycoprotein-mediated multidrug resistance and the lipid phase of the cell membrane. Eur J Biochem (FEBS) 2000; 267: 277-94.

[45] Osterberg T, Norinder U. Theoretical calculation and prediction of P-glycoprotein-interacting drugs using MolSurf parametrization and PLS statistics. Eur J Pharm Sci 2000; 10: 295-303.

[46] Gombar VK, Polli JW, Humphreys JW, *et al*. Predicting P-Glycoprotein Substrates by a Quantitative Structure-Activity Relationship Model. J Pharm Sci 2004; 93: 957-68.

[47] Xue Y, Yap CW, Sun LZ, *et al*. Prediction of P-Glycoprotein Substrates by a Support Vector Machine Approach. J Chem Inf Comput Sci 2004; 44: 1497-1505.

[48] Ekins S, Kim RB, Leake BE, *et al*. Application of Three-Dimensional Quantitative Structure-Activity Relationships of P-Glycoprotein Inhibitors and Substrates. Mol Pharmacol 2002; 61: 974-81.

[49] Penzotti JE, Lamb ML, Evensen E, *et al*. A Computational Ensemble Pharmacophore Model for identifying substrates of P-Glycoprotein. J Med Chem 2002; 45: 1737-40.

[50] Pajeva IK, Wiese M. Pharmacophore Model of Drugs Involved in P-Glycoprotein Multidrug Resistance: Explanation of Structural Variety (Hypothesis*)*. J Med Chem 2002; 45: 5671-86.

[51] Cianchetta G, Singleton RW, Zhang M, *et al*. A Pharmacophore Hypothesis for P-Glycoprotein Substrate Recognition Using GRIND-Based 3D-QSAR. J Med Chem 2005; 48: 2927-35.

[52] Wang YH, Li Y, Yang SL, *et al*. Classification of Substrates and Inhibitors of P-Glycoprotein Using Unsupervised Machine Learning Approach. J Chem Inf Model 2005; 45: 750-57.

[53] Bakken GA, Jurs PC. Classification of Multidrug-Resistance Reversal Agents Using Structure-Based Descriptors and Linear Discriminant Analysis. J Med Chem 2000; 43: 4534-41.

[54] Estrada E, Patlewicz G, Chamberlain M, *et al*. Computer-aided knowledge generation for understanding skin sensitiza-tion mechanisms: the TOPS-MODE approach. Chem Res Toxicol 2003; 16: 1226-35.

[55] Estrada E, Patlewicz G, Gutierrez Y. From knowledge generation to knowledge archive: A general strategy using TOPS-MODE with DEREK to formulate new alerts for skin sensitisation. J Chem Inf Comput Sci 2004; 4: 688-98.

[56] Estrada E, Molina E. Automatic extraction of structural alerts for predicting chromosome aberrations of organic compounds. J Mol Graphics Model 2006; 25, 275-88.

[57] Estrada E. Spectral moments of the edge adjacency matrix in molecular graphs. 1. Definition and applications to the prediction of physical properties of alkanes. J Chem Inf Comput Sci 1996; 36: 844-49.

[58] Estrada E. Spectral moments of the edge adjacency matrix in molecular graphs. 2. Molecules containing heteroatoms and QSAR applications. J Chem Inf Comput Sci 1997; 37: 320-28.

[59] Estrada E. Spectral moments of the edge adjacency matrix in molecular graphs. 3. Molecules containing cycles. J Chem Inf Comput Sci 1998; 38: 23-27.

[60] Estrada E, Peña A, García-Domenech R. Designing sedative/hypnotic compounds from a novel substructural graph-theoretical approach. J Comput-Aided Mol Des 1998; 12: 583-95.

[61] Estrada E, Uriarte E, Montero A, *et al*. A novel approach to the rational selection and design of anticancer compounds. J Med Chem 2000; 43: 1975-85.

[62] Estrada E, Uriarte E. Recent advances on the role of topological indices in drug discovery research. Curr Med Chem 2001; 8: 1573-88.

[63] Ramu N, Ramu A. Circumvention of Adriamycin Resistance by Dipyridamole Analogues: A Structure-Activity Relationship Study. Int J Cancer 1989; 43: 487-91.

[64] Ramu A, Ramu N. Reversal of multidrug resistance by phenothiazines and structurally related compounds. Cancer Chemother Pharmacol 1992; 30: 165-73.

[65] Ramu A, Ramu N. Reversal of multidrug resistance by bis-(phenylalkyl)amines and structurally related compounds. Cancer Chemother Pharmacol 1994; 34: 423-30.

[66] Manthena V, Sateesh K, Panchagnula R. Functional Role of P-Glycoprotein in Limiting Intestinal Absorption of Drugs: Contribution of Passive Permeability to P-Glycoprotein Mediated Efflux Transport. Mol Pharm 2005; 2: 12-21.

[67] Estrada E. Edge adjacency relationships and a novel topological index related to molecular volume. J Chem Inf Comput Sci 1995; 35: 31-33.

[68] Wang R, Gao Y, Lai L. Calculating partition coefficient by atom-additive method. Perspect Drug Discovery Des 2000; 19: 47-66.

[69] Ertl P, Rhode B, Selzer P. Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. J Med Chem 2000; 43: 3714-17.

[70] Ghose AK, Crippen GM. Atomic physicochemical parameters for three-dimensional-structure-directed quantitative structure-activity relationships. 2. Modeling dispersive and hydrophobic interactions. J Chem Inf Comput Sci 1987; 27: 21-35.

[71] Bondi A. van der Waals volumes and radii, J Phys Chem 1964; 68: 441-51.

[72] Gasteiger J, Marsilli M. A new model for calculating atomic charges in molecules. Tetrahedron Lett 1978; 34: 3181-4.

[73] Estrada E, Molina E. Novel local (fragment-based) topological molecular descriptors for QSPR/QSAR and molecular design. J Mol Graphics Model 2001; 20: 54-64.

[74] Randić M. Resolution of ambiguities in structure-property studies by use of orthogonal descriptors. J Chem Inf Comput Sci 1991; 31: 311-20.

[75] Randić M. Orthogonal molecular descriptors. New J Chem 1991; 15: 517- 25.

[76] Randić M. Correlation of enthalpy of octanes with orthogonal connectivity indices. J Mol Struct (Theochem) 1991; 233: 45-59.

[77] Lucˇić B, Nikolić S, Trinajstić N, *et al*. The structure-property models can be improved using the orthogonalized descriptors. J Chem Inf Comput Sci 1995; 35: 532-8.

[78] Klein DJ, Randić M, Babić D, *et al*. Hierarchical orthogonalization of descriptors. Int J Quant Chem 1997; 63: 215-22.

[79] Estrada E, Perdomo I, Torres J. Combination of 2D-, 3D-connectivity and quantum chemical descriptors in QSPR: Complexation of alpha- and beta-cyclodextrin with benzene derivatives. J Chem Inf Comput.Sci 2001; 41: 1561-8.

[80] Callaghan R, Riordan JR. Collateral sensitivity of multidrug resistant cells to narcotic analgesics is due to effects on the plasma membrane. Biochim Biophys Acta 1995; 1236: 155-62.

[81] McNicholas JM. Inhibitors of abc drug transporters in microbial cells. USA: US Patent & Trademark Office, 2004.

[82] Perloff MD, Stormer E, von Moltke LL, *et al*. Rapid assessment of P-glycoprotein inhibition and induction *in vitro*. Pharm Res 2003; 20: 1177-83.

[83] Keogha JP, Kuntab JR. Development, validation and utility of an *in vitro* technique for assessment of potential clinical drug-drug interactions involving P-glycoprotein. Eur J Pharm. Sci 2006; 27: 543-54.

[84] Bradley M, Waller CL. Polarizability fields for use in three-dimensional quantitative structure-activity relationship (3D-QSAR). J Chem Inf Comput Sci 2001; 41: 1301-7.

[85] Hansch C, Steinmetz WE, Leo AJ, *et al*. On the role of polarizability in chemical-biological interactions. J Chem Inf Comput Sci 2003; 43: 120-5.

[86] Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982; 143: 29-36.

[87] Povoa P. Coelho L, Almeida E, *et al*. C-reactive protein as a marker of infection in critically ill patients. Clin Microbiol Infect 2005; 11: 101-8.

[88] Pajeva I, Wiese M. Human P-Glycoprotein Pseudoreceptor Modeling: 3D-QSAR Study on Thioxanthene Type Multidrug Resistance Modulators. Quant Struct Act Relat 2001; 20: 130-8.

[89] Pajeva I, Wiese M. Molecular Modeling of Phenothiazines and Related Drugs As Multidrug Resistance Modifiers: A Comparative Molecular Field Analysis Study. J Med Chem 1998; 41: 1815-26.

[90] Friche E, Beck W. Molecular pharmacology of reversal of multidrug resistance and its clinical implication. In *Multidrug Resistance in Cancer Cells*; Gupta S, Tsuruo T, Eds.; Wiley: New York, 1996; 361-74.

[91] Ford JM, Prozialeck WC, Hait WN. Structural features determining activity of phenothiazine and related drugs for inhibition of cell growth and reversal of multidrug resistance. Mol Pharmacol 1989; 35: 105-15.

Received: May 31, 2010 **Accepted: June 30, 2010**