

## A new topological descriptors based model for predicting intestinal epithelial transport of drugs in caco-2 cell culture.

Yovani Marrero Ponce, Miguel A. Cabrera Pérez, Vicente Romero Zaldivar, Humberto González Díaz, Francisco Torrens  
Department of Pharmacy, Faculty of Chemical-Pharmacy. Central University of Las Villas, Santa Clara, Villa Clara, Cuba;  
Department of Drug Design, Chemical Bioactive Center. Central University of Las Villas, Santa Clara, Villa Clara, Cuba;  
Faculty of Informatics. University of Cienfuegos, Cienfuegos, Cuba; Institut Universitari de Ciència Molecular, Universitat de València, Burjassot (València), Spain

Received 20 February 2004, Revised 26 March 2004, Accepted 14 April 2004, Published 29 June 2004

**Abstract Purpose:** Quantitative Structure-Permeability Relationships (QSPeR) of the intestinal permeability across the (Caco-2) cells monolayer could be obtained by the application of new molecular descriptors. **Method:** A novel topologic-molecular approach to computer molecular design (*TOMOCOMD-CARDD*) has been used to estimate the intestinal-epithelial transport of drug in Caco-2 cell culture. **Results:** The Permeability Coefficients in Caco-2 cells (P) for 33 structurally diverse drugs were well described using quadratic indices of the molecular pseudograph's atom adjacency matrix as molecular descriptors. A quantitative model that discriminates the high-absorption compounds from those with moderate-poor absorption was obtained for the training data set, showing a global classification of 87.87%. In addition, two QSPeR models, through a multiple linear regression, were obtained to predict the P [apical to basolateral (AP-BL) and basolateral to apical (BL-AP)]. A leave-*n*-out and leave-*one*-out cross-validation procedure revealed that the discriminant and regression models respectively, had a good predictability. Furthermore, others 18 drugs were selected as a test set in order to assess the predictive power of the models and the accuracy of the final prediction was similar to achieve for the data set. Besides, the use of both regression models, in a combinative way, is possible to predict the Permeability Directional Ratio (PDR, BL-AP/AP-BL) value. The found models were used in virtual screening of drug intestinal permeability and a relationship between calculated P and percentage of human intestinal absorption for several compounds was established. Furthermore, this approximation permits us to obtain a good explanation of the experiment based on the molecular structural features.

**Corresponding Author:** Yovani Marrero Ponce, Department of Pharmacy, Faculty of Chemical-Pharmacy, Central University of Las Villas, Santa Clara 54830, Villa Clara, Cuba. yovanimp@qf.uclv.edu.cu

**Conclusions:** These results suggest that the proposed method is able to predict the P values and it proved to be a good tool for studying the oral absorption of drug candidates during the drug development process.

### INTRODUCTION

The estimation of human oral absorption of new drug candidates in the early stage of the drug discovery process is a useful tool in the lead-compound selection (1-2). Several *in vitro* approaches, based on cell cultures, have been used in this area over the last few years. Among the cell culture models, the Caco-2 cell line is the most widely employed and it has been investigated like a potential *in vitro* model for drug absorption and metabolism studies (3-4). In this sense, considering the similarity of Caco-2 cell with the small intestine enterocytes and their capacity to express carrier-mediated transport systems and typical small intestinal enzymes (3-4), the permeability coefficient across Caco-2 cell monolayer (P) is increasingly used to estimate the oral absorption of new chemical entities (5-7).

Nevertheless, the permeability of compounds that are transported via carrier-mediated absorption is lower than that obtained through the human small intestine. In addition, these carcinoma colon cells have poor permeability for hydrophilic compounds, which can pass through this barrier by paracellular route (through the intercellular space) (8). Furthermore, due to cancerous origin of this cell line, they over-express the P-glycoprotein with the consequently lower permeability's in the absorptive direction (9). Finally, the long culture period with the consequent high-research cost and the difficulty to extrapolate these *in vitro* results with the *in vivo* situation are considered as main practical limitations.

In order to accelerate the growing of Caco-2 cell monolayer a rapid culture system has been used ( $P_{\text{accelerate}}$ ) (10).

On the other hand, the theoretical approach appears to be a good alternative to prediction of human absorption for new drug candidates obtained by combinatorial chemistry methodologies (11-13), avoiding significant failure in late stage of the drug-discovery process (14). In the last few years, graph-theoretical methods have become one of the most important tools for quantifying molecular structure. These theoretical strategies have emerged as a promising solution to the efficient search of new lead compounds (15) and it has been very useful to elucidate quantitative structure-property (QSPR) and quantitative structure-activity (QSAR) relationships. Trying to demonstrate the importance and relevance of the theoretical methods in the determination of the drugs absorption, the main objectives of the present work were, to find a quantitative model that discriminates, within the data set, the high-absorption compounds from those with moderate-poor absorption; and to develop quantitative relationships between the structure, expressed by the quadratic indices of the molecular pseudograph's atom adjacency matrix (16-18), and the P in both membrane directions, using a multiple linear regression method. Finally, to use both obtained models on virtual screening of drug intestinal permeability.

## THEORETICAL APPROACH

The general principles of the quadratic indices of the "molecular pseudograph's atom adjacent matrix" for small-to-medium sized organic compounds have been explained in some detail elsewhere (17, 18). However, an overview of this approach will be given.

The molecular vector ( $X$ ) is constructed in order to calculate the molecular quadratic indices for a molecule where the components of this vector are numeric values, which represent a certain atomic property.

These properties characterize each kind atom within the molecule. Such properties can be the electronegativity, density, atomic radii, and so on.

If a molecule consists of  $n$  atoms (vector of  $\mathfrak{R}^n$ ), then the  $k^{\text{th}}$  total quadratic indices,  $q_k(x)$  are calculated as a

quadratic form ( $q: \mathfrak{R}^n \rightarrow \mathfrak{R}$ ) in canonical basis as shown in Eq. 1,

$$q_k(x) = \sum_{i=1}^n \sum_{j=1}^n {}^k a_{ij} X_i X_j \quad (1)$$

where,  ${}^k a_{ij} = {}^k a_{ji}$  (symmetric square matrix),  $n$  is the number of atoms of the molecule and  $X_1, \dots, X_n$  are the coordinates of the molecular vector ( $X$ ) in a system of basis vectors of the  $\mathfrak{R}^n$ . One can choose the basis vectors the coordinates of the same vector will be different. The values of the coordinates depend thus in an essential way on the choice of the basis. The so-called canonical ('natural') bases,  $e_j$  denote the  $n$ -tuple having 1 in the  $j^{\text{th}}$  position and 0's elsewhere. In the canonical bases, the coordinates of any vector  $X$  coincide with the components of this vector. For that reason, those coordinates can be considered as weights (atom-labels) of the vertices of the pseudograph.

The coefficients  ${}^k a_{ij}$  are the elements of the  $k^{\text{th}}$  power of the matrix  $M$  ( $G$ ) of the molecular pseudograph ( $G$ ). Here,  $M(G) = M = [a_{ij}]$ , denote the matrix of  $q_k(x)$  with respect to the natural basis. In this matrix  $n$  is the number of vertices (atoms) of  $G$  and the elements  $a_{ij}$  are defined as follows:

$$\begin{aligned} a_{ij} &= P_{ij} \text{ if } i \neq j \text{ and } \exists e_k \in E(G) \\ &= L_{ii} \text{ if } i = j \\ &= 0 \text{ otherwise} \end{aligned} \quad (2)$$

where,  $E(G)$  represents the set of edges of  $G$ .  $P_{ij}$  is the number of edges between vertices  $v_i$  and  $v_j$  and  $L_{ii}$  is the number of loops in  $v_i$ .

Given that  $a_{ij} = P_{ij}$ , the elements  $a_{ij}$  of this matrix represent the number of bonds between an atom  $i$  and other  $j$ . The matrix  $M^k$  provides the number of walks of length  $k$  that links the vertices  $v_i$  and  $v_j$ . For this reason, each edge in  $M^1$  represents 2 electrons belonging to the covalent bond between atoms (vertices)  $v_i$  and  $v_j$ ; e.g. the inputs of  $M^1$  are equal to 1, 2 or 3 when appears simple, double or triple bonds between vertices  $v_i$  and  $v_j$ , respectively. On the other hand, molecules containing aromatic rings with more than one canonical structure are represented like a pseudograph. It happens for substituted aromatic compounds such as pyridine, naphthalene, quinoline, and so on, where the presence  $\text{PI}(\pi)$  electrons is accounted by means of loops in each atom of the aromatic ring. Conversely, aromatic rings having only one canonical struc-

ture, such as furan, thiophene and pyrrol are represented like a multigraph.

On the other hand, the defining equation (1) for  $q_k(x)$  may be written as the single matrix equation:

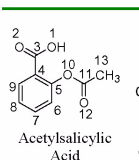
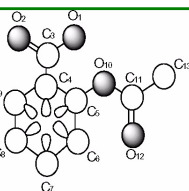
$$q_k(x) = [X_1 \quad \dots \quad X_n] \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix}^k \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} \quad (3)$$

or in the more compact form,

$$q_{kL}(x) = \sum_{i=1}^m \sum_{j=1}^m {}^k a_{ijL} X_i X_j \quad (4)$$

where  $[X]$  is a column vector (a  $nx1$  matrix) of the coordinates of  $X$  in the canonical base of  $\mathfrak{R}^n$ ,  $[X]^t$  the transpose of  $[X]$  (a  $1xn$  matrix) and  $M^k$  the  $k^{th}$  power of the matrix  $M$  of the molecular pseudograph  $G$  (mathematical quadratic form's matrix). In Table 1, the calculation of six quadratic indices for acetylsalicylic acid is exemplified.

**Table 1: Definition and calculation of six ( $k = 0-5$ ) total quadratic indices of the molecular pseudograph's atom adjacency matrix of the molecule of acetylsalicylic acid.**

 <p>Acetylsalicylic Acid</p>	 <p>Molecular Pseudograph (G) (Suppressed Hydrogen Atom)</p>	<p><math>X = [O_1, O_2, C_3, C_4, C_5, C_6, C_7, C_8, C_9, O_{10}, C_{11}, O_{12}, C_{13}]</math> Molecular Vector: <math>X \in \mathfrak{R}^{13}</math></p> <p>In the definition of the <math>X</math>, as molecular vector, the chemical symbol of the element is used to indicate the corresponding electronegativity value. That is: if we write <math>O</math> it means <math>\chi(O)</math>, oxygen Mulliken electronegativity or some atomic property, which characterizes each atom in the molecule. So, if we use the canonic bases of <math>\mathfrak{R}^{13}</math>, the coordinates of any vector <math>X</math> coincide with the components of that molecular vector</p> <p><math>[X]^t = [3.17 \ 3.17 \ 2.63 \ 2.63 \ 2.63 \ 2.63 \ 2.63 \ 2.63 \ 2.63 \ 3.17 \ 2.63 \ 3.17 \ 2.63]</math></p> <p><math>[X]^t =</math> transposed of <math>[X]</math> and it means the vector of the coordinates of <math>X</math> in the canonical base of <math>\mathfrak{R}^{13}</math> (a <math>1 \times n</math> matrix) <math>[X]</math>: vector of coordinates of <math>X</math> in the canonical base of <math>\mathfrak{R}^{13}</math> (a <math>n \times 1</math> matrix)</p>
$q_0(x) = \sum_{i=1}^n \sum_{j=1}^n {}^0 a_{ij} X_i X_j = [X]^t M^0 [X] = 102.4472$		$M^1(G) = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$ <p><math>M^1(G)</math>: Molecular Pseudograph's Atom Adjacency Matrix</p>
$q_1(x) = \sum_{i=1}^n \sum_{j=1}^n {}^1 a_{ij} X_i X_j = [X]^t M^1 [X] = 268.8912$		
$q_2(x) = \sum_{i=1}^n \sum_{j=1}^n {}^2 a_{ij} X_i X_j = [X]^t M^2 [X] = 373.5982$		
$q_3(x) = \sum_{i=1}^n \sum_{j=1}^n {}^3 a_{ij} X_i X_j = [X]^t M^3 [X] = 2566.8034$		
$q_4(x) = \sum_{i=1}^n \sum_{j=1}^n {}^4 a_{ij} X_i X_j = [X]^t M^4 [X] = 8381.1414$		
$q_5(x) = \sum_{i=1}^n \sum_{j=1}^n {}^5 a_{ij} X_i X_j = [X]^t M^5 [X] = 25593.612$		

In addition to a total quadratic indices computed for the whole-molecule a local-fragment (atom and atom-type) formalism can be developed. These descriptors are termed local quadratic indices of the “molecular

pseudograph's atom adjacency matrix”,  $q_{kL}(x)$ . The  $q_{kL}(x)$  are graph-theoretical invariant for a given fragment  $F_R$  (connected subgraph) within a specific pseudograph  $G$ . The definition of these descriptors is as follows:

$$q_{kL}(x) = \sum_{i=1}^m \sum_{j=1}^m {}^k a_{ijL} X_i X_j \quad (5)$$

where  $m$  is the number of atoms of the fragment of interest and  ${}^k a_{ijL}$  is the element of the file “ $i$ ” and column “ $j$ ” of the matrix  $M^k_L = M^k(G, F_R) [q_{kL}(x) = q_k(x, F_R)]$ . This matrix is extracted from the  $M^k$  matrix and contains the information referred to the vertices of the specific molecular fragments (FR) and of the molecular environment. The matrix  $M^k_L = [{}^k a_{ijL}]$  with elements  ${}^k a_{ijL}$  is defined as follows:

$${}^k a_{ijL} = \begin{cases} {}^k a_{ij} & \text{if both } v_i \text{ and } v_j \text{ are vertices contained within } F_R \\ \frac{1}{2} {}^k a_{ij} & \text{if } v_i \text{ or } v_j \text{ is contained within } F_R \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

with the  ${}^k a_{ij}$  being the elements of the  $k^{th}$  power of  $M$ . These local analogues can also be expressed in matrix form by the expression:

$$q_{kL}(x) = [X]^t M^k_L [X] \quad (7)$$

Note that for every partitioning of a molecule into  $Z$  molecular fragment there will be  $Z$  local molecular fragment matrices. That is to say, if a molecule is partitioned into  $Z$  molecular fragments, the matrix  $M^k$  can be partitioned into  $Z$  local matrices  $M^k_L, L = 1, \dots, Z$  and the  $k^{th}$  power of matrix  $M$  is exactly the sum of the  $k^{th}$  power of the local (“molecular fragment”)  $Z$  matrices,

$$M^k = \sum_{L=1}^Z M^k_L \quad (8)$$

or in the same way as  $M^k = [{}^k a_{ij}]$  where,

$${}^k a_{ij} = \sum_{L=1}^Z {}^k a_{ijL} \quad (9)$$

and the total quadratic indices is the sum in the quadratic indices of the  $Z$  molecular fragments,

$$q_k(x) = \sum_{L=1}^Z q_{kL}(x) \quad (10)$$

Any local quadratic index has a particular meaning, especially for the first values of  $k$ , where the information about the structure of the fragment  $F_R$  is contained.

High values of  $k$  are in relation with the environment information of the fragment  $F_R$  considered inside the molecular pseudograph (G).

Atom and atom-type quadratic indices are specific case of local quadratic indices (for  $F_R = \text{atom or atom-type}$ ). In this sense, the  $k^{\text{th}}$  atom-type quadratic indices are calculated by summing the  $k^{\text{th}}$  atom quadratic indices of all atoms of the same atom type in the molecule.

In the atom-type quadratic indices formalism, each atom in the molecule is classified into an atom-type (fragment), such as heteroatoms, heteroatoms H-bonding acceptor (O, N and S), halogens, aliphatic carbon chain, aromatic atoms (aromatic rings), and so on. For all data sets, including those with a common molecular scaffold as well as those with very diverse structure, the  $k^{\text{th}}$  atom-type quadratic indices provide much useful information.

In any case, whether a complete series of indices is considered, a specific characterization of the chemical structure is obtained (whole structure or fragment), which is not repeated in any other molecule. The generalization of the matrices and descriptors to "superior analogues" is necessary for the evaluation of situations where only one descriptor is unable to bring a good structural characterization (19). These local indices can also be used together with total indices as variables of QSAR and QSPeR models for properties or activities that depend more on a region or a fragment than on the whole molecule.

## METHODS

### *TOMOCOMD-CARDD Approach*

*TOMOCOMD* is an interactive program for molecular design and bioinformatics research (16-18). The program is composed by four subprograms, each one of them dealing with drawing structures (drawing mode) and calculating 2D and 3D molecular descriptors (calculation mode). The modules are named CARDD (Computed-Aided 'Rational' Drug Design), CAMPS (Computed-Aided Modelling in Protein Science),

CANAR (Computed-Aided Nucleic Acid Research) and CABPD (Computed-Aided Bio-Polymers Docking). In this paper, we outline salient features concerned with only one of these subprograms: CARDD. This subprogram was developed based on a user-friendly philosophy without *prior* knowledge of programming skills.

The calculation of total and local quadratic indices for any organic molecule (or any drug-like compounds) was implemented in the *TOMOCOMD-CARDD* software (16). The main steps for the application of this method in QSAR/QSPeR can be briefly resumed as follows:

1. Draw the molecular pseudographs for each molecule of the data set, using the software-drawing mode. This procedure is carried out by a selection of the active atom symbol belonging to different groups of the periodic table,
2. Use appropriated atom weights in order to differentiate the molecular atoms. In this work, we used as atomic property the Mulliken electronegativity (20) for each kind of atom,
3. Compute the total and local quadratic indices of the molecular pseudograph's atom adjacency matrix. They can be carried out in the software calculation mode, which you can select the atomic properties and the family descriptor previously to calculate the molecular indices. This software generate a table in which the rows correspond to the compounds and columns correspond to the total and local quadratic indices or any others family molecular descriptors implemented in this program,
4. Find a QSPeR/QSAR equation by using statistical techniques, such as multilinear regression analysis (MRA), Neural Networks (NN), Linear Discrimination Analysis (LDA), and so on. That is to say, we can find a quantitative relation between P values and the quadratic indices having, for instance, the following appearance,

$$P = a_0 q_0(x) + a_1 q_1(x) + a_2 q_2(x) + \dots + a_k q_k(x) + c \quad (11)$$

where  $*q_k(x)$  [or  $q_{kL}(x)$ ] is the  $k^{\text{th}}$  total [or local] qua-

dratic indices, and the  $a_k$ 's are the coefficients obtained by the linear regression analysis.

- Test the robustness and predictive power of the QSPeR/QSAR equation by using internal and external cross-validation techniques,
- Develop a structural interpretation of obtained QSPeR/QSAR model using quadratic indices as molecular descriptors.

The descriptors used in order to achieve the theoretical models were the following:

(1)  $q_k(x)$  and  $q_k^H(x)$  are the  $k^{th}$  total quadratic indices calculated using the  $k^{th}$  power of the matrices  $[M^k(G)]$  of the molecular pseudograph (G) considering and not considering hydrogen atoms respectively.

(2)  ${}^E q_{kL}(x)$  [or  ${}^E q_{kL}^H(x)$ ],  ${}^A q_{kL}(x)$  [or  ${}^A q_{kL}^H(x)$ ] and  ${}^H q_{kk}(x)$  are the  $k^{th}$  local quadratic indices calculated using a  $k^{th}$  power of the local matrices  $[M^k_L(G, Fi)]$  of the molecular pseudograph (G) not considering (or considering) hydrogen atoms for heteroatoms (S, N, O), aromatic systems and hydrogen-bonding heteroatoms (S, N, O), respectively.

**Permeability Data**

The experimental data set, used in the present study, was composed by 33 structurally diverse drugs, which were experimentally studied by Liang et al (21).

These authors used a traditionally and accelerate Caco-2 cell permeability model and the experimental values of P (AP–BL) and P (BL–AP) are depicted in Tables 2 y 3, respectively. The P values of the test set of 18 drugs were taken from a study developed by Yazdani and co-workers (6). The compounds used by Liang et al. were taken from reference 6. Both studies were carried out under the same experimental conditions. The selected data set for 'in silico' permeability studies included model compounds with absorption by paracellular and transcellular diffusion, carrier-mediated absorption as well as substrates for carrier-mediated secretion via P-glycoprotein. These compounds had a molecular weight between 60 and 515 amu and their molecular charge was variable at pH 7.4 (6).

**Table 2: Experimental (traditional or accelerated model) and calculated values of the Caco-2 cell permeability coefficients (AP–BL) of 33 compounds as well as residuals of the regression and cross-validation.**

No Compound	$\times 10^6$ (cm/s)						
	$P_{traditional}^a$ (obsd)	$P_{accelerated}^a$	$P_{traditional}^b$ (calc)	$\Delta^c$	$CV-\Delta^d$	$P_{P_{obsd}/P_{calc}}$	
1 Aminopyrine	36.5	65.2	29.78	6.72	8.25	1.8	0.8
2 Caffeine	30.8	51.9	25.45	5.35	7.81	1.7	0.8
3 Nevirapine	30.1	66.7	21.77	8.33	9.43	2.2	0.7
4 Phenytoin	26.7	53.6	19.31	7.39	8.90	2.0	0.7
5 Testosterone	24.9	61	23.44	1.46	1.68	2.4	0.9
6 Phencyclidine	24.7	50.6	21.84	2.86	3.16	2.0	0.9
7 Desipramine	24.4	39.3	23.16	1.24	1.39	1.6	0.9
8 Progesterone	23.7	39	24.55	-0.85	-1.03	1.6	1.0
9 Clonidine	21.8	46.2	26.45	-4.65	-6.32	2.1	1.2
10 Propranolol	21.8	34.3	12.63	9.17	10.34	1.6	0.6
11 Chlorpromazine	19.9	32.5	25.22	-5.32	-6.05	1.6	1.3
12 Meloxicam	19.5	39.5	16.93	2.57	25.83	2.0	0.9
13 Nicotine	19.4	54.6	18.91	0.49	0.53	2.8	1.0
14 Pindolol	16.7	31.2	3.77	12.93	15.57	1.9	0.2
15 Telmisartan	15.1	21.1	18.02	-2.92	-11.96	1.4	1.2
16 Hydrocortisone	14	22.5	14.47	-0.47	-0.55	1.6	1.0
17 Timolol	12.8	25.4	3.24	9.56	12.29	2.0	0.3
18 Scopolamine	11.8	20.9	10.34	1.46	1.69	1.8	0.9
19 Dopamine	9.33	17.5	15.41	-6.08	-7.33	1.9	1.7
20 Labetalol	9.31	15.3	7.71	1.60	2.16	1.6	0.8
21 Bremazocine	8.02	19.1	14.67	-6.65	-7.92	2.4	1.8
22 Urea	4.56	7.78	11.80	-7.24	-8.70	1.7	2.6
23 Uracil	4.24	6.86	13.89	-9.65	-10.69	1.6	3.3
24 Cimetidine	1.37	2.29	5.54	-4.17	-4.60	1.7	4.0
25 Methyl scopolamine	0.69	1.56	7.76	-7.07	-8.85	2.3	11.2
26 Hydrochlorothiazide	0.51	0.93	-2.54	3.05	4.63	1.8	-5.0
27 Acebutolol	0.1	1.81	4.75	-4.24	-4.66	3.5	9.3
28 Ranitidine	0.49	1.17	-4.10	4.59	5.91	2.4	-8.4
29 Pirenzepine	0.44	1.34	11.56	-11.12	-11.74	3.0	26.3
30 Mannitol	0.38	1.24	-1.01	1.39	1.81	3.3	-2.7
31 Sulfasalazine	0.3	0.92	5.62	-5.32	-7.12	3.1	18.7
32 Acyclovir	0.25	2.13	5.47	-5.22	-5.96	8.5	21.9
33 Chlorothiazide	0.19	1.11	-0.60	0.79	1.17	5.8	-3.1

Test Set							
NoCompounds	$P_{OBS}^a$	$P_{CAL}^b$	No Compounds	$P_{OBS}^a$	$P_{CAL}^b$	No Compounds	$P_{OBS}^a$ $P_{CAL}^b$
34 Griseofulvin	36.60	52.62	40 Salicylic acid	22.00	23.76	46 Zidovudine	6.93 -14.30
35 Piroxicam	35.60	32.55	41 Warfarin	21.10	40.20	47 Nadolol	3.88 7.80
36 Diazepam	33.40	26.55	42 Indomethacin	20.40	27.62	48 Sucrose	1.71 -26.62
37 Alprenolol	25.30	15.21	43 Estradiol	16.90	21.45	49 Atenolol	0.53 3.92
38 Metoprolol	23.70	9.66	44 Dexamethasone	12.2	24.67	50 Terbutaline	0.47 8.49
39 Corticosterone	21.20	17.07	45 Acetylsalicylic acid	9.09	24.84	51 Ganciclovir	0.38 0.67

<sup>a</sup>From Ref. (21). <sup>b</sup>Calculated with the Eq. (13). <sup>c</sup>Residual, defined as  $P_{traditional} (obsd) - P_{traditional} (calc)$ . <sup>d</sup>Residual of the cross-validation.

**Table 3: Experimental (by either traditional or accelerated model) and calculated values of the Caco-2 cell permeability coefficients (BL–AP) of 17 compounds as well as residuals of the regression and cross-validation.**

No Compounds	$\times 10^6$ (cm/s)						
	$P_{traditional}^a$ (obsd)	$P_{accelerated}^a$	$P_{traditional}^b$ (calc)	$\Delta^c$	$CV-\Delta^d$	$P_{P_{obsd}/P_{calc}}$	
2 Caffeine*	27.7	59.8	--	--	--	2.16	--
5 Testosterone	39.1	66.9	25.85	13.2	15.64	1.71	0.66
8 Progesterone	20.3	40.6	25.96	-5.66	-6.65	2.00	1.28
9 Clonidine	27.1	41	24.18	2.92	3.28	1.51	0.89
10 Propranolol	33.2	32.2	28.26	4.94	5.74	0.97	0.85
15 Telmisartan	42.5	35.4	40.83	1.67	4.86	0.83	0.96
16 Hydrocortisone	15.9	28.8	12.49	3.41	3.74	1.81	0.79
18 Scopolamine	14.2	19.5	11.69	2.51	2.74	1.37	0.82
21 Bremazocine	19.1	20.3	25.80	-6.70	-7.62	1.06	1.35
22 Urea	7.91	9.84	17.28	-9.37	-10.74	1.24	2.18
24 Cimetidine	1.67	1.94	1.26	0.41	0.49	1.16	0.75
25 Methyl scopolamine	1.78	2.31	10.03	-8.25	-9.07	1.30	5.64
26 Hydrochlorothiazide	1.46	1.21	0.37	1.09	1.66	0.83	0.25
27 Acebutolol	10.70	6.22	10.12	0.58	0.63	0.58	0.95
28 Ranitidine	3.49	1.36	-4.99	8.48	11.40	0.39	-1.43
29 Pirenzepine	2.12	1.73	10.62	-8.50	-9.26	0.82	5.01
33 Chlorothiazide	2.64	3.03	3.42	-0.78	-1.24	1.15	1.30

\*Outlier. <sup>a</sup>From Ref. (21). <sup>b</sup>Calculated with the Eq. (14). <sup>c</sup>Residual, defined as  $P_{traditional} (obsd) - P_{traditional} (calc)$ . <sup>d</sup>Residual of the cross-validation.



### Statistical Analysis

All statistical analyses were developed with the STATISTICA 5.5 (22). The classification analysis was done using the Linear Discriminant Analysis (LDA). The quality of the model was determined examining the statistics parameter of multivariable comparison of the regression and by the cross-validation procedure (leave-*n*-out). The classification trees module was used to carry out leave-*n*-out cross validation procedures.

The linear multiple regression analysis (LMR) was developed to obtain the quantitative models between chemical structure and P values determined by traditional and accelerated methods.

The statistics quality of these models was evaluated taking into consideration the statistics graphics, the statistics parameters of multivariable comparison of the regression and the cross-validation procedure (leave-one-out).

## RESULTS AND DISCUSSION

### Developing the Discrimination Function

Linear discriminant analysis (LDA) was employed with the aim of developing a function that discriminates between high and moderate-poor absorbed compounds. For this purpose, this data set was split into two subsets according to the quantitative value of P (AP-BL) (23): the first group was composed by 18 compounds (high-absorption group;  $P(\text{AP-BL}) \geq 10 \times 10^{-6} \text{cm/s}$ ) and the second one by 15 compounds with moderate-poor absorption ( $P(\text{AP-BL}) < 10 \times 10^{-6} \text{cm/s}$ ). The range selection for permeability coefficient in Caco-2 cells is a bottleneck whether a correlation with the human absorption is searched. Several classification methods have been described in the literature, (6, 24-26) where the inter-laboratory and experimental variability is considered. Nevertheless, if all the approaches reported in the literature are analyzed, it can be stated that a value of P greater than  $10 \times 10^{-6} \text{cm/s}$  will classify good-absorption compounds (70-100%). Taking into consideration the reported selection approaches, a second group with moderate-poor absorption ( $< 10 \times 10^{-6} \text{cm/s}$ ) was selected, although in this range a high variability is appreciated when the human oral absorption is predicted from the P values (6, 24-26).

From a practical perspective the established boundary, assure that classified compounds have a good absorption profile. The best classification model found, by a forward-stepwise variable selection procedure, together with the statistical parameters, is shown below:

$$f_{(H,M,P)} = 4.748 - 0.176^{H1} q_{2L}(x) + 4.423 \times 10^{-2H1} q_{2L}(x) - 9.335 \times 10^{-H1} q_{2L}(x) - 0.287^P q_{1L}^{H1}(x) + 0.024^P q_{3L}^{H1}(x) + 6.99 \times 10^{-3} q_{6L} q_{6L}(x) \quad (12)$$

$N = 33$        $\lambda = 0.40$        $D^2 = 5.955$        $F(6,26) = 6.3985$        $p < 0.0003$

where, N is the number of compounds,  $\lambda$  is Wilks' coefficient, F is the Fisher ratio,  $D^2$  is the squared Mahalanobis distance and *p*-value is the significance level. The Wilks'  $\lambda$  parameter for overall discrimination can take values in the range of 0 (perfect discrimination) to 1 (no discrimination). The Mahalanobis distance indicates the separation between the respective groups. This model classified correctly 80.00% of compounds with moderate-poor absorption properties and a 94.44% of compounds with high absorption. The global classification for the data set was 87.87%. Only four compounds were classified bad, one of them with high absorption was classified as a moderate-poor absorption drug (3.03%, false negative compound) and three belonging to the moderate-poor absorption group were classified as high-absorption drugs (9.09%, false positive compounds).

From a practical point of view and in the development of the classifier model, the prediction of false negatives is considered more important because they are compounds that will be rejected for their poor predicted intestinal absorptions and therefore they will never be evaluated experimentally, and their true absorption would never be discovered. On the contrary, the false positive compounds eventually will be detected.

In Table 4 are shown the results of classification and a posteriori probabilities for the 33 compounds of the data set. The predictability of the model obtained by LDA was assessed through a leave-one-out (LOO) cross-validation procedure. In this methodology, the model was built after removing one compound and the resulting model was used to predict the property of the one removed. This was repeated to obtain a prediction for every compound. Using this approach, the model classified correctly 80.00% and 94.44% of compounds from the training set that belong to the moderate-poor and high-absorption group, respectively.

The global classification of the cross-validation was 87.87%.

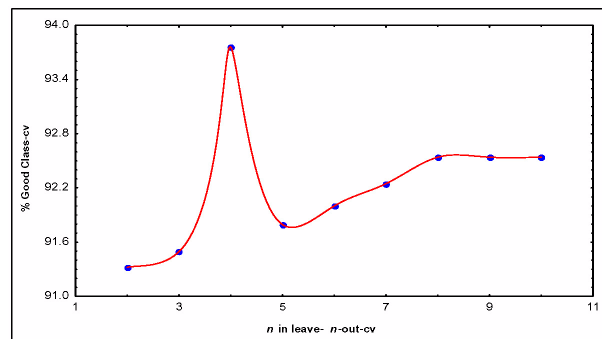
**Table 4: Results of Compounds Classification for the Data and External Prediction Set (AP–BL).**

No Compounds	Class <sup>a</sup>	Prob <sup>b</sup>	p-cv <sup>c</sup>	No Compounds	Class <sup>a</sup>	Prob <sup>b</sup>	p-cv <sup>c</sup>
<i>High absorption group</i>							
1 Aminopyrine	H	97.44	96.82	12 Meloxicam	H	97.29	96.82
2 Caffeine	H	74.13	70.61	13 Nicotine	H	97.52	96.93
3 Nevirapine	H	96.71	96.2	14 Pindolol	H	54.95	55.61
5 Testosterone	H	97.82	97.43	15 Telmisartam	H	99.89	99.85
6 Phencyclidine	H	99.46	99.31	16 Hydrocortisone	H	62.39	61.96
7 Desipramine	H	97.61	97.18	17 Timolol	H	68.71	67.1
8 Progesterone	H	99.81	99.75	18 Scopolamine	H	57.7	55.25
9 Clonidine	H	97.29	96.95	21 Bremazocine	M-P	89.35	88.61
10 Propranolol	H	94.05	93.4	23 Uracil	M-P	57.72	55.92
11 Chlorpromazine	H	99.33	99.14	27 Acebutolol	M-P	56.62	56.86
<i>Moderate-poor absorption group</i>							
4 Phenytoin	H	64.87	64.1	28 Ranitidine	M-P	98.17	97.98
19 Dopamine	M-P	92.34	91.86	29 Pirenzepine	M-P	71.91	72.98
20 Labetalol	M-P	99.37	99.21	30 Mannitol	M-P	99.47	99.35
22 Urea	M-P	94.84	94.61	31 Sulfasalazine	M-P	93.57	93.36
24 Cimetidine	M-P	76.78	75.53	32 Acyclovir	M-P	99.84	99.81
25 Methyl scopolamine	M-P	64.03	66.14	33 Chlorothiazide	M-P	98.4	98.25
26 Hydrochlorothiazide	M-P	99.81	99.76				
<i>Test Set</i>							
No Compounds	Class <sup>a</sup>	Prob <sup>b</sup>	No Compounds	Class <sup>a</sup>	Prob <sup>b</sup>		
34 Griseofulvin	H	99.56	43 Estradiol	H	84.09		
35 Piroxicam	H	98.78	44 Dexamethasone	H	83.52		
36 Diazepam	H	95.3	45 Acetylsalicylic acid	M	57.06 <sup>*</sup>		
37 Alprenolol	H	88.96	46 Zidovudine	M	98.73		
38 Metoprolol	H	90.91	47 Nadolol	M	69.25		
39 Corticosterone	H	83.17	48 Sucrose	M	99.99		
40 Salicylic acid	H	79.08 <sup>*</sup>	49 Atenolol	M	90.58		
41 Warfarin	H	93.93	50 Terbutaline	M	92.70		
42 Indomethacin	H	94.39	51 Ganciclovir	M	99.98		

<sup>a</sup>Observed classification: H (high-absorption group) and M-P (moderate-poor absorption group). <sup>b</sup>Probability calculated. <sup>c</sup>Probability calculated LOO cross-validation. <sup>d</sup>Probability calculated for the test set. <sup>\*</sup>Misclassified Compounds for the test set.

For a more exhaustive testing of the predictive power of the found model a leave- $n$ -out (LnO) cross validation procedures was carried out using the classification tree module (22). The selected conditions for the validation procedure were discriminant-based linear combination as split method, prune on misclassification error as stopping rule and the same prior probabilities than in Eq. 12. Once the selected conditions are applied in the classification trees' module, Eq. 12 is obtained and a LnO procedure can be developed varying the folding parameter of the cross validation. This model shown a 91.3, 91.5, 93.8, 91.8, 92.0 and 92.2% of good global classification when  $n$  varied from 2 to 7 respectively in the LnO cross validation procedures. The model was stabilized around 92.5% when  $n$  was > 7 (see Figure 1). In addition, to assess the predictive power of the model an external prediction set of 18 drugs was used, where the percentage of good global classification was 88.88%. If we considered the data set and the test set together (*full set*) the percentage of good classification was 88.23%. At the bottom of Table

4 appear the obtained results for the external prediction set. As it can be seen, in both series, the predictability and robustness of the theoretical model was demonstrated.



**Figure 1: Result obtained to assess the predictive power of the discrimination model (Eq. 12) using LnO Cross-Validation procedure. The model (Eq. 12) was stabilized around 92.5% when  $n$  was > 7.**

From the *full* data set, only 6 compounds were classified badly. Two drugs (salicylic acid and phenytoin) were classified as false negative and four as false positive compounds (bremazocine, uracil, acebutolol and Acetylsalicylic acid). The high percentage of misclassified compounds belongs to the moderate-poor absorption group. In this set, they were included compounds with P values lower than  $10 \times 10^{-6}$ . This is a logic result due to the fact that in this group there are some compounds with high variability in relation to the human absorption values and it is not always possible an extrapolation between human absorption and the P values ( $P < 10$ ); for example a drug like acebutolol ( $P = 0.51$ ) has a 90% of human absorption (27). In addition, others compounds such as acetylsalicylic acid ( $P = 9.09$ ) and bremazocine ( $P = 8.02$ ) have P values close to the selected limit value ( $10 \times 10^{-6}$ ) and in the first case the low percentage of classification (57.06%) can be explained when the human absorption value is considered (100%) (6).

### The Regression Models

With the aim of predicting the P values, we develop two quantitative models that relate the quadratic indices of the molecules with the P (AP–BL) and P (BL–AP). The best linear regression models for P were obtained by a forward stepwise procedure; the equation and the statistical parameters are shown below:

$$P(\text{AP} \rightarrow \text{BL}) = 28.3222 - 2.619 \times 10^{-4} q_{\text{H}}^{\text{H}}(x) + 2.1 \times 10^{-5} q_{\text{H}}^{\text{H}}(x) - 0.0555^{\text{E}} q_{\text{H}}^{\text{H}}(x) - 2.204 \times 10^{-8} q_{\text{H}}^{\text{H}}(x) - 4.8761 \times 10^{-7} q_{\text{H}}^{\text{H}}(x) + 0.03^{\text{E}} q_{\text{H}}^{\text{H}}(x) \quad (13)$$

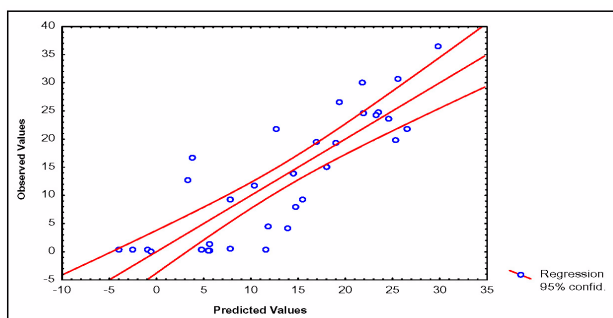
N = 33 R = 0.8417 F(6,26) = 10.534 s = 6.6438 s<sub>CV</sub> = 8.5654 p < 0.00001

$$P(\text{BL} \rightarrow \text{AP}) = 32.44068 - 0.30871^{\text{E}} q_{\text{H}}^{\text{H}}(x) + 7.97 \times 10^{-8} q_{\text{H}}^{\text{H}}(x) \quad (14)$$

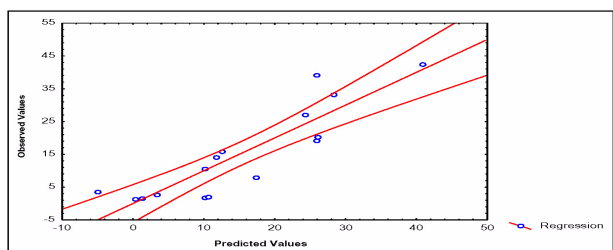
N = 16 R = 0.8896 F(2,13) = 24.647 s = 6.8420 s<sub>CV</sub> = 7.3052 p < 0.00004

where, R is the multiple regression coefficient, s the standard deviation of the regression, s<sub>CV</sub> are the standard deviation of the LOO cross-validation procedure; F is the Fisher ratio at the 95% confidence level and p-value is the significance level.

In Tables 2 and 3, the values of experimental and calculated permeability coefficients for the data set are given, and in the Figures 2 and 3, the existing linear relationships between them are depicted. In the development of the quantitative model for description of P (BL-AP) of the data set, one compound was detected as statistical outliers (Caffeine). Outliers' detection was carried out by using the following standard statistical tests: residual, standardized residuals, studentized residual and Cooks' distance (28).



**Figure 2: Correlation between experimental and calculated P (AP-BL) values of 33 compounds of the data set.**



**Figure 3: Correlation between experimental and calculated P (BL-AP) values of 16 compounds of the data set. Caffeine was detected as statistical outlier. For this reason, this compound was excluded of the statistical analysis.**

Several researchers have explored QSPeR involving Caco-2 cell permeability. In these studies, some types of molecular descriptors have been introduced, where size and hydrogen-bonding descriptors (13), polar surface area (PSA) (8, 29, 30), Molsurf-derived descriptors (31), MO-calculation (11) and using membrane-interaction analysis (32) are included. These QSPeR models have predicted the P values with a reasonable accuracy, although the numbers of compounds in the data sets have been limited. For example, Fujiwara *et al.* considered quadratic and interactive terms (11) in the equations and the correlation coefficients were 0.74 and 0.76, respectively. In the same paper, these authors increased the correlation coefficient up to 0.790 using a neural network. In addition, van de Waterbeemd *et al.* obtained an R-value for P between 0.513 and 0.884 (13). In other paper, Ren and Lien (33) developed a QSAR analysis where an adequate regression coefficient value, for the same data set of 51 compounds used in this study, was obtained (0.79). Finally, in a recently study developed by Kulkarni *et al.* (32), about prediction of P values, 6 predictive models were obtained using Multidimensional Linear Regression (MLR) and the R values were between 0.86 and 0.92; but in this case only 74% from the original data set (6) was chosen.

### Interpretation of QSPeR Models

Up to now, it is known that the absorption is influenced by a different kind of interactions. Several studies have demonstrated that the permeability coefficient, measured by a transport through Caco-2 monolayer cell cultures, is correlated with lipophilicity (6, 12, 13, 33), while others emphasize on the role of hydrogen-bonding capacity or charge (7, 8, 12, 13). A paradigm of structure-permeability relationship has been expressed as (13):

$$\text{Permeability} = f(\text{lipophilicity, molecular size, H-bonding capacity, charge}) \quad (15)$$

As it can be observed, in the discriminant and the regression models, the included variables are very close to the factors that influence on the P values. These factors are related with the structural features of molecules. For example in Eq. 12, the variables  $^{\text{H}}q_{2\text{L}}(x)$ ,  $^{\text{H}}q_{4\text{L}}(x)$  and  $^{\text{H}}q_{7\text{L}}(x)$  are connected with the hydrogen atoms as donors, while the  $^{\text{E}}q_{1\text{L}}^{\text{H}}(x)$  and  $^{\text{E}}q_{3\text{L}}^{\text{H}}(x)$  variables contain information about the number of hydro-



gen acceptors and the charge of molecules. All of them are related with the total hydrogen bond capacity.  $P$  negatively depends on these descriptors. The values of these molecular indices increase with the rise of the numbers of heteroatoms and the hydrogen bond to heteroatoms in the molecules. For this reason, we can say that these molecular descriptors have a negative contribution to  $P$ . These are a logical results because increasing the number of heteroatoms and the hydrogen bond to heteroatoms in the molecules decrease the permeability across the biological membrane. This effect is rather close to the molecular lipophilicity decrease and the possibility of the molecule of ionization and to obtain a charge. The charge factor is related with the negative charge of the biological membrane (34). This observation is supported by a study developed by Ren et al., where the same full set used in the present study. First, a low regression coefficient ( $R = 0.749$ ) was evidenced when anionic, cationic and neutral compounds (the full set), were studied using the net charge of molecules like a descriptor. When these compounds were divided into three subgroups, namely neutral, cationic and anionic compounds, much better correlation coefficients ( $R = 0.968, 0.915$  and  $0.931$ , respectively) were obtained (33).

Other descriptor with a positive contribution to the discriminant function is  $q_{\sigma}q_{\sigma}(x)$ . This variable contains information about the molecular weight and consequently of the size of molecules. For this reason, although, the number of heteroatoms is increased (negative contribution to the permeability coefficient) the quadratic influence of molecular size (descriptor) should increase the permeability of molecules. In addition, these properties (molecular weight, size), H-bonding and charge are components of lipophilicity (13). For each property there are limited ranges as it is established in the Rule-of-5 (1), but anyone is independent (35).

Taking into consideration the above mentioned approach it should be considered that successful drug candidates will be characterized by an optimal range of values for H-bonding, lipophilicity and size (36) and for this reason, compounds with extreme positive values of these properties could have a marked negative effect on permeability across the biological membrane (13).

As it can be seen, in the case of the linear regression model there are some influences of the local and the total descriptor of different orders, in similar way to the classification function. In the Eq.13, it appears the variable  $^Aq_{14L}(x)$  that indicates the presence and size of aromatic systems. In order to describe the  $P$  ( $AP \rightarrow BL$ ) and  $P$  ( $BL \rightarrow AP$ ) the descriptors selected by the regression process, in both equations, were different. This fact suggests that the transport process is different in two directions, explaining that some compounds are potential substrate of cellular efflux pumps (21). During the last few years, the Caco-2 cell model has been used to evaluate whether new chemical entities have possible cellular efflux pumps. For compounds that are cellular efflux pump substrates the  $P$  ( $BL \rightarrow AP$ ) value will be higher than  $P$  ( $AP \rightarrow BL$ ) (21) and for compounds that are not substrates the ratio between this two values should be near to 1. This ratio is called the Permeability Directional Ratio (PDR,  $BL \rightarrow AP / AP \rightarrow BL$ ). For this reason through the use of both regression models, in a combinative way, it is possible to predict the PDR value. This theoretical approach appears like a new alternative to the study of cellular efflux pump.

#### *Virtual Screening and Correlation with 'in vivo' Data*

One of the most important aspects of any quantitative structure-permeability model is its ability to predict the studied  $P$  for any compound not included in the data (or training) set. Virtual screening has emerged as an interesting alternative to the handling and screening of large databases in order to find a reduced set of new potential drug candidates (37-39). In the present study, we simulated a virtual search of  $P$  ( $AP \rightarrow BL$ ) values by using the discriminant function (Eq. 12) and regression equation (Eq. 13) obtained through the *TOMOCOMD-CARDD* approach. In Table 5, the Caco-2 cell permeability data for 134 structurally diverse compounds, obtained from different sources are summarized (32, 33, 40-50). Sometimes they were obtained  $P$  experimental values from two or more sources, existing significant variability in these values. Several researchers have demonstrated inter-laboratory differences for Caco-2 cell permeability studies (8, 51).

The evaluation results of these compounds are given in Table 5.

**Table 5: Result of the Virtual Screening for 134 Structurally Diverse Compounds.**

Compounds	Ref. <sup>a</sup>	P <sub>Obs</sub> <sup>b</sup>	P <sub>Cal</sub> <sup>c</sup>	Class <sup>d</sup>	Compounds	Ref. <sup>a</sup>	P <sub>Obs</sub> <sup>b</sup>	P <sub>Cal</sub> <sup>c</sup>	Class <sup>d</sup>
Acebutolol	42	4.47	4.75	56.62 <sup>M</sup>	Oxprenolol	42	66.07	10.15	92.51 <sup>H</sup>
	6	0.51				54	120.00		
Acebutolol ester <sup>f</sup>	42	77.62	2.75	90.45 <sup>H</sup>	Oxprenolol ester <sup>f</sup>	42	97.72	8.15	97.53 <sup>H</sup>
Acetylsalicylic acid	40	2.40	24.84	57.06 <sup>M</sup>	Phencyclidine	6	24.55	21.84	99.45 <sup>H</sup>
	26	30.90			Phenytol	6	26.92	19.31	64.86 <sup>M</sup>
	6	9.12			Pindolol	6	16.60	3.77	54.94 <sup>H</sup>
Acyclovir	26	2.00	5.47	99.84 <sup>M</sup>	Pirenzepine	6	0.44	11.56	71.91 <sup>M</sup>
	6	0.25			Piroxicam	6	35.48	32.55	98.78 <sup>H</sup>
Alprenolol	40	40.74	15.21	88.96 <sup>H</sup>		50	46.25		
	7	40.74			Practolol	40	0.89	1.89	53.34 <sup>M</sup>
	42	75.86				7	0.89		
	6	25.12				12	1.38		
	46	38.30				46	0.92		
	47	242.00				47	3.50		
Alprenolol ester <sup>f</sup>	42	107.15	13.23	95.81 <sup>H</sup>	Prazocin	26	43.65	5.13	63.30 <sup>M</sup>
Aminopyrine	6	36.31	29.78	97.44 <sup>H</sup>	Progesterone	25	78.23	24.55	99.81 <sup>H</sup>
Artemisinin	43	30.40	11.42	76.30 <sup>H</sup>		6	23.70		
Artesunate	43	4.00	5.74	71.34 <sup>H</sup>	Propranolol	40	41.69	12.63	94.04 <sup>H</sup>
Atenolol	40	0.20	3.92	90.58 <sup>M</sup>		7	41.69		
	7	0.20				42	83.18		
	44	0.47				26	27.54		
	6	0.52				6	21.88		
	12	0.23				43	11.20		
	43	0.13				46	43.03		
	24	4				24	14.8		
	48	1.16			Sumatriptan	26	3.02	-5.43	88.34 <sup>M</sup>
	50	0.19			Propranolol ester <sup>f</sup>	42	104.71	10.61	98.02 <sup>H</sup>
	46	0.20			Quinidine	26	20.42	13.33	90.36 <sup>H</sup>
	47	1.00			Ranitidine	6	0.49	-4.10	98.17 <sup>M</sup>
Azithromycin	26	1.05	-42.87	73.43 <sup>M</sup>	Salicylic acid	40	12.02	23.76	79.08 <sup>M</sup>
Benzylpenicillin	26	1.95	5.39	61.77 <sup>M</sup>		6	21.88		
Betaxolol	42	95.50	6.40	82.57 <sup>H</sup>	Scopolamine	6	11.75	10.34	57.70 <sup>H</sup>
Betaxolol ester <sup>f</sup>	42	95.50	4.38	94.39 <sup>H</sup>	Sucrose	6	1.70	-26.62	99.99 <sup>M</sup>
Tenidap	26	51.29	22.78	63.05 <sup>H</sup>		48	0.71		
Bremazocine	6	7.94	14.67	89.35 <sup>H</sup>	Sulfasalazine	40	0.13	5.62	93.56 <sup>M</sup>
Caffeine	26	50.50	25.45	74.13 <sup>H</sup>		6	0.30		
	6	30.90				6	0.30		
	48	84			Taurocholic acid	26	34.67	-19.30	99.26 <sup>M</sup>
	24	21.4			Ibuprofen	26	52.50	24.11	84.92 <sup>H</sup>
<b>Compounds</b>	<b>Ref.<sup>a</sup></b>	<b>P<sub>Obs</sub><sup>b</sup></b>	<b>P<sub>Cal</sub><sup>c</sup></b>	<b>Class<sup>d</sup></b>	<b>Compounds</b>	<b>Ref.<sup>a</sup></b>	<b>P<sub>Obs</sub><sup>b</sup></b>	<b>P<sub>Cal</sub><sup>c</sup></b>	<b>Class<sup>d</sup></b>
Terbutaline	40	0.38	8.49	92.70 <sup>M</sup>	Indomethacin	6	25.12	27.62	94.39 <sup>H</sup>
	48	1.04			Lactulose	47	0.27	-24.84	99.99 <sup>M</sup>
Chlorpromazine	6	19.95	25.22	99.33 <sup>H</sup>	Testosterone	7	51.80	23.44	97.81 <sup>H</sup>
Clonidine	26	30.20	26.45	97.29 <sup>H</sup>		26	72.27		
	6	21.88				6	24.9		
Chlorothiazide	6	0.19	-0.60	98.40 <sup>M</sup>		48	44.5		
Corticosterone	40	54.95	17.07	83.17 <sup>H</sup>		12	58.88		
	6	21.38			Timolol	42	44.67	3.24	68.70 <sup>H</sup>
	7	54.5				6	12.88		
Desipramine	26	21.38	23.16	97.61 <sup>H</sup>	Chloramphenicol	26	20.42	4.29	99.75 <sup>M</sup>
	6	24.55			Timolol ester <sup>f</sup>	42	79.43	-5.75	56.07 <sup>H</sup>
Dexamethasone	40	12.59	24.67	83.52 <sup>H</sup>	Trovafoxacin	26	30.20	42.18	93.58 <sup>H</sup>
	48	26.92			Uracil	6	4.27	13.89	57.72 <sup>H</sup>
	26	23.44			Urea	6	4.57	11.80	94.84 <sup>M</sup>
	6	12.30			Valproic acid	26	48.00	19.24	80.02 <sup>H</sup>
Dexamethasone-b-D-glucoside	48	0.44	0.40	93.10 <sup>M</sup>	Warfarin	40	38.02	40.20	93.93 <sup>H</sup>
Dexamethasone-b-D-glucuronide	48	1.15	-0.97	94.06 <sup>M</sup>		6	20.89		
Diazepam	26	70.79	26.55	95.31 <sup>H</sup>		12	53.70		
	6	33.11			Zidovudine	6	6.92	-14.30	98.73 <sup>M</sup>
	47	756.00			Ziprasidone	26	12.30	14.32	88.25 <sup>H</sup>
Dopamine	6	33.4			Cephalexin	43	0.18	12.49	98.97 <sup>M</sup>
Doxorubicin	6	9.33	15.41	92.34 <sup>M</sup>	Gly-Pro	50	6.10	6.33	95.31 <sup>M</sup>
Erythromycin	26	0.16	24.67	99.90 <sup>M</sup>	D-glucose	50	17.53	-0.60	99.68 <sup>M</sup>
Estradiol	6	3.72	-44.98	80.42 <sup>M</sup>	L-Phenylalanine	50	18.37	13.45	66.89 <sup>M</sup>
Felodipine	7	22.91	45.06	80.85 <sup>H</sup>	Ketoprofen	50	23.15	30.25	90.95 <sup>H</sup>
Fluconazole	26	29.51	4.18	60.60 <sup>H</sup>	Furosemide	50	0.29	-10.80	99.94 <sup>M</sup>
Ganciclovir	25	2.69	0.67	99.98 <sup>M</sup>	Sulpiride	47	0.39	-1.39	99.80 <sup>M</sup>
	6	0.38			SB 209670	47	8.80	18.73	82.91 <sup>M</sup>
	6	0.38			SB 217242	47	70.00	22.18	81.46 <sup>M</sup>
	6	36.31	52.62	99.56 <sup>H</sup>	Cimetidine	44	0.50	5.54	76.78 <sup>M</sup>
Gliseofulvin	7	0.91	-14.27	99.41 <sup>M</sup>		26	3.09		
H216/44	46	1.14				6	1.37		
	46	1.14				50	0.35		
Hydrochlorothiazide	44	1.51	-2.54	99.81 <sup>M</sup>	Hydrocortisone	7	21.38	14.47	62.38 <sup>H</sup>
	6	0.51				25	35.40		
Oxazepam	47	246.00	23.37	63.27 <sup>H</sup>		6	15.85		
Nordazepam	47	307.00	22.81	90.50 <sup>H</sup>		49	44.67		
Metolazone	47	6.10	13.92	95.20 <sup>M</sup>		48	12.9		

Compounds	Ref. <sup>a</sup>	P <sub>Obs</sub> <sup>b</sup>	P <sub>Cal</sub> <sup>c</sup>	Class <sup>d</sup>	Compounds	Ref. <sup>a</sup>	P <sub>Obs</sub> <sup>b</sup>	P <sub>Cal</sub> <sup>c</sup>	Class <sup>d</sup>
Labetalol	6	10.00	7.71	99.36 <sup>M</sup>	Amiloride	47	0.78	-9.54	99.96 <sup>M</sup>
Mannitol	40	0.18	-1.01	99.47 <sup>M</sup>	Alfentanil	47	310.00	-1.19	60.86 <sup>M</sup>
	25	3.24			Coumarin	12	77.62	34.07	98.79 <sup>H</sup>
	26	0.65			Theophylline	12	44.67	18.46	39.91 <sup>H</sup>
	6	0.38			Epinephrine	12	0.95	11.58	98.19 <sup>M</sup>
	12	2.63			Guanoxan	12	19.50	1.91	69.59 <sup>H</sup>
	50	0.12			Guanabenz	12	72.44	9.74	59.08 <sup>H</sup>
	24	0.5			Lidocaine	12	61.66	13.86	81.35 <sup>H</sup>
	48	0.17			Tiarcilast	12	12.59	30.01	57.07 <sup>H</sup>
	49	0.83			Nitrendipine	12	16.98	34.93	90.07 <sup>M</sup>
Meloxicam	6	19.95	16.93	97.29 <sup>H</sup>	Fleroxacin	12	15.49	41.49	63.85 <sup>H</sup>
Methanol	25	131.83	23.91	88.32 <sup>H</sup>	Diltiazem	12	48.98	19.02	85.14 <sup>H</sup>
Methotrexate	26	1.20	24.19	99.99 <sup>M</sup>	Verapamil	12	26.30	33.00	97.58 <sup>H</sup>
Methylscopolamine	6	0.79	7.76	64.02 <sup>M</sup>	Mibefradil	12	13.49	2.98	97.14 <sup>H</sup>
Metoprolol	40	26.92	9.66	90.91 <sup>H</sup>	Bosentan	12	1.05	7.74	99.84 <sup>M</sup>
	7	26.92			Proscillaridin	12	0.63	8.27	61.10 <sup>M</sup>
	50	18.80			Ceftriaxone	12	0.13	3.81	99.99 <sup>M</sup>
	46	26.71			Remikiren	12	0.74	-26.00	90.99 <sup>M</sup>
	47	92.00			Saquinavir	12	0.55	20.36	68.61 <sup>H</sup>
Nadolol	6	3.89	7.80	69.25 <sup>M</sup>	Olsalazine	40	0.11	21.76	99.81 <sup>M</sup>
	24	4.50			Glycine	26	80	18.16	90.99 <sup>H</sup>
Naloxone	26	28.18	27.36	87.84 <sup>H</sup>	Amoxicillin	48	0.33	-4.31	99.72 <sup>M</sup>
Naproxen	25	74.13	22.31	73.47 <sup>H</sup>	Enalaprilat	48	0.62	0.74	65.48 <sup>M</sup>
Nevirapine	6	30.20	21.77	96.70 <sup>H</sup>	Lisinopril	48	0.05	-5.96	99.52 <sup>M</sup>
Nicotine	6	19.50	18.91	97.51 <sup>H</sup>	Gabapentin	49	4.33	14.43	95.01 <sup>M</sup>
D-Phe-L-Pro	26	44.3	13.31	83.16 <sup>H</sup>		49	1.50		
BVaraU	24	4	6.39	99.82 <sup>M</sup>	Raffinose	47	0.05	-51.23	99.99 <sup>M</sup>
Pravastatin	24	2.3	2.76	97.56 <sup>M</sup>	Sildenafil	47	87.00	13.91	83.83 <sup>M</sup>
L-Glutamine	48	0.85	1.43	99.14 <sup>M</sup>	Antipyrine	50	47.23	26.15	98.99 <sup>H</sup>
SQ-29852	24	0.02	-0.58	97.81 <sup>M</sup>		47	2.15		
Foscarnet	47	0.05	6.25	96.34 <sup>M</sup>	Ciprofloxacin	47	1.90	34.94	71.32 <sup>H</sup>
Telmisartan	6	15.14	18.02	99.88 <sup>H</sup>	Imipramine	26	14.13	23.80	98.66 <sup>H</sup>
	6	0.47							
	50	0.97							

References where were collected the P values. <sup>b</sup>Permeability coefficient: P (AP-*BL*) x 10<sup>6</sup> cm/s, obtained from literature. <sup>c</sup>Permeability coefficient: P (AP-*BL*) x 10<sup>6</sup> cm/s, obtained using regression model (Eq. 13). <sup>d</sup>Classification and probability obtained using discriminant function (Eq. 12): <sup>H</sup>[high-absorption group (P (AP-*BL*) ≥ 10x10<sup>6</sup> cm/s)] and <sup>M</sup>[moderate-poor absorption group (P (AP-*BL*) < 10x10<sup>6</sup> cm/s)]. <sup>f</sup>O-cyclopropate carboxylic acid ester.

As can be seen in this Table, most of the 134 evaluated compounds are adequately predicted, although it should be considered that collected data are influenced by variations in cell culture conditions such as passage number, type of medium, day in culture, as well as the experimental conditions used for their measurement. It is obvious that from the results the quality of the predictions corroborates the predictive power of the models found and justified their use in the prediction of this important property. Besides, this is not a fortuitous result due to the data set used in this study including absorption model compounds such as drugs with paracellular passive absorption (e.g. atenolol, furosemide, mannitol, terbutaline), transcellular passive absorption (e.g. antipyrine, ketoprofen, metoprolol, piroxicam), active transport processes (carrier-mediated up-take, e.g. Gly-Pro and L-phenylalanine) and carried mediated secretion via P-glycoprotein (e.g. talinolol, cimetidine, acebutolol, ranitidine, chlorothiazide).

On the other hand, the ‘*in silico*’ estimated intestinal permeability could be used as a predictor of the true fraction of the absorbed drug.

The theoretical relationship between the fraction of drug absorbed (Fa) and permeability has been described by Amidon et al. (52):

$$Fa = (1 - e^{-P \cdot 10^{-6}})100[\%] \quad (16)$$

When the absorbed dose fraction, from human studies, is compared with the predictive P, a good relationship between the theoretical and observed values is obtained.

The following Table 6 a, b demonstrates this relation for several compounds used in the study.

**Table 6: a Caco-2 cell permeability coefficients calculate using discriminant and regression models and percent GI absorption in human from the different reports for compounds with moderate-poor absorption.**

Compounds	Ref. <sup>a</sup>	P <sub>Obs</sub> <sup>b</sup>	P <sub>Calc</sub> <sup>c</sup>	Class <sup>d</sup>	%Abs <sup>e</sup>	Compounds	Ref. <sup>a</sup>	P <sub>Obs</sub> <sup>b</sup>	P <sub>Calc</sub> <sup>c</sup>	Class <sup>d</sup>	%Abs <sup>e</sup>
<i>Compounds included in training data set</i>											
Caffeine	26	50.5	25.45	74.13 <sup>H</sup>	100	Propranolol	7	41.9	12.63	94.04 <sup>H</sup>	90
	6	30.9		100			26	27.5			90
	48	84		100			6	21.8			90
	24	21.4		100			43	11.2			95
Diazepam	26	70.97	26.55	95.31 <sup>H</sup>	100		24	14.8			90
	6	33.4		100			46	43.03			
Corticosterone	7	54.5	17.07	83.17 <sup>H</sup>	100		42	83.18			
	6	21.38		100		Scopolamine	6	11.8	10.34	57.70 <sup>H</sup>	100
Hydrocortisone	7	21.5	14.47	62.38 <sup>H</sup>	89	Telmisartan	6	15.1	18.02	99.88 <sup>H</sup>	90
	25	35.4		80		Dexamethasone	7	12.5	24.67	83.52 <sup>H</sup>	100
	6	14.00		89			26	23.4			100
	49	44.67		95			48	26.92			
	48	12.9		80, 95			6	12.3			
Testosterone	7	51.8	23.44	97.81 <sup>H</sup>	100	Desipramine	26	21.6	23.16	97.61 <sup>H</sup>	100
	26	72.27		100			6	24.4			>95
	6	24.9		100		Progesterone	25	78.93	24.55	99.81 <sup>H</sup>	100
	48	44.5		100			6	23.7			
	12	58.88				Nicotine	6	19.4	18.91	97.51 <sup>H</sup>	100
<i>Compounds included in screening data set</i>											
Glycine	26	80	18.16	90.99 <sup>H</sup>	100	Naproxen	25	74.17	22.31	73.47 <sup>H</sup>	100
Ibuprofen	26	52.50	24.11	84.92 <sup>H</sup>	100	Tenidap	26	51.2	22.78	63.05 <sup>H</sup>	90
Glycine	26	80	18.16	90.99 <sup>H</sup>	100	Noloxone	26	28.2	27.36	87.84 <sup>H</sup>	91
D-Phe-L-Pro Valproic acid	26	44.3	13.31	83.16 <sup>H</sup>	100	Felodipine	7	22.7	45.06	80.85 <sup>H</sup>	100
	26	48.00	19.24	80.02 <sup>H</sup>	100	Imipramine	26	14.1	23.8	98.66 <sup>H</sup>	100

**Table 6: b Caco-2 cell permeability coefficients calculate using discriminant and regression models and percent GI absorption in human from the different reports for compounds with high absorption**

Compounds	Ref. <sup>a</sup>	P <sub>Obs</sub> <sup>b</sup>	P <sub>Calc</sub> <sup>c</sup>	Class <sup>d</sup>	%Abs <sup>e</sup>	Compounds	Ref. <sup>a</sup>	P <sub>Obs</sub> <sup>b</sup>	P <sub>Calc</sub> <sup>c</sup>	Class <sup>d</sup>	%Abs <sup>e</sup>
<i>Compounds included in training data set</i>											
Atenolol	40	0.2	3.92	90.58 <sup>M</sup>		Sucrose	48	0.71	-26.62	99.99 <sup>M</sup>	42
	24	4		50			6	1.7			
	48	1.16		40-70		Ganciclovir	25	2.69	0.67	99.98 <sup>M</sup>	8
	7	0.2		50			6	0.38			3
	6	0.52		50		Acyelovir	26	2	5.47	99.84 <sup>M</sup>	30
	44	0.47					6	0.25			20
	12	0.23				Mannitol	40	0.18	-1.01	99.47 <sup>M</sup>	16
	43	0.13					25	3.24			17
	50	0.19					26	0.65			16
	46	0.2					6	0.38			16
	47	1					12	2.63			
Cimetidine	26	3.09	5.54	76.78 <sup>M</sup>	62		50	0.12			
	6	1.37		95			24	0.5			16
	44	0.5					48	0.17			5, 16
	50	0.35					49	0.83			5
Nadolol	6	3.89	7.8	69.25 <sup>M</sup>		Ranitidine	6	0.49	-4.1	98.17 <sup>M</sup>	50
	24	4.50				Terbutaline	6	0.47	8.49	92.70 <sup>M</sup>	73
Sulphasalazine	40	0.13	5.62	93.56 <sup>M</sup>	13		48	1.04			25-80
	6	0.30		13			40	0.38			
							50	0.97			
<i>Compounds included in screening data set</i>											
L-Glutamine	48	0.85	1.43	99.14 <sup>M</sup>	60-90	Enalaprilat	48	0.62	0.74	65.48 <sup>M</sup>	<10
Lisinopril	48	0.05	-5.96	99.52 <sup>M</sup>	25	Sumatriptan	26	3.02	-5.43	88.34 <sup>M</sup>	55
Doxorubicin	26	0.16	24.67	99.90 <sup>M</sup>	5	Pravastatin	24	2.3	2.76	97.56 <sup>M</sup>	34
Olsalazine	40	0.11	21.76	99.81 <sup>M</sup>	2	Gabapentin	49	4.33	14.43	95.01 <sup>M</sup>	74
BVaraU	24	4	6.39	99.82 <sup>M</sup>	82		49	1.50			36
Benzylpenicilin	26	1.95	5.39	61.77 <sup>M</sup>	30	Methotrexate	26	1.2	24.19	99.99 <sup>M</sup>	20

<sup>a</sup>References where were collected the P values. <sup>b</sup>Permeability coefficient: P (AP-BL) x10<sup>6</sup>cm/s, obtained from literature. <sup>c</sup>Permeability coefficient: P (AP-BL) x10<sup>6</sup>cm/s, obtained using regression model (Eq. 13). <sup>d</sup>Classification and probability obtained using discriminant function (Eq. 12): <sup>H</sup>[high-absorption group (P (AP-BL) ≥ 10x10<sup>6</sup>cm/s)] and <sup>M</sup>[moderate-poor absorption group (P (AP-BL) < 10x10<sup>6</sup>cm/s)]. <sup>e</sup>Human fraction absorbed.

## CONCLUSIONS

The total and local quadratic indices appear to be a promising structural invariant. Using these molecular indices and statistical techniques, a function discriminant that allows the correct classification of 87.87% of the compounds in the data set was developed. Using multiple regressions, two QSPeR models were obtained for the description and determination of P (AP-BL) and P (BL-AP) transport across monolayer of intestinal epithelial (Caco-2) cell. A LnO and LOO cross-validation procedure revealed that the discriminant and regression models respectively, had a fairly good predictability. Furthermore, other 18 drugs were selected as a test set (external prediction set) in order to assess the predictive power of the models. The P values of the test-set compounds were predicted with the same accuracy as the compounds of the data set. The use of both regression models, in a combinative way, is

possible to predict the Permeability Directional Ratio (PDR, BL-AP/ AP-BL) value. The obtained models were used in virtual screening for drug intestinal permeability obtaining good relation with in vivo human absorption data. The 'in silico' methodology, used in this study, shows how to provide an excellent alternative to the experimental models for rank-ordering compounds by reducing time and cost. Moreover, this approximation permits us to obtain significant interpretation of the experiment result in terms of the structural features of molecules. The application of the present method to the prediction of pharmacokinetic properties for several classes of organic compounds is now in progress and will be the subject of a future publication.

## ACKNOWLEDGEMENTS

We sincerely thank Drs. Eric J. Lien, Mehran Yazdani, A. J. Hopfinger, Mitsuru Hashida and Han van de Waterbeemd for providing some manuscript reprints from their works, which significantly contribute to the development of this paper. In addition, the authors thank the anonymous referees for their useful comments, which contributed to an improved presentation of these results.

## REFERENCES

- [1] Lipinski C.A.; Lombardo, F.; Dominy, B.W.; Feeney, P.J., Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev*, 23:3-25, 1997.
- [2] Anonymous, Waiver of in vivo bioavailability and bioequivalence studies for immediate-release solid oral dosage forms based on a biopharmaceutics classification system. 2000. Available from [http://www.fda.gov/cder/OPS/BCS\\_guidance.htm](http://www.fda.gov/cder/OPS/BCS_guidance.htm)
- [3] Artusson, P., Cell cultures as models for drug absorption across the intestinal mucosa. *Crit Rev Ther Carrier Syst*, 8:305-330, 1991.
- [4] Quaroni, A.; Hochman, J., Development of intestinal cell culture models for drug transport and metabolism studies. *Adv Drug Deliv Rev*, 22:3-52, 1996.
- [5] Hidalgo, I. J.; Raub, T.J.; Borchardt, R.T., Characterization of the human colon carcinoma cell line (Caco-2) as a model system for intestinal permeability. *Gastroenterology*, 96:736-749, 1989.
- [6] Yazdani, M.; Glynn, S.L.; Wright, J.L.; Hawi, A., Correlating partitioning and Caco-2 cell permeability of structurally diverse small molecular weight compounds. *Pharm Res*, 15:1490-1494, 1998.
- [7] Artursson, P.; Karlsson, J., Correlation between oral drug permeability coefficients in human intestinal epithelial (Caco-2) cells. *Biochem Biophys Res Commun*, 175:880-885, 1991.
- [8] Artursson, P.; Palm, K.; Luthman, K., Caco-2 monolayer in experimental and theoretical predictions of drug transport. *Adv Drug Deliv Rev*, 22:67-84, 1996.
- [9] Anderle, P.; Niederer, E.; Rubas, W.; Hilgendorf, C.; Spahn-Langguth, P., P-glycoprotein (P-gp) mediated efflux in caco-2 cell monolayers: The influence of culturing condition and drug exposure on P-gp expression levels. *J Pharm Sci*, 87:757-762, 1998.
- [10] Lentz, K., Hayashi, J.; Polli, J.E., Development of a more rapid culture system for Caco-2 monolayers. *Pharm Sci*. 1:S456, 1998.
- [11] Fujiwara, S.I.; Yamashita, F.; Hashida, M., Prediction of Caco-2 cell permeability using a combination of MO-calculation and neural network. *Int J. Pharm*, 237:95-105, 2002.
- [12] Camenisch, G.; Alsenz, J.; van de Waterbeemd, H.; Folkers, G., Estimation of permeability by passive diffusion through Caco-2 cell monolayer using the drugs' lipophilicity and molecular weight. *Eur J Pharm Sci*, 6:313-319, 1998.
- [13] van de Waterbeemd, H.; Camenisch, G., Estimation of Caco-2 cell permeability using calculated molecular descriptors. *Quant Struct-Act Relat*, 15:480-490, 1996.
- [14] Clark, E.; Pickett, D. S., Computational methods for the prediction of drug-likeness'. *Drug Disc Today*. 5:49-58, 2000.
- [15] Ruovray, D., Taking a short cut to drug design. *New Sci*, 35-38, 1993.
- [16] Marrero, Y.; Romero, V. *TOMOCOMD* software, version 1.0, 2002, Central University of Las Villas. *TOMOCOMD* (*TOP*ological *MO*lecular *COM*puter *Design*) for Windows, version 1.0 is a preliminary experimental version; in future a professional version will be obtained upon request to Y. Marrero: [yovanimp@qf.uclv.edu.cu](mailto:yovanimp@qf.uclv.edu.cu) or [ymarrero77@yahoo.es](mailto:ymarrero77@yahoo.es)
- [17] Marrero, Y., Quadratic indices of the "molecular pseudograph's atom adjacency matrix". Total and local definition and applications to the prediction of physical properties of organic compounds. *Molecules*, 8:687-726, 2003. <http://www.mdpi.org>
- [18] Marrero, Y.; Cabrera, M.A.; Siverio, D.; Romero, V.; Ofori, E.; Montero, L. A. Total and local quadratic indices of the "molecular pseudograph's atom adjacency matrix". Application to prediction of caco-



- 2 permeability of drugs. *Int. J. Mol. Sci*, 4:512-536, 2003. www.mdpi.org/ijms/
- [19] Randi I., M., Generalized molecular descriptors. *J Math Chem*, 7:155-168, 1991.
- [20] Cotton, F. A., *Advanced Inorganic Chemistry*, Ed Revolucionaria, Havana, Cuba, 1970.
- [21] Liang, E.; Chessic, K.; Yazdaniyan, M., Evaluation of an accelerated caco-2 cell permeability model. *J Pharm Sci*, 89:336-345, 2000.
- [22] STATISTICA version. 5.5, StatSoft, Inc. 1999.
- [23] Chaturveldi, P.R.; Deker, C.J.; Odinecs, A. Prediction of pharmacokinetic properties using experimental approaches during early drug discovery. *Curr Opin Chem Biol*, 5:452-463, 2001.
- [24] Chong, S.; Dando, S.A.; Morrison, R. Evaluation of biocoat intestinal epithelium differentiation environment (accelerated cultured caco-2 cells) as an absorption-screening model with improved productivity. *Pharm Res*, 14:1835-1837, 1997.
- [25] Rubas, W.; Jezyk, N.; Grass, G.M. Comparison of the permeability characteristics of a human colonic epithelial (Caco-2) cell line to colon of rabbit, monkey and dog intestine and human drug absorption. *Pharm Res*, 10:113-118, 1993.
- [26] Yee, S. In vitro permeability across caco-2 cells (colonic) can predict in vivo (small intestinal) absorption in man-fact or myth. *Pharm Res*, 14:763-766, 1997.
- [27] Jack, D.B. *Handbook of Clinical Pharmacokinetic Data*, Macmillan Publishers Ltd, pp 25-85, 1992.
- [28] Belsey, D.A.; Kuh, E.; Welsch, R.E., *Regression Diagnostics*. Wiley, New York, 1980.
- [29] van de Waterbeemd, H. and Kansy, M., Hydrogen-bonding capacity permeability using calculated molecular descriptors. *Quant. Struct.-Act Relat*. 15:480-490, 1992.
- [30] Krarup, H.; Christensen, T.I.; Hovgaard, L.; Frokjaer, S., Predicting drug absorption from molecular surface properties based on molecular dynamics simulations. *Pharm Res*. 15:972-978, 1998.
- [31] Norinder, U.; Osterber, T.; Artursson, P., Theoretical calculation and prediction of caco-2 cell permeability using MolSurf parameterization and PLS statistics. *Pharm Res*. 14:1786-1791, 1997.
- [32] Kulkarmi, A.; Han, Y.; Hopfinger, J., Predicting caco-2 cell permeation coefficients of organic molecules using membrane-interaction QSAR analysis. *J Chem Inf Comput Sci*, 42:331-342, 2002.
- [33] Ren, S. and Lien, E.J. Caco-2 cell permeability vs human gastro-intestinal absorption: QSPR analysis. *Prog Drug Res*, 54:3-23, 2000.
- [34] Conradi, R.A.; Buton, P.S.; Borcjhardt, R.T., In Pliska, V.; Testa, B.; van de Waterbeemd, H., (Eds.). *Lipophilicity in Drug Action and Toxicology*, VCH, Weinheim, 223-252, 1996.
- [35] van de Waterbeemd, H.; Smith, D.A.; Jones, B.C., Lipophilicity in pK desing: Methyl, ethyl, futile. *J Comput-Aided Mol Des*, 15:273-286, 2001.
- [36] Stenberg, P.; Luthman, K.; Artursson, P., Virtual screening of intestinal drug permeability. *J Contr Rel*, 65:231-243, 2000.
- [37] Walters, W.P.; Stahl, M.T.; Murcko, M.A., Virtual screening-an overview. *Drug Disc Today*, 3:160-178, 1998.
- [38] Drie, J.H.V. and Lajinees, M.S., Approaches to virtual library design. *Drug Disc Today*, 3:274-283, 1998.
- [39] de Juli'an-Ortiz, J.V.; Gálvez, J.; Muñoz-Collado, C.; García-Domenech, R.; Gimeno-Cardona, C., Virtual combinatorial syntheses and computational screening of new potential anti-herpes compounds. *J Med Chem*, 42:3308-3314, 1999.
- [40] Artursson, P., Epithelial transport of drugs in cell culture. I: A model for studying the passive diffusion of drugs over intestinal absorptive (Caco-2) cells. *J Pharm Sci*, 79:476-482, 1990.
- [41] Haeberlin, B.; Rubas, W.; Nolen III, H.; Friend, D.R., In vitro evaluation of dexamethasone- $\beta$ -D-glucuronide for colon-specific drug delivery. *Pharm. Res*, 10:1553-1562, 1993.
- [42] Hovgaard, L.; Brøndsted, H.; Buur, A.; Bundgaard, H., Drug delivery studies in Caco-2 monolayers. Synthesis, hydrolysis, a transport of O-cyclopropane carboxylic acid ester produgs of various  $\beta$ -blocking agents. *Pharm Res*, 12:387-397, 1995.
- [43] Augustijns, P.; D' Hulst, A.; Daele, J.V.; Kinget, R., Transport of artemisinin and sodium artesnate in Caco-2 intestinal epithelial cell. *J Pharm Sci*. 85:577-579, 1996.
- [44] Collett, A.; Sims, E.; Walker, D.; He, Y.L.; Ayrton, J.; Rowland, M.; Warhurst, G. Comparison of HT29-18-C<sub>1</sub> and Caco-2 cell lines as models for studying intestinal paracellular drug absorption. *Pharm Res*.13:216-221, 1996.
- [45] Walter, E.; Janich, S.; Roessler, B.J.; Hilfinger, J.M.; Amidon, G.L., HT29-MTX/Caco-2 cocultures as an in vitro model for the intestinal epithelium: In vitro-in vivo correlation with permeability data from rats and humans. *J Pharm Sci*, 85:1070-1076, 1996.
- [46] Artursson, P. and Magnusson, C., Epithelial transport of drugs in cells culture. II: Effect of extracellular calcium concentration on the paracellular transport of drugs of different lipophilicities across

- monolayers of intestinal epithelial (Caco-2) cells. *J Pharm Sci.* 79:595-600, 1990.
- [47] Stenberg, P.; Norinder, U.; Luthman, K.; Artursson, P., Experimental and computational screening models for the prediction of intestinal drug absorption. *J. Med. Chem.* 44:1927-1937, 2001.
- [48] Grès, M.; Julian, B.; Bourrié, M.; Meunier, V.; Roques, C.; Berger, M.; Boulenc, X.; Berger, Y.; Fabre, G., Correlation between oral drug absorption in humans and apparent drug permeability in TC-7 Cells, a human epithelial intestinal cell line: Comparison with the parental Caco-2 cell line. *Pharm Res*, 15:726-733, 1998.
- [49] Stewar, B.H.; Chan, O.H.; Lu, R.H.; Reyner, E.L.; Schmid, H.L.; Hamilton, H.W.; Steinbaugh, B.A.; Taylor, M.D. Comparison of intestinal permeabilities determined in multiple in vitro and in situ models: Relationships to absorption in humans. *Pharm Res.* 12 :693-699, 1995.
- [50] Hilgendorf, C.; Spahn-Langguth, H.; Regardh, C.G.; Lipka, E.; Amidon, G.L.; Langguth, P., Caco-2 versus Caco-2/HT29-MTX co-cultured cell lines: Permeability via diffusion, inside-and outside-directed carrier-mediated transport. *J Pharm Sci*, 89:63-75, 2000.
- [51] Augustijns, P.; D' Hulst, A.; Daele, J.V.; Kinget, R., Transport of artemisinin and sodium artesnate in Caco-2 intestinal epithelial cell. *J Pharm Sci.* 85:577-579, 1996.
- [52] Amidon, G.L.; Sinko, P.J.; Fleisher, D., Estimating human oral fraction dose absorbed: a correlation using rat intestinal membrane permeability for passive and carrier-mediated compounds. *Pharm Res.* 5:651-654, 1988.