

Internet Electronic Journal of **Molecular Design**

February 2005, Volume 4, Number 2, Pages 124–150

Editor: Ovidiu Ivanciuc

Prediction of Intestinal Epithelial Transport of Drug in (Caco–2) Cell Culture from Molecular Structure using *in silico* Approaches During Early Drug Discovery

Yovani Marrero Ponce,^{1,2} Miguel A. Cabrera Pérez,² Vicente Romero Zaldivar,³
Marival Bermejo Sanz,⁴ Dany Siverio Mota,¹ and Francisco Torrens⁵

¹ Department of Pharmacy, Faculty of Chemical–Pharmacy. Central University of Las Villas, Santa Clara, 54830, Villa Clara, Cuba

² Department of Drug Design, Chemical Bioactive Center. Central University of Las Villas, Santa Clara, 54830, Villa Clara, Cuba

³ Faculty of Informatics. University of Cienfuegos, Cienfuegos, Cuba

⁴ Department of Pharmacy and Pharmaceutical Technology, University of València, Dr. Moliner 50, E–46100 Burjassot (València), Spain

⁵ Institut Universitari de Ciència Molecular, Universitat de València, Dr. Moliner 50, E–46100 Burjassot (València), Spain

Received: December 12, 2003; Revised: September 26, 2004; Accepted: October 7, 2004; Published: February 28, 2005

Citation of the article:

Y. M. Ponce, M. A. C. Pérez, V. R. Zaldivar, M. B. Sanz, D. S. Mota, and F. Torrens, Prediction of Intestinal Epithelial Transport of Drug in (Caco–2) Cell Culture from Molecular Structure using *in silico* Approaches During Early Drug Discovery, *Internet Electron. J. Mol. Des.* **2005**, 4, 124–150, <http://www.biochempress.com>.

Prediction of Intestinal Epithelial Transport of Drug in (Caco–2) Cell Culture from Molecular Structure using *in silico* Approaches During Early Drug Discovery

Yovani Marrero Ponce,^{1,2,*} Miguel A. Cabrera Pérez,² Vicente Romero Zaldivar,³ Marival Bermejo Sanz,⁴ Dany Siverio Mota,¹ and Francisco Torrens⁵

¹ Department of Pharmacy, Faculty of Chemical–Pharmacy. Central University of Las Villas, Santa Clara, 54830, Villa Clara, Cuba

² Department of Drug Design, Chemical Bioactive Center. Central University of Las Villas, Santa Clara, 54830, Villa Clara, Cuba

³ Faculty of Informatics. University of Cienfuegos, Cienfuegos, Cuba

⁴ Department of Pharmacy and Pharmaceutical Technology, University of València, Dr. Moliner 50, E–46100 Burjassot (València), Spain

⁵ Institut Universitari de Ciència Molecular, Universitat de València, Dr. Moliner 50, E–46100 Burjassot (València), Spain

Received: December 12, 2003; Revised: September 26, 2004; Accepted: October 7, 2004; Published: February 28, 2005

Internet Electron. J. Mol. Des. 2005, 4 (2), 124–150

Abstract

Motivation. The high interest in the prediction of the intestinal absorption for new chemical entities is generated by the increasing rate in the synthesis of compounds by combinatorial chemistry and the extensive cost of the traditional evaluation methods.

Method. Novel molecular descriptors have been applied to estimate the intestinal epithelial transport of drug in Caco–2 cell culture. Total and local (atom and atom–type) quadratic indices used in this study were calculated by TOMOCOMD–CARDD software. Linear Discriminant Analysis (LDA) was used to obtain a quantitative model that discriminates the high absorption compounds ($P \geq 8 \times 10^{-6}$ cm/s) from those with moderate–poor absorption ($P < 8 \times 10^{-6}$ cm/s). A data set of 134 diverse structure drugs and two series of drugs–like compounds (12 compounds) were used as training and test set, respectively. In addition, Multiple Linear Regression (MLR) has been carried out to derive QSPeR models. All statistical analyses were performed with the STATISTICA software package.

Results. The obtained LDA model classified correctly 81.13% of compounds with moderate–poor absorption properties and the 96.30% of compounds with high absorption, showing a global good classification of 90.30% in the training set. The model showed a high Matthews' correlation coefficient (MCC = 0.80). Internal and external validation processes to demonstrate the robustness and predictive power of the obtained model were carried out. In this sense, the model classified correctly 87.31% (MCC = 0.73) in the leave–one–out cross–validation procedure. The discriminant model was also assessed by a 10–fold full cross–validation (removing approximately 13 compounds in each cycle, 85.82% of good classification), yielding a MCC of 0.70. Also this model shown an 87.5, 85.6, 84.7, 85.0, 85.3, 83.5, 84.1, 86.2, 85.9 and 85.9% of global good classification when n varied from 2 to 11 in the leave– n –out cross validation procedure. The model was stabilized around 85.9%

* Correspondence author; E–mail: ymarrero77@yahoo.es.

when n was > 9 . In addition, a data set of 7 HIV protease inhibitors (4 linear peptidomimetic and 3 new cyclic urea) and 5 new 6-fluoroquinolones derivatives was used as external test set. The LDA-QSPeR model achieved a MCC of 0.71 (83.33% correct prediction) in this study. This approach permits us to obtain a good explanation of the experiment based on the molecular structural features, evidencing the main role of H-bonding and size properties in permeability process. Finally, the model developed was used in the virtual screening of 241 drugs with the percentage of human intestinal absorption (Abs %) values reported. A relationship between the predicted permeability coefficients in Caco-2 and the Abs % (145 compounds with good data quality) was established, with a percentage of good relation greater than 82 %. A comparison with results derived from other three theoretical studies shown a quite satisfactory behavior of the present method.

Conclusions. All these results shown that total and local (atom and atom-type) quadratic indices can successfully predict intestinal permeability and suggest that the proposed methodology will be a good tool for studying the oral absorption of drug candidates during the drug development process.

Keywords. Oral drug absorption; Caco-2 cell permeability coefficient; TOMOCOMD-CARDD approach; quadratic indices; QSPeR; quantitative structure-permeability relationships; QSAR; quantitative structure-activity relationships.

1 INTRODUCTION

During the last few years, the role of biopharmaceutical properties, in the drug discovery research, has been increased [1]. Oral bioavailability is one of these biopharmaceutical components that have been widely studied; due to the oral administration is one of the most important routes for its convenience, low cost and high patient compliance rates. The estimation of oral absorption for new drug candidates is very useful in the early stage of the drug discovery process [2,3].

In order to obtain a rapid estimation of human absorption, in high throughput screening (HTS) [4], many cell culture models has been investigated as potential *in vitro* models for drug absorption and metabolism studies [5,6]. Among them, Caco-2 monolayer is the most advanced *in vitro* model due to this cell line expresses several of the biological membrane properties [5-7]. The apical to basolateral permeability coefficient across Caco-2 cell monolayer [$P_{\text{Caco-2}}(\text{AP} \rightarrow \text{BL})$] is increasingly used to estimate oral absorption of new chemical entities (NCEs) [7-10]. Nevertheless, Caco-2 cell models have several disadvantages [11-13], being the long culture periods (21-24 day culture times) the major practical shortcoming of this approximation, with consequently extensive cost. Intestinal permeability can be considered as a predictor of the true fraction absorbed. The theoretical relationship between the fraction of absorbed drug (Fa) and permeability has been described by Amidon *et al.* [14]:

$$\text{Fa} = (1 - e^{-A_{\text{peff}} \times 10^{-6}}) 100[\%] \quad (1)$$

A good correlation between the extent of oral drug absorption in humans and rates of transport across the Caco-2 cell monolayers was obtained by Artursson and Karlsson [8]. However, there are several examples of application of Caco-2 cell models for prediction or correlation with human intestinal absorption where the obtained results for Caco-2 cell permeability coefficients are influenced by the inter-laboratory differences [11,12]. In this sense, in the literature there are several reports about use of Caco-2 cell permeability in the prediction of human absorption. Yazdanian *et al.* [7] reported that compounds with $P_{\text{Caco-2}}$ values less than 0.4×10^{-6} cm/s exhibited

very poor oral absorption, whereas compounds with $P_{\text{Caco-2}}$ values greater than 7×10^{-6} cm/s had excellent oral absorption. In others papers, Artursson *et al.* [8] and Rubas *et al.* [15] reported that compounds with $P_{\text{Caco-2}}$ values over 1×10^{-6} and 70×10^{-6} cm/s were completely absorbed in human, respectively. In addition, Chong *et al.* [16] concluded that compounds with $P_{\text{Caco-2}}$ greater than 1×10^{-6} cm/s would have acceptable absorption in humans (>20%) and Yee [17] determined that only compounds with $P_{\text{Caco-2}}$ greater than 10×10^{-6} cm/s were well-absorbed in humans (70–100%). Based on these criterions Chaturvedi *et al.* [18] has suggested that compounds with the following apparent permeability coefficients: $<1 \times 10^{-6}$ cm/s, $1-10 \times 10^{-6}$ cm/s, $>10 \times 10^{-6}$ cm/s can be classified as poorly (0–20%), moderately (20–70%) and well (70–100%) absorbed drugs.

1.1 Caco–2 Cells, Physic–Chemical Properties for “*in vitro*–Oral Absorption” and development Quantitative structure Permeability Relationships (QSPeR) Studies

Currently, it is known that the oral absorption is influenced by a different kind of interactions. In some studies has been demonstrated that permeability coefficients measured for transport through Caco–2 monolayer cell cultures are correlated with lipophilicity [7,9,19,20], while in others is discussed the role of hydrogen bonding or charge [8,9,11,19]. Waterbeemd *et al.* have used a function, which represent permeability–physicochemical property relationship [19]:

$$\text{Permeability} = f(\text{lipophilicity, molecular size, H-bonding capacity, charge}) \quad (2)$$

However, charge is included in lipophilicity when the distribution coefficient (log D) instead of partition coefficient (log P) is used. Furthermore, molecular size and H–bonding capacity are components of lipophilicity. In this sense, for each property there are limited ranges as is established in the rule–of–5 [2,21].

The tree main reasons for clinical failure of NCEs are lack of efficacy, toxicity and unfavorable pharmacokinetic properties [18]. The significant failure rate of drug candidates in late stage development is driving the need for predictive tools that can eliminate inappropriate compounds before substantial time and money is invested in testing [22,23]. Theoretical approaches appears to be a good alternative to *in silico* prediction of human absorption for new drug candidates obtained by combinatorial chemistry methodologies [24–32]. Therefore, is expected an increasing use in the estimation of absorption parameters from NCEs potentiality actives using quantitative structure property relationship (QSPR) methods during the drug discovered and development process. In the drug discovery research, several computationally calculated properties have been used to assess the oral absorption potential for drug candidates [32–39]. In this sense, some researchers have explored QSPeR studies involving Caco–2 cell permeability. Some types of molecular descriptors have been introduced in these studies, where are including size and hydrogen–bonding descriptors [19], polar surface area (PSA) [11,40,41], Molsurf–derived descriptors [42], MO–calculation [43], and using membrane–interaction analysis [44]. Fujiwara *et al.* [43] obtained a regression coefficient of 0.790

using neural network and correlation coefficients of 0.74 and 0.76, applying linear multiple regressions (LMR). Waterbeemd *et al.* using a representative set of molecular weight and various H-bonding descriptors obtained 13 models applying LMR and PLS analysis. Besides, in this paper these authors obtained a correlation coefficient of 0.88 ($s = 0.52$), where two principal components, were used as variables in the linear regression models [19]. In other paper, Ren and Lien [20] developed a QSAR analysis where an adequate regression coefficient value (0.79), for a data set of 51 compounds, was obtained. Finally, a recently study about prediction of $P_{\text{Caco-2}}$ was carried out by Kulkarni *et al.* [44] where 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 [7] was selected. Almost all these studies were carried out over a limited experimental data.

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 for new lead compounds [45] and have been very useful in elucidating quantitative structure-property and quantitative structure-activity relationships. These studies have become an important area of research in computational chemistry [46,47]. Currently there exist a large number of molecular descriptors that can be used in QSPR studies [48–50]. The topological indices (TIs) are based on the two-dimensional topological structure of molecules and have structural information of the planar molecular, without consideration any physico-chemical molecular feature [51]. Recently, several TIs have been defined and tested in QSAR models [52–63]. In this sense, one of the present authors have developed a novel method called TOPological MOlecular COMputer Design (TOMOCOMD) [64–66]. It calculates several families of topologic molecular descriptors. One of these families has been defined as quadratic indices in analogy to the quadratic mathematical forms. The quadratic indices of the “molecular pseudograph atom adjacent matrix” have been used in QSPR studies [64,65].

Considering the previously mentioned the aims of the present work were: to obtain a classification model, for a large and heterogeneous data set compiled, that permit the identification of molecules with poor-moderate and high absorption from their molecular structure, using total and local quadratic indices as molecular descriptors. In second place, to interpret, in structural terms, the obtained models and to identify the driving force that leading the absorption process through biological membranes. Later, to assess the predictive power of the model found using an external prediction set of 12 compounds and a leave-one, 10-fold, and (n)-out cross-validation procedure. Finally, to simulate a virtual screening experiment with the obtained model, in order to find a relationship between the predicted permeability coefficients in Caco-2 cell and the human absorption.

2 COMPUTATIONAL METHODS

2.1 Molecular Descriptors for QSPeR Analysis

The general principles of the quadratic indices have been explained in some detail elsewhere [64, 65]. However, an overview of this approach will be given. For a given molecule composed of n atoms, the “molecular vector” (\mathbf{X}) is constructed and the k^{th} total quadratic indices, $q_k(x)$ are calculated as quadratic:

$$q_k(x) = \sum_{i=1}^n \sum_{j=1}^n {}^k a_{ij} x_i x_j \quad (3)$$

where n is the number of atoms of the molecule and x_1, \dots, x_n are the coordinates or components of the “molecular vector” (\mathbf{X}) in a system of canonical basis vectors of \mathfrak{R}^n . The components of the “molecular” vector are numeric values, which can be considered as weights (atom–labels) for the vertices of the pseudograph. Certain atomic properties (electronegativity, density, atomic radii, etc) can be used with this propose.

The coefficients ${}^k a_{ij}$ are the elements of the k^{th} power of the symmetric square matrix $\mathbf{M}(G)$ of the molecular pseudograph (G) and are defined as follows:

$$\begin{aligned} a_{ij} &= P_{ij} \text{ if } i \neq j \text{ and } \exists e_k \in E(G) \\ &= L_{ii} \text{ if } i = j \\ &= 0 \text{ otherwise} \end{aligned} \quad (4)$$

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

Eq. (3) for $q_k(x)$ can be written as the single matrix equation:

$$q_k(\mathbf{x}) = \mathbf{X}^t \mathbf{M}^k \mathbf{X} \quad (5)$$

where \mathbf{X} is a column vector (a $nx1$ matrix), \mathbf{X}^t the transpose of \mathbf{X} (a $1xn$ matrix) and \mathbf{M}^k the k^{th} power of the matrix \mathbf{M} of the molecular pseudograph G (mathematical quadratic form matrix).

In addition to total quadratic indices, computed for the whole–molecule, local–fragment (atom and atom–type) formalisms can be developed. These descriptors are termed local quadratic indices, $q_{kL}(x)$ [64,65]. The definition of these descriptors is as follows:

$$q_{kL}(x) = \sum_{i=1}^m \sum_{j=1}^m {}^k a_{ijL} x_i x_j \quad (6)$$

where m is the number of atoms of the fragment of interest and ${}^k a_{ijL}$ is the element of the row “ i ” and column “ j ” of the matrix \mathbf{M}_{L}^k . This matrix is extracted from the \mathbf{M}^k matrix and contains the information referred to the vertices of the specific molecular fragments and also of the molecular

environment.

The matrix $\mathbf{M}^k_L = [{}^k a_{ijL}]$ with elements ${}^k a_{ijL}$ is defined as follows:

$$\begin{aligned} {}^k a_{ijL} &= {}^k a_{ij} \text{ if both } v_i(x_i) \text{ and } v_j(x_j) \text{ are atoms contained within the molecular fragment} \\ &= 1/2 {}^k a_{ij} \text{ if } v_i(x_i) \text{ or } v_j(x_j) \text{ is an atom contained within the molecular fragment but not both} \\ &= 0 \text{ otherwise} \end{aligned} \quad (7)$$

These local analogues can also be expressed in matrix form by the expression:

$$\mathbf{q}_{kL}(x) = \mathbf{X}^t \mathbf{M}^k_L \mathbf{X} \quad (8)$$

Note that the above scheme follows the spirit of a Mulliken population analysis. Also note that for every partitioning of a molecule into Z molecular fragment there will be Z local molecular fragment matrices. In this case, if a molecule is partitioned into Z molecular fragments, the matrix \mathbf{M}^k can be partitioned into Z local matrices \mathbf{M}^k_L , $L = 1, \dots, Z$, and the k^{th} power of matrix \mathbf{M} is exactly the sum of the k^{th} power of the local Z matrices:

$$\mathbf{M}^k = \sum_{L=1}^Z \mathbf{M}^k_L \quad (9)$$

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

$$\mathbf{q}_k(x) = \sum_{L=1}^Z \mathbf{q}_{kL}(x) \quad (10)$$

Atom and atom-type quadratic indices are specific cases of local quadratic indices. In this sense, the k^{th} atom-type quadratic indices are calculated by summing the k^{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, H-bonding to heteroatoms (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^{th} atom-type quadratic indices provide important 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 analogs” is necessary for the evaluation of situations where only one descriptor is unable to bring a good structural characterization [67]. These local indices can also be used together with total indices as variables of QSAR and QSPR models for properties or activities that depend more on a region or a fragment than on the whole molecule.

2.2 TOMOCOMD–CARDD Software

TOMOCOMD is an interactive program for molecular design and bioinformatics research [66]. 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 Modeling 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.

The calculation of total and local quadratic indices for any organic molecule (or any drug–like compounds) was implemented in the TOMOCOMD–CARDD software [66]. The main steps for the application of this method in QSAR/QSPR 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 [68] for each kind of atom,
- (3) Compute the total and local quadratic indices of the molecular pseudograph 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 QSPR/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 a property P and the quadratic indices having, for instance, the following appearance,

$$P = a_0q_0(x) + a_1q_1(x) + a_2q_2(x) + \dots + a_kq_k(x) + c \quad (11)$$

where P is the measurement of the property, $q_k(x)$ [or $q_{kL}(x)$] is the k^{th} total [or local] quadratic indices, and the a_k 's are the coefficients obtained by the linear regression analysis.

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

Table 1. Caco-2 cell permeability for 134 structurally diverse compounds

Compounds	Ref	P _{OBS} ^a	Compounds	Ref	P _{OBS} ^a	Compounds	Ref	P _{OBS} ^a
Acebutolol	69	4.47	Telmisartan	7	15.14	Sulphasalazine	70	0.13
	7	0.51		7	0.47		7	0.3
Acebutolol ester	69	77.62		76	0.97		7	0.3
Acetylsalicylic acid	70	2.4	Terbutaline	70	0.38	Taurocholic acid	17	34.67
	17	30.9		75	1.04	Ibuprophen	17	52.5
	7	9.12	Chlorpromacine	7	19.95	Amiloride	72	0.78
Aciclovir	17	2	Clonidine	17	30.2	Alfentanil	72	310
	7	0.25		7	21.88	Cumarin	9	77.62
Alprenolol	70	40.74	Chlorotiazide	7	0.19	Theophylline	9	44.67
	8	40.74	Corticosterone	70	54.95	Epinephrine	9	0.95
	69	75.86		7	21.38	Guanoxan	9	19.5
	7	25.12		8	54.5	Guanabenz	9	72.44
	71	38.3	Desipramine	17	21.38	Lidocaine	9	61.66
	72	242		7	24.55	Tiactrilast	9	12.59
Alprenolol ester	69	107.15	Dexamethasone	70	12.59	Nitrendipine	9	16.98
Aminopyrine	7	36.31		75	26.92	Fleroxacin	9	15.49
Artemisinin	73	30.4		17	23.44	Diltiazem	9	48.98
Artesunate	73	4		7	12.3	Verapamil	9	26.3
Atenolol	70	0.2	Dexamethasone-β-D-glucoside	75	0.44	Mibefradil	9	13.49
	8	0.2	Dexamethasone-β-D-glucuronide	75	1.15	Bosentan	9	1.05
	74	0.47	Diazepam	17	70.79	Proscillaridin	9	0.63
	7	0.52		7	33.11	Ceftriaxone	9	0.13
	9	0.23		72	756	Remikiren	9	0.74
	73	0.13		7	33.4	Squinavir	9	0.55
	16	4	Dopamine	7	9.33	Olsalazine	70	0.11
	75	1.16	Doxorubicin	17	0.16	Glycine	17	80
	76	0.19	Erithromycin	17	3.72	Amoxicillin	75	0.33
	71	0.2	Estradiol	7	19.95	Enaprilate	75	0.62
	72	1	Felodipine	8	22.91	Lisinopril	75	0.05
Azithromycin	17	1.05	Fluconazole	17	29.51	Gabapentin	77	4.33
Penicillin G	17	1.95	Ganciclovir	15	2.69		77	1.5
Betaxolol	69	95.5		7	0.38	Raffinose	72	0.05
Betaxolol ester	69	95.5		7	0.38	Sildenafil	72	87
Tenidap	17	51.29	Gliseofulvin	7	36.31	Antipyrine	76	47.23
Bremazocine	7	7.94	H216/44	8	0.91		72	2.15
Caffeine	17	50.5		71	1.14	Ciprofloxacin	72	1.9
	7	30.9	Hydrochlorothiazid	74	1.51	Imipramine	17	14.13
	75	84		7	0.51	Indomethacin	7	25.12
	16	21.4	Oxazepam	72	246	Lactulose	72	0.27
Labetalol	7	7	Nordazepam	72	307	Testosterone	8	51.8
Mannitol	70	0.18	Metolazone	72	6.1		7	24.9
	15	3.24	Oxprenolol	69	66.07		75	44.5
	17	0.65		72	120		9	58.88
	7	0.38	Oxprenolol ester	69	97.72	Timolol	69	44.67
	9	2.63	Phencyclidine	7	24.55		7	12.88
	76	0.12	Phenytoin	7	26.92	Timolol ester	69	79.43
	16	0.5	Pindolol	7	16.6	Trovafloxacin	17	30.2
	75	0.17	Pirenzepine	7	0.44	Uracil	7	4.27
	77	0.83	Piroxicam	7	35.48	Urea	7	4.57
Meloxicam	7	19.95		76	46.25	Valproic acid	17	48
Methanol	15	131.83	Practolol	70	0.89	Warfarin	70	38.02
Methotrexate	17	1.2		8	0.89		7	20.89
Methylscopolamine	7	0.79		9	1.38		9	53.7
Metoprolol	70	26.92		71	0.92	Ziduvudine	7	6.92
	8	26.92		72	3.5	Ziprasidone	17	12.3

Table 1. (Continued)

Compounds	Ref	$P_{OBS.}^a$	Compounds	Ref	$P_{OBS.}^a$	Compounds	Ref	$P_{OBS.}^a$
	76	18.8	Prazocin	17	43.65	Cephalexin	73	0.18
	71	26.71	Progesterone	15	78.23	Gly-Pro	76	6.1
	72	92		7	23.7	D-glucose	76	17.53
Nadolol	7	3.89	Propranolol	70	41.69	L-Phenylalanine	76	18.37
	16	4.5		8	41.69	Ketoprofen	76	23.15
Naloxone	17	28.18		69	83.18	Furosemide	76	0.29
Naproxen	15	74.13		17	27.54	Sulpiride	72	0.39
Nevirapine	7	30.2		7	21.88	SB 209670	72	8.8
Nicotine	7	19.5		73	11.2	SB 217242	72	70
D-Phe-L-Pro	17	44.3		71	43.03	Cimetidine	74	0.5
BvaraU	16	4		16	14.8		17	3.09
Pravastatin	16	2.3	Sumatriptan	17	3.02		7	1.37
L-Glutamine	75	0.85	Propranolol ester	69	104.71		76	0.35
SQ-29852	16	0.02	Quinidine	17	20.42	Hydrocortisone	8	21.38
Foscarnet	72	0.05	Ranitidine	7	0.49		15	35.4
Sucrose	7	1.7	Salicylic acid	70	12.02		7	15.85
	75	0.71		7	21.88		77	44.67
Chloramphenicol	17	20.42	Scopolamine	7	11.75		75	12.9

^a Permeability coefficient: $P_{Caco-2}(AP \rightarrow BL) \times 10^{-6}$ cm/s, obtained from diverse source

The following descriptors were calculated:

(1) $q_k(x)$ and $q_k^H(x)$ are the k^{th} total quadratic indices considering and not considering H-atoms in the molecular pseudograph (G), respectively.

(2) $q_{KL}(x_E)$ and $q_{KL}^H(x_E)$ are the k^{th} local (atom-type = heteroatoms: S, N, O) quadratic indices considering and not considering H-atoms in the molecular pseudograph (G), respectively.

(3) $q_{KL}^H(x_{E-H})$ are the k^{th} local (atom-type = H-atoms bonding to heteroatoms: S, N, O) quadratic indices considering H-atoms in the molecular pseudograph (G).

3 PERMEABILITY DATA

A data set of 134 structurally diverse compounds (training set) was compiled from several sources [7–9,15–17,69–77]. Experimental values of $P_{Caco-2}(AP \rightarrow BL)$ are illustrated in Table 1. The data set used for *in silico* permeability studies included compounds with a diverse molecular weight and net charge. Also were included model compounds with to different absorption mechanism. The external test set (prediction set) was also selected from literature [78,79].

4 DATA ANALYSIS

A linear discrimination analysis (LDA) that discriminates the high from the moderate–poor absorbed compounds was employed to develop a simple linear QSPeR between structure and Caco–2 cell permeability coefficients [80,81]. For this purpose, this data set was split into two subsets according to a boundary quantitative value of P_{Caco-2} (8×10^{-6} cm/s). This value of P_{Caco-2}

was fixed taking into consideration the experimental results reported in the literature and the wide inter-laboratory variability [11,12]. In the developing of the classification function the absorption degree was encoded by a dummy variable $f_{(H-M,P)}$. The values of 1 and -1 were assigned to high and moderate-poor absorption compounds, respectively. The statistical analyses were carried out with the STATISTICA software [80]. The quality of the model was determined examining the statistics parameter of multivariable comparison (Wilk's λ statistic, Mahalanobis distance, Fisher ratio F, the corresponding p -level as well as the proportion between the cases and variables in the equation) and several internal and external validation tests.

In this sense, the validation of the model was carried out by a leave-one-out cross-validation procedure. Also a full (10%) cross-validation test of the model was investigated. From the general data set (134 chemicals), 9 groups of 13 observations and 1 group of 17 cases were randomly selected ten times. Each group was left out (leave-group-out, LGO) and that group predicted by the model developed from the remaining observations. In this way, every observation was left out once and its value predicted. Afterward, the classification trees module was used to carry out the leave- n -out cross-validation routine. In addition, to assess the predictive power of the model an external test set of 12 (7 anti-VIH and 5 6-fluoroquinolones derivatives) drugs-like compounds were used [78,79]. Finally, the calculation of percentages of global good classification (accuracy), sensibility, specificity, positive and negative predictive values and of Matthews correlation coefficient in all validation experiments permit us to carry out the assessment of the model.

5 DEVELOPING OF THE DISCRIMINATION FUNCTION

In order to develop the LDA, the data was conformed by 81 compounds with high absorption ($P \geq 8 \times 10^{-6}$ cm/s) and 53 compounds with moderate-poor absorption ($P < 8 \times 10^{-6}$ cm/s). The best discrimination model found, by a forward-stepwise variable selection procedure, together with the statistical parameters, is shown below:

$$\begin{aligned} f_{(H-M,P)} = & 4.1493 + 0.07622q_0(x) - 0.00475q_2(x) - 0.00462q_2^H(x) + 2.08 \times 10^{-7}q_{10}^H(x) \\ & - 0.1099q_{3L}(x_{E-H}) + 0.03735q_{4L}(x_{E-H}) - 9.6 \times 10^{-5}q_{8L}(x_{E-H}) - 0.02427q_{1L}^H(x_E) \end{aligned} \quad (12)$$

$N = 134 \quad \lambda = 0.48 \quad D^2 = 4.518 \quad F(8.125) = 16.878 \quad p < 0.0001$

where N is the number of compounds, λ is Wilk coefficient, F is the Fisher ratio, D^2 is the squared Mahalanobis distance and p -value is the significance level. The Wilks λ parameter can takes values in the range of 0 (perfect discrimination) to 1 (no discrimination) and the Mahalanobis distance indicates the separation between the respective groups. It shows whether the model has an appropriate discriminatory power for differentiating between the two respective groups. The classification of cases was carried out by means of the posterior classification probabilities. Using the Mahalanobis distances to do the classification, we can now derive probabilities. The probability that a case belongs to a particular group is basically proportional to the Mahalanobis distance from

that group centroid. In summary, the posterior probability is the probability, based on our knowledge of the values of others variables, that the respective case belongs to a particular group.

This model classified correctly 81.13% (negative predictive value) of compounds with moderate–poor absorption properties and the 96.30% (positive predictive value) of compounds with high absorption. The global good classification for the data set was 90.30% (accuracy). This model showed a high Matthews' correlation coefficient (MCC) of 0.80. MCC quantified the strength of the linear relation between the molecular descriptors and the classifications [82]. These result and the two most commonly used operating characteristics of “diagnostic” tests (sensitivity and specificity) are depicted in Table 2. In Table 3 are shown the results of classification and a posteriori probabilities for 134 compounds of the training data set.

Table 2. Overall measures of accuracy obtained in the training and predictive sets, and in the jackknife (leave–1, 10 fold, and *n*–out cross–validation) tests for the obtained model

Series	Matthews Corr. Coefficient	Accuracy (%)	Sensitivity (%)	Specificity (%)	Predictive Value (+) (%)	Predictive Value (–) (%)
Training Set	0.80	90.30%	88.64	93.48	96.30	81.13
Predictive Set	0.71	83.33	85.71	80.00	85.71	80.00
Jackknife (cross–validation) tests						
Leave–1–out	0.73	87.31	87.21	87.50	92.59	79.25
Leave–10 fold–out	0.70	85.82	86.9	84.00	90.12	79.25

To assess the predictability of the classification model, Eq. (12), a leave–one–out cross–validation was carried out. This methodology systematically removed one data point at a time from the data set. A discriminant model was then constructed on the basis of this reduced data set and subsequently used to predict the removed data point. This procedure was repeated until a complete set of predicted classification was obtained. Using this approach, the model classified correctly 87.31% (MCC = 0.73) in the leave–one–out cross–validation procedure (see Table 2).

Then, the reliability of the model was tested by a 10–fold full cross–validation test. For each group of observations left out (10% of the whole data set, approximately 13 compounds), a model was developed from the remaining 90% of the data. This process was carried out ten times on ten unique subsets. The statistical results are also depicted in Table 2.

For a more exhaustive testing of the predictive power of the model found a leave–*n*–out cross validation procedures was carried out using the classification tree module. The selected conditions for the validation procedure were the following: 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 were applied to the module of classification tree, the Eq. (12) was obtained and varying the folding parameter of the cross validation, a leave–*n*–out routine can be developed. This model shown an 87.5, 85.6, 84.7, 85.0, 85.3, 83.5, 84.1, 86.2, 85.9 and 85.9% of

global good classification when n varied from 2 to 11 in the leave- n -out cross validation procedures. The model was stabilized around 85.9% when n was > 9 (see Figure 1).

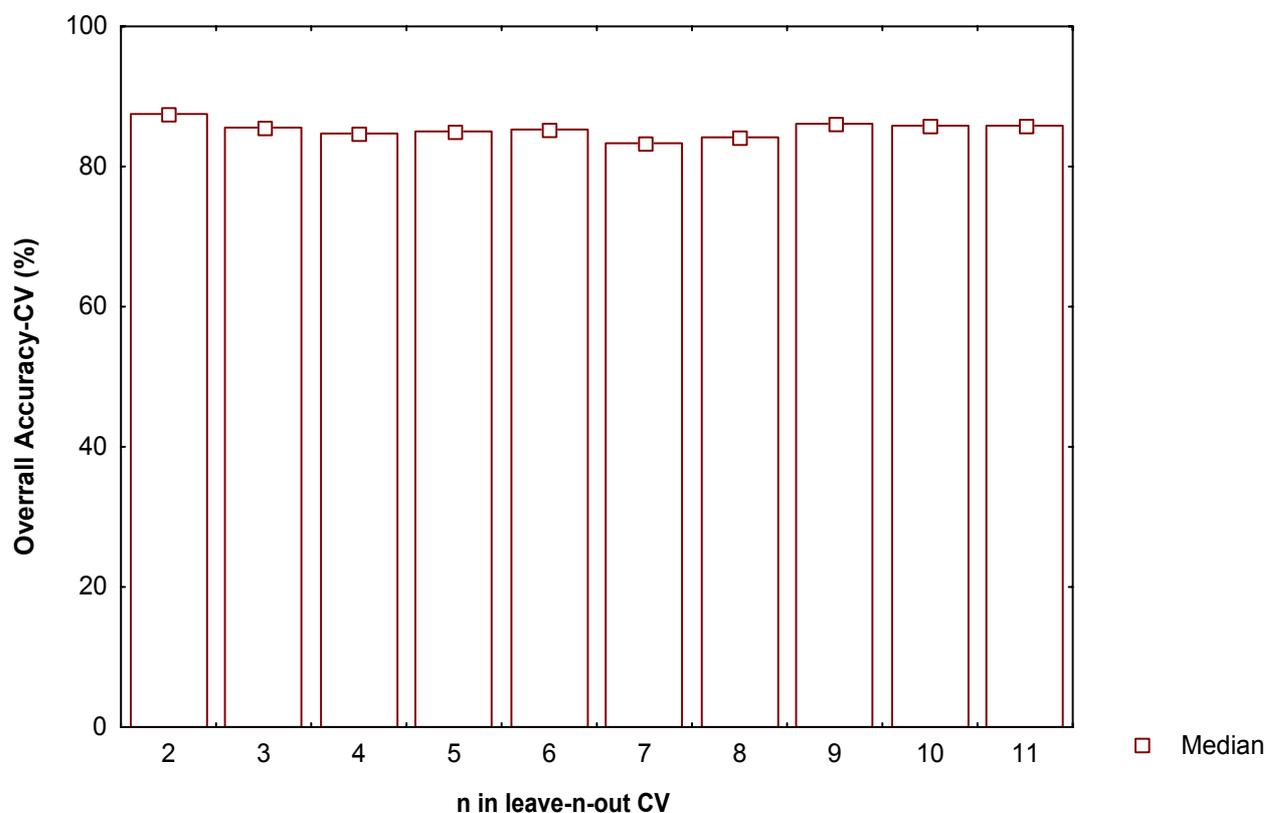


Figure 1. Behavior of the total percentages of good classification in different n -fold cross-validation analysis.

The most important criterion for the quality of the discriminant model is based on the statistics for the prediction set. In this way, the predictive power of the model was also tested through an external test set of 12 drugs-like compounds. In this sense, we have selected two series of compounds, previously reported as antiviral and antimicrobial agents, which will be evaluated by model (12) as high and moderate-poor absorbed compounds. The first set of drugs is composed for 4 linear peptidomimetic and 3 new cyclic urea HIV protease inhibitors [78]. The second set is conformed by 5 new 6-fluoroquinolone derivatives [79]. At the bottom of Table 3 appear the results for the external prediction set. As can be seen, in both series, the predictability and robustness of the theoretical model were demonstrated. Only 2 compounds were bad classified, one in each test group (Amprenavir and CNV 97100 with $P_{\text{Caco-2}}$ of 21.6 and 3.6, respectively). The LDA-QSPeR model achieved a MCC of 0.71 (83.33% correct prediction) in this test set. These values, as well as the sensitivity, specificity and positive and negative predictive values, remained stable in all (leave-1, 10 fold, and n -out cross-validation) Jackknife procedures, a fact that indicates an acceptable level of predictability.

Table 3. Results of the classification of compounds in the data set and the external test set through the discriminant function obtained using total and local quadratic indices as molecular descriptors

Compounds	Prob ^a	Prob-cv ^b	Compounds	Prob ^a	Prob-cv ^b	Compounds	Prob ^a	Prob-cv ^b
Training set								
High absorption group (H)								
Acebutolol ester	0.54	0.52	Meloxicam	0.66	0.61	Valproic acid	0.88	0.87
Acetylsalicylic acid	0.89	0.89	Methanol	0.94	0.94	Warfarin	0.98	0.98
Alprenolol	0.92	0.91	Metoprolol	0.91	0.91	Ziprasidone	0.91	0.90
Alprenolol ester	0.91	0.91	Naloxone	0.94	0.94	D-glucose	0.18*	0.14*
Aminopyrine	0.97	0.97	Naproxen	0.98	0.98	L-Phenylalanine	0.82	0.80
Artemisinin	0.93	0.92	Nevirapine	0.87	0.87	Ketoprofen	0.98	0.98
Betaxolol	0.84	0.84	Nicotine	0.97	0.97	SB 209670	0.96	0.95
Betaxolol ester	0.89	0.88	Oxprenolol	0.83	0.83	SB 217242	0.98	0.98
Bremazocine	0.94	0.94	Oxprenolol ester	0.88	0.88	Sildenafil	0.73	0.70
Caffeine	0.91	0.90	Phencyclidine	0.98	0.98	Oxazepam	0.97	0.96
Chloramphenicol	0.51	0.49*	Phenytoin	0.68	0.65	Nordazepam	0.96	0.96
Chlorpromazine	0.99	0.99	Pindolol	0.69	0.66	Antipyrine	0.98	0.98
Clonidine	0.86	0.86	Piroxicam	0.79	0.76	Alfentanil	0.81	0.78
Corticosterone	0.71	0.69	Prazocin	0.83	0.82	Cumarin	0.99	0.99
Desipramine	0.97	0.97	Progesterone	0.97	0.97	Theophylline	0.75	0.72
Dexamethasone	0.73	0.72	Propranolol	0.96	0.96	Guanoxan	0.70	0.63
Diazepam	0.99	0.99	Propranolol ester	0.97	0.97	Guanabenz	0.82	0.78
Dopamine	0.65	0.63	Quinidine	0.96	0.96	Lidocaine	0.89	0.89
Estradiol	0.93	0.93	Salicylic acid	0.80	0.79	Tiacrilast	0.89	0.88
Felodipine	0.98	0.98	Scopolamine	0.87	0.87	Nitrendipine	0.95	0.95
Fluconazole	0.88	0.87	Taurocholic acid	0.00*	0.00*	Fleroxacin	0.98	0.98
Gliseofulvin	0.99	0.99	Telmisartan	1.00	1.00	Diltiazem	0.97	0.97
Hydrocortisone	0.51	0.49*	Tenidap	0.84	0.83	Verapamil	0.99	0.99
Ibuprofen	0.94	0.94	Testosterone	0.93	0.93	Mibefradil	0.96	0.96
Imipramine	0.99	0.99	Timolol	0.60	0.57	Squinavir	0.68	0.23*
Indomethacin	0.98	0.98	Timolol ester	0.59	0.53	Glycine	0.84	0.81
Labetalol	0.16*	0.11*	Trovafloxacin	0.92	0.91	D-Phe-L-Pro	0.59	0.59
Moderate-poor absorption group (M-P)								
Acebutolol	0.79	0.77	Nadolol	0.37*	0.33*	Lactulose	1.00	0.99
Aciclovir	0.97	0.97	Olsalazine	0.83	0.81	Foscarnet	0.97	0.97
Artesunate	0.08*	0.05*	Pirenzepine	0.27*	0.10*	Ciprofloxacin	0.23*	0.21*
Atenolol	0.58	0.55	Practolol	0.45*	0.43*	Amiloride	0.97	0.94
Azithromycin	0.96	0.94	Ranitidine	0.67	0.62	Epinephrine	0.69	0.66
Penicilina G	0.67	0.66	Sucrose	1.00	1.00	Bosentan	0.67	0.57
Chlorothiazide	1.00	1.00	Sulphasalazine	0.97	0.96	Proscillaridin	0.56	0.51
Cimetidine	0.40*	0.33*	Sumatriptan	0.86	0.85	Ceftriaxone	1.00	0.99
Dexamethasone								
-β-D-glucoside	0.93	0.92	Terbutaline	0.60	0.56	Remikiren	0.99	0.99
Dexamethasone								
-β-D-glucuronide	0.96	0.96	Uracil	0.03*	0.02*	Gabapentin	0.25*	0.11*
Doxorubicin	0.71	0.62	Urea	0.47*	0.38*	BVaraU	0.94	0.94
Erythromycin	0.99	0.99	Ziduvudine	0.99	0.99	Pravastatin	0.88	0.87
Ganciclovir	0.96	0.96	Cephalexin	0.88	0.87	Amoxicillin	0.96	0.96
H216/44	0.22*	0.16*	Gly-Pro	0.63	0.61	Enaprilate	0.51	0.49*
Hidrochlorothiazide	0.97	0.95	Furosemide	1.00	1.00	Lisinopril	0.97	0.96
Mannitol	0.95	0.87	Sulpiride	0.99	0.99	SQ-29852	0.91	0.90
Metthotrexate	1.00	1.00	Raffinose	1.00	1.00	Glutamine	0.82	0.79
Methylscopolamine	0.85	0.75	Metolazone	0.94	0.94			
External test set								
(Virtual Screening Simulation of anti-VIH compounds)								
Compounds	Ref.	P Obs. ^c	Class ^d	Prob ^a	F (%) ^e			
DMP450	78	36.8	H	0.56	NA			
DMP850	78	12.4	H	0.54	NA			

Table 3. (Continued)

Compounds	Ref.	P Obs. ^c	Class ^d	Prob ^a	F (%) ^e	
DMP851	78	5.2	M–P	0.68	NA	
Indinavir	78	6.0	M–P	0.76	65	
Ritonavir	78	3.9	M–P	0.99	60–80	
Nelfinavir	78	3.4	M–P	0.80	20–80	
Amprenavir	78	21.6	M–P	0.68*	NA	
(Virtual Screening Simulation of 6–fluoroquinolones)						
Compounds	Ref.	P Obs. ^c	Class ^d	Prob ^a	P(BL–AP) x10 ⁻⁶ cm/s	Efflux Ratio P _{BL–AP} /P _{AP–BL}
CNV 97100	79	3.6	H	0.81*	14.0	3.88
CNV 97101	79	21.8	H	0.93	25.4	1.17
CNV 97102	79	16.8	H	0.93	21.1	1.25
CNV 97103	79	15.2	H	0.92	17.5	1.15
CNV 97104	79	13.9	H	0.92	17.8	1.28

^a Probability calculated for each subset. ^b Probability calculated for each subset using leave–one–out cross–validation procedure. ^c Permeability coefficient: $P_{\text{Caco-2}}(\text{AP} \rightarrow \text{BL}) \times 10^{-6} \text{ cm/s}$, obtained from references [78,79]. ^d Results of the classification of anti–VIH compounds and 6–fluoroquinolones derivatives obtained from Eq. (12); H: High absorption group ($P \geq 8 \times 10^{-6} \text{ cm/s}$) and M–P: moderate–poor absorption group ($P < 8 \times 10^{-6} \text{ cm/s}$). ^e Oral bioavailability in human (%). Incorrect classifications are marked with (*).

In addition, to demonstrate the true merit of this approach, we developed a direct comparison with other approaches. In connection, three virtual experiments were carried out: (1) with the data set of 17 diverse structural compounds studied by van de Waterbeemd *et al.* [19], where were used the lipophilicity, molecular size (shape), and hydrogen bonding descriptors, (2) with the data set of 30 structurally diverse drugs studied by Kulkarni *et al.* [44], where in this case the membrane–solute interaction descriptors and intermolecular dissolution and solvation descriptors of the solute, were applied and, (3) with the data set of 51 diverse compounds studies by Ren and Lien [20] where several physicochemical properties were introduced. The regression results for the different data sets, using the total and local quadratic indices, and the comparison with the reported models of $P_{\text{Caco-2}}$ are given in Table 4.

In these equations R is the correlation coefficient, R_{CV} the cross–validated (leave–one–out) correlation coefficient, s the standard deviation of the regression, and F a measure for the statistical significance of the regression models. The statistical parameters show a high statistical quality of the developed models using total and local quadratic indices. In the first virtual experiment, the statistical parameters of the Eq. (13) are higher than obtained by van de Waterbeemd *et al.* [19] whose used two principal component as variables for the description of $\text{Log } P_{\text{Caco-2}}$. In the second comparison, the statistics parameters obtained with our approach [Eq. (15) and Eq. (16)] are lightly lower than obtained by Kulkarni *et al.* [44]. Finally, the obtained models with our approach for the same data set of 51 compounds reported by Ren and Lien [20] [Eq. (17) and Eq. (18)] and when these compounds were splitted into three subgroups, namely neutral [NC: Eq. (19) and Eq. (20)], cationic [CC: Eq. (21) and Eq. (22)] and anionic [AC: Eq. (23)] compounds had better statistical parameters than obtained by these authors.

Table 4. Regression results for different data sets, using the total and local quadratic indices, and the comparison with the reported models of $P_{\text{Caco-2}}$

Eq. or source	Models	<i>n</i>	<i>R</i>	<i>R_{CV}</i>	<i>s</i>	<i>F</i>
First virtual experiment: 17 structural diverse compounds [19]						
Eq. (13)	$\text{Log } P_{\text{Caco-2}} = -4.553 - 0.00247^{\text{H}} q_{5\text{L}}^{\text{H}}(x) + 0.001618 q_0^{\text{H}}(x)$	17	0.920	0.860	0.436	38.692
van de Water–beemd <i>et al.</i> [19]	Eq. (4) in Ref. [19]	17	0.884	0.836	0.520	25.2
Second virtual experiment: 30 structural diverse compounds [44]						
Eq. (15)	$P_{\text{Caco-2}} = 31.2842 - 0.72^{\text{H}} q_{0\text{L}}^{\text{H}}(x) - 5.7841 \times 10^{-3} q_{5\text{L}}^{\text{H}}(x) + 0.017693^{\text{E}} q_{4\text{L}}^{\text{E}}(x)$	30	0.877	0.841	4.974	28.737
Eq. (16)	$P_{\text{Caco-2}} = 31.42089 - 1.5847^{\text{H}} q_{0\text{L}}^{\text{H}}(x) - 5.438 \times 10^{-3} q_{5\text{L}}^{\text{H}}(x) + 0.01617^{\text{E}} q_{4\text{L}}^{\text{E}}(x) + 0.12^{\text{H}} q_{3\text{L}}^{\text{H}}(x)$	30	0.887	0.839	4.874	22.959
Kulkarni <i>et al.</i> [44]	Eq. (4) in Ref. [44]	30	0.894	0.860	–	–
Kulkarni <i>et al.</i> [44]	Eq. (5) in Ref. [44]	30	0.905	0.866	–	–
Third virtual experiment: Full Data Set (51 structural diverse compounds) [20]						
Eq. (17)	$\text{Log } P_{\text{Caco-2}} = 1.849 - 3.326 \times 10^{-5} q_7^{\text{H}}(x) + 7.48 \times 10^{-4} q_{3\text{L}}(x) - 0.031^{\text{E}} q_{1\text{L}}(x) + 1.906 \times 10^{-3} q_{3\text{L}}^{\text{E}}(x) - 1.309 \times 10^{-4} q_{7\text{L}}(x) + 1.485 \times 10^{-8} q_{14\text{L}}(x)$	51	0.800	0.761	0.471	13.068
Eq. (18)	$\text{Log } P_{\text{Caco-2}} = 1.907 + 4.12 \times 10^{-6} q_7^{\text{H}}(x) + 9.423 \times 10^{-4} q_{3\text{L}}(x) - 0.0387^{\text{E}} q_{1\text{L}}(x) + 2.25 \times 10^{-3} q_{3\text{L}}^{\text{E}}(x) - 1.12 \times 10^{-4} q_{7\text{L}}(x) + 1.478 \times 10^{-8} q_{14\text{L}}(x)$	49	0.849	0.75	0.414	18.107
Ren and Lien [20]	Eq. (6) in Ref. [20]	51	0.797	–	0.465	19.98
Ren and Lien [20]	Eq. (8) in Ref. [20]	50	0.749	–	0.506	14.40
Split Data Set [20]						
Eq. 19 (NC)	$\text{Log } P_{\text{Caco-2}} = 0.809 - 0.0224^{\text{H}} q_{2\text{L}}^{\text{H}}(x) + 1.222 \times 10^{-3} q_1^{\text{H}}(x)$	17	0.915	0.854	0.324	35.768
Eq. 20 (NC)	$\text{Log } P_{\text{Caco-2}} = 1.335 - 0.0718^{\text{H}} q_{2\text{L}}^{\text{H}}(x) + 0.0183^{\text{H}} q_{3\text{L}}^{\text{H}}(x) + 1.1069 \times 10^{-3} q_1^{\text{H}}(x) - 4.202 q_2(x)$	17	0.967	0.894	0.219	43.830
Ren and Lien [20]	Eq. (11) in Ref. [20]	17	0.968	–	0.217	44.54
Eq. (21) (CC)	$\text{Log } P_{\text{Caco-2}} = 1.0249 - 0.137^{\text{E}} q_{0\text{L}}^{\text{H}}(x) + 2.875 \times 10^{-3} q_1(x) + 0.0997^{\text{E}} q_{0\text{L}}(x)$	26	0.833	0.784	0.438	16.618
Eq. (22) (CC)	$\text{Log } P_{\text{Caco-2}} = 1.463 - 0.0184^{\text{E}} q_{1\text{L}}(x) + 4.315 \times 10^{-8} q_{10\text{L}}(x) + 7.416 \times 10^{-3} q_1(x) - 3.367 \times 10^{-3} q_0^{\text{H}}(x) - 1.433 q_2(x) - 7.272 \times 10^{-3} q_{2\text{L}}(x)$	23	0.936	0.894	0.299	19.053
Ren and Lien [20]	Eq. (17) in Ref. [20]	26	0.901	–	0.352	22.60
Ren and Lien [20]	Eq. (16) in Ref. [20]	25	0.915	–	0.325	35.87
Eq. (23) (AC)	$\text{Log } P_{\text{Caco-2}} = 1.186 - 0.115^{\text{H}} q_{3\text{L}}(x) + 0.359^{\text{H}} q_{2\text{L}}(x)$	8	0.975	0.94	0.172	48.591
Ren and Lien [20]	Eq. (20) in Ref. [20]	8	0.931	–	0.284	16.38

6 DRIVING FORCES FOR PERMEABILITY OF DRUG IN CACO–2 CELL

As can be observed, in the discriminant model, the included variables are very close to the factors that influence on the permeability values [see Eq (2)]. These factors are related with the structural features of molecules. For example, in Eq. (12), the variables $q_{3\text{L}}(x_{\text{E-H}})$, $q_{4\text{L}}(x_{\text{E-H}})$ and $q_{8\text{L}}(x_{\text{E-H}})$ are connected with the hydrogen atoms as donors, while the $q_{1\text{L}}^{\text{H}}(x_{\text{E}})$ variable contain information about the number of hydrogen acceptors and the charge of molecules. All of them are related with the total hydrogen bond capacity. If the total contribution of these descriptors is analyzed, the obtained values are negative (not shown data) which are logical results due to when the number of heteroatoms and the hydrogen atom bond to heteroatoms in the molecules is increased, the permeability across the biological membrane decrease. This effect is very close with the molecule lipophilicity decrease and the possibility of molecule ionization. The charge factor is related with the negative charge of biological membrane [83]. This observation is supported in a study developed by Ren *et al.* [20]. Firstly, a low regression coefficient ($R = 0.749$) was evidenced when anionic, cationic and neutral compounds (the full set), using the net charge of molecules like a descriptor, were studied. Once these compounds were separated into three subgroups (neutral, cationic and anionic compounds), much better correlation coefficients ($R = 0.968$, 0.915 and 0.931 , respectively) were obtained [20]. Other descriptor with a positive contribution to the permeability

coefficient is $q_0(x)$. This variable contains information about the molecular weight and consequently of the molecule size. 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. Taking into consideration the above mentioned approach, should be considered that successful drug candidates will be characterized by an optimal range of values for H-bonding, lipophilicity and size [84] and for this reason, compounds with extreme positive values of these properties could have a marked negative effect on permeability across biological membrane [19].

7 VIRTUAL SCREENING

Several QSAR studies have shown their importance in the prediction of human intestinal absorption [32,85-88] and the so-called rule-of-5 has proved very popular as a rapid screen for compounds that are likely to be poorly absorbed [2]. In the present study was simulated a virtual search to predict the absorption profile of 241 compounds [88], using the discriminant function [Eq. (12)] obtained. The aim of this approach is to evaluate the capacity of human absorption prediction from the classification model, in high and moderate-poor, for drug absorption in Caco-2 cells [Eq. (12)]. In this sense, some compounds included in the model obtaining (training or test set) were also used in this screening. As the compounds selected for the virtual screening were obtained from different sources, only the first 145 compounds (data of best quality, classified as OK and Good by Abraham *et al.* [88]) should be used to bring a better comparative criterion, nevertheless the rest of the compounds can be evaluated but their human absorption values (Abs %) were not comparatively reliable [88]. These experimental values and the evaluation results of these compounds are depicted in Table 5. In this Table we give the posterior probabilities values calculated from the Mahalanobis distance [P(H) and P(M-P)] for each compounds; $\Delta P\% = [P(H) - P(M-P)] \times 100$, where P(H) is the probability that the equation classifies a compound with $P_{Caco-2} \geq 8 \times 10^{-6}$ cm/s. Conversely, P(M-P) is the probability that the model classifies a compound with $P_{Caco-2} < 8 \times 10^{-6}$ cm/s. This values ($\Delta P\%$) takes positives values when $P(H) > P(M-P)$ and negative, otherwise. Therefore, when $\Delta P\%$ is positive (negative) the compounds were classified with High (Moderate-Poor) absorption profile.

As can be seen in Table 5, of the 123 compounds with high human absorption [Abs % ≥ 70 ; see column eighth with absorption data (or average values) chosen from literature], the Caco-2 cell model, Eq. (12), classified 103 in the high absorption group ($P_{Caco-2} \geq 8 \times 10^{-6}$ cm/s) for a 83.74% of correspondence between the *in vitro* values of permeability (Caco-2) and the reported human absorption values. For the 22 compounds with moderate-poor human absorption (Abs % < 70), 17 compounds were well classified, Eq. (12), for a 77.27% of correspondence with the human reported data. The global percentage of good extrapolation (from *in vitro* to *in vivo*) was 82.76%.

Table 5. Results of the virtual screening of 241 drugs. Permeability coefficient from model [Eq. (12)] and observed human absorption and bioavailability from literature

Compounds	Prob H ^a	Prob M–P ^b	ΔP% ^c	%Abs. ^d	%Abs. ^e	%Bio. ^f	%Abs. ^g
1–Cisapride	0.65	0.35	30.20	100		100	100
2–Valproic acid	0.88	0.12	75.35	100	~100	90(68–100)	100
3–Salicylic acid	0.80	0.20	60.62	100	100		100
4–Diazepam	0.99	0.01	97.79	97–100	100		100
5–Sudoxicam	0.67	0.33	33.14		100		100
6–Glyburide	0.02	0.98	–96.46				100
7–Gallopamil	0.99	0.01	98.20		~100	15	100
8–Mexiletine	0.97	0.03	94.07		100	88	100
9–Nefazodone	0.93	0.07	85.33		100	15–23	100
10–Naproxen	0.98	0.02	96.25	94–99	100	99	99
11–Lamotrigine	0.78	0.22	56.90	70		98	98
12–Tolmesoxide	0.99	0.01	98.03	100		85	98
13–Disulfiran	0.91	0.09	82.59		91		97
14–Torasemide	0.03	0.97	–94.36			96	96
15–Metoprolol	0.91	0.09	82.94	95–100	>90	50	95
16–Naloxone	0.00	1.00	–100.00				91
17–Terazocin	0.79	0.21	57.18	91	~100	90	90
18–Sulindac	0.00	1.00	–100.00		90		90
19–Sultopride	0.82	0.18	63.56	100	~100		89
20–Topiramate	0.00	1.00	–99.12			81–95	86
21–Tolbutamide	0.09	0.91	–82.98				85
22–Propiverine	0.99	0.01	97.96		84		84
23–Digoxin	0.02	0.98	–96.95			67	81
24–Mercapto ethane sulfonic acid	0.75	0.25	50.49				77
25–Cimetidine	0.60	0.40	20.22	62–98		60	64
26–Furosemide	0.00	1.00	–99.26	61	61	61	61
27–Metformin	0.40	0.60	–19.57			50–60	53
28–Rimiterol	0.41	0.59	–17.69				48
29–Cymarin	0.46	0.54	–8.44		47		47
30–Ascorbic Acid	0.80	0.20	59.87				35
31–Fosfomycin	0.08	0.92	–84.11				31
32–Fosmidomycin	1.00	0.00	99.91		30		30
33–k–Strophanthoside	0.99	0.01	98.20		16		16
34–Adefovir	0.00	1.00	–99.57	12		12	16
35–Acarbose	0.00	1.00	–100.00		1–2.		2
36–Ouabain	0.97	0.03	94.07		1.4		1.4
37–Kanamycin	0.00	1.00	–99.99				1
38–Lactulose	0.00	1.00	–99.53	0.6	0.6		0.6
39–Camazepan	0.99	0.01	98.53	99		100	100
40–Indomethacin	0.98	0.02	96.62	100		100	100
41–Levomorgestrel	0.97	0.03	93.03			100	100
42–Tenoxicam	1.00	0.00	99.91			100	100
43–Theophylline	0.75	0.25	49.05	96		100	100
44–Oxatomide	0.91	0.09	81.18	100			100
45–Desipramine	0.97	0.03	93.61	95–100	>95	40	100
46–Fenclofenac	0.97	0.03	93.22	100			100
47–Imipramine	0.99	0.01	98.67	95–100	>95	22–67	100
48–Lormetazepan	1.00	0.00	99.26	100	100	80	100
49–Diclofenac	0.91	0.09	81.58	100		90	100
50–Granisetron	0.92	0.08	84.75	100	100		100
51–Testosterone	0.93	0.07	86.27	100	100		100
52–Caffeine	0.91	0.09	82.30	100	100		100
53–Corticosterone	0.71	0.29	41.91	100	100		100
54–Ethinyl estradiol	0.97	0.03	93.07	100	~100	59	100
55–Isoxicam	0.73	0.27	45.81		100		100

Table 5. (Continued)

Compounds	Prob H ^a	Prob M–P ^b	ΔP% ^c	%Abs. ^d	%Abs. ^e	%Bio. ^f	%Abs. ^g
56–Lornoxicam	1.00	0.00	99.90		100		100
57–Nicotine	0.97	0.03	94.91	100	100		100
58–Ondansetron	0.99	0.01	97.89	100	100	60	100
59–Piroxicam	0.79	0.21	58.00	100	100		100
60–Verapamil	0.99	0.01	98.20	100	>90	10–52.	100
61–Progesterone	0.97	0.03	94.07	91–100	91		100
62–Stavudine	0.29	0.71	–41.62			100	100
63–Toremifene	1.00	0.00	99.89			100	100
64–Cypoterone acet.	0.99	0.01	98.19			100	100
65–Praziquantel	0.96	0.04	92.64		100		100
66–Cicaprost	0.69	0.31	38.42		100		100
67–Aminopyrine	0.97	0.03	93.79	100			100
68–Nordazepam	0.96	0.04	92.58	99		99	99
69–Carfencillin	0.06	0.94	–87.04	100			99
70–Prednisolone	0.61	0.39	21.45	99		70–100	99
71–Propranolol	0.96	0.04	91.58	90–100	>90	30	99
72–Viloxazine	0.91	0.09	82.45	100	~100	61–98	98
73–Warfarin	0.98	0.02	96.88	98	~100	93–98	98
74–Atropine	0.88	0.12	76.86		90		98
75–Minoxidil	0.78	0.22	55.17		95		98
76–Clofibrate	0.98	0.02	95.78	96		95–99	97
77–Trimethoprim	0.75	0.25	50.53	97		92–102	97
78–Venlafaxine	0.95	0.05	90.49	92			97
79–Antipyrine	0.98	0.02	95.26	100	~100	97	97
80–Bumetanide	0.01	0.99	–98.81	100	100	~100	96
81–Trapidil	0.92	0.08	84.09			96	96
82–Fluconazole	0.88	0.12	76.60	95–100		>90	95
83–Sotalol	0.05	0.95	–89.61	95	~100	90–100	95
84–Codeine	0.98	0.02	96.99	95		91	95
85–Flumazenil	0.98	0.02	97.00	95	>95	16	95
86–Ibuprofen	0.94	0.06	88.38	100			95
87–Labetalol	0.16	0.84	–68.70	90–95	>90	33	95
88–Oxprenolol	0.83	0.17	66.92	97	90	50	95
89–Practolol	0.55	0.45	10.51	95	~100		95
90–Timolol	0.60	0.40	19.08	72	>90	75	95
91–Alprenolol	0.92	0.08	83.35	93–96	>93		93
92–Amrinone	0.74	0.26	48.93		93		93
93–Ketoprofen	0.98	0.02	95.23	100	~100	>92	92
94–Hydrocortisone	0.51	0.49	2.85	89–95	84–95		91
95–Betaxolol	0.84	0.16	68.82	90	90	80–89	90
96–Ketorolac	0.98	0.02	96.14	100	Well	80–100	90
97–Meloxicam	0.66	0.34	31.42	90		90	90
98–Phenytoin	0.68	0.32	35.01	90	90	90	90
99–Amphetamine	0.89	0.11	77.18				90
100–Chloramphenicol	0.51	0.49	1.33	90		80	90
101–Felbamate	0.09	0.91	–82.38		90–95	102	90
102–Nizatidine	0.29	0.71	–41.42	99		>90	90
103–Alprazolam	0.99	0.01	98.83			80–100	90
104–Tramadol	0.96	0.04	91.50			65–75	90
105–Nisoldipine	0.96	0.04	92.96				89
106–Oxazepam	0.97	0.03	93.08	97	~100	92.8	89
107–Tenidap	0.84	0.16	68.23	90		89	89
108–Dihydrocodeine	0.99	0.01	97.80			20	88
109–Felodipine	0.98	0.02	96.10	100	100	16	88
110–Nitrendipine	0.95	0.05	89.75			23	88
111–Saccharin	0.52	0.48	3.68	97	88		88

Table 5. (Continued)

Compounds	Prob H ^a	Prob M–P ^b	ΔP% ^c	%Abs. ^d	%Abs. ^e	%Bio. ^f	%Abs. ^g
112–Mononidine	0.63	0.37	25.07			88	87
113–Bupropion	0.92	0.08	84.73	87		87	87
114–Pindolol	0.69	0.31	37.54	92–95	>90	87	87
115–Lamivudine	0.12	0.88	–76.60			86–88	85
116–Morphine	0.96	0.04	91.17	100	~100	20–30	85
117–Lansoprazole	1.00	0.00	99.89			85	85
118–Oxyfedrine	0.90	0.10	80.63			85	84
119–Captopril	0.51	0.49	1.18	77	71	62	84
120–Bromazepam	0.95	0.05	89.68	84		84	84
121–Acetylsalicylic acid	0.89	0.11	78.05				82
122–Sorivudine	0.06	0.94	–88.69	82	82	61	82
123–Methylprednisolone	0.01	0.99	–97.13	82		82	82
124–Mifobate	0.90	0.10	80.95				81
125–Flecainide	0.94	0.06	88.15			81	81
126–Quinidine	0.96	0.04	92.41	80	81	81	81
127–Piroximone	0.58	0.42	15.56			81	80
128–Acebutolol	0.21	0.79	–58.53	90	90	50	80
129–Ethambutol	0.23	0.77	–53.61		75–80		80
130–Acetaminophen	0.80	0.20	60.27	80–100	80	68.95	80
131–Dexamethasone	0.73	0.27	46.17	92–100		80	80
132–Guanabenz	0.82	0.18	63.42	75			80
133–Isoniazid	0.60	0.40	19.97				80
134–Omeprazole	0.92	0.08	84.85				80
135–Methadone	0.99	0.01	98.82			80	80
136–Fanciclovir	0.49	0.51	–2.53			77	77
137–Metolazone	0.06	0.94	–88.71	64	62–64		64
138–Fenoterol	0.43	0.57	–13.23		60		60
139–Nadolol	0.63	0.37	26.80	20–35	34	34	57
140–Atenolol	0.42	0.58	–16.90	50–54	50	50	50
141–Sulpiride	0.01	0.99	–98.62	36		30	44
142–Metaproterenol	0.56	0.44	11.12		44	10	44
143–Famotidine	0.00	1.00	–99.95			37–45	28
144–Foscarnet	0.03	0.97	–93.98	17	17(12–22)		17
145–Cidofovir	0.32	0.68	–35.32			<5	3
146–Isradipine	0.97	0.03	94.44	92	90–95	17	92
147–Terbutaline	0.40	0.60	–20.16	60–73	50–73	16	62
148–Reproterol	0.08	0.92	–84.97		60		60
149–Lincomycin	0.04	0.96	–92.92		20–35		28
150–Streptomycin	0.00	1.00	–100.00		poor		1
151–Fluvastatin	0.95	0.05	90.94	100	>90	19–29	100
152–Urapidil	0.70	0.30	40.00			68	78
153–Propylthiouracil	0.67	0.33	34.07	75		76(53–88)	76
154–Recainam	0.33	0.67	–33.16				71
155–Cycloserine	0.83	0.17	66.46				73
156–Hydrochlorothiazide	0.03	0.97	–93.76	67–90	65–72		69(65–72)
157–Pirbuterol	0.21	0.79	–57.49				60
158–Sumatriptan	0.14	0.86	–72.95	55–75	>57	14	57
159–Amiloride	0.03	0.97	–93.08				50
160–Mannitol	0.05	0.95	–89.03	16–26			16
161–Ganciclovir	0.04	0.96	–92.95	3–3.8	3	3	3
162–Neomycin	0.00	1.00	–100.00				1

Table 5. (Continued)

Compounds	Prob H ^a	Prob M–P ^b	ΔP% ^c	%Abs. ^d	%Abs. ^e	%Bio. ^f	%Abs. ^g
163–Raffinose	0.00	1.00	–99.98	0.3			0.3
164–Phenglutarimide	0.83	0.17	65.50	100			100
165–Bornaprine	0.79	0.21	58.67	100			100
166–D–Phe–L–Pro	0.59	0.41	18.69	100			100
167–Scopolamine	0.87	0.13	74.35	90–100			95
168–Naloxone	0.94	0.06	87.42	91			91
169–Ziprasidone	0.91	0.09	81.69	60			60
170–Guanoxan	0.70	0.30	40.08		50		50
171–Netivudine	0.13	0.87	–74.27		28		28
172–Gentamicin–C1	0.00	1.00	–99.90	0	poor		poor
Zwitterionic drugs							
173–Cefadroxil	0.04	0.96	–91.09			100	100
174–Ofloxacin	0.98	0.02	95.45			100	100
175–Pefloxacin	0.99	0.01	97.61			100	100
176–Cephalexin	0.12	0.88	–75.82	98	100		100
177–Loracarbef	0.22	0.78	–56.73	100	100		100
178–Glycine	0.84	0.16	67.41	100			100
179–Amoxicillin	0.04	0.96	–92.76	94		93	93
180–Tiagabine	0.92	0.08	83.85			90	90
181–Telmisartan	1.00	0.00	99.28	90	rapid	43	90
182–Trovafoxacin	0.92	0.08	83.44	88		88	88
183–Acrivastine	0.97	0.03	94.20	88			88
184–Nicotinic acid	0.88	0.12	75.94				88
185–Levodopa	0.36	0.64	–27.61	100	80–90	86	86
186–Cefatrizine	0.01	0.99	–98.51			75	75
187–Ampicilin	0.10	0.90	–80.30				62
188–Vigabatrin	0.66	0.34	32.52				58
189–Tranexamic acid	0.54	0.46	8.32	55			55
190–Eflurnithine	0.95	0.05	89.86				55
191–Metyldopa	0.50	0.50	0.16		41		41
192–Ceftriaxone	0.00	1.00	–99.77	1	1		1
193–Distigminebromide	0.87	0.13	74.89			47	8
194–Ziduvudine	0.01	0.99	–98.72	100	100	63	100
195–Ximoprofen	0.80	0.20	59.41	100		98	98
196–Clonidine	0.86	0.14	72.24	85–100	100	75–95	95
197–Viomycin	0.00	1.00	–100.00				85
198–Ceftizoxime	0.01	0.99	–97.70				72
199–Capreomycin	0.00	1.00	–100.00				50
200–AAFC	0.46	0.54	–7.46		32		32
201–Bretvlum tosilate	0.98	0.02	96.45	23		23	23
Dose-limited, dose-dependent, and formulation-dependent drugs							
202–Spironolactone	0.93	0.07	87.00		>73		73
203–Etoposide	0.83	0.17	65.54	50		50(25–75)	50(25–75)
204–Cefetamet pivoxil	0.02	0.98	–95.34			47	47
205–Cefuroximeaxetil	0.15	0.85	–70.25	36		36–58	44(36–52)
206–Azithromycin	0.04	0.96	–91.43	35–37		37	37
207–Fosinopril	0.48	0.52	–3.48		36	25–29	36
208–Pravastatin	0.12	0.88	–75.90	34	34	18	34
209–Cyclosporin	0.00	1.00	–99.99	35		10–60	28(10–65)
210–Bromocriptine	0.18	0.82	–63.87	28	28	6	28
211–Doxorubicin	0.29	0.71	–42.87	5	trace	5	12(0.7–23)

Table 5. (Continued)

Compounds	Prob H ^a	Prob M–P ^b	ΔP% ^c	%Abs. ^d	%Abs. ^e	%Bio. ^f	%Abs. ^g
212–Cefuroxime	0.01	0.99	–98.58				1
213–Iothalamate sodium	0.15	0.85	–69.64	1.9	1.9		1.9
214–Sulphasalazine	0.03	0.97	–94.60	12–13.			59(56–61)
215–Benazepril	0.72	0.28	43.63	37	>37		≥37
216–Lisinopril	0.03	0.97	–93.39	25	25	25–50	28(25–50)
217–Esalaprilat	0.19	0.81	–62.46	9–10.	10–40.		25(10–40)
218–Anfotericina	0.00	1.00	–100.00	5	poor		3(2–5)
219–Aztreonam	0.00	1.00	–99.96		<1	<1	1
220–Mibefradil	0.96	0.04	92.12			37–109	69(37–100)
221–Ranitidine	0.33	0.67	–34.45	50–61		50(39–88)	64(39–88)
222–Chlorotiazide	0.00	1.00	–99.32	13–56			49(36–61)
223–Aciclovir	0.03	0.97	–94.91	20–30		15–30	23(15–30)
224–Norfloxacin	0.77	0.23	53.66	35	30–40	~70	71
225–Metthotrexate	0.00	1.00	–99.35	20–100	100		70(53–83)
226–Gabapentin	0.75	0.25	50.14	50	well	60A(36–64)	59(43–64)
227–Prazocin	0.83	0.17	66.16	100		44–69	86(77–95)
228–Olsalazine	0.17	0.83	–65.21	2.3		2.3	24(17–31)
Drugs expected to have higher absorption							
229–Ciprofloxacin	0.77	0.23	53.94	69–100		69	≥69
230–Ribavirin	0.02	0.98	–95.20			33	≥33
231–Pafenolol	0.21	0.79	–57.69			28	≥29
232–Azosemide	0.00	1.00	–99.54			10	≤10
233–Xamoterol	0.12	0.88	–75.51			5	≥5
234–Enalapril	0.50	0.50	0.99	66	60–70	29–50	66(61–71)
235–Phenoxymethyl penicillin	0.19	0.81	–61.16	45	45(31–60)		59(49–68)
236–Gliclazide	0.08	0.92	–84.02				≥65
237–Benzylpenicillin	0.33	0.67	–34.22	30	15–30		≥30
238–Thiacetazone	0.38	0.62	–23.02				≥20
239–Lovastatin	0.86	0.14	72.39	30	31		≥10
240–Cromolym sodium	0.55	0.45	10.94				≥0.4
241–Erythromycin	0.03	0.97	–93.89	35		35	≥35

Results of the classification (Probability calculated for each subset) of compounds obtained from Eq. (12); ^a H: High absorption group ($P \geq 8 \times 10^{-6}$ cm/s), and ^b M–P: moderate–poor absorption group ($P < 8 \times 10^{-6}$ cm/s). ^c ΔP% = [P(High absorption group) – P(moderate–poor group)]x100. ^d The data used for QSAR studies was taken from Clark [85] and Wessel [32], Palm [86], Yazdaniyan [7], Yee [17], and Chiou [87]. ^e Absorption data obtained from the original and review literature. ^f Bioavailability or absolute bioavailability of oral administration. ^g Absorption data (or average values) chosen in the reference 88 based on the analysis of literature.

Compounds from **146** to **172**, according to the Abraham *et al.* [88] classification, were considered as uncertain and unchecked data. For the 11 compounds with high Abs % values, 10 were well classified with a high Caco–2 cells “*in vitro*” permeability (Eq. 12), and only two, with moderate–poor human absorption were bad classified. The global good classification was 88.89%.

In the same sense the group of zwitterionic drugs (20), reported by Abraham *et al.* [88], were also analyzed. For this kind of drugs our model only showed a 50% of correct correspondence between “*in vitro*” classification and “*in vivo*” results. For compounds **193–201** (group of missing fragments, according Abraham *et al.* [88]) were bad predicted with our model more than a half of compounds for a 44.4% of correspondence. For the group with dose limited, dose–dependent and

formulation-dependent drugs the correspondence between *in vitro* permeability and the human absorption values was 81.5%. Finally, for the analysis of drugs with expected high absorption, according to Abraham *et al.* [88], were not reported a value or average for the human absorption. Nevertheless, if the data from the fifth, sixth and seventh columns of data in Table 5 are considered our model explained 84.61% of correspondence. If the compounds **146-241** are considered as an only great group, where less realistic data of Abs% are reported, the percentage of correct correspondence between “*in vitro*” permeability data [Caco-2 cells, predicted by Eq. (12)] and the human absorption is 73.78% (70/95). This group has a less percentage of correspondence than the first 145 compounds previously analyzed.

Considering the full set (241 compounds) the model [Eq. (12)] showed a 78.84% of explanation of the human absorption values, which is a logic result considering the structure variability and the biological property.

On the other hand has been widely reported in the literature the influence of transport mechanism in the prediction of this biopharmaceutical property, for example: Methotrexate is absorbed by a carrier-mediated process, Zidovudine is absorbed by active transport, Amoxicillin and Cefatrizine are absorbed via dipeptide carrier system and in the Etoposide case it is suggested that its distribution into the brain is partially controlled by an active transport process [85]. Also Cefadroxil, Digoxin and Cepahalexin were compounds with known active transport [89]. Other compounds with the same skeleton pattern of cephalosporins (Cefatrizine and Ceftizoxime), cardiotonic glycosides (Ouabain) and antiviral nucleoside analogues (Stavudine, Lamivudine, Sorivudine) appears bad classified (uncorrelation between the permeability predicted in Caco-2 cells and the human absorption values), suggesting a active transport system for these drugs. In addition, in the case of the Viomycin, with an appropriate intestinal absorption (Abs % = 85), it has a molecular weight value of 685 g mol^{-1} (>500), similar with those drugs with poor intestinal absorption, for what it could be suggested that this compound can be actively transported, as was pointed out by Egan *et al.* [90] in the case of Rifampicin.

It is obvious that from these results the quality of the predictions assessing the predictive power of the models found and justified its use in the prediction of this important biopharmaceutical property. Also, this is not a fortuitous result due to the data set used in this study including any sort of absorption model compounds.

8 CONCLUDING REMARKS

Computer-aided molecular design has become in a very important tool in the development of novel chemical compounds to be used in different areas of human life. The focus of modern drug discovery is now not simply on the pharmacological activity, but also on seeking favorable absorption, distribution, metabolism, and excretion properties [88,91,92]. The growth in drug

discovery of combinatorial chemistry methods, where large numbers of candidate compounds are synthesized and screened in parallel for *in vitro* pharmacological activity, has dramatically increased the demand for rapid and efficient models for estimating human absorption. Thus, the continuous definition of novel molecular descriptors that could explain different biological properties by means of QSAR is necessary. Consequently, we have developed LDA model that could permit us to predict by fast *in silico* screening, the intestinal permeability of chemical and outline preliminary conclusions about possible human intestinal absorption profile. This result demonstrated that total and local quadratic indices appear to be a very promising structural invariant and was able to produce an adequate model for the correct classification of the intestinal permeability for structural diverse drugs. Acceptable efficiency and a fairly good predictability was found in the prediction of absorption, especially if we take into account the variety of the selected data set and the simplicity of the calculations used, which are in relation with the computational feasibility of the TOMOCOMD–CARDD method. This approach permits that these indices can be applied to large sets of NCEs synthesized via combinatorial chemistry approach. Furthermore, this approximation permits to obtain significant interpretation of the experiment result in terms of the structural features of molecules. Some works, where the potential of this method in the prediction of several biological properties for different classes of organic compounds have been proved, are now in progress and will be published in a forthcoming paper.

Acknowledgment

We wish thank to Eric J. Lien, Mehran Yazdanian, A. J. Hopfinger and Han van de Waterbeemd for providing reprints of their work, which significantly contribute to the development of this paper. The authors also acknowledge the anonymous referees for their valuable comments that increased the quality of the manuscript.

5 REFERENCES

- [1] C. Zhu, L. Jiang, T. M. Chen, and K. K. Hwang, A Comparative Study of Artificial Membrane Permeability Assay for High Throughput Profiling of Drug Absorption Potential, *Eur. J. Med. Chem.* **2002**, *37*, 399–407.
- [2] C. A. Lipinski, F. Lombardo, B. W. Dominy, and P. J. Feeney, Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings, *Adv. Drug Deliv. Rev.* **1997**, *23*, 3–25.
- [3] 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
- [4] A. D. Rodríguez, Rational High Throughput Screening in Preclinical Drug Metabolism, *Med Chem Res.* **1998**, *8*, 422–433.
- [5] P. Artursson, Cell Cultures as Models for Drug Absorption Across the Intestinal Mucosa, *Cri. Rev. Ther. Carrier Syst.* **1991**, *8*, 305–330.
- [6] A. Quaroni and J. Hochman, Development of Intestinal Cell Culture Models for Drug Transport and Metabolism Studies, *Abv. Drug Del. Rev.* **1996**, *22*, 3–52.
- [7] M. Yazdanian, S. L. Glynn, J. L. Wright, and A. Hawi, Correlating Partitioning and Caco–2 Cell Permeability of Structurally Diverse Small Molecular Weight Compounds, *Pharm. Res.* **1998**, *15*, 1490–1494.
- [8] P. Artursson and J. Karlsson, Correlation between Oral Drug Permeability Coefficients in Human Intestinal

- Epithelial (Caco-2) Cells, *Biochem. Biophys. Res. Commun.* **1991**, *175*, 880–885.
- [9] G. Camenisch, J. Alsenz, H. van de Waterbeemd, and G. Folkers, Estimation of Permeability by Passive Diffusion Through Caco-2 Cell Monolayer Using the Drugs' Lipophilicity and Molecular Weight, *Eur. J. Pharm. Sci.* **1998**, *6*, 313–319.
- [10] E. Walter, T. Kissel, M. Reers, G. Dickneite, D. Hoffmann, and W. Stueber, Transepithelial Transport Properties of Peptidomimetic Thrombin Inhibitors in Monolayers of a Human Intestinal Cell Line (Caco-2) and Their Correlation to in vivo Data, *Pharm. Res.* **1995**, *12*, 360–365.
- [11] P. Artursson, K. Palm, and K. Luthman, Caco-2 Monolayer in Experimental and Theoretical Predictions of Drug Transport, *Adv. Drug Deliv. Rev.* **1996**, *22*, 67–84.
- [12] F. Delie and W. A. Rubas, Human Colonic Cell Line Sharing Similarities with Enterocytes as a Model to Examine Oral Absorption. *Crit. Rev. Ther. Drug Carrier Syst.* **1997**, *14*, 221–286.
- [13] P. Anderle, E. Niederer, W. Rubas, C. Hilgendorf, and P. Spahn-Langguth, 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.* **1998**, *87*, 757–762.
- [14] G. L. Amidon, P.J. Sinko, and D. Fleisher, Estimating Human Oral Fraction Dose Absorbed: A Correlation Using Rat Intestinal Membrane Permeability for Passive and Carrier-Mediated Compounds, *Pharm. Res.* **1988**, *5*, 651–654.
- [15] W. Rubas, N. Jezyk, G. M. Grass, 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.* **1993**, *10*, 113–118.
- [16] S. Chong, S. A. Dando, and R. Morrison, Evaluation of Biocoat Intestinal Epithelium Differentiation Environment (Accelerated Cultured Caco-2 cells) as an Absorption Screening Model with Improved Productivity, *Pharm Res.* **1997**, *14*, 1835–1837.
- [17] S. Yee, In Vitro Permeability Across Caco-2 Cells (Colonic) can Predict in Vivo (Small Intestinal) Absorption in Man-Fact or Myth, *Pharm. Res.* **1997**, *14*, 763–766.
- [18] P. R. Chaturvedi, C. J. Dekker, and A. Odinecs, Prediction of Pharmacokinetic Properties Using Experimental Approaches During Early Drug Discovery, *Curr. Opin. Chem. Biol.* **2001**, *5*, 452–463.
- [19] H. van de Waterbeemd and G. Camenisch, Estimation of Caco-2 cell Permeability Using Calculated Molecular Descriptors, *Quant. Struct.-Act. Relat.* **1996**, *15*, 480–490.
- [20] S. Ren and E. J. Lien, Caco-2 Cell Permeability vs Human Gastro-Intestinal Absorption: QSPR Analysis, *Progr. Drug Res.* **2000**, *54*, 3–23.
- [21] H. van de Waterbeemd, D. A. Smith, and B. C. Jones, Lipophilicity in PK Design: Methyl, Ethyl, Futile, *J. Comp.-Aided Mol. Des.* **2001**, *15*, 273–286.
- [22] M. Sugawara, Y. Takekuma, H. Yamada, M. Kobayashi, K. Iseki, and K. Miyazaki, A General Approach for the Prediction of the Intestinal Absorption of Drugs: Regression Analysis Using the Physicochemical Properties and Drug-Membrane Electrostatic Interaction, *J. Pharm. Sci.* **1998**, *87*, 960–966.
- [23] D.E. Clark and D. S. Pickett, Computational Methods for the Prediction of 'Drug-Likeness', *Drug Disc. Today.* **2000**, *5*, 49–58.
- [24] J. B. Dressman, G. L. Amidon, and D. Fleisher, Absorption Potential: Estimating the Fraction Absorbed for Orally Administered Compounds, *J. Pharm. Sci.* **1985**, *74*, 588–589.
- [25] H. W. Hamilton, B. B. Steinbaugh, A. H. Stewart, O. H. Chan, H. L. Schmid, R. Schroeder, M. J. Ryan, J. Keiser, M. D. Taylor, C. J. Blankley, J. S. Kalttenbronn, J. Wright, and J. Hicks, Evaluation of Physicochemical Parameters Important to the Oral Bioavailability of Peptide-Like Compounds: Implications for the Synthesis of Renin Inhibitors, *J. Med. Chem.* **1995**, *38*, 1446–1455.
- [26] K. Palm, K. Luthman, A. L. Ungell, G. Strandlund, and P. Artursson, Correlation of Drug Absorption with Molecular Surface Properties, *J. Pharm. Sci.* **1996**, *85*, 32–39.
- [27] M. H. Abraham, H. S. Chadha, and R. C. Mitchell, Hydrogen Bonding. 33. Factors that Influence the Distribution of Solutes between Blood and Brain, *J. Pharm. Sci.* **1994**, *83*, 1257–1268.
- [28] S. C. Basak, B. D. Gute, and E. R. Drewes, Predicting Blood-Brain Transport of Drugs: A Computational Approach, *Pharm. Res.* **1996**, *13*, 775–778.
- [29] R. O. Potts and R. H. Guy, A Predictive Algorithm for Skin Permeability: The Effects of Molecular Size and Hydrogen Bond Activity, *Pharm. Res.* **1995**, *12*, 1628–1663.

- [30] F. Yoshida and J. G. Topliss, Unified Model for the Corneal Permeability of Related and Diverse Compounds with Respect to Their Physico-chemical Properties, *J. Pharm. Sci.* **1996**, *85*, 819–823.
- [31] J. V. S. Gobburu and W. H. Shelver, Quantitative Structure-Pharmacokinetic Relationships (QSPR) of Beta Blockers Derived Using Neural Networks, *J. Pharm. Sci.* **1995**, *84*, 862–865.
- [32] M. D. Wessel, P. C. Jurs, J. W. Tolan, S. M. Muskal, Prediction of Human Intestinal Absorption of Drug Compounds from Molecular Structure, *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 726–735.
- [33] C. W. Andrews, L. Bennett, and L. X. Yu, Predicting Human Oral Bioavailability of a Compound: Development of a Novel Quantitative-Structure-Bioavailability Relationship, *Pharm. Res.* **2000**, *17*, 639–644.
- [34] M. R. Feng, X. L. Richard, R. Brown, and A. Hutchaleelaha, Allometric Pharmacokinetic Scaling: Towards the Prediction of Human Oral Pharmacokinetics, *Pharm. Res.* **2000**, *17*, 410–418.
- [35] L. J. Lesko, M. Rowland, C. C. Peck, T. F. Blaschke, Optimizing the Science of Drug Development: Opportunities for Better Candidate Selection and Accelerated Evaluation in Humans, *Pharm. Res.* **2000**, *17*, 1335–1344.
- [36] J. Breitzkreutz, Prediction of Intestinal Drug Absorption Properties by Three-Dimensional Solubility Parameters, *Pharm. Res.* **1998**, *15*, 1370–1375.
- [37] J. A. Platts, M. H. Abraham, A. Hersey, and D. Butina, Estimation of Molecular Linear Free Energy Relationship Descriptors. 4. Correlation and Prediction of Cell Permeation, *Pharm. Res.* **2000**, *17*, 1013–1018.
- [38] R. J. Riley, A. J. Parker, S. Trigg, and C. N. Manners, Development of a Generalized, Quantitative Physicochemical Model of CYP3A4 Inhibition for Use in Early Drug Discovery, *Pharm. Res.* **2001**, *18*, 652–655.
- [39] S. Ekins and J. Rose, In silico ADME/Tox: the State of the Art, *J. Mol. Graph. Modell.* **2002**, *20*, 305–309.
- [40] H. van de Waterbeemd and M. Kansy, Hydrogen-Bonding Capacity Permeability Using Calculated Molecular Descriptors, *Quant. Struct.-Act. Relat.* **1992**, *15*, 480–490.
- [41] H. Krarup, T. I. Christensen, L. Hovgaard, and S. Frokjaer, Predicting Drug Absorption from Molecular Surface Properties Based on Molecular Dynamics Simulations, *Pharm. Res.* **1998**, *15*, 972–978.
- [42] U. Norinder, T. Osterber, and P. Artursson, Theoretical Calculation and Prediction of Caco-2 Cell Permeability Using MolSurf Parameterization and PLS Statistics, *Pharm. Res.* **1997**, *14*, 1786–1791.
- [43] S. I. Fujiwara, F. Yamashita, M. Hashida, Prediction of Caco-2 Cell Permeability Using a Combination of MO-Calculation and Neural Network, *Inter. J. Pharm.* **2002**, *237*, 95–105.
- [44] A. Kulkarni, Y. Han, J. Hopfinger, Predicting Caco-2 Cell Permeation Coefficients of Organic Molecules Using Membrane-Interaction QSAR Analysis, *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 331–342.
- [45] D. Ruovray, Taking a Short Cut to Drug Design, *New Sci.* **1993**, 35–38.
- [46] A. R. Katritzky, V. S. Lobanov, M. Karelson, QSPR: The Correlation and Quantitative Prediction of Chemical and Physical Properties from Structure, *Chem. Soc. Rev.* **1995**, 279–287.
- [47] A. Katritzky, U. Maran, V. S. Lobanov, and M. Karelson, Structurally Diverse Quantitative Structure-Property Relationship Correlations of Technologically Relevant Physical Properties, *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1–18.
- [48] M. Karelson, *Molecular Descriptors in QSAR/QSPR*, John Wiley & Sons, New York, 2000.
- [49] R. Todeschini, V. Consonni, *Handbook of Molecular Descriptors*, Wiley VCH, Weinheim, Germany, 2000.
- [50] *Topological Indices and Related Descriptors in QSAR and QSPR*, Eds. J. Devillers, & A.T. Balaban, Gordon and Breach Science; Amsterdam, The Netherlands, 1999.
- [51] E. Estrada and E. Uriarte, Recent Advances on the Role of Topological Indices in Drug Discovery Research, *Curr. Med. Chem.* **2001**, *8*, 1699–1714.
- [52] *QSPR/QSAR Studies by Molecular Descriptors*, Ed. M. V. Diudea, Nova Science, Huntington, N.Y., 2001.
- [53] O. Ivanciuc, T. Ivanciuc, D. Cabrol-Bass, and A. T. Balaban, Evaluation in Quantitative Structure-Property Relationship Models of Structural Descriptors Derived from Information-Theory Operators, *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 631–643.
- [54] T. Balaban, D. Mills, O. Ivanciuc, and S. C. Basak, Reverse Wiener Indices, *Croat. Chem. Acta.* **2000**, *73*, 923.
- [55] O. Ivanciuc, T. Ivanciuc, D. J. Klein, W. A. Seitz, and A. T. Balaban, Wiener Index Extension by Counting Even/Odd Graph Distances, *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 536–549.
- [56] O. Ivanciuc, Building-Block Computation of the Ivanciuc-Balaban Indices for the Virtual Screening of Combinatorial Libraries, *Internet Electron. J. Mol. Des.* **2002**, *1*, 1–9, <http://www.biochempress.com>.
- [57] I. Rios-Santamarina, R. García-Domenech, J. Cortijo, P. Santamaria, E. J. Morcillo, J. Gálvez, Natural

- Compounds with Bronchodilator Activity Selected by Molecular Topology, *Internet Electron. J. Mol. Des.* **2002**, *1*, 70–79, <http://www.biochempress.com>.
- [58] D. J. G. Marino, P. J. Peruzzo, E. A. Castro, A. A. Toropov, QSAR Carcinogenic Study of Methylated Polycyclic Aromatic Hydrocarbons Based on Topological Descriptors Derived from Distance Matrices and Correlation Weights of Local Graph Invariants, *Internet Electron. J. Mol. Des.* **2002**, *1*, 115–133, <http://www.biochempress.com>.
- [59] O. Ivanciuc, QSAR Comparative Study of Wiener Descriptors for Weighted Molecular Graphs, *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1412–1422.
- [60] E. Estrada, Spectral Moment of Edge Adjacency Matrix in Molecular Graphs. I. Definition and Application to the Prediction of Physical Properties of Alkanes, *J. Chem. Inf. Comp. Sci.* **1996**, *36*, 846–849.
- [61] E. Estrada, N. Guevara, and I. Gutman, Extension of Edge Connectivity Index. Relationships to Line Graph Indices and QSPR Applications, *J. Chem. Inf. Comp. Sci.* **1998**, *38*, 428–431.
- [62] E. Estrada, Edge-Connectivity Indices in QSPR/QSAR Studies. 2. Accounting for Long-Range Bond Contributions, *J. Chem. Inf. Comp. Sci.* **1999**, *39*, 1042–1048.
- [63] E. Estrada and I. A Gutman, Topological Index Based on Distances of Edges of Molecular Graphs, *J. Chem. Inf. Comp. Sci.* **1996**, *36*, 850–853.
- [64] Y. Marrero, Total and local Quadratic Indices of the “Molecular Pseudograph’s Atom Adjacency Matrix”: Applications to the Prediction of Physical Properties of Organic Compounds, *Molecules.* **2003**, *8*, 687–726. <http://www.mdpi.org>.
- [65] Y. Marrero, M. A. Cabrera, V. Romero, E. Ofori, and L. A. Montero, 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.* **2003**, *4*, 512–536. www.mdpi.org/ijms/
- [66] Y. Marrero and V. Romero, TOMO-COMD software; Central University of Las Villas, **2002**. **TOMO-COMD** (**T**Opological **M**Olecular **C**OMputer **D**esign) 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; ymarrero77@yahoo.es
- [67] M. Randić, Generalized Molecular Descriptors, *J. Math. Chem.* **1991**, *7*, 155–168.
- [68] F. A. Cotton, *Advanced Inorganic Chemistry*, Ed. Revolucionaria, Cuba, pp. 103.
- [69] L. Hovgaard, H. Brøndsted, A. Buur, and H. Bundgaard, Drug Delivery Studies in Caco-2 Monolayers. Synthesis, Hydrolysis, and Transport of O-cyclopropane Carboxylic Acid Ester Prodrugs of Various β -blocking Agents, *Pharm Res.* **1995**, *12*, 387–397.
- [70] P. Artursson, 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.* **1990**, *79*, 476–482.
- [71] P. Artursson, Magnusson, Epithelial Transport of Drugs in Cell 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.* **1990**, *79*, 595–600.
- [72] P. Stenberg, U. Norinder, K. Luthman, and P. Artursson, Experimental and Computational Screening Models for the Prediction of Intestinal Drug Absorption, *J. Med. Chem.* **2001**, *44*, 1927–1937.
- [73] P. Augustijns, A. D’Hulst, J. V. Daele, R. Kinget, Transport of Artemisinin and Sodium Artesinate in Caco-2 Intestinal Epithelial Cell, *J. Pharm. Sci.* **1996**, *85*, 577–579.
- [74] A. Collett, E. Sims, D. Walker, Y-L. He, J. Ayrton, M. Rowland, and G. Warhurst, Comparison of HT29-18-C₁ and Caco-2 Cell Lines as Models for Studying Intestinal Paracellular Drug Absorption, *Pharm Res.* **1996**, *13*, 216–221.
- [75] M. Grès, B. Julian, M. Bourrié, V. Meunier, C. Roques, M. Berger, X. Boulenc, Y. Berger, and G. Fabre, 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.* **1998**, *15*, 726–733.
- [76] C. Hilgendorf, H. Spahn-Langguth, C. G. Regardh, E. Lipka, G. L. Amidon, and P. Langguth, Caco-2 versus Caco-2/HT29-MTX Co-cultured Cell Lines: Permeability via Diffusion, Inside- and Outside-Directed Carrier-Mediated Transport, *J Pharm Sci.* **2000**, *89*, 63–75.
- [77] B. H. Stewar, O. H. Chan, R. H. Lu, E. L. Reyner, H. L. Schmid, H. W. Hamilton, B. A. Steinbaugh, and M. D. Taylor, Comparison of Intestinal Permeabilities Determined in Multiple *in vitro* and *in situ* Models: Relationships to Absorption in Humans, *Pharm. Res.* **1995**, *12*, 693–699.

- [78] B. J. Aungst, N. H. Nguyen, J. P. Bulgarelli, and K. Oates–Lenz, The Influence of Donor and Reservoir Additives on Caco–2 Permeability and Secretory Transport of HIV Protease Inhibitors and other Lipophilic Compounds, *Pharm. Res.* **2000**, *17*, 1175–1180.
- [79] A. Ruiz–García, H. Lin, J. M. Plá–Delfina, M. Hu, Kinetic Characterization of Secretory Transport of a New Ciprofloxacin Derivative (CNV97100) Across Caco–2 Cell Monolayers, *J Pharm Sci.* **2002**, *12*, 2511–2519.
- [80] STATISTICA ver. 5.5, Statsoft, Inc. 1999.
- [81] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, Prentice–Hall, Englewood Cliffs, N.J., 1998.
- [82] P. Baldi, S. Brunak, Y. Chauvin, C. A. Andersen, and H. Nielsen. Assessing the Accuracy of Prediction Algorithms for Classification: an Overview. *Bioinformatics.* **2000**, *16*, 412–424.
- [83] R. A. Conradi, P. S. Buton, and R. T. Borejhardt; in: *Lipophilicity in Drug Action and Toxicology*, Eds. V. Pliska, B. Testa, H. van de Waterbeemd, VCH, Weinheim, 1996, pp. 223–252.
- [84] P. Stenberg, K. Luthman, and P. Artursson, Virtual Screening of Intestinal Drug Permeability, *J. Contr. Rel.* **2000**, *65*, 231–243.
- [85] D. E. Clark, Rapid Calculation of Polar Molecular Surface Area and Its Application to the Prediction of Transport Phenomena. 1. Prediction of Intestinal Absorption, *J. Pharm Sci.* **1999**, *88*, 807–814.
- [86] K. Palm, P. Stenberg, K. Luthman, and P. Artursson, Polar Molecular Surface Properties Predict and Intestinal Absorption of Drugs in Humans, *Pharm. Res.* **1997**, *14*, 568–571.
- [87] W. L. Chiou and A. Barve, Linear Correlation of the Fraction of Oral Dose Absorbed of 64 Drugs Between Humans and Rats, *Pharm Res.* **1998**, *15*, 1792–1795.
- [88] Y. H. Zhao, J. Le, M. H. Abraham, A. Hersey, P. J. Eddershaw, C. N. Luscombe, D. Boutina, G. Beck, B. Sherborne, I. Cooper, and J. A. Platts, Evaluation of Human Intestinal Absorption Data and Subsequent Derivation of a Quantitative Structure–Activity Relationship (QSAR) with the Abraham Descriptors, *J. Pharm. Sci.* **2001**, *90*, 749–784.
- [89] L. Z. Benet, S. Øie, and J. B. Schwartz, Design and Optimisation of Dosage Regiments; Pharmacokinetic Data; in: *Pharmacological Basis of Therapeutics*, 9th ed., Eds. J. G. Hardman, L. E. Limbird, and A.G. Gilman, McGraw–Hill, New York, 1996, pp. 1707–1793.
- [90] W. J. Egan, K. M. Merz, and J. J. Baldwin, Prediction of Drug Absorption Using Multivariate Statistics. *J. Med. Chem.* **2000**, *43*, 3867–3877.
- [91] K. Lauritsen, L. S. Laursen, and J. Raskmadsen, Clinical Pharmacokinetics of Drugs Used in the Treatment of Gastrointestinal Diseases Part (I), *Clin Pharmacokinet.* **1990**, *19*, 11–31.
- [92] M. C. Walker and P. N. Patsalos, Clinical Pharmacokinetics of New Antiepileptic Drugs, *Pharmacol Ther.* **1995**, *67*, 351–384.