molecular pharmaceutics

articles

Computational Models for Identifying Potential P-Glycoprotein Substrates and Inhibitors

Patrizia Crivori,*,† Benedetta Reinach,‡ Daniele Pezzetta,‡ and Italo Poggesi†

Prediction and Modeling, Preclinical Profiling, Preclinical Development, Nerviano Medical Sciences, viale Pasteur 10, 20014 Nerviano, Italy

Received September 1, 2005

Abstract: Multidrug resistance mediated by ATP binding cassette (ABC) transporters such as P-glycoprotein (P-gp) represents a serious problem for the development of effective anticancer drugs. In addition, P-gp has been shown to reduce oral absorption, modulate hepatic, renal, or intestinal elimination, and restrict blood−brain barrier penetration of several drugs. Consequently, there is a great interest in anticipating whether drug candidates are P-gp substrates or inhibitors. In this respect, two different computational models have been developed. A method for discriminating P-gp substrates and nonsubstrates has been set up based on calculated molecular descriptors and multivariate analysis using a training set of 53 diverse drugs. These compounds were previously classified as P-gp substrates or nonsubstrates on the basis of the efflux ratio from Caco-2 permeability measurements. The program Volsurf was used to compute the compounds' molecular descriptors. The descriptors were correlated to the experimental classes using partial least squares discriminant analysis (PLSD). The model was able to predict correctly the behavior of 72% of an external set of 272 proprietary compounds. Thirty of the 53 previously mentioned drugs were also evaluated for P-gp inhibition using a calcein-AM (CAM) assay. On the basis of these additional P-gp functional data, a PLSD analysis using GRIND-pharmacophore-based descriptors was performed to model P-gp substrates having poor or no inhibitory activity versus inhibitors having no evidence of significant transport. The model was able to discriminate between 69 substrates and 56 inhibitors taken from the literature with an average accuracy of 82%. The model allowed also the identification of some key molecular features that differentiate a substrate from an inhibitor, which should be taken into consideration in the design of new candidate drugs. These two models can be implemented in a virtual screening funnel.

Keywords: P-glycoprotein; structure−property relationships; SPR; 3D pharmacophore, in silico screening; P-gp substrates; P-gp inhibitors; Volsurf; GRIND; Caco-2 permeability, efflux ratio; calcein-AM assay; CAM assay

Introduction

P-glycoprotein (P-gp) belongs to a superfamily of ATP binding cassette (ABC) transporters that acts as energy-dependent efflux pump, transporting out of cells a wide variety of compounds.[1,2] P-gp encoded by the MDR1 (multidrug resistance-1) gene in humans and the mdr1 (also

* To whom correspondence should be addressed. Mailing address: Prediction and Modeling, Nerviano Medical Sciences, Viale Pasteur, 10, 20014 Nerviano, Italy. E-mail: patrizia.crivori@nervianoms.com. Tel: +39 0331 581476. Fax: +39 0331 581105.

† Prediction and Modeling, Preclinical Development.

‡ Preclinical Profiling, Preclinical Development.

(1) Ambudkar, S. V.; Dey, S.; Hrycyna, C. A.; Ramachandra, M.; Pastan, I.; Gottesman, M. M. Biochemical, cellular, and pharmacological aspects of the multidrug transporter. *Annu. Rev. Pharmacol. Toxicol.* **1999**, *39*, 361−398.

(2) Borst, P.; Evers, R.; Kool, M.; Wijnholds, J. A family of drug transporters: the multidrug resistance-associated proteins. *J. Natl. Cancer Inst.* **2000**, *92*, 1295−1302.

10.1021/mp050071a CCC: $33.50
Published on Web 12/08/2005

called mdr1b) and mdr3 (also called mdr1a) in rodents is composed of two homologous halves, each containing a trasmembrane region of six α-helices involved in the efflux of xenobiotics and a cytosolic domain responsible for ATP binding and hydrolysis.[3] Highly expressed in many cancer cells, P-gp is one of the major mechanisms of cancer resistance to chemotherapy.[4] This transporter is also expressed in many normal tissues such as intestine, liver, kidney, lung, and endothelia of brain, testis, and placenta, consistent with its role as a natural detoxification system. By this activity it can have a profound impact on the pharmacokinetics and pharmacodynamics of many drugs.[1,2] In particular, P-gp has been shown to limit oral absorption, modulate hepatic, renal, or intestinal elimination, and restrict central nervous system entry of certain drugs.[5−7] In addition, because of its broad substrate specificity, P-gp mediated drug−drug interactions may occur when substrates, inducers, and inhibitors are coadministered.[8] For example, P-gp inhibitors lead to an increase in the systemic exposure and tissue distribution of coadministered P-gp substrate drugs that could cause serious adverse effects.[8] In light of these finding, in vitro assays and in silico models for predicting P-gp substrates or inhibitors have been recognized to be valuable tools during early phases of drug development. Several in vitro screens for identifying potential substrates and inhibitors are currently available and include bidirectional permeability across cell monolayers, accumulation of fluorescent substrates, and ATPase activation.[9−12] Although it is widely accepted that P-gp protein has multiple binding sites for its ligands, detailed 3D information is still lacking. In the absence of high-resolution X-ray P-gp structures,[13] structure−function relationships have been mainly derived, with the exception of a few studies that relied on homology models,[14,15] from the analysis of compounds interacting with P-gp. Using different sets of molecules and various in vitro assays to measure P-pg activities, a number of structure−property relationships (SPR) have been developed to elucidate the physicochemical properties characterizing the P-gp substrates.[16−22] More detailed 3D pharmacophoric hypotheses for P-gp substrates and inhibitors have also been proposed.[23−26]

(3) Dahl, S. G.; Sylte, I.; Ravna, A. W. Structures and models of transporter proteins. *J. Pharmacol. Exp. Ther.* **2004**, *309*, 853−860.

(4) Gottesman, M. M.; Pastan, I. Biochemistry of multidrug resistance mediated by the multidrug transporter. *Annu. Rev. Biochem.* **1993**, *62*, 385−427.

(5) Benet, L. Z.; Izumi, T.; Zhang, Y.; Silverman, J. A.; Wacher, V. J. Intestinal MDR transport proteins and P-450 enzymes as barriers to oral drug delivery. *J. Controlled Release* **1999**, *62*, 25−31.

(6) Varma, M. V. S.; Sateesh, K.; Panchagnula, R. Functional role of P-glycoprotein in limiting intestinal absorption of drugs: contribution of passive permeability to P-glycoprotein meditated efflux transport. *Mol. Pharm.* **2005**, *2*, 12−21.

(7) Schinkel, A. H.; Jonker, J. W. Mammalian drug efflux transporters of the ATP binding cassette (ABC) family: an overview. *Adv. Drug Delivery Rev.* **2003**, *55*, 3−29.

(8) Lin, J. H. Drug-drug interaction mediated by inhibition and induction of P-glycoprotein. *Adv. Drug Delivery Rev.* **2003**, *55*, 53−81.

(9) Zhang, Y.; Bachmeier, C.; Miller, D. W. In vitro and in vivo models for assessing drug efflux transporter activity. *Adv. Drug Delivery Rev.* **2003**, *55*, 31−51.

(10) Hollo, Z.; Homolya, L.; Davis, C. W:; Sarkadi, B. Calcein accumulation as a fluorometric functional assay of the multidrug resistance transporter. *Biochim. Biophys. Acta* **1994**, *1191*, 384−388.

(11) Homolya, L.; Hollo, Z.; Germann, U. A.; Pastan, I.; Gottesman, M. M.; Sarkadi, B. Fluorescent cellular indicators are extruded by the multidrug resistance protein. *J. Biol. Chem.* **1993**, *268*, 21493−21496.

(12) Tiberghien, F.; Loor, F. Ranking of P-glycoprotein substrates and inhibitors by a Calcein-AM fluorometry screening assay. *Anti-Cancer Drugs* **1996**, *7*, 568−578.

(13) Rosenberg, M. F.; Kamis, A. B.; Callaghan, R.; Higgins, C. F.; Ford, R. C. Three-dimensional Structures of the Mammalian Multidrug Resistance P-glycoprotein Demonstrate Major Conformational Changes in the Transmembrane Domains upon Nucleotide Binding. *J. Biol. Chem.* **2003**, *278*, 8294−8299.

(14) Stenham, D. R.; Campbell, J. D.; Sanson, M. S. P.; Higgins, C. F.; Kerr, I. D.; Linton, K. J. An Atomic detail model for the human ATP binding cassette transporter P-glycoprotein derived from disulfide cross-linking and homology modeling. *FASEB J.* **2003**, *17*, 2287−2289.

(15) Pajeva, I. K.; Globisch, C.; Wiese, M. Structure−Function Relationships of Multidrug Resistance P-Glycoprotein. *J. Med. Chem.* **2004**, *47*, 2523−2533.

(16) Litman, T.; Zeuthen, T.; Skovsgaard, T.; Stein, W. D. Structure-activity relationships of P-glycoprotein interacting drugs: kinetic characterization of their effects on ATPase activity. *Biochim. Biophys. Acta* **1997**, *1361*, 159−168.

(17) Seelig, A. A general pattern for substrate recognition by P-glycoprotein. *Eur. J. Biochem.* **1998**, *251*, 252−261.

(18) Seelig, A.; Landwojtowicz, E. Structure-activity Relationship of P-glycoprotein substrates and modifiers. *Eur. J. Pharm. Sci.* **2000**, *12*, 31−40.

(19) Osterberg, T.; Norinder, U. Theoretical calculation and prediction of P-glycoprotein-interacting drugs using MolSurf parametrization and PLS statistics. *Eur. J. Pharm. Sci.* **2000**, *10*, 295−303.

(20) Stouch, T. R.; Gudmundsson, O. Progress in Understanding the structure-activity relationships of P-glycoprotein *Adv. Drug Delivery Rev.* **2002**, *54*, 315−328.

(21) Gombar, V. K.; Polli, J. W.; Humphreys, J. W.; Wring, S. A.; Serabjit-Singh, C. S. Predicting P-Glycoprotein Substrates by a Quantitative Structure-Activity Relationship Model. *J. Pharm. Sci.* **2004**, *93*, 957−968.

(22) Xue, Y.; Yap, C. W.; Sun, L. Z.; Cao, Z. W.; Wang, J. F.; Chen, Y. Z. Prediction of P-Glycoprotein Substrates by a Support Vector Machine Approach. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1497−1505.

(23) Ekins, S.; Kim, R. B.; Leake, B. E.; Dantzig, A. H.; Schuetz, E. G.; Lan, L. B.; Yasuda, K.; Shepard, L.; Winter, M. A.; Schuetz, J. D.; Wikel, J. H.; Wrighton, S. A. Application of Three-Dimensional Quantitative Structure−Activity Relationships of P-Glycoprotein Inhibitors and Substrates. *Mol. Pharmacol.* **2002**, *61*, 974−981.

(24) Penzotti, J. E.; Lamb, M. L.; Evensen, E.; Grootenhuis, P. D. J. A Computational Ensemble Pharmacophore Model for identifying substrates of P-Glycoprotein. *J. Med. Chem.* **2002**, *45*, 1737−1740.

(25) Pajeva, I. K.; Wiese, M. Pharmacophore Model of Drugs Involved in P-Glycoprotein Multidrug Resistance: Explanation of Structural Variety (Hypothesis). *J. Med. Chem.* **2002**, *45*, 5671−5686.

Wang et al.[27] recently published a model based on an unsupervised machine learning approach for classifying potential P-gp substrates and inhibitors. Despite the difficulties in developing computational models due to broad substrate specificity, multiple binding P-gp sites and different modulator mechanisms (i.e. competitive, noncompetitive, alteration of cell membrane lipids, etc.),[1] the availability of virtual screens for discriminating substrates, nonsubstrates, and inhibitors would be helpful in the design of new drugs. In particular, these models can be used for assessing the potential for drug−drug interactions of new candidates. Furthermore, they could also be applied for designing P-gp inhibitors acting selectively on cancer cells in order to reverse multidrug resistance and improve the efficacy of chemotherapy. In addition, P-gp inhibitors could also be developed to enhance intestinal absorption and/or increase brain penetration of P-gp substrate drugs.[28,29]

The aim of the present study was first to develop a computational model for discriminating P-gp substrates from nonsubstrates, based on Caco-2 efflux ratio values. For this purpose, a PLSD model based on Volsurf descriptors was generated.[30]

In addition, to explore more in detail the 3D pharmacophoric features that mainly differentiate the substrates with poor or no inhibitory activity in the range of the therapeutically relevant concentrations from inhibitors having no evidence of significant transport, a PLSD model based on GRIND descriptors[31] was developed combining the experimental results of a calcein-AM (CAM) assay[10−12] with the Caco-2 efflux ratio data. It has to be considered that the different inhibition mechanisms as well as the presence of multiple potential binding P-gp sites can make difficult the interpretation of the in vitro data. For this reason the information obtained from both assays was combined and used in a qualitative manner.[32,33]

Both models, validated using external data sets, can be implemented as virtual screens in early phases of drug discovery.

Experimental Methods

Materials. Human colon adenocarcinoma (Caco-2) cells were obtained from the American Type Culture Collection (Rockville, MD) at passage 20, and used between passages 28 and 43. LoVoDx cells are derived from the human colon adenocarcinoma LoVo cell line by exposure to increasing concentrations of doxorubicin and maintained under the same culture conditions.[34] The resistance is due to the overexpression of MDR1 phenotype. An increase of other drug transporter proteins, such as multidrug resistance-associated protein (MRP) and lung resistance-related protein (LRP), has also been described in the literature.[35−37] This cell line is available in the Nerviano Medical Sciences Cell Bank, Nerviano, Italy. All cell culture medium and reagents were from Gibco BRL (Grand Island, NY). Transwell chambers for cell culture were obtained from Becton Dickinson (1.0 μm BD Falcon HTS 24-Multiwell Insert System, Bedford, MA). Vybrant multidrug resistance assay kit and calcein-AM were purchased from Molecular Probes (Eugene, OR). All test compounds and reagents used were of analytical grade.

Cell Cultures. Caco-2 cells, originating from a human colorectal carcinoma, were maintained in tissue culture flasks in Dulbecco's modified Eagle medium (DMEM pH 7.4), supplemented with 1% L-glutamine, 1% MEM (nonessential amino acids), and 10% fetal calf serum (FCS) at 37 °C and 10% CO_2 in 95% relative humidity. Three weeks before the

(26) Cianchetta, G.; Singleton, R. W.; Zhang, M.; Wildgoose, M.; Giesing, D.; Fravolini, A.; Cruciani, G.; Vaz, R. J. A Pharmacophore Hypothesis for P-Glycoprotein Substrate Recognition Using GRIND-Based 3D-QSAR. *J. Med. Chem.* **2005**, *48*, 2927−2935.

(27) Wang, Y.-H.; Li, Y.; Yang, S.-L.; Yang, L. Classification of Substrates and Inhibitors of P-Glycoprotein Using Unsupervised Machine Learning Approach. *J. Chem. Inf. Comput. Sci.* **2005**, *45*, 750−757.

(28) Abbara, C.; Rouchon, C.; Hosten, B.; Farinotti, R.; Bonhomme-Faivre, L. Enhanced oral bioavailability of paclitaxel by recombinant interleukin-2 in mice with murine Lewis lung carcinoma. *Drug Metab. Drug Interact.* **2004**, *20*, 219−231.

(29) Kemper, E. M.; Cleypool, C.; Boogerd, W.; Beijnen, J. H.; Van Tellingen, O. The influence of the P-glycoprotein inhibitor zosuquidar trihydrochloride (LY335979) on the brain penetration of paclitaxel in mice. *Cancer Chemother. Pharmacol.* **2004**, *53*, 173−178.

(30) Cruciani, G.; Crivori, P.; Carrupt, P.-A.; Testa, B. Molecular fields in quantitative structure-permeation relationships: The VolSurf approach. *J. Mol. Struct. (THEOCHEM)* **2000**, *503*, 17−30.

(31) Pastor, M.; Cruciani, G.; McLay, I.; Pickett, S.; Clementi, S. GRid-INdependent Descriptors (GRIND): A Novel Class of Alignment-Independent Three-Dimensional Molecular Descriptors. *J. Med. Chem.* **2000**, *43*, 3233−3243.

(32) Polli, J. W.; Wring, S. A.; Humphreys, J. E.; Huang, L.; Morgan, J. B.; Webster, L. O.; Serabjit-Singh, C. S. Rational use of in vitro P-glycoprotein assays in drug discovery. *J. Pharmacol. Exp. Ther.* **2001**, *299*, 620−628.

(33) Schwab, D.; Fischer, H.; Tabatabaei, A.; Poli, S.; Huwyler, J. Comparison of in Vitro P-Glycoprotein Screening Assays: Recommendations for Their Use in Drug Discovery. *J. Med. Chem.* **2003**, *46*, 1716−1725.

(34) Grandi, M.; Geroni, C.; Giuliani, F. C. Isolation and characterization of a human colon adenocarcinoma cell line resistant to doxorubicin. *Br. J. Cancer* **1986**, *54*, 515−518.

(35) Rivoltini, L.; Colombo, M. P.; Supino, R.; Ballinari, D.; Tsuruo, T.; Parmiani, G. Modulation of multidrug resistance by verapamil or mdr1 anti-sense oligodeoxynucleotide does not change the high susceptibility to lymphokine-activated killers in mdr-resistant human carcinoma (LoVo) line. *Int. J. Cancer* **1990**, *46*, 727−732.

(36) Fanciulli, M.; Bruno, T.; Giovannelli, A.; Gentile, F. P.; Di Padova, M.; Rubiu, O.; Floridi, A. Energy metabolism of human LoVo colon carcinoma cells: correlation to drug resistance and influence of lonidamine. *Clin. Cancer Res.* **2000**, *6*, 1590−1597.

(37) Meschini, S.; Calcabrini, A.; Monti, E.; Del Bufalo, D.; Stringaro, A.; Dolfini, E.; Arancia, G. Intracellular P-glycoprotein expression is associated with the intrinsic multidrug resistance phenotype in human colon adenocarcinoma cells. *Int. J. Cancer* **2000**, *87*, 615−628.

permeability experiment, the cells were split and placed on permeable cell culture inserts (1.0 μm BD Falcon HTS 24-multiwell insert system) at a density of 10^5 cells/cm^2. The medium, consisting of DMEM pH 7.4 supplemented with 1% L-glutamine, 1% MEM, penicillin (100 units/mL), streptomycin (100 μg/mL), and 10% FCS, was changed every other day. The LoVoDx cell line were grown in monolayer in Ham's F12 medium (Gibco BRL) supplemented with penicillin (100 units/mL), streptomycin (100 μg/mL), and 10% FCS, at 37 °C and 5% CO_2 in 95% relative humidity.

Sample Preparation. A 10 mM stock solution of every single compound was prepared in DMSO. Test samples for experiments were prepared by diluting the stock solution in Hanks' balanced salt solution (HBSS, pH 7.4).

Caco-2 Permeability Assay. The day of the experiment, cell monolayers were washed twice and preincubated with 37 °C HBSS containing 25 mM Hepes (Gibco BRL) at pH 7.4. To ensure that monolayers were confluent and tight junctions were intact, the transepithelial electric resistance (TEER) was measured using the EVOM equipment (Endohm, WPI, Germany). Only monolayers with a TEER in the range of 400−650 Ω/cm^2 were used. Permeability studies were performed in triplicate in two transport directions, absorption (a → b) and secretion (b → a). Fresh donor solution containing 10 μM test compound in HBSS was added to either the apical (a) or the basolateral (b) (donor) side, while drug-free HBSS was placed on the receiver side. The 24-transwell plates were placed on a heated plate shaker to provide the right temperature (37 °C) and agitation rate. After 120 min, the buffers from the receiving and donor chambers were collected and aliquots were analyzed via liquid chromatography with tandem mass spectrometry detection (LC−MS/MS). Monolayer integrity was checked at the end of the experiment, measuring the lucifer yellow (50 μM) P_{app} using a Packard BF1000 FluoroCount system (Meriden, CT).

CAM Assay. For inhibition experiments, approximately 80000 LoVoDx cells/well were seeded into flat-botton 96-well plates (white 96-well microplate with clear bottom, Packard, Meriden, CT). On day 2 after plating, the culture medium was removed and the cell monolayers were washed with the incubation buffer (HBSS containing 25 mM Hepes, pH = 7.4). The assay was performed as described by Tiberghien and Loor[12] using 50 μM verapamil as reference compound. Briefly, triplicate monolayers were preincubated for 30 min at 37 °C in 5% CO_2 either with 50 μM test compounds or verapamil or with blank incubation buffer. The incubation with 0.5 μM calcein-AM (Molecular Probes, Eugene, OR) was performed in the same conditions of the preincubation. The intracellular accumulation of calcein due to inhibition of P-gp was monitored every 5 min, over 1 h, by measurement of the fluorescence developed using a Packard BF1000 FluoroCount plate reader (excitation wavelength of 485 nm, emission at 530 nm).

Liquid Chromatography−Mass Mass (LC−MS/MS) Analysis. The LC system comprised an HTS PAL autosampler, 100 μL syringe (CTC Analytics, Zwingen, Switzerland), 1100 binary pump, and oven (Agilent Technologies, Palo Alto, CA). The detection was performed using an API2000 triple quadrupole mass spectrometer with built-in switching valve, Turbo ion spray source and Analyst and Automaton softwares (AB/MDS-Sciex, Foster City, CA). The chromatographic conditions were as follows: analytical guard column SB-C8 4.6 × 12.5 mm 5 μm (Zorbax−Agilent Technologies) used as analytical column. Mobile phase A: 95% ammonium formate 10 mM pH 4.0 + 5% acetonitrile. Mobile phase B: 5% ammonium formate 10 mM pH 4.0 + 95% acetonitrile. Until the first 0.35 min of the run time, the eluent, containing incubation buffer salts, is discarded. After this, the built-in divert valve is switched, sending all the eluent to the source, until 1.45 min. Total run to run time was 2 min. The Analyst software was used for peak integration.

MS/MS. Samples were analyzed in multiple reaction monitoring (MRM) mode. The turbo ion spray source was set at 4800 V, with purified air as desolvation gas at 45 psi and turbo gas at 80 psi, purified nitrogen as curtain gas at 35 psi and as collision gas, and a temperature of 400 °C. Orifice (declustering) and collision energies related to any specific compound were automatically retrieved from an Access database file, where optimal MS/MS conditions for all the compounds tested were previously stored by using Automaton and Analyst software. The Automaton process uses three flow injections: two injections to select the best polarity and declustering potential and one injection for selecting collision energy. The concentration used for Automaton was 100 μM in 50% A/50% B mobile phase for each compound, in a stream of 150 μL/min of 50% A/50% B mobile phase, with the same voltages reported above for the turbo ion spray source, which was set at 300 °C.

Caco-2 Permeability Data Analysis. The permeability coefficients (P_{app}) were calculated in cm/s using the following equation (eq 1):

$$P_{app} = \frac{dC}{dt}\frac{V_r}{AC_0} \quad \text{[cm/s]} \tag{1}$$

where C_0 is the initial concentration, dC/dt is the flux across the monolayer, V_r is the volume of the receiver compartment, and A is the membrane surface area.

Efflux ratio was determined dividing the basolateral to apical P_{app} by the apical to basolateral P_{app} (eq 2).

$$\text{efflux ratio} = \frac{\text{basolateral to apical } P_{app}}{\text{apical to basolateral } P_{app}} \tag{2}$$

CAM Data Analysis. Average fluorescence values for blank wells were subtracted from all wells. All test fluorescence values were compared as a ratio to 50 μM verapamil standard to show percent inhibition (eq 3).

$$\%\ \text{P-gp inhibition} = \frac{\text{av fluorescence test compound} - \text{av fluorescence blank}}{\text{av fluorescence verapamil} - \text{av fluorescence blank}} \times 100 \tag{3}$$

A compound was determined to inhibit P-gp when an inhibition greater than or equal to 15% was reached.

Theoretical Calculations

Molecular Structures. When available, the bidimensional structures (2D) of the 53 and 125 drugs reported in Tables 1 and 2 were extracted from National Cancer Institute, MDL Drug Data Report, and Comprehensive Medicinal Chemistry databases; the remaining structures were manually sketched using the program Cerius2.[38] The structures of the 272 proprietary compounds were extracted from our internal database. All molecules, modeled in their neutral form, were converted into three-dimensional (3D) structures using the program Corina.[39] From the 3D structures, molecular descriptors were calculated using Volsurf[40] and Almond[41] programs. All these calculations were carried out on a SGI Octane2 workstation.

Molecular Descriptors. Volsurf Descriptors. VolSurf is a molecular modeling software[40] that generates 2D molecular descriptors from the 3D molecular interaction field (MIF) on GRID maps. GRID is a computational procedure for the determination of energetically favorable binding sites between a probe and all the atoms in a molecule virtually inserted in a 3D GRID map.[42,43] The VolSurf method[30] is simple to apply and is specifically designed to calculate descriptors relevant to pharmacokinetic properties.[44,45] It has been demonstrated that most VolSurf descriptors are only marginally influenced by conformational sampling.[30,44] This is probably due to the peculiarity of the GRID force field, which allows the hydrogens and lone pairs of a molecule to be flexible and move in the close 3D space.[42,43] In this work, VolSurf calculation produced 94 descriptors (grid spacing 0.5 Å) using the water (OH2), the hydrophobic (DRY) and the carbonyl oxygen (O) probes. These well-known descriptors refer to molecular size and shape, hydrophilic and hydrophobic regions, hydrophilic and lipophilic balance, and hydrogen bond donor/acceptor capacity.

GRIND-Pharmacophore-Based Descriptors. The generation of the alignment-independent GRIND descriptors using the Almond program[41] involves different steps that are reported in detail by Pastor et al.[31] Briefly, the Almond procedure initially extracts from each MIF a fixed number of intense favorable (negative) energies of interaction (nodes). In principle, the selection of the nodes is based on two criteria: (1) the intensity of the field at a node and (2) the mutual node−node distances between the chosen nodes. The optimization of these two criteria can be tuned by giving to each of them a different relative weight. The second step in the Almond procedure involves the transformation of the extracted nodes in autocorrelograms and cross-correlograms using a method known as maximum auto- and cross-correlation (MACC-2).[46] This method computes the products of normalized interaction energy for each pair of nodes extracted at the previous step. The normalization is performed as follows: each field energy is normalized by dividing the value of the energy of the interaction by a constant representing a theoretical maximum of energy for this probe. The distance between the nodes is binned into a discrete number of categories. In each category only the highest product of node−node interaction energy is stored, while the others are discarded. In the current calculation, the following parameters were applied: the maximum number of extracted nodes for all the fields was set to 150, 35% of the weight was assigned to the distance criterion, and a smoothing window width of 0.8 grid unit was used for the node−node distance range. A total of six correlograms, containing 84 variables each, were obtained for each compound. Three are autocorrelograms obtained by analyzing node−node interactions belonging to the same MIF (i.e., DRY−DRY, N1−N1, O−O). The remaining three are cross-correlograms obtained by analyzing node−node interactions belonging to different pairs of MIFs (i.e., DRY−N1, DRY−O, N1−O). Each correlogram is a compact representation of the geometrical relationship between energetically favorable regions of the MIFs.

Partial Least Squares Discriminant (PLSD) Analysis. PLSD,[47] implemented in both Volsurf and Almond programs, were applied to identify correlations between the calculated descriptors and P-gp related classes. The Volsurf descriptors centered and scaled to unit variance were used for PLSD analysis.

Results and Discussion

A Volsurf-Based Model for Discriminating between P-gp and Non-P-gp Substrates. A set of 53 drugs were classified as P-gp and non-P-gp substrates according to the

(38) Cerius^2, version 4.10; available from Accelrys Inc., San Diego, CA (www.accelrys.com).

(39) CORINA, version 3.2; available from Molecular Networks, GmbH, Computerchemie, Erlangen, Germany (www.mol-net.de).

(40) Volsurf, version 4.0; available from Molecular Discovery Ltd., London, U.K. (www.moldiscovery.com).

(41) Almond, version 3.2.; available from Molecular Discovery Ltd., London, U.K. (www.moldiscovery.com).

(42) Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849−857.

(43) GRID, version 22; available from Molecular Discovery Ltd., London, U.K. (www.moldiscovery.com).

(44) Crivori, P.; Cruciani, G.; Carrupt, P.-A.; Testa, B. Predicting blood−brain barrier permeation from three-dimensional molecular structure. *J. Med. Chem.* **2000**, *43*, 2204−2216.

(45) Crivori, P.; Zamora, I.; Speed, B.; Orrenius, C.; Poggesi, I. Model based on GRID-derived descriptors for estimating CYP3A4 enzyme stability of potential drug candidates. *J. Comput.-Aided Mol. Des.* **2004**, *18*, 155−166.

(46) Clementi, M.; Clementi, S.; Clementi, S.; Cruciani, G.; Pastor, M.; Nilsson, J. E. Chemometric detection of binding sites of 7TM receptors. In *Molecular Modeling and Prediction of Bioactivity*; Gundertofte, K., Jorgensen, F. S., Eds.; Kluwer Academic/Plenum Publishers: New York, 2000; pp 207−212.

(47) Wold, S.; Johansson, E.; Cocchi, M. PLS-partial least-squares projections to latent structures. In *3D QSAR in Drug Design-Theory Methods and Applications*; Kubyni, H., Ed.; Escom Science Publishers: Leiden, 1993; pp 523−550.

Table 1. Experimental Caco-2 Bidirectional Permeability (P_{app}), Efflux Ratio (ER), and Percentage of Inhibitory Activity versus Verapamil Using a CAM Assay for the 53 Drugs

		P_{app}(a→b) (10^{-6} cm/s)		P_{app}(b→a) (10^{-6} cm/s)			calcein-AM assay
no.	name	av[a]	SD[b]	av	SD	ER[c]	% inhibn vs verapamil[d]
1	astemizole	7.0	0.9	3.9	0.3	0.6	123
2	atenolol	0.2	0.4	0.6	0.5	2.5	−6
3	β-estradiol	1.2	0.5	1.2	0.7	1.0	nd[e]
4	caffeine	52.4	8	30.2	2.8	0.6	−6
5	carbamazepine	52	14	40.2	15.6	0.8	−1
6	chloroquine	2.8	3.5	4.5	4.7	1.6	−1
7	chlorpromazine	19.1	3.2	12.1	3.2	0.6	74
8	cimetidine	1.5	0.5	6.9	1.0	4.7	−1
9	clonidine	67.7	4.8	36	3.7	0.5	nd
10	clozapine	45.2	3.3	31.9	6.0	0.7	38
11	corticosterone	45.4	3.0	34.7	5.7	0.8	nd
12	cyclosporin A	0.2	0.2	2.6	1.1	11.2	nd
13	desipramine	43.4	7.5	24.5	3.2	0.6	nd
14	digoxin	2.5	0.5	12.8	3.7	5.1	−2
15	diltiazem	44	4.2	37.8	4.1	0.9	51
16	doxorubicin	0.1	0.1	0.8	0.9	7	nd
17	enoxacin	2.9	0.9	5.2	0.7	1.8	nd
18	erytromycin	0.5	0.5	8.8	1.9	16.3	−2
19	etoposide	0.1	0.1	3.4	1.0	24.9	−2
20	haloperidol H	19.7	1.1	15.9	1.4	0.8	28
21	hydrocortisone	10.5	2.9	12.6	5.7	1.2	nd
22	imipramine	39.7	8.5	27.3	5.9	0.7	nd
23	isoxicam	5.0	4.6	17.7	13.4	3.5	nd
24	ketoconazole	8.5	4.4	23.6	5.6	2.8	129
25	lansoprazole	37.3	6	30.6	5.6	0.8	22
26	levofloxacin	4.2	1.6	9.7	1.7	2.3	−4
27	lomefloxacin	4.0	1.7	6.4	1.9	1.6	nd
28	loperamide	22	4.6	27.5	5.6	1.3	nd
29	mebendazole	31.1	6.4	22.9	14.8	0.7	27
30	methotrexate	0.1	0.2	0.3	0.2	1.9	−4
31	neostigmine	0.7	0.2	0.8	0.4	1.1	−5
32	nicardipine	16.8	1.1	12.1	0.2	0.7	154
33	nifedipine	60.3	12.2	63.4	12.1	1.1	22
34	norfloxacin	0.8	0.2	3.1	0.7	4.0	nd
35	ofloxacin	3.6	0.4	7.0	0.5	2.0	nd
36	omeprazole	47.1	3.4	47.3	6.1	1.0	16
37	piroxicam	32	11.3	37.2	8.4	1.2	nd
38	prazosin	15.1	7.5	35.4	6.6	2.3	27
39	prednisolone	4.9	1.0	17.2	6.4	3.5	−4
40	progesterone	39.7	16.6	39.7	1.6	1.0	nd
41	promazine	41.8	2.2	25.6	1.3	0.6	nd
42	quinidine	28.4	12.9	37	1.4	1.3	124
43	quinine	31.8	6.4	33.7	4.0	1.1	48
44	ranitidine	1.0	0.3	2.9	0.5	3.0	2
45	rhodamine 123	0.3	0.2	11.3	0.2	36.5	nd
46	tamoxifen	0.8	1.1	0.3	0.4	0.4	112
47	paclitaxel	0.2	0.3	14	6.2	91.1	nd
48	tenoxicam	10.6	5.8	27.7	0.6	2.6	nd
49	terfenadine	0.7	1.0	1.7	1.3	2.4	nd
50	testosterone	1.9	0.0	1.7	0.0	0.9	nd
51	trimethoprim	25.2	6.7	28.7	3.7	1.1	0
52	verapamil	45.6	8.5	43.1	4.5	0.9	100
53	vinblastine	0.4	0.2	68.6	67.7	161	nd

[a] Average of triplicate determinations. [b] Standard deviation. [c] Drug was considered as P-gp substrate when it showed an ER >2. Enoxacin and methotrexate, with an ER close to 2, were also considered as P-gp substrates. [d] Drug was considered as P-gp inhibitor when it showed % inhibition ≥15. [e] Not determined.

Table 2. Compounds Names[a] and Actual[b] and Predicted[c] Classes for the Test Set Analyzed in the PLSD GRIND-Based Model

	class			class	
name	actual	predicted	Name	actual	predicted
actinomycin D	S[d]	S	topotecan	S	S
aldosterone	S	S	trazodone	S	I
amoxicillin	S	S	vincristine	S	S
amprenavir	S	S	vindesine	S	S
bisantrene	S	S	vindoline	S	S
bunitrolol	S	S	vinorelbine	S	S
catharanthine	S	S	8-geranyl-chrysin	I	I
ciprofloxacin	S	S	8-geranyl-dehydrosilybin	I	S
clotrimazole	S	I[e]	8-prenyl-chrysin	I	I
colchicines	S	S	8-prenyl-dehydrosilybin	I	S
daunarubicin	S	S	8-prenyl-galagin	I	I
debrisoquine	S	S	apigenin	I	S
delavirdine	S	S	atorvastatin	I	S
dexamethasone	S	S	atovaquone	I	I
dibucaine	S	S	azacyclonol	I	I
digitoxin	S	S	azelastine	I	I
docetaxel	S	S	benidipine	I	I
domperidone	S	S	broussochalcone A	I	S
efavirenz	S	S	carvedilol	I	I
elitriptan	S	I	chalcone	I	I
endosulfan	S	S	chlorzoxane	I	I
epirubicin	S	S	chrysin	I	I
farnesol	S	S	cyproheptadine	I	I
fexofenadine	S	S	dehydrosilybin	I	S
Hoechst33342	S	S	dexrazoxane	I	S
hydroxyrubicin	S	S	doxazosin	I	I
indinavir	S	S	felodipine	I	I
iodipine	S	S	fentanyl	I	I
irinotecan	S	S	flavone	I	I
levodopa	S	S	fluoxetine	I	I
loratadine	S	I	fluphenazine	I	I
melphalan	S	S	fluvoxamine	I	I
methyldigoxin	S	S	galagin	I	S
methylprednisolone	S	S	josamycin	I	S
mithramycin	S	S	lidocaine	I	I
mitoxantrone	S	S	loratadine	I	I
monesin	S	S	maprotiline	I	I
morphine	S	S	mefloquine	I	I
morphine-6-glucuronide	S	S	methadone	I	I
nevirapine	S	S	mifepristone	I	S
nortriptyline	S	S	NSC665333	I	I
NSC328426	S	S	NSC667739	I	I
NSC339281	S	S	NSC676590	I	I
NSC359449	S	S	NSC676591	I	I
NSC630176	S	S	NSC676597	I	S
NSC640085	S	I	NSC676599	I	I
NSC66490	S	S	NSC676600	I	I
ondansetron	S	I	NSC68075	I	I
paclitaxel	S	S	pantoprazole	I	I
pafenolol	S	S	piperine	I	I
phenobarbital	S	S	probenecid	I	I
phenoxazine	S	S	promethazine	I	I
phenytoin	S	S	propafenone	I	I
podophyllotoxin	S	S	quercetin	I	S
propiconazole	S	I	sertraline	I	I
puromycin	S	S	simvastatin	I	I
rapamycin	S	S	spironolactone	I	S
rifampin	S	S	sufentanil	I	I
risperidone	S	I	thioridazine	I	I
sparfloxacin	S	S	trifluoperazine	I	I
tacrolimus	S	S	trimipramine	I	I
talinolol	S	S	valinomycin	I	S
teniposide	S	S			

[a] The total number of compounds disagrees with that reported in the original list taken from ref 27. In fact, the compounds in common with our training set were removed. In addition, molecules listed twice or listed as both substrate and inhibitor in the original reference were removed. Moreover, the structures of five substrates and five inhibitors were not found. [b] Classification taken from ref 27. [c] Predicted classes by the PLSD GRIND-based model. [d] Substrate. [e] Inhibitor.

Table 3. Accuracy Analysis for the Training Set (Repredicted classes) of the P-gp Substrate/Nonsubstrate Model Developed Using Physicochemical (i.e., Volsurf) Descriptors

	predicted class				
actual class	no. of P-gp substrates	no. of nonsubstrates	% interclass accuracy	% precision	% overall accuracy
no. of P-gp substrates	19	3	86.4	86.4	88.7
no. of nonsubstrates	3	28	90.3	90.3	

efflux ratio (ER) of the basolateral to apical (P_{app}(b→a)) and to the apical to basolateral (P_{app}(a→b)) transports across a Caco-2 cell monolayer. Although the expression of P-gp protein was monitored by flow cytometry analysis (data not shown), the Caco-2 cell line can also express different levels of other ABC transporters. Consequently it is likely that some of the efflux ratios analyzed could be a result of multiple transporter mechanisms.

All compounds with an ER greater than or equal to 2 were classified as P-gp substrates. Two compounds, i.e., enoxacin and methotrexate, with an ER close to 2 were also considered as P-gp substrates. The Caco-2 permeability data and efflux ratios are reported in Table 1. The training set including 22 P-gp and 31 non-P-gp substrates, respectively, was modeled by partial least squares discriminant analysis using 94 Volsurf descriptors. The final PLSD model, after leave-one-out cross-validation, had two significant latent variables (LVs), and the PLSD t1−t2 score plot is shown in Figure 1. Several parameters were calculated to evaluate the statistical significance of the model developed. The model showed an overall accuracy of 89% (calculated as the total number of compounds correctly assigned to each class divided by the total number of compounds), interclass accuracy of 86% for the P-gp substrate class (calculated as the ratio between true positives and true positives plus false negatives), and 90% for the non-P-gp substrate class (calculated as the ratio between true negatives and true negatives plus false positives) and a precision of 86% for the P-gp substrate (calculated as the ratio between true positives and true positives plus false positives) and 90% for the non-P-gp substrate (calculated as the ratio between true negatives and true negatives plus false negatives) classes. These parameters are summarized in Table 3. The number of compounds utilized to develop the model was too small to be divided into a training and test sets; therefore, 272 compounds characterized in terms of ER from Caco-2 permeability measurements were extracted from our internal database to be used as external test set. Since the low permeability of molecules poses difficulties to the accurate determination of ER, all selected compounds were highly permeable in the apical to basolateral Caco-2 measurement (i.e., permeability greater than 15×10^{-6}cm/s). One hundred fifteen of the 272 compounds with an experimental ER greater than 2 were classified as P-gp substrates while the remaining 157 were classified as non-P-gp substrates. The projections of the P-gp substrates and nonsubstrates on the PLSD model t1−t2 score plot, reported in Figure 2a and Figure 2b, respectively, clearly show the ability of the model to discriminate between the two classes of compounds. The model correctly classified 72% of the entire evaluation set with an interclass accuracy of 61% for the P-gp substrates and 81% for the nonsubstrates. All statistical results for this test set are reported in Table 4. This model, which is easily reproducible, appears a valuable in silico screening tool that allows selecting compounds not potentially transported by P-glycoprotein. In addition, the projection and, consequently, the analysis of the relative positioning of synthesized or virtual libraries on the t1−t2 score plot allows a fast selection or prioritization of the most promising one.

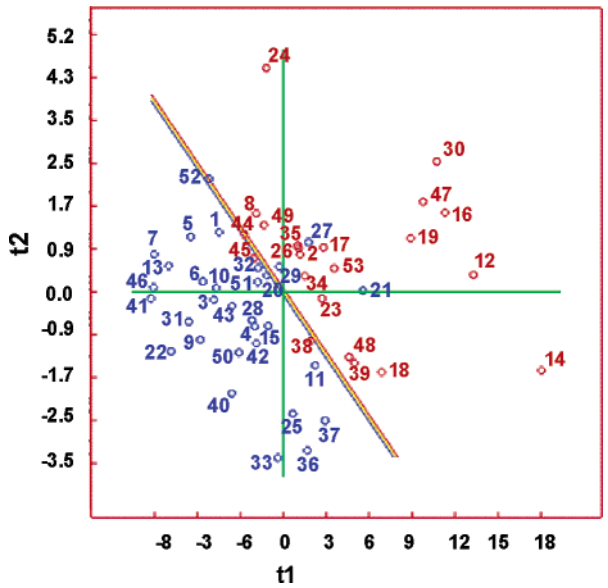

Figure 1. PLSD t1−t2 score plot derived from the analysis of the Volsurf descriptors calculated for the training set. Red and blue numbers refer to P-gp substrates and nonsubstrates, respectively. The numbers correspond to the compound names listed in Table 1.

Since the Volsurf descriptors are easily interpretable, it was straightforward to identify the most relevant physicochemical properties characterizing the P-gp substrates of this study. The plot of coefficients, which shows the contribution of all Volsurf descriptors to explain the PLSD model, is reported in Figure 3. Variables with the highest positive and negative values are the most significant to explain the P-gp substrate characteristics. Descriptors related to the size, shape, and flexibility of the molecule such as molecular surface (S), globularity (G), and elongation (Elon) (see Figure 3 and Table 5) were directly correlated with P-gp substrates. Other authors[16,19] observed a similar relationship between ATPase activity and surface area. Wide hydrophilic regions and their relative spatial arrangements (descriptors W1−W3; Iw1−Iw3, Emin3, d12, d13, and d23 in Figure 3 and Table 5)

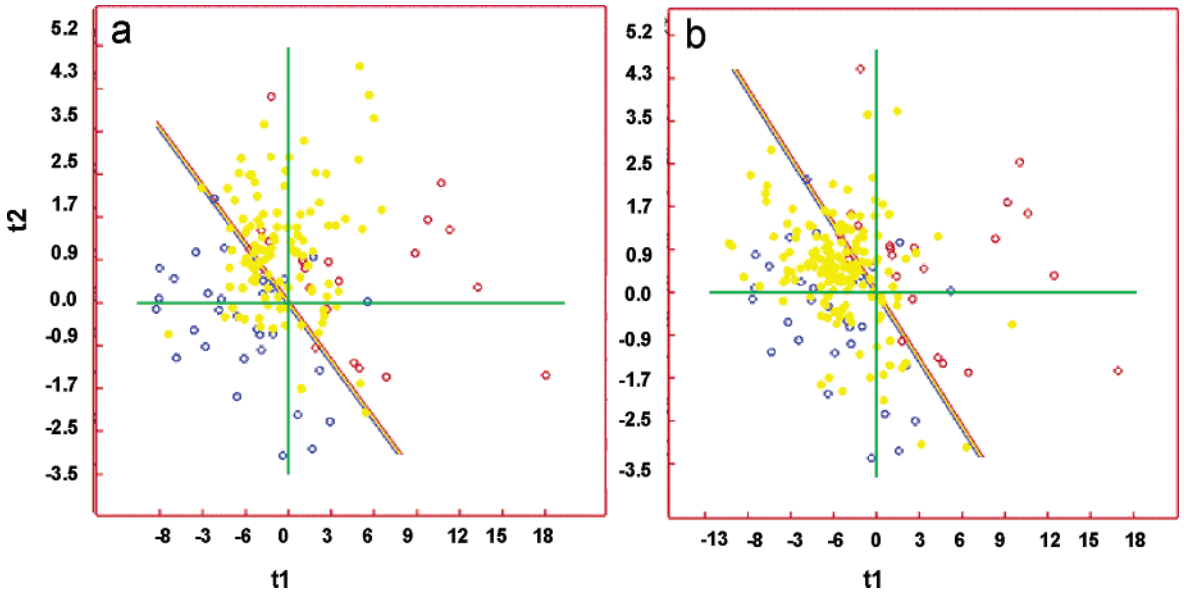

Figure 2. Projection of the test set including (a) 115 P-gp substrates and (b) 157 nonsubstrates (both indicated as yellow circles) on the PLSD-Volsurf model t1−t2 score plot. Red (substrates) and blue (nonsubstrates) circles represent the training set.

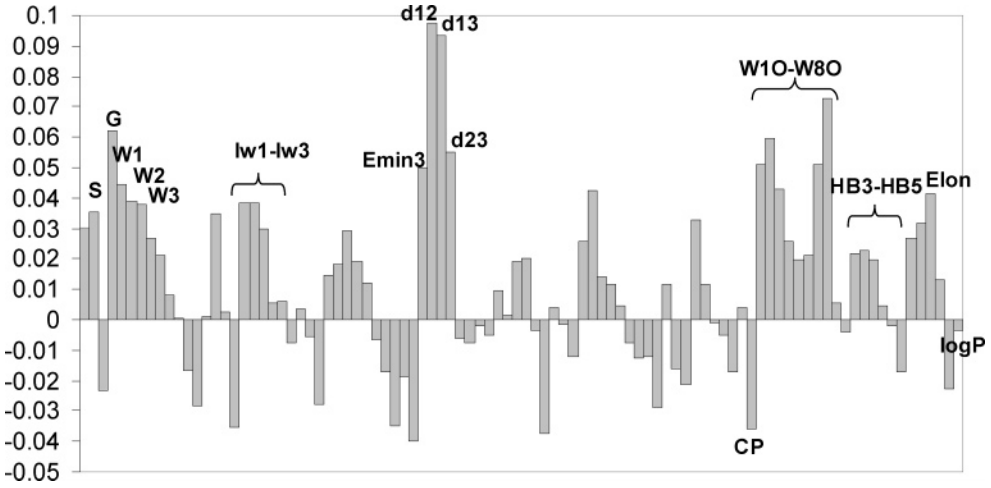

Figure 3. PLSD coefficient plot for all descriptors used in the P-gp substrate/nonsubstrate model. The most important descriptors for identifying P-gp substrates are highlighted.

Table 4. Accuracy Analysis for the Test Set Used To Validate the P-gp Substrate/Nonsubstrate Model

	predicted class				
actual class	no. of P-gp substrates	no. of nonsubstrates	% interclass accuracy	% precision	% overall accuracy
no. of P-gp substrates	70	45	60.9	70.0	72.4
no. of nonsubstrates	30	127	80.9	73.8	

and H-bonding potential (W1O−W8O and HB3−HB5 descriptors representing H-bond donor and acceptor regions, respectively) were characteristic of the P-gp substrates, as already reported in previous studies.[16−19,22,26] Conversely, a low value of critical packing (CP) contributes to explain P-gp substrate characteristics. As shown in Figure 3, the octanol/water partition coefficient (log *P*) had a very marginal role in discriminating between P-gp substrates and nonsubstrates. This was already pointed out by different authors.[16,26] However, an appropriate hydrophobic−hydrophilic balance of P-gp substrates is probably a prerequisite for membrane partitioning and binding to P-gp as it was already underlined by Seelig and Landwojtowicz.[18]

A GRIND-Based Descriptor Model for Discriminating between P-gp Substrates and Inhibitors. Thirty compounds of the 53 drugs listed in Table 1 were also evaluated for P-gp inhibition using a CAM assay. Their relative potency as P-gp inhibitors compared to the verapamil reference

Table 5. Definitions of the Most Important Volsurf Descriptors to Explain P-gp Substrate Characteristics

	Descriptors Obtained from the Water Interaction Fields
V	volume of the water molecule interaction field at 0.2 kcal/mol energy level
S	surface of the water interaction field at 0.2 kcal/mol energy level
G	globularity: ratio between the surface (S) and the surface of a sphere with the same volume (V)
W1−W8	volumes of the water molecule interaction fields at eight different energy levels: −0.2, −0.5, −1.0, −2.0, −3.0, −4.0, −5.0, and −6.0 kcal/mol
Iw1−Iw8	integy moments: distances between the center of mass of the molecule and the center of the hydrophilic regions calculate at the same 8 energy levels as W1−W8
Emin1, Emin2, Emin3	energy values for the three lowest energy minima
d12, d13, d23	distances between the three lowest energy minima
	Descriptors Obtained from the Water and DRY Interaction Fields
CP	critical packing describes a ratio between the hydrophobic and hydrophilic parts of a molecule
	Descriptors Obtained from the Carbonyl and Water Interaction Fields
W1O−W8O	volumes of the carbonyl molecule interaction fields at eight different energy levels: −0.2, −0.5, −1.0, −2.0, −3.0, −4.0, −5.0, and −6.0 kcal/mol
HB1−HB8	hydrogen bond donor capability; differences between the volumes of the water interaction fields and the carbonyl interaction fields at 8 different energy levels: −0.2, −0.5, −1.0, −2.0, −3.0, −4.0, −5.0, and −6.0 kcal/mol
	Descriptor Obtained Directly from the Molecular Structure
Elon	elongation, the maximum extension a molecule could reach if properly stretched

compound are reported in Table 1. Combining the information obtained with Caco-2 and CAM measurements, two sets of compounds including 9 P-gp substrates and 14 inhibitors, respectively, were selected for modeling purposes. Compounds without any P-gp interacting properties or having both P-pg inhibitory and substrate characteristics were not included into the analysis. The rationale was to elucidate the 3D-pharmacophoric features of substrates having poor or no inhibitory activity and inhibitors having no evidence of significant transport.

As opposed to Volsurf, the GRIND descriptors used in this analysis were dependent on the molecular conformation selected. This is a generic limitation of the 3D-QSAR methods that cannot be overcome in the absence of information of the most probable bioactive conformations. However, on the basis of our experience with GRIND descriptors, this can be partially overcome if the same conformational analysis and/or 3D converter is consistently applied. In this work, the 3D geometries of the compounds analyzed were obtained automatically from their 2D structures using the program Corina, as described in Theoretical Calculations. This method produced extended conformations of reasonable low energy for all the compounds. Although some of the studied drugs contained protonatable groups at physiological pH, all molecules were modeled in their neutral form. This choice was based on indications from Pajeva and Wiese[25] that found similar pharmacophore patterns for either neutral or protonated P-gp interacting compounds. The relationship between the P-gp substrate and inhibitor classes and the 504 GRIND descriptors was determined using the PLSD analysis. After filtering out the descriptors having no variability in the two compound classes, the number of variables decreased to 434. The model obtained, after leave-one-out cross-validation, had two significant LVs. The excellent discrimination between substrates and inhibitors is shown in the PLSD t1−t2 score plot (Figure 4). The robustness of the model was confirmed by predicting an external set of 125 drugs that included 68 substrates and 56 inhibitors taken from the literature[27] (Table 2). The comparisons between predicted and actual classes for this test set are summarized in Table

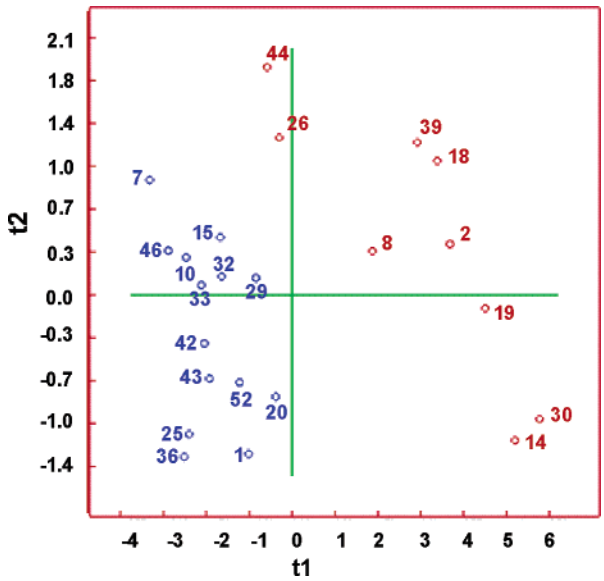

Figure 4. PLSD t1−t2 score plot derived from the analysis of the GRIND descriptors calculated for the training set. Red and blue numbers refer to P-gp substrates and inhibitors, respectively. The numbers correspond to the compound names listed in Table 1.

Table 6. Accuracy Analysis of the Test Set Used To Validate the P-gp Substrate/Inhibitor Model Developed Using GRIND Descriptors

	predicted class				
actual class	no. of P-gp substrates	no. of nonsubstrates	% interclass accuracy	% precision	% overall accuracy
no. of P-gp substrates	61	8	88.4	81.3	82.4
no. of nonsubstrates	14	42	75.0	84.0	

2. The overall accuracy was 82%, and the accuracy in predicting the substrates (88%) was better compared to that obtained for the inhibitors (75%) All relevant results are reported in Table 6.

In order to identify which 3D pharmacophoric features differentiate inhibitors (i.e., variables with the lowest negative coefficients) from substrates (i.e., variables with the highest positive coefficients), the PLSD coefficient plot (Figure 6) was analyzed more in-depth. The descriptors of the DRY−DRY autocorrelogram (i.e., 11−28 and 11−29 in Figure 5) represent the optimal distances (around 11.5 Å) that should separate the hydrophobic regions of a P-gp inhibitor. P-gp inhibitors were also characterized by high values of the N1−N1 descriptors (33−19 and 33−20 in Figure 5), which were related to the presence of favorable interacting regions placed 8 Å apart around two H-bond acceptor groups. Favorable interacting regions around a hydrophobic region and a H-bond acceptor group separated by a distance of 4.0 Å also differentiated P-gp inhibitors (cross-correlogram DRY−N1: 13−7, 13−9, and 13−10 descriptors in Figure 5) from substrates. A graphical representation of all these 3D-pharmacophoric features for astemizole inhibitor as reference compound is reported in Figure 6. The key elements of this pharmacophore are similar, to some extent, to those proposed by Ekins et al.[23] In fact, they obtained four pharmacophores

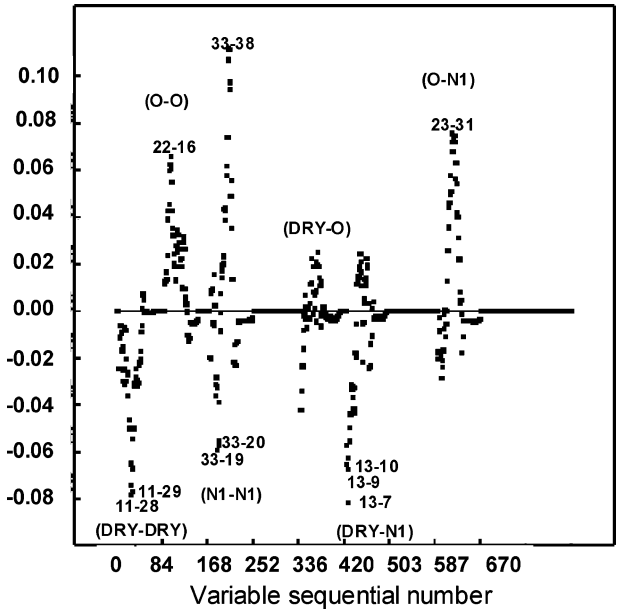

Figure 5. PLSD coefficient plot for all GRIND descriptors used in the P-gp substrate/inhibitor model. The most important descriptors to explain P-gp substrates and inhibitors are highlighted.

derived from different sets of molecules and in vitro data that contained combinations of H-bond acceptors, H-bond donors, hydrophobes, and ring aromatic features.[23]

The most important descriptors for P-gp substrates were mainly related to H-bonding potential properties. The analyzed compounds were characterized by intense O−O (i.e., the regions around H-bond donor groups, see PLS coefficient 22−16 in Figure 5), N1−N1 (i.e., the regions around the H-bond acceptor groups; see PLS coefficient 33−38 in Figure 5), and O−N1 (i.e., the regions around the H-bond donor and acceptor groups; see PLS coefficient 23−31 in Figure 5) interaction energies placed respectively 6.5, 15, and 12.5 Å apart. A picture summarizing the relevant pharmacophoric features for the P-gp substrate cimetidine is reported in Figure 7.

Cianchetta et al.[26] recently published a quantitative model, based on a combination of GRIND and Volsurf descriptors, developed using a different set of P-gp substrates. Although the two models were not directly comparable, a key recognition element, i.e., two hydrogen-bond acceptors groups around 11.5[26]−15 Å apart (since the Almond descriptors are highly dependent on the selected conformations, a tolerance for the observed distances identified in the two models should be taken into consideration) was found in both studies. To some extent, this finding reinforces the reliability of our model.

As already discussed, the 3D geometries considered in this analysis could not accurately represent the bioactive conformations. Afzelius et al.[48] proposed a method based on

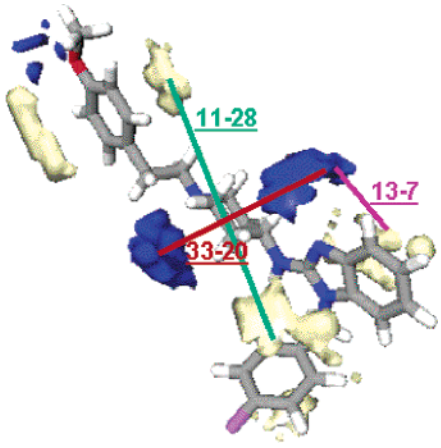

Figure 6. A graphical representation of the most important 3D-pharmacophoric GRIND features for the astemizole inhibitor, as reference compound. The colored areas around the molecule are the GRID MIF calculated with the DRY (yellow) and N1 (blue) probes.

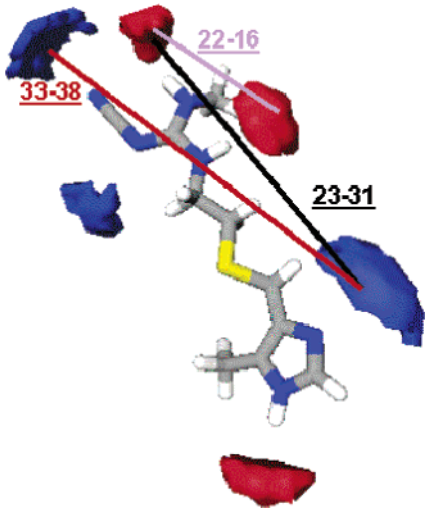

Figure 7. A graphical representation of the most important 3D-pharmacophoric GRIND features for the cimetidine substrate, as reference compound. The colored areas around the molecule are the GRID MIF calculated with the O (red) and N1 (blue) probes.

flexible GRID molecular interaction fields for calculating conformer-independent GRIND descriptors. This approach is currently under evaluation in our group in order to assess the sensitivity of the reported model to the 3D conformation selected.

It is also important to underline that the experimental data used for developing such a model were unable to provide strong indications on the recognition site and mechanism (i.e., competitive, noncompetitive inhibition, etc.) of molecular P-gp interaction. However, a common set of pharmacophoric features was identified for both P-gp substrates and P-gp inhibitors, respectively. Thus, a common mechanism of action was assumed to characterize each group of the analyzed compounds. Without any additional experimental information, hypotheses on the fine mechanisms cannot be drawn. Although we recognize the limitation of this model, this in-depth understanding of the main 3D pharmacophoric features that discriminate the P-gp substrates from inhibitors appears valuable for the design of new drug candidates.

Conclusions

The computational model for discriminating P-gp substrates and nonsubstrates, developed using various structurally diverse drugs and calculated physicochemical descriptors, was both predictive and robust, as indicated by the results obtained with an external validation set. The model interpretation through PLSD coefficient analysis was also in good agreement with the known molecular features influencing the interactions with P-gp. The model, easily reproducible, can be used as a virtual screening tool in early discovery programs for assisting chemists in the selection of the most promising libraries to be further characterized.

The GRIND-based model was built in order to elucidate the pharmacophore patterns that mainly differentiate the P-gp substrates having poor or no inhibitory activity from inhibitors having no evidence of significant transport. Also in this case the robustness of the model was showed by the evaluation of an external set of 125 drugs. For both classes of compounds, the identification of the most relevant functional groups and their appropriate spatial arrangement provided valuable information that would be helpful in the design of new candidate drugs.

Both models suffered from the statistical limitations characteristic of the conventional classification methods, i.e., the presence of uncertainty in predictions close to the border separating classes; nevertheless, they showed in general a good predictive performance and acceptable classification errors.

In conclusion, a combined and rational use of in vitro assays and in silico tools gives a competitive advantage in early discovery phases for predicting and understanding the potential pharmacological and pharmacokinetic implications of P-gp interacting compounds. When more specific assays, targeting individual P-gp sites and modulator mechanisms, as well as more detailed 3D structural information on P-gp protein become available, it will be possible to develop computational models addressing the fine mechanisms of these interactions. These models will become useful in the later stage of drug discovery.

MP050071A

(48) Afzelius, L.; Zamora, I.; Masimirembwa, C. M.; Karlen, A.; Andersson, T. B.; Mecucci, S.; Baroni, M.; Cruciani, G. Conformer- and Alignment-Independent Model for Predicting Structurally Diverse Competitive CYP2C9 Inhibitors. *J. Med. Chem.* **2004**, *47*, 907−914.