Available online at www.sciencedirect.com
ScienceDirect
International Journal of Pharmaceutics 343 (2007) 98–105
INTERNATIONAL JOURNAL OF PHARMACEUTICS
www.elsevier.com/locate/ijpharm

Development of an *in silico* model for predicting efflux substrates in Caco-2 cells

Litai Zhang *, Praveen V. Balimane, Stephen R. Johnson, Saeho Chong

Bristol-Myers Squibb Company PRI, Route 206 & Provinceline Road, P.O. Box 4000, Princeton, NJ 08543, USA

Received 13 March 2007; received in revised form 4 May 2007; accepted 8 May 2007
Available online 18 May 2007

Abstract

P-glycoprotein (P-gp) is an ATP dependent efflux transporter protein that has been demonstrated to play a critical role in affecting the absorption, metabolism, elimination and toxicity (ADMET) characteristics of a large number of marketed drugs. Therefore, it is important to evaluate whether or not compounds of interest are likely to interact with P-gp and/or other efflux transporters. An *in silico* efflux substrate (potential substrate of P-gp and or other transporters) classification model has been developed based on *in vitro* bi-directional Caco-2 cell permeability and five descriptors, using 14 marketed drugs and >100 discovery compounds synthesized at Bristol-Myers Squibb PRI. The model suggests that efflux substrates tend to contain electron deficient aromatic rings, are highly branched, and most contain tertiary nitrogen. This model demonstrated ~80% predictability of both non-substrates and substrates from a training set of 125 compounds. For a validation set of 46 compounds the predictability was ~72% for non-substrates and ~89% for substrates. The model has the potential to be used both as a filter for library designs to identify potential efflux substrates in early discovery as well as a primary screening methodology to identify the efflux substrate potential of drug candidates.
© 2007 Elsevier B.V. All rights reserved.

Keywords: P-gp substrates; Efflux; *In silico* model; Caco-2 cell; Drug discovery

1. Introduction

Recent reports have put the final price tag for bringing a drug to the market at a staggering US$ 1.7 billion dollars with an estimated research time running into multiple years (FDA, 2004). Keeping in mind the tremendous amount of time, resources and money that goes into bringing new drugs to the market, it is imperative for the pharmaceutical industry to constantly look for smarter ways of doing research. Some recent efforts to reduce cost and expedite the drug discovery cycle include combinatorial chemistry, proteomics, genomics, robotics and miniaturization. In addition to these technological advances, new drug design efforts incorporate a parallel approach to drug discovery where the pharmacological efficacy is screened in parallel to the initial ADMET profiling of compounds, providing more information for selecting compounds with superior quality for further development. One of the cornerstones of such an approach, however, is the availability of highly accurate, cost effective and high throughput screening techniques that can provide fast and reliable read-outs on the developability characteristics of discovery compounds. Such screening techniques can help select compounds with a greater probability of succeeding in the clinic and also provide guidance to the medicinal chemists in designing better compounds. Thus, the task of screening discovery compounds for biopharmaceutical properties (e.g. solubility, intestinal permeability, metabolic stability, etc.) in a most efficient manner has become a major challenge facing the industry.

Despite innovations in novel drug delivery systems in the recent past, the oral route still remains the preferred route of administration by virtue of its convenience and better compliance. For a compound to be developed as a successful oral medicine it should have sufficient pharmacological potency coupled with adequate structural and bio-pharmaceutical attributes to reach the site of action. Amongst the biopharmaceutical properties that need to be considered in early discovery, permeability assessment and P-glycoprotein (P-gp) interaction studies can be two of the most critical properties in determining the fate of a compound (Kim, 2002; Lin, 2003). Transport of drug substances across the intestinal membrane is a complex and dynamic

* Corresponding author. Tel.: +1 609 252 3872; fax: +1 609 252 6030.
E-mail address: Litai.zhang@bms.com (L. Zhang).

0378-5173/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijpharm.2007.05.017

process. It includes the passage of compounds across various functional pathways in parallel. Passive transport occurs through the cell membrane of enterocytes (transcellular) or via the tight junctions between the enterocytes (paracellular). Various influx and efflux mechanisms (via carriers and transporters) play a key role in the disposition and efficacy of the compound. A drug efflux transporter such as P-gp is known to be a major determinant of absorption, distribution and elimination of a wide variety of drugs (Polli et al., 1999; Matheny et al., 2001; Sababi et al., 2001; Lin, 2003; Lin and Yamazaki, 2003) P-g p is known to limit the oral absorption of drugs such as cyclosporin and taxol; it can limit entry of drugs such as HIV protease inhibitors into brain/CNS; and it can actively facilitate excretion of drugs via biliary and urinary routes. Since P-gp interactions of a drug can play such a pivotal role in dictating their pharmacokinetics, increasing efforts are being made in early discovery and development to identify compounds that can potentially interact with P-gp.

There are literature reports of various *in vitro* and *in vivo* models that are used for assessing P-gp interactions with test compounds (Adachi et al., 2001; Polli et al., 2001; Yamazaki et al., 2001; Perloff et al., 2003; Balimane et al., 2006). *In vitro* assays such as ATPase assay, rhoadmine-123 uptake assay, calcein AM uptake assay, cell based bi-directional assay, radio-ligand binding assay along with *in vivo* models such as transgenic (knockout mice) and mutant animal models are most commonly used. Several *in vitro* models, though capable of high throughput screening due to automation/miniaturization, are typically not functional since they provide only binding potential with P-gp. *In vivo* animal models, on the other hand, are more predictive and functional in nature but are not practical because of their high cost and limited throughput. Because of these limitations, the cell based bi-directional permeability assays are currently the method of choice for P-gp substrate identification in drug discovery labs (Polli et al., 2001; Perloff et al., 2003; Balimane et al., 2004).

Concerted efforts by several investigators (Seelig, 1998; Ekins et al., 2002; Penzotti et al., 2002; Stouch and Gudmundsson, 2002; Gombar et al., 2004; Cianchetta et al., 2005) have resulted in different levels of success with regards to *in silico* predictability of P-gp interaction of discovery compounds. Researchers have demonstrated correlations between physicochemical properties, molecular weight, polar surface area, H-bonding capabilities and the compounds propensity to interact with P-gp (Wang et al., 2003; Gombar et al., 2004; Xue et al., 2004; Wang et al., 2005). Development of a successful *in silico* model that could accurately predict whether or not a discovery compound will interact with efflux transporters (P-gp and or other's) as a substrate can significantly enhance drug discovery efforts. It would provide valuable insights into intestinal absorption, tissue distribution, brain penetration and liver/biliary elimination of test compounds. These criteria can certainly help select the optimal compounds for development and save tremendous amount of resources and cost.

This paper describes the development and validation of an *in silico* classification model for efflux based on *in vitro* bi-directional Caco-2 cell permeability, using 14 marketed drugs and >100 discovery compounds synthesized at Bristol-Myers Squibb PRI.

2. Materials and methods

Caco-2 cells (passage #17) were obtained from the American Type Culture Collection (Rockville, MD). Dulbecco's modified eagle's medium, non-essential amino acids, L-glutamine and antibiotics were purchased from JHR Biosciences (Lenexa, KS). Fetal bovine serum was obtained from Hyclone Lab. Inc. (Logan, Utah). HTS-Transwell® inserts (surface area: 0.33 cm^2 for 24 well) with a polycarbonate membrane (0.4 μm pore size) were purchased from Costar (Cambridge, MA). Hank's balanced salt solution (HBSS) and *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid (HEPES) were purchased from Sigma Chemical Co. (St. Louis, MO). All solvents were analytical grade. ^{3}H-Digoxin, ^{14}C Mannitol and ^{14}C Taxol were obtained from Perkin Elmer Life Sciences (Boston, MA). All other test compounds were obtained from Sigma Chemical Co. (St. Louis, MO).

2.1. Caco-2 cell culture procedure

Caco-2 cells were seeded onto filter membrane at a density of ∼100,000 cells/cm^2. The cells were grown in culture medium consisting of Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum, 1% non-essential amino acids, 1% L-glutamine, 100 U/mL penicillin-G, and 100 μg/mL streptomycin. The culture medium was replaced every 2 days and the cells were maintained at 37 °C, 95% relative humidity, and 5% CO_2. Permeability studies were conducted with the monolayers cultured for approximately 21 days with the cell passage numbers between 50 and 80. Physiologically and morphologically well-developed Caco-2 cell monolayers with TEER values greater than 400 ohm × cm^2 were used for the studies reported in this manuscript. Radiolabeled mannitol was used as a control to insure the integrity of cell monolayer in all permeability experiments.

2.2. Caco-2 cell bi-directional efflux substrate assay

The transport medium used for the bi-directional studies was modified HBSS buffer containing 10 mM HEPES. The pH of both the apical and basolateral compartments was 7.4. Prior to all experiments, each monolayer was washed twice with buffer and Trans-Epithelial Electrical Resistance (TEER) was measured to ensure the integrity of the monolayers. The concentration of test compounds was typically 50 μM in this assay. The bi-directional permeability studies were initiated by adding an appropriate volume of buffer containing test compound to either the apical (for apical to basolateral transport; A to B) or basolateral (for basolateral to apical transport; B to A) side of the monolayer. Volume of the apical and basolateral compartment was maintained at 0.2 and 0.6 mL, respectively. The monolayers were then placed in an incubator for 2 h at 37 °C. Samples were taken from both the apical and basolateral compartment at the end of the 2 h period and the concentrations of test compound were analyzed by a

HPLC-UV assay. Permeability coefficient (Pc) was calculated according to the following equation:

$$\mathrm{Pc} = \frac{\mathrm{d}A}{\mathrm{d}tSC_\mathrm{o}},$$

where Pc is permeability in nm/s, dA/dt the flux of the test compound across the monolayer (nmole/s), S the surface area of the cell monolayer, and C_o is the initial concentration (50 μM or measured starting concentration) in the donor compartment.

Caco-2 cells were used for developing this *in silico* model since it is the most widely used cell line in the pharmaceutical industry for the last 2 decades. Regulatory agency (FDA, 2006) as well as several investigators have supported the use of Caco-2 for performing efflux studies since they have adequate expression of several efflux transporter proteins. Caco-2 cells are known to express several efflux transporters such as P-gp, breast cancer resistance protein (BCRP), multi-drug resistance protein (MRP2), etc. (Taipalensuu et al., 2001; Englund et al., 2006; Seithel et al., 2006) Thus, the bi-directional permeability studies in Caco-2 cells is used to identify overall efflux substrates rather than solely P-gp substrates. A compound was considered to be an efflux substrate if the ratio of its apparent permeability in the B to A direction to that in the A to B direction was >2. It was considered to be a non-substrate if the ratio was ≤2. A cut-off value of 2 was selected based on the historical dataset within our labs as well as guidance from the literature (Polli et al., 2001). All data used for the *in silico* modeling were from studies performed in triplicate with mass balance of >60% in both directions. Mannitol (paracellular probe) and digoxin (classical P-gp substrate) were included as quality controls in every test run to confirm the performance of the Caco-2 cells with respect to the integrity of cell monolayer and efflux transporter functionality.

2.3. In silico model development

After omitting compounds that had recovery (mass balance) less than 60% in either the A to B or B to A Caco-2 cell permeability assays, 113 substrates and 58 non-substrates were available for constructing the model for efflux substrate recognition. Selecting compounds with adequate recovery was critical to insure that only the high quality data with minimum non-specific binding were incorporated in the model development. A total of 157 compounds were selected from 23 internal research programs in 10 different therapeutic areas. The remaining 14 compounds were marketed drugs representing a diverse chemical space with a wide range of physicochemical and pharmacological properties. The physicochemical properties of the

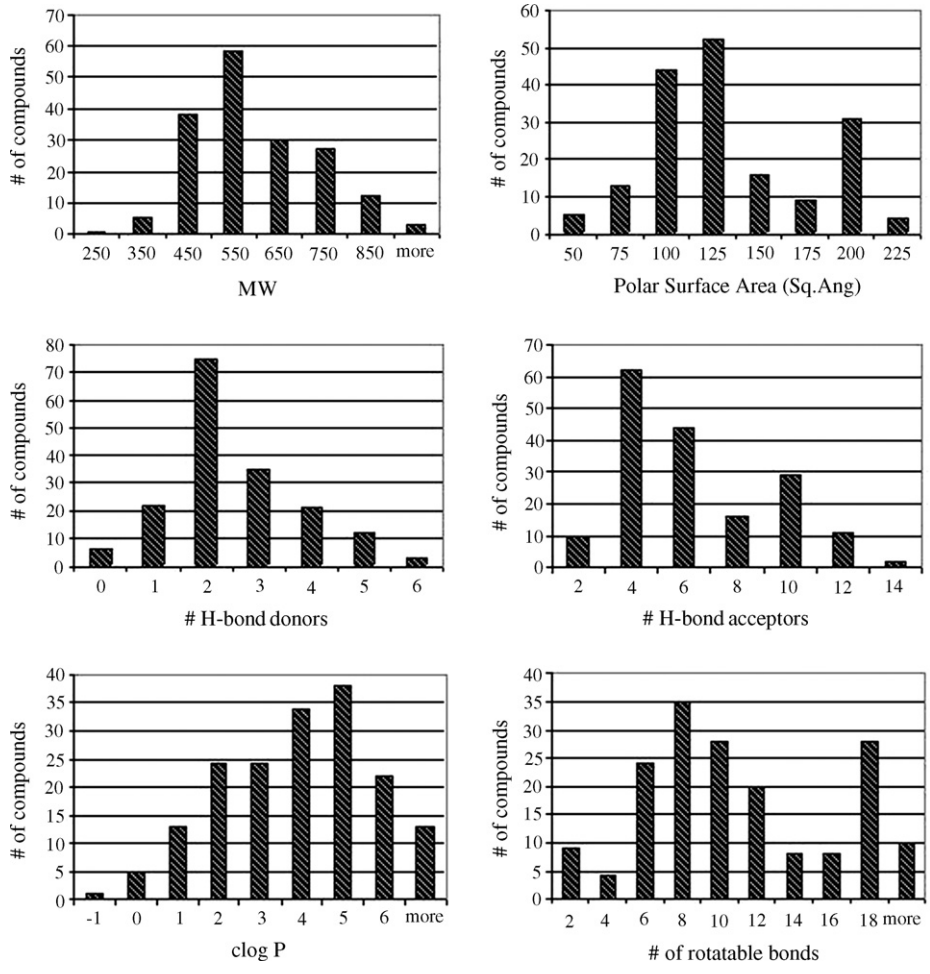

Fig. 1. Expanded rule of five parameters.

171 structures are depicted in Fig. 1 as an extended rule of five (Lipinski et al., 1997).

SMILES notation of 171 compounds were converted to 3D structures using Cerius^2 (Accelrys, Inc., San Diego, CA), then minimized using the Merck Molecular Force Field (MMFF) in Maestro (Schrodinger, LLC, New York). The energy minimized structures were stored in an SD file format then imported into a Cerius^2 study Table. Ninety-nine 2D and 3D structural descriptors were initially computed for the training and validation sets. These included Cerius^2 spatial, electro-topological, and physicochemical parameters, various fragment keys based on Alogp atom types. Eighty-nine descriptors were used to build classification trees after disregarding descriptors with zero variance. Pair-wise correlations were performed to ensure the orthogonality of descriptors used for the final model.

To ensure uniform coverage of substrates and non-substrates in both training and validation sets, the 113 substrates and 58 non-substrates were clustered separately using Daylight fingerprints with group average clustering and a Tanimoto similarity coefficient of 0.8. There were 66 substrates clusters, 46 of the 66 were singletons, the average number of compounds for the rest 19 clusters was 3 except the first cluster which consisted 26 compounds. There were 39 non-substrates clusters, 29 of the 39 were singletons, the average number of compounds for the rest 10 clusters was 3. Twenty compounds were selected randomly from the first cluster of the substrates to represent this cluster. The training set was made of those 20 compounds plus cluster centers of both substrates and non-substrates; the rest of the compounds were designated as the validation set. In total 125 compounds were used for training the model and 46 compounds were used for validating the model.

The recursive partitioning (RP) method (Breiman et al., 1984; Hawkins et al., 1997; van Rhee et al., 2000; Mazzatorta et al., 2004) implemented in Cerius^2 was used to develop and validate a classification tree to separate substrates from non-substrates. The RP method recursively partitions data according to a relationship between the X and Y values, creating a tree of partitions. It finds a set of cuts or groupings of X values that best predict a Y value. It does this by exhaustively searching all possible cuts or groupings. These splits (or partitions) of the data are done recursively forming a tree of decision rules until the desired fit is reached. RP parameters have to be optimized to build a balanced tree predicting both classes in the training set evenly. The parameters listed in Table 1 were used for the final RP model.

Table 1
Parameters used for the final recursive partitioning model

Name of Parameters	Values
Type of weighting	Classes
Score splits	Gini impurity
Pruning factor	3
Minimum number of samples per node	6
Limit knots per variable	40
Maximum Lookahead node	0
Maximum tree depth	10

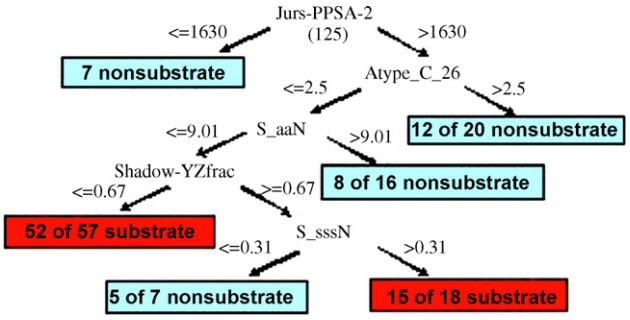

Fig. 2. Decision tree for the training set compounds.

The resulting efflux classification model is based on five descriptors, and is depicted for the training set in Fig. 2. The selected features include physicochemical, shape, and atom-type descriptors. The prediction of validation set is depicted in Fig. 3. Each split is based on a descriptor value, which partitions compounds in order to separate substrates from non-substrates. For example, the first split separates 7 out of 32 non-substrates on the basis of descriptor Jurs-PPSA-2. This node is a pure node containing only non-substrates. The Jurs-PPSA-2 is the total charge weighted positive surface area, calculated as:

$$\text{Jurs-PPSA-2} \equiv \left(\sum_{i}^{N_{\text{pos}}} \text{SA}_i\right) \times \left(\sum_{j}^{N_{\text{pos}}} q_j\right)$$

where SA_i is the solvent-accessible surface area of the ith atom with a positive partial charge, q_j is the charge on the jth atom with a positive partial charge, both summed over all atoms in the molecule with a partial positive charge. Atype-C-26 is a count of aromatic carbons bonded to a heteroatom with a single bond. This carbon is typically partially positive charged. Shadow-YZfrac is the area of the molecule projected onto the YZ-plane divided by the area of the bounding box of the molecule in this same plane. The area of the molecule in the YZ plane represents the minimum aspect ratio of the molecule. Once normalized to the size of the bounding box, this descriptor essentially determines the compactness of the molecule in the YZ plane. S_aaN is an electrotopological state key descriptor of a N bonded to two other atoms with aromatic bonds, while the S_sssN descriptor is for a tertiary N. Histograms of these descriptors are shown in Fig. 4.

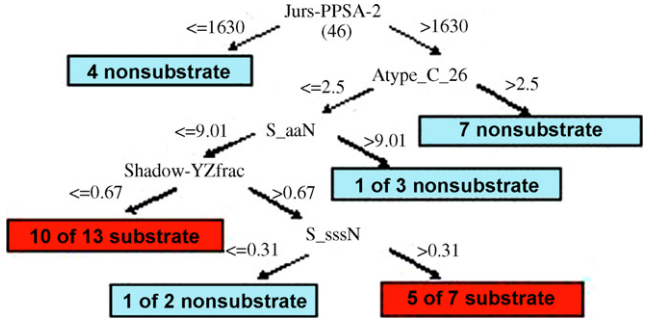

Fig. 3. Decision tree for the validation set compounds.

3. Results

As shown in Fig. 2, the resulting RP model resembles a tree with 6 leaves, or terminal nodes. Only two nodes result in prediction that a compound is a substrate for efflux.

Based on the model, efflux substrates generally tend to be electron deficient (Jurs-PPSA-2 > 1630) and have few substituted phenyl rings (atype-c-26 ≤ 2.5), no more than two aromatic nitrogens (S_aaM ≤ 9.01). Additionally, efflux substrates have either (1) a sparsely filled minimum aspect ratio (e.g. a compound with significant branching as opposed to a fairly linear molecule), or (2) at least one tertiary nitrogen. The model also highlights compounds with two or more aromatic nitrogens (S_aaN > 9.01) as a group with significant number of substrates, but with no structural features that clearly identifies them. The model suggests that efflux substrates are likely to contain electron deficient aromatic rings, are likely to be highly branched, and most of them contain tertiary nitrogens.

It is worth highlighting that this interpretation is drawn from the descriptors used in the model as well as the data used to train and validate the model. Consequently, the resulting rules may have significant limitations that will only become apparent with continued forward application. This future application, however, must stay within the confines of the structural descriptors used in this report. Fig. 4 shows the distributions of the descriptors of the model. Users should be aware when new compounds are represented by descriptors that are outside of these bounds, or in sparsely populated regions of this descriptor space, as these extrapolations are likely to have significant prediction errors.

Figs. 5 and 6 summarize the prediction for the selected RP model for the training and validation sets. This RP model based on five descriptors correctly predicted 80% of both non-substrates and substrates from a training set of 125 compounds. Viewed another way, 89% (67 of 75) of the compounds predicted to be substrates by the model were actually substrates. Only 64% (32 of 50) of the compounds predicted to be non-substrates were actually non-substrates. This difference may have ramifications in how the model should be applied in real time. Specifically, the model appears to identify substrates more reliably than non-substrates.

For a validation set of 46 compounds, 72% of non-substrates and 89% of substrates were correctly predicted with this model. All 14 marketed drugs are predicted correctly by the model. The experimental and predicted classifications of the 14 marketed drugs are listed in Table 2. Higher predictability of ∼89% observed for predicting the efflux substrates in contrast to lower (though, quite acceptable) predictability of ∼72% for non-substrates of the validation set can be explained by

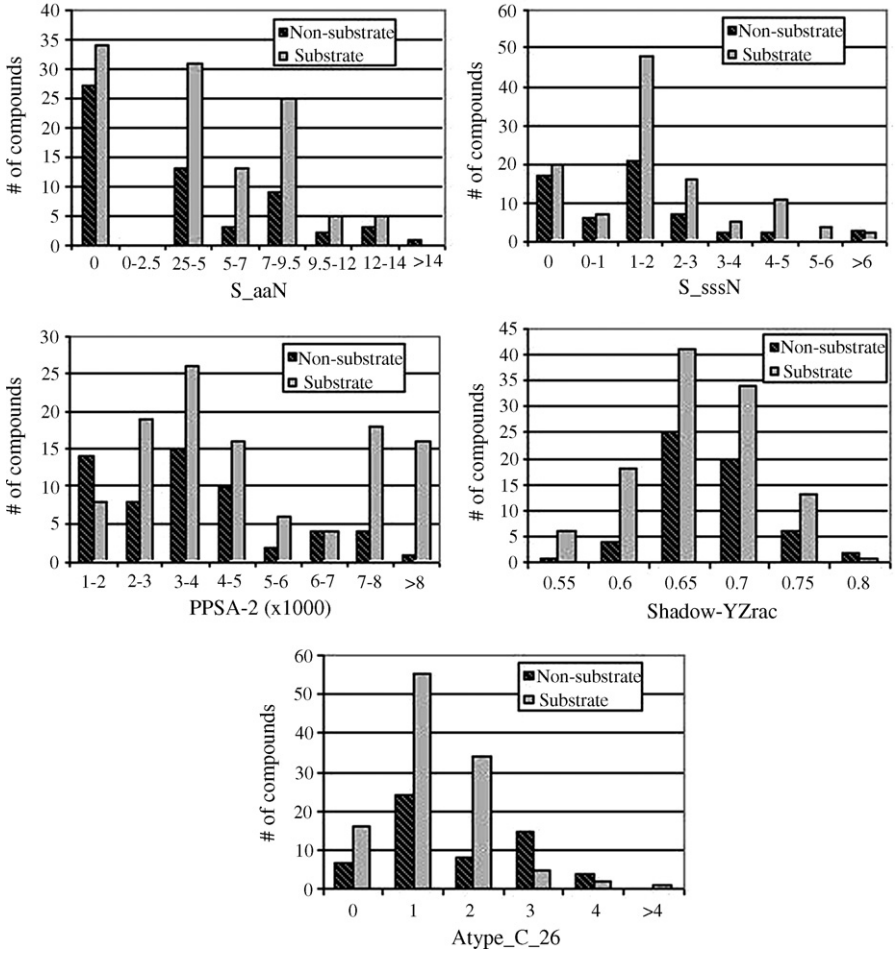

Fig. 4. Training set histograms of features used in efflux model.

Training Data	Predicted non-substrates	Predicted substrates	Total
Observed non-substrates	32	8	40
% correct prediction	80		
Observed substrates	18	67	85
% correct prediction		79	
Total	50	75	125

Fig. 5. Confusion matrix for the training set.

the nature of the *in vitro* data used to develop the *in silico* model.

There are some special scenarios under which there is a possibility that compounds that are efflux/P-gp substrates might get erroneously identified as non-substrates by the *in vitro* assay. First, P-gp is a transporter protein that can be saturated and for high affinity (low K_m) compounds, the efflux transport could be over-whelmed at study concentrations (50 μM) falsely identifying a compound as a non-substrate. Second, test compounds that are completely impermeable (transcellular passive permeability is negligible) often demonstrate baseline permeability in both directions (efflux ratio would be ∼1) but they cannot be conclusively classified to be non-substrates. Third, compounds with very high intrinsic permeability (passive permeability value high in both direction) might again demonstrate efflux ratio lower than 2 despite being efflux/P-gp substrates. Thus, the slightly lower predictability observed in accurately identifying the non-substrates might be a function of the quality of experimental data rather than the caliber of the *in silico* model. Verapamil, a known efflux/P-gp susbtrates, is listed as a non-substrate via both experimental and predicted mode in Table 2. Since the Caco-2 cell studies were all performed with test compounds at 50 μM, verapamil, a high permeability compound, saturates the efflux transporters leading to a high Pc value in both direction resulting in a net efflux ratio of ∼1. Under optimized conditions (starting concentrations in single digit μM) verapamil would demonstrate a net efflux ratio of greater than 3 and would be labeled as a substrate. However, under present experimental conditions, it gets identified as a non-substrate. Since the objective of this paper was to develop a practical *in silico* model based on data that are normally going to be available in research settings, we have considered permeability data only from 50 μM and not preferentially selected any data at lower concentrations.

Validation Data	Predicted non-substrates	Predicted substrates	Total
Observed non-substrates	13	5	18
% correct prediction	72		
Observed substrates	3	25	28
% correct prediction		89	
Total	16	30	46

Fig. 6. Confusion matrix for the validation set.

Table 2
Experimental and predicted classification of marketed drugs

Drugs	Experimental	Predicted
Digoxin	Substrate	Substrate
Etoposide[a]	Non-substrate	Non-substrate
Indinavir	Substrate	Substrate
Indomethacin	Non-substrate	Non-substrate
Mannitol	Non-substrate	Non-substrate
Methotrexate	Non-substrate	Non-substrate
Metoprolol	Non-substrate	Non-substrate
Probenecid	Non-substrate	Non-substrate
Rhodamine 123	Substrate	Substrate
Dexamethasone	Substrate	Substrate
Taxol	Substrate	Substrate
Verapamil	Non-substrate	Non-substrate
Vincristine	Substrate	Substrate
Vinblastine	Substrate	Substrate

[a] Etoposide demonstrated an efflux ratio of ∼1.5 and was considered to be a non-substrate (since ratio was <2).

An important consideration in model development is an understanding of the variability of the data being used to construct the model. The experimental error may lead to incorrect categorization of compounds in the training set. Historical data from this assay have demonstrated that variability of data from the bidirectional Caco-2 cell experiment used here is generally between 2 and 5 fold, with compounds showing higher ratio having variability in the higher end of that range. These projected error rates were used in a Monte Carlo simulation to gauge the effect of the quality of the data used to derive the *in silico* model. The data were perturbed by a normal distribution of error, within an error rate based on the Caco-2 cell B-A/A-B ratio. The perturbed data was then used to classify compounds as substrates or non-substrates. The agreement between the categorization using the experimental values and the categorization using the perturbed data is calculated. This process was repeated for 1000 iterations, building a distribution of correlations. This process results in the cumulative distribution plot shown in Fig. 7 that indicates the median percent agreement ∼85% based on the expected error. This value is likely the upper bound on the performance of an *in silico* model. Indeed, it is probably overly optimistic as it does not account for the atypical errors of greater than 20-fold.

When this *in silico* efflux model was applied to predict a set of 25 compounds tested after the model was developed, the overall classification was 84% accuracy. All nine non-substrates were predicted to be non-substrates. Of the substrates, 63% were correctly predicted as substrates. The model performed reasonably well considering two substrates had subtle structure modifications of non-substrates (CF_3 instead of Cl). Addition-

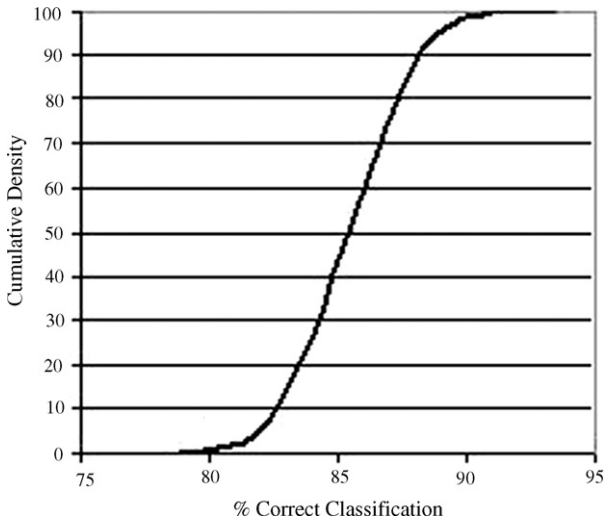

Fig. 7. Monte Carlo simulation results for the maximum correct classification based on assay variability.

ally, two other substrates were structurally very dissimilar to the training data used to develop the efflux model as indicated by the fact that the most similar training set compounds had a Tanimoto similarity coefficient of less than 0.5. Another prediction was performed with 115 newly tested compounds, 75% substrates were predicted as substrates, 72% non-substrates predicted as non-substrates. These predictions show the generality of the model to predict substrates/non-substrates. The failure to predict subtle structure changes of a particular chemotype indicated that the model should not be used as a local model for structural modifications. The model should be used in the early stage of drug discovery to screen compounds for efflux substrates.

4. Discussion

Due to the variety of modeling techniques, and datasets employed, comparing the model described here with others presented in the literature is a difficult task. The approach we have described here focuses on a combination of physicochemical properties (PPSA-2), shape (SHDW-YZfrac), and pharmacophoric atom types (atype-C-26, S_aaN, S_sssN). The approach employed by Cianchetta et al. (2005) de-emphasized the significance of diffusion related physicochemical properties, probably because the dataset excluded compounds shown to be non-substrates in the bidirectional Caco-2 cell assay. Instead, they focused only on compounds with a B-A/A-B permeability ratio in Caco-2 cells greater than 1 which may have significantly reduced the impact on diffusion related considerations. Ekins et al. (2002) derived a substrate pharmacophore from an overlay of two known substrates, upon which they overlayed a third compound to match. Xue et al. (2004) employ a support vector machine in their modeling. Several of the descriptors they highlight are similar to those presented here. However, the coefficients are not disclosed, making it impossible to interpret the degree to which a descriptor makes it more/less likely to be a efflux/P-gp substrate.

5. Conclusions

A highly predictive *in silico* model based on experimental data and five descriptors capable of assessing the efflux substrate (P-gp and or other's) potential of discovery compounds was developed and validated. The model highlights important structural features for efflux substrate recognition. The predictive *in silico* model developed here affords a primary screening tool for discovery scientists to identify potential efflux substrates. With the increased understanding of the critical role played by various transporter proteins in affecting the pharmacokinetics of test compounds, there has been a renewed focus to identify drug-transporter interactions as early as possible in discovery cycle. *In vitro* cell-based models to identify efflux substrates are routinely utilized by discovery organizations but the cell-based assays require valuable resources, time, and cost (Marino et al., 2005). Additionally, the cell-based models often lack the high-throughput potential that is essential to effectively provide timely feedback to synthesis groups for Structure Activity Relationship (SAR) efforts. Utilization of a predictive *in silico* model as a "primary" screening tool can be invaluable in providing guidance to medicinal chemistry groups regarding the efflux interaction potential of novel compounds. Such efforts can guide the chemotypes away from potential efflux substrates and help lead the compound libraries towards more "developable" characteristics. Thus, *in silico* models such as the one discussed in this manuscript can be used effectively as a primary filter for library design and screening discovery compounds for efflux interaction potential.

References

Adachi, Y., Suzuki, H., Sugiyama, Y., 2001. Comparative studies on in vitro methods for evaluating in vivo function of MDR1 P-gp. Pharm. Res. 18, 1660–1668.

Balimane, B.V., Han, Y.-H., Chong, S., 2006. Current industrial practices of assessing permeability and P-glycoprotein interaction. AAPS 8, E1–E13.

Balimane, P.V., Patel, K., Marino, A., Chong, S., 2004. Utility of 96 well Caco-2 cell system for increased throughput of P-gp screening in drug discovery. Eur. J. Pharm. Biopharm. 58, 99–105.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, a.C.J., 1984. Classification and Regression Trees.

Cianchetta, G., Singleton, R.W., Zhang, M., Wildgoose, M., Giesing, D., Fravolini, A., Cruciani, G., Vaz, R.J., 2005. A Pharmacophore hypothesis for P-glycoprotein substrate recognition using GRIND-based 3D-QSAR. J. Med. Chem. 48, 2927–2935.

Ekins, S., Kim, R.B., Leake, B.F., Dantzig, A.H., Schuetz, E.G., Lan, L.-B., Yasuda, K., Shepard, R.L., Winter, M.A., Schuetz, J.D., Wikel, J.H., Wrighton, S.A., 2002. Application of three-dimensional quantitative structure-activity relationships of P-glycoprotein inhibitors and substrates. Mol. Pharmacol. 61, 974–981.

Englund, G., Rorsman, F., Ronnblom, A., Karlbom, U., Lazorova, L., Grasjo, J., Kindmark, A., Artursson, P., 2006. Regional levels of drug transporters along the human intestinal tract: co-expression of ABC and SLC transporters and comparison with Caco-2 cells. Eur. J. Pharm. Sci. 29, 267–277.

FDA, 2004. Challenges and Opportunity on the critical path to new medical products. FDA Report.

FDA, 2006. Guidance for Industry: Drug interaction studies-Study design, data analysis, and implications for dosing and labeling. FDA Report.

Gombar, V.K., Polli, J.W., Humphreys, J.E., Wring, S.A., Serabjit-Singh, C.S., 2004. Predicting P-glycoprotein substrates by a quantitative structure-activity relationship model. J. Pharm. Sci. 93, 957–968.

Hawkins, D.M., Young, S.S., Rusinko III, A., 1997. Analysis of large structure-activity data set using recursive partitioning. Quant. Struct. Act. Relat. 16, 296–302.

Kim, R., 2002. Transporter and xenobiotic disposition. Toxicology 181–182, 291–297.

Lin, J., 2003. Drug-drug interaction mediated by inhibition and induction of P-glycoprotein. Adv. Drug Deliv. Rev. 55, 53–81.

Lin, J., Yamazaki, M., 2003. Role of P-glycoprotein in pharmacokinetics. Clin. Pharmacokinet. 42, 59–98.

Lipinski, C.A., Lombardo, F., Dominy, B.W., Feeney, P.J., 1997. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv. Drug Deliv. Rev. 23, 3–25.

Marino, A., Yarde, M., Patel, H., Chong, S., Balimane, P., 2005. Validation of the 96 well Caco-2 cell culture model for high throughput permeability assessment of discovery compounds. Int. J. Pharm. 297, 235–241.

Matheny, C., Lamb, M., Brouwer, K., Pollak, G., 2001. Pharmacokinetic and pharmacodynamic implications of P-gp modulation. Pharmacotherapy 21, 778–796.

Mazzatorta, P., Benfenati, E., Lorenzini, P., Vighi, M., 2004. QSAR in ecotoxicity: an overview of modern classification techniques. J. Chem. Inf. Comput. Sci. 44, 105–112.

Penzotti, J.E., Lamb, M.L., Evensen, E., Grootenhuis, P.D.J., 2002. A computational ensemble pharmacophore model for identifying substrates of P-glycoprotein. J. Med. Chem. 45, 1737–1740.

Perloff, M., Stromer, E., von Moltke, L., Greenblatt, D., 2003. Rapid assessment of P-gp inhibition and induction in vitro. Pharm. Res. 20, 1177–1183.

Polli, J., Jerrett, J., Studenberg, J., Humphreys, J., Dennis, S., Brower, K., Wooley, J., 1999. Role of P-gp on CNS disposition of amprenavir, an HIV protease inhibitor. Pharm. Res. 16, 1206–1212.

Polli, J., Wring, S., Humphreys, J., Huang, L., Morgan, J., Webster, L., Serabjit-Singh, C., 2001. Rational use of in vitro P-gp assays in drug discovery. J. Pharmacol. Exp. Ther. 299, 620–628.

Sababi, M., Borga, O., Bengtsson, U., 2001. The role of P-gp in limiting intestinal absorption of digoxin in rats. Eur. J. Pharm. Sci. 14, 21–27.

Seelig, A., 1998. How does P-glycoprotein recognize its substrates? Int. J. Clin. Pharmacol. Ther. 36, 50–54.

Seithel, A., Karlsson, J., Hilgendorf, C., Bjorquist, A., Ungell, A., 2006. Variability in mRNA expression of ABC- and SLC-transporters in human intestinal cells: comparison between human segments and Caco-2 cells. Eur. J. Pharm. Sci. 28, 291–299.

Stouch, T.R., Gudmundsson, O., 2002. Progress in understanding the structure-activity relationships of [drugs interacting with] P-glycoprotein. Adv. Drug Del. Rev. 54, 315–328.

Taipalensuu, J., Tornblom, H., Lindberg, G., Einarsson, C., Sjoqvist, F., Melhus, H., Garberg, P., Sjostrom, B., Lundgren, B., Artursson, P., 2001. Correlation of gene expression of ten drug efflux proteins of the ATP-binding cassette transporter family in normal human jejunum and in human intestinal epithelial Caco-2 cell monolayers. J. Pharmacol. Exp. Ther. 299, 164–170.

van Rhee, A.M., Stocker, J., Printzenhoff, D., Creech, C., Wagoner, P.K., Spear, K.L., 2000. Retrospective analysis of an experimental high-throughput screening data set by recursive partitioning. J. Comb. Chem. 3, 267–277.

Wang, R.B., Kuo, C.L., Lien, L.L., Lien, E.J., 2003. Structure-activity relationship: Analyses of p-glycoprotein substrates and inhibitors. J. Clin. Pharm. Ther. 28, 203–228.

Wang, Y.-H., Li, Y., Yang, S.-L., Yang, L., 2005. Classification of substrates and inhibitors of P-glycoprotein using unsupervised machine learning approach. J. Chem. Inform. Model. 45, 750–757.

Xue, Y., Yap, C.W., Sun, L.Z., Cao, Z.W., Wang, J.F., Chen, Y.Z., 2004. Prediction of P-glycoprotein substrates by a support vector machine approach. J. Chem. Inform. Comp. Sci. 44, 1497–1505.

Yamazaki, M., Neway, W., Ohe, T., Chen, I., Rowe, J., Hochman, J., Chiba, M., Lin, J., 2001. In vitro substrate identification studies for P-gp mediated transport: Species difference and predictability of in vivo results. J. Pharm. Exp. Ther. 296, 723–735.