In Silico Prediction of Membrane Permeability from Calculated Molecular Parameters

Hanne H. F. Refsgaard,*,† Berith F. Jensen,§ Per B. Brockhoff,|| Søren B. Padkjær,‡ Mette Guldbrandt,† and Michael S. Christensen†

Departments of Drug Metabolism and Protein Structure, Novo Nordisk A/S, Novo Nordisk Park, DK-2760 Måløv, Denmark, Department of Natural Sciences, Royal Veterinary and Agricultural University, Thorvaldsensvej 40, DK-1871 Frederiksberg C, Denmark, and Informatics and Mathematical Modeling, Richard Petersens Plads, Building 321, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark

Received May 5, 2004

A data set consisting of 712 compounds was used for classification into two classes with respect to membrane permeation in a cell-based assay: (0) apparent permeability (P_{app}) below 4×10^{-6} cm/s and (1) P_{app} on 4×10^{-6} cm/s or higher. Nine molecular descriptors were calculated for each compound and Nearest-Neighbor classification was applied using five neighbors as optimized by full cross-validation. A model based on five descriptors, number of flex bonds, number of hydrogen bond acceptors and donors, and molecular and polar surface area, was selected by variable selection. In an external test set of 112 compounds, 104 compounds were classified and 8 compounds were judged as "unknown". Among the 104 compounds, 16 were misclassified corresponding to a misclassification rate of 15% and no compounds were falsely predicted in the nonpermeable class.

Introduction

Poor absorption, distribution, metabolism, and/or excretion (ADME) properties are some of the main reasons for terminating the development of chemical drug candidates.[1] This acknowledgment has resulted in development and application of a wide range of in vitro screening tests for classification of compounds with respect to ADME properties during the early drug discovery process. Increasingly, generated in vitro data are used for developing predictive models leading the way toward in silico ADME methods, reflected in a large number of recent reviews and meetings on in silico ADME methodology.[1−5] The application of in silico technology offers considerable potential for reducing the number of experimental studies required for compound selection and for improving the success rate.

In the selection of orally bioavailable drug candidates it is important to predict the absorption properties of new compounds. Absorption of a compound is driven by its solubility and membrane permeability. Intestinal wall active transports and intestinal wall metabolic events can also influence the absorption and oral bioavailability of a compound, but such events are beyond the scope of this study. Here we will focus on in silico prediction of membrane permeability by passive diffusion, and we have based our model on experimental data from two in vitro models of intestinal absorption: Transport experiments in MDCK (Madin-Darby canine kidney) Strain I and Caco-2 (human colon carcinoma) cells. Drug permeability in cell cultures has been shown to be a useful predictor of drug absorption in vivo.[6−9]

It is generally accepted that physicochemical descriptors of drug molecules can be useful for predicting drug absorption. Lipinski et al.[10] proposed "the rule of five" for a preliminary estimation of a compound's absorption on the basis of molecular weight, lipophilicity and the number of hydrogen bond donor and acceptor atoms in the molecule. Different groups have tried to predict permeability from calculated properties, and Winiwarter et al.[11] report that the best model of passive intestinal permeability use the following variables: number of hydrogen bond donors and polar surface area either alone or combined with a lipophilicity descriptor. Raevsky et al.[12] showed that the best descriptor of human intestinal absorption of 32 passive transported compounds is the sum of hydrogen bond acceptor and donor values, characterizing the total ability of a compound to form hydrogen bonds. Of 57 molecular descriptors, Agatonovic-Kustrin et al.[13] found that lipophilicity, conformational stability, and intermolecular interactions (polarity and hydrogen bonding) were the best descriptors of intestinal absorption. Faassen et al.[14] reports good agreement between predicted oral absorption properties based on water/octanol partition coefficient, polar surface area, and molecular weight and measured Caco-2 permeabilities. Additionally, PSA has been applied as a computational filter for membrane permeability.[15,16]

We here present an in silico prediction methodology for membrane permeability classification based on a large number of in vitro permeability data obtained in the same laboratory and nine calculated molecular descriptors. Furthermore, we bring information on structural fragments that are of importance for the permeability of compounds.

* Corresponding author. E-mail: HaRe@novonordisk.com. Tel: + 45 44 43 03 67. Fax: + 45 44 66 39 39.
† Department of Drug Metabolism, Novo Nordisk A/S.
‡ Department of Protein Structure, Novo Nordisk A/S.
§ Royal Veterinary and Agricultural University.
|| Technical University of Denmark.

10.1021/jm049661n CCC: $30.25
Published on Web 01/14/2005

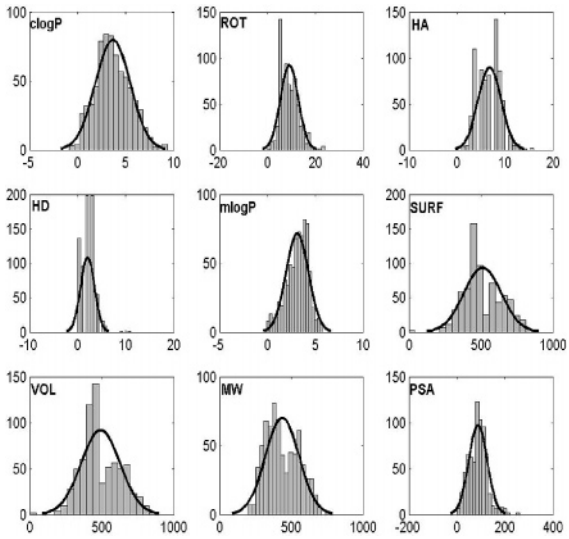


Figure 1. Distribution of the nine descriptors: clogP (water/octanol partition coefficient), mlogP (Moriguchi partition coefficient), ROT (number of flex bonds), HA and HD (number of hydrogen bond acceptors and donors), SURF and PSA (molecular and polar surface area), VOL (molecular volume), and MW (molecular weight).

Results

The Calibration Set. A total of 712 compounds were tested for membrane permeation in one of two cell lines: MDCK Strain I and Caco-2. In the MDCK Strain I transport experiment, apparent permeability, P_{app}, was measured for all compounds in the absorptive direction (AP to BL). Using the Caco-2 cell monolayers, both absorptive and secretory P_{app} was addressed in order to investigate for polarized transport, e.g. active efflux of the drug substance due to P-glycoprotein (P-gp) transport. Experimental data from Caco-2 cells were only included in the calibration data set if the compound was not subjected to polarized transport. Thus, only compounds revealing passive transcellular transport and only the absorptive (AP-BL) P_{app} values were included in the calibration data set. The calibration data set was grouped into two classes with respect to membrane permeation in the cell assays: (0) 380 compounds had P_{app} below 4×10^{-6} cm/s and (1) 332 had P_{app} on 4×10^{-6} cm/s or higher.

The 712 compounds were from 19 Novo Nordisk discovery projects and had been tested over a 2-year period. The complete structures of the compounds cannot be disclosed at this time because they are in the discovery stage at Novo Nordisk. To address the structural diversity of the 712 compounds, a clustering analysis using 166 MACCS structural keys as descriptors was performed. Of the resulting 437 clusters, the 5 highest populated clusters contained 48, 30, 23, 16, and 13 members, respectively; the next 94 clusters contained 2−10 members and 338 clusters were singletons.

Distribution for the nine calculated descriptors is shown in Figure 1, and the distribution was nearly Gaussian like that for clogP (water/octanol partition coefficient) and mlogP (Moriguchi partition coefficient), but more irregular for the other seven descriptors. The distribution curves for SURF (molecular surface area), VOL (molecular volume), and MW (molecular weight) appeared to be bimodal. The mean values for the nine calculated molecular descriptors are given in Table 1 for the two permeability classes. For all the descriptors, except for mlogP, the mean values were higher for the compounds in the nonpermeable group.

Table 1. Calculated Molecular Descriptors and Experimental Apparent Permeability, P_{app}, Values for Compounds in the Calibration Data Set[a]

	class 0 (380 compounds)		class 1 (332 compounds)		total (712 compounds)	
descriptor	mean	std. dev.	mean	std. dev.	mean	std. dev.
clogP	4	2.2	3	1.2	4	1.8
ROT	11	4.0	8	2.4	9	3.7
HA	8	2.2	6	2.1	7	2.4
HD	3	1.2	1	1.1	2	1.4
mlogP	3	1.2	3	1.1	3	1.2
SURF	567	142	447	74	511	130
VOL	551	146	425	78	492	135
MW	492	117	373	78	437	117
PSA	107	35	65	28	87	38
P_{app}	0.5	1.0	33	25	16	24

[a] Units and calculation of each descriptor: clogP (water/octanol partition coefficient), mlogP (Moriguchi partition coefficient), ROT (number of flex bonds), HA and HD (number of hydrogen bond acceptors and donors), SURF and PSA (molecular and polar surface area), VOL (molecular volume), and MW (molecular weight) are given in the Experimental Section. Permeability class: (0) P_{app} below 4×10^{-6} cm/s and (1) P_{app} on 4×10^{-6} cm/s or higher.

The correlations between the nine descriptors are given in Table 2, and the three descriptors, VOL, SURF, and MW, were highly correlated. The model with the highest Cohen's Kappa value including only one of the three descriptors was selected. The model was based on the following five descriptors: SURF, number of flex bonds (ROT), number of hydrogen bond acceptors (HA) and donors (HD), and polar surface area (PSA). The Cohen's Kappa value was at a maximum (0.72) for $k = 5$, meaning that the optimal prediction was carried out by using the five nearest neighbors, see Table 3 for the classification results with $k = 5$.

All 712 compounds were decomposed into ring fragments and functional groups as described in the Experimental Section. Selected structural groups for which we found statistical difference in frequency between the permeability classes is shown in Table 4. The results suggest that compounds containing one benzene ring or an amide group were more likely to penetrate membranes. Structural moieties such as biphenyl, naphthalene, diketopiperazine, and carboxylic acid had higher frequencies in compounds in the nonpermeable class (0). Inclusion of the six structural fragments as descriptors did not improve the prediction so the presented classification results are without the structural descriptors.

The calibration data was investigated for outliers by finding the Mahalanobis distance from each compound to the 5th nearest neighbor compound. Most distances (696) were less than 10, nine distances were between 10 and 20, four were between 20 and 50, and three were above 50. Any exclusion of these possible outliers with distances above 10 increased the cross-validation error rate, and it was decided to maintain all compounds in the calibration set.

The Test Sets. A total of 112 new compounds which did not belong to the calibration set were tested for membrane permeation in MDCK Strain I cells. Fur-

Table 2. Correlation Coefficients for the Nine Descriptors and for Apparent Permeability, P_{app}

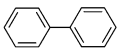
	P_{app}	clogP	ROT	HA	HD	mlogP	SURF	VOL	MW	PSA
P_{app}	1	−0.048	−0.268	−0.475	−0.558	−0.139	−0.392	−0.404	−0.479	−0.511
clogP	−0.048	1	0.431	−0.163	−0.068	0.581	0.566	0.571	0.511	−0.209
ROT	−0.268	0.431	1	0.420	0.386	−0.017	0.762	0.739	0.682	0.445
HA	−0.475	−0.163	0.420	1	0.636	−0.008	0.524	0.524	0.662	0.836
HD	−0.558	−0.068	0.386	0.636	1	−0.083	0.412	0.415	0.501	0.826
mlogP	−0.139	0.581	−0.017	−0.008	−0.083	1	0.309	0.323	0.350	−0.198
SURF	−0.392	0.566	0.762	0.524	0.412	0.309	1	0.996	0.905	0.469
VOL	−0.404	0.571	0.739	0.524	0.415	0.323	0.996	1	0.914	0.452
MW	−0.479	0.511	0.682	0.662	0.501	0.350	0.905	0.914	1	0.526
PSA	−0.511	−0.209	0.445	0.836	0.826	−0.198	0.469	0.452	0.526	1

Table 3. Cross-Validation Classification of the Calibration Data Set[a]

	predicted class 0	predicted class 1	not predicted	sum
lab test class 0	287	63	30	380
lab test class 1	20	294	18	332
sum	307	357	48	712

[a] Permeability class: (0) P_{app} below 4×10^{-6} cm/s and (1) P_{app} on 4×10^{-6} cm/s or higher.

Table 4. Occurrence and Frequency of Ring Fragments and Functional Groups in the Calibration Data Set[a]

Structure group	# class 0	# class 1	F(0)	F(1)
	124	228	0.66	1.39***
	62	7	1.68	0.22***
	49	5	1.70	0.20***
	92	188	0.61	1.44***
	111	0	1.87	0.00***
	57	1	1.84	0.04***

[a] The frequency (F) was calculated as described in the method section. Permeability class: (0) P_{app} below 4×10^{-6} cm/s and (1) P_{app} on 4×10^{-6} cm/s or higher. ***Significant difference on a 0.001 test level from the frequency of the structural group in class 0.

thermore, the nine molecular descriptors were calculated for each compound and investigated for outliers by finding the Mahalanobis distance from each compound to the 5th nearest neighbor compound in the calibration set. None of these 112 distances were above 13, so the test set compounds were all well within the space of the calibration set, and none were excluded in advance.

Mean values for ROT, HA, HD, SURF, and PSA for the 112 compounds are given in Table 5, and the classification table of the test set is given in Table 6. Among the 112 compounds, 8 compounds, corresponding to 7%, had a maximal posterior weight less than 0.60, and the permeability class for these compounds was not predicted. The 8 compounds which were not classified did not share structure patterns or values for the calculated parameters that could explain the low posterior weights. Among the 104 compounds there were classified, 16 were misclassified corresponding to a misclassification rate of 15%. Among the 53 compounds predicted into the nonpermeable class, no compounds were misclassified (Table 6).

To validate our Caco-2 and MDCK cell models, six reference drugs were tested and the obtained *P*app values are listed in Table 7. Three highly permeable drugs (metoprolol, propranolol, and verapamil) and three low permeable drugs (ranitidine, furosemide, and chloramphenicol) from the list provided in the FDA Guidance based on the Biopharmaceutical Classification System (BCS)[21] were selected for this study. The results were in agreement with earlier published data both in vitro (P_{app} in Caco-2) and in vivo (fraction of dose absorbed in human) as depicted in Table 7. The five descriptors, ROT, HA, HD, SURF, and PSA, were calculated for each drug, and the descriptors were applied for classification of permeability. The permeability class was predicted correct for metoprolol, propranolol, verapamil, furosemide, and chloramphenicol. Ranitidine was misclassified as permeable (Table 7).

Table 5. Calculated Molecular Descriptors and Experimental Apparent Permeability, P_{app}, Values for 112 External Compounds[a]

	class 0 (73 compounds)		class 1 (39 compounds)		total (112 compounds)	
descriptor	mean	std. dev.	mean	std. dev.	mean	std. dev.
ROT	11	3.4	7	2.4	10	3.7
HA	7	1.7	6	1.3	7	1.6
HD	3	1.4	1	1.0	2	1.5
SURF	583	131	425	83	528	139
PSA	96	25	67	16	85	26
P_{app}	0.3	0.7	35	27	12	23

[a] Calculation of each descriptor and units are given in the Experimental Section. Permeability class: (0) P_{app} below 4×10^{-6} cm/s and (1) P_{app} on 4×10^{-6} cm/s or higher.

Table 6. Classification of 112 External Compounds from Five Nearest Neighbors in the Calibration Data Set[a]

	predicted class 0	predicted class 1	not predicted	sum
lab test class 0	53	16	4	73
lab test class 1	0	35	4	39
sum	53	51	8	112

[a] Permeability class: (0) P_{app} below 4×10^{-6} cm/s and (1) P_{app} on 4×10^{-6} cm/s or higher.

Discussion

Absorption is driven by solubility and permeability of the compound, as well as interactions with transporters and metabolizing enzymes in the gut wall. Oral bioavailability in turn depends on a superposition of two processes: absorption and first-pass metabolism in liver and gut. The present investigation was limited to passive diffusion of compounds through cell monolayers to determine molecular properties of importance for permeability.

Physicochemical parameter-based estimation methods of permeability are attractive because of their high

Table 7. Apparent Permeability (P_{app}) and Predicted Permeability Class of Six Reference Drugs[a]

compound	$P_{app} \times 10^{-6}$ (cm/s) in MDCK[c]	$P_{app} \times 10^{-6}$ (cm/s) in Caco-2[c]	reported Caco-2 $P_{app} \times 10^{-6}$ (cm/s)	%F[d]	predicted permeability class
		High Permeability Drugs			
metoprolol	37.78 ± 0.76	18.21 ± 0.79	43.4 ± 0.7[17]	95[18]	1
propranolol	34.36 ± 2.26	12.63 ± 0.51	33.9 ± 1.8[17]	100[18]	1
verapamil	16.60 ± 0.37	7.20 ± 0.26	15.8 ± 1.2[17]	100[18]	1
		Low Permeability Drugs			
ranitidine	0.27 ± 0.03	ND[e]	1.24 ± 0.25[17]	50[19]	1
furosemide	0.27 ± 0.01	0.39 ± 0.16	0.03 ± 0.00[17]	variable[20]	0
chloramphenicol[b]	3.15 ± 0.37	ND	-	-	0

[a] Permeability class: (0) P_{app} below 4×10^{-6} cm/s and (1) P_{app} on 4×10^{-6} cm/s or higher. [b] Not listed in the FDA Guidance.[21] [c] Some of the P_{app} data were presented in the poster: Taub et al. Utilization of optimized in vitro and *in silico* methodologies for solubility, permeability, and efflux transport studies in discovery-based compound screening. Benzon Symposium, September 2001, Copenhagen, Denmark. [d] Fraction (%) of dose absorbed in human as reported in the literature. [e] ND: not determined.

throughput capacity and possible prediction of permeability category prior synthesis, but methods lacking information of real physiological conditions can be vulnerable to false predictions. The method presented here only misclassified 15% in an external data set, and no compounds were falsely predicted to belong to the nonpermeable group.

Poor solubility of compounds can cause problems with compound precipitation in the cell assays, giving false permeability results. LogP is a crucial factor governing passive membrane partitioning; an increase in logP enhances permeability while reducing solubility. Therefore, it is important to include logP as a molecular descriptor at absorption screens as suggested by Lipinski et al.[10] For ionizable molecules, the effective lipophilicity at physiological pH will not be the same as its intrinsic lipophilicity, and the distribution coefficient (logD) will be different from logP. The model is likely to improve if logD was included as a calculated property as seen by Winiwarter et al.[11] However, Raevsky et al.[12] did not find a correlation between logD and membrane penetration for 32 compounds. Furthermore, it is of interest that the model presented in this study did not include the lipophilicity descriptors: clogP and mlogP.

Hydrogen bonding has been identified as an important parameter for describing drug permeability as seen in several studies,[11−13] and this is in agreement with our findings. In a recent study Winiwarter et al.[22] found that hydrogen bond acceptor descriptors were less important for prediction of human intestinal permeability.

The calibration set was structural diverse evaluated by clustering analysis of MACCS structural keys. However, some bias toward common structure series within projects was seen. We show evidence that one benzene ring and one amide are structural groups in compounds which penetrate membranes. Biphenyl, naphthalene, diketopiperazine, and carboxylic acid were shown to be present more frequently in compounds belonging to the nonpermeable group. An explanation for the poor permeability could be poor solubility of compounds containing biphenyl and naphthalene ring systems causing the compounds to precipitate under the experimental conditions used. Aromatic ring structures, like biphenyl and naphthalene, are known to be recognized by P-gp,[23,24] and substrates for efflux transporters would have limited permeability. However, in the present study compounds subjected to polarized transport, e.g. active efflux due to P-gp transport, were not included in the calibration set. Compounds containing a carboxylic acid group will be charged at pH 6.5−7.4 where the transport experiments were performed, and that could limit the membrane penetration of the compounds. It has earlier been observed that acids generally have better oral bioavailability,[25] stressing the importance of pH in in vitro permeability assays.

The apparent permeability was tested for three highly permeable drugs (metoprolol, propranolol, and verapamil) and two low permeable drugs (ranitidine and furosemide) from the list in the FDA Guidance based on the Biopharmaceutical Classification System.[21] The results were in agreement with published data both in vitro[17] and in vivo.[18−20] Furthermore, the permeability class was predicted for the six reference drugs, and the predictions were correct for metoprolol, propranolol, verapamil, furosemide, and chloramphenicol. However, ranitidine was misclassified as permeable. A possible explanation for the false prediction of the permeability class for ranitidine could be the ranitidines presumable affinity for cellular efflux pumps as described in Yazdanian et al.[17] and Lee et al.[26]

Prediction of permeability of compounds through Caco-2 and MDCK Strain I cells are important because they are good models for human intestinal absorption.[6−9] But another central question for preclinical studies is whether the compound is absorbed in rat. Caco-2 permeability data are shown to correlate well with in vivo data both from rat and human.[27,28] In a large study based on over 1100 drug candidates Veber et al.[29] reports that the descriptors molecular flexibility as well as either polar surface area or total hydrogen bond count (sum of donors and acceptors) correlate with oral bioavailability in rat. These findings are in agreement with our model based on in vitro data.

A growing consensus is that the in silico predictions are as predictable as data obtained using in vitro tests, with the advantage that much less investment in technology, resources, and time is needed. In addition, and of critical importance, it is possible to screen before synthesis. The early assessment of ADME properties will help to select the best candidates for development, as well as to reject candidates with a low probability of success.

Conclusions

An in silico methodology for prediction of membrane permeability was developed based on data from transport experiments in cell cultures for 712 compounds and the five calculated molecular parameters, number of flex

bonds, number of hydrogen bond acceptors and donors, and molecular and polar surface area. In an external test set of 112 compounds 93% of the compounds could be classified in one of two permeability classes: (0) P_{app} below 4×10^{-6} cm/s or (1) P_{app} on 4×10^{-6} cm/s or higher. The misclassification rate was 15%, and no compounds were misclassified in the nonpermeable group. The presented in silico method would be a valuable tool in the drug discovery process to select the molecules with the greatest change of success prior to synthesis.

Experimental Section

Materials. Caco-2 cells were obtained from the American Type Culture Collection (Manassas, VA). MDCK Strain I cells were kindly donated by Dr. Weihomogeneity, Chiang Shen (Los Angeles, CA). All cell culture reagents were purchased from Life Technologies (Høje Taastrup, Denmark), unless otherwise noted. [^{14}C]Mannitol was purchased from Amersham International (Buckinghamshire, UK), and [^{14}C]testosterone, and [^{3}H]verapamil were purchased from NEN Life Science Products (Boston, MA). Sodium taurocholic acid, metoprolol, propranolol, verapamil, ranitidine, furosemide, and chloramphenicol were purchased from Sigma-Aldrich (St. Louis, MO). Hanks's Balanced Salt Solution (HBSS) and HEPES were purchased from Invitrogen, Gibco (Denmark). Ultima Gold scintillion fluid was obtained from Packard BioScience (Groningen, The Netherlands). All analytical solvents were of HPLC grade and purity.

Cell Culture. For drug transport experiments, Caco-2 cells at passage 47 were seeded in culture flasks and passaged in Dulbecco's Modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum, penicillin−streptomycin (100 U/ml and 100 μg/mL, respectively), 1% l-glutamine, and 1% nonessential amino acids. The cells (between passages 48−82) were seeded onto tissue culture-treated Transwells (Costar, NY) at a density of 10^5 cells/cm^2. Transport experiments were subsequently performed on days 26−28 after seeding. The MDCK Strain I cells (passages 14−30) were seeded onto 1 cm^2 polycarbonate filter Transwells (Costar, NY) at a density of 2.2×10^4 cells/cm^2. Culture conditions for the MDCK cells were maintained as has been described previously,[30,31] and confluent cell monolayers were obtained 5 to 7 days postseeding. Caco-2 and MDCK cell monolayer cultures were grown in an atmosphere of 5% CO_2−95% O_2 at 37 °C. Growth media were replaced every other day. Transepithelial electrical resistance (TEER) was measured in Ω cm^2, at 37 °C, using an epithelial voltohmmeter (Millicellers; Millipore, Billerica, MA). Following subtraction of background TEER, i.e., the resistance exhibited by the filter alone, the mature Caco-2 cell monolayer exhibited a TEER > 600 Ω cm^2 and the MDCK cell monolayer exhibited a TEER > 2000 Ω cm^2 prior to use in transport experiments.

Drug Permeability Studies. The integrity of the Caco-2 and MDCK monolayer was evaluated via measurement of [^{14}C]mannitol (0.4 μCi/mL) apparent permeability, P_{app}, and monitoring the change in TEER over the course of a 1 h experiment. In each experiment, [^{3}H]testosterone (0.2 μCi/ml) was used as a positive control for passive transcellular transport. In the Caco-2 transport experiments, [^{3}H]verapamil (0.2 μCi/ml) was used as an internal control for the expression and function of the P-gp efflux transporter. Prior to the experiment, growth media was removed from the mature Caco-2 and MDCK monolayers and the monolayers were rinsed once with 37 °C HBSS. All drug transport experiments were carried out using a 10 μM or 50 μM solution of the test compound. The apical (AP) volume was 400 μL and the basolateral (BL) volume was 1200 μL. In the MDCK transport experiments, test compounds were added in the AP compartment at pH 6.5 and 200 μL was removed from the BL compartment at 15, 30, and 60 min after which the BL compartment was replenished with 200 μL of fresh, preheated buffer. In the Caco-2 transport experiments, test compounds were added to either the AP or BL (donor) compartment at pH 7.4 and samples were taken from the opposite (receiver) compartment at 15, 30, and 60 min. The receiver compartment was replenished with fresh, preheated buffer. At 60 min, a 200 μL sample was taken from the donor compartment in order to establish the concentration of compound at the end of the experiment. In the P_{app} studies using the MDCK cell monolayers the HBSS contained 5 mM sodium taurocholate in the AP compartment. All experiments were conducted at 37 °C. Filters and cells containing the radioactive controls were rinsed, excised, and then counted using liquid scintillation chromatography while samples containing drug were analyzed using high-performance liquid chromatography (HPLC) or by LC/MS using a MDS Sciex API 365 triple quadrupole mass spectrometer (Toronto, Canada).

HPLC and LC/MS Analysis. A reversed-phase HPLC method was used, applying a Zorbax RX-C_8 column (Agilent, Palo Alto, CA) and a gradient buffer system. A Waters pump, controller, auto sampler, and photodiode array detector were used. Mobile phase A consisted of 100% Milli-Q water, and mobile phase B consisted of 100% acetonitrile; both mobile phases contained 0.05% trifluoroacetic acid. For LC/MS analysis, a Phenomenex Prodigy column (Torrance, CA) was used. Mobile phase A consisted of 95% Milli-Q water, 5% methanol, and mobile phase B consisted of 5% Milli-Q water, 95% methanol; both mobile phases contained 0.1% formic acid.

Calculation of Permeability. For both Caco-2 and MDCK transport experiments, the accumulated amount of drug appearing in the BL (and AP for Caco-2) compartment over time, dQ/dt, was used to calculate the P_{app} using the following equation: $P_{app} = \mathrm{d}Q/\mathrm{d}t \times 1/(AC_0)$, where A is the area of the filter (1 cm^2) and C_0 is the initial concentration in the donor compartment. Depending on the rate of AP to BL or BL to AP, transport for the different compounds, sink conditions, as defined by >80% of administered compound remaining in the donor compartment, may not be maintained over the course of the 1 h experiment. P_{app} values were therefore calculated using the slope of the steady-state rate constant dQ/dt before sink conditions were no longer present. Thus, the potential for passive BL to AP or AP to BL (backward) diffusion was eliminated in the above calculations, providing a more reliable estimate of absorptive P_{app}. All experiments were performed in at least triplicate, and data are expressed as mean ± standard deviation.

Structural Fragments. To address the structural diversity of the 712 compounds in the calibration dataset, a clustering analysis using 166 MACCS structural keys as descriptors and a Tanimoto coefficient of 0.85 was performed.[32]

Two decomposition schemes were developed to define molecular equivalence classes.[33] The first converted a H-depleted molecular graph into a list of ring fragments by removing all atoms and associated bonds not in a ring. The second used standard definitions for donors and acceptors[34] and included their beta neighborhood into the associated fragment. Overlapping fragments were considered as one union fragment. For both decomposition schemes, each H-depleted molecular graph were transformed into fragments and represented in a canonical line notation similar to SLN[35] as illustrated in Figure 2 for the drug mizolastine, an antihistamine. All algorithms were implemented as Cheshire scripts.[36] It was subsequent feasible to apply standard pivoting techniques in a spreadsheet on records associating compound id, fragment string, and permeability class.

The frequency of a fragment in a permeability class was calculated as ((number of compounds containing the fragment in a permeability class) times (total number of compounds)) divided with ((total number of compounds containing the fragment) times (number of compounds in the permeability class)). The frequencies of a fragment in a permeability class were statistically compared by standard binomial techniques.

Calculated Molecular Descriptors. Molecular properties for molecules were calculated from SD files in the software Sybyl 6.6[37] − mostly via built-in functions, but in some cases using an external program.

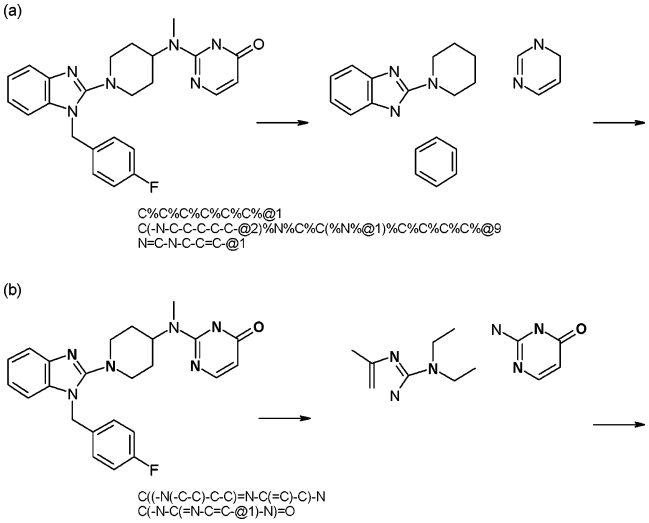


Figure 2. Decomposition of mizolastine with respect to the ring scheme (a) and the donor/acceptor scheme (b).

The properties calculated were the following:

mlogP: a rough, but robust estimate of octanol/water partition coefficient. Originally devised by Moriguchi et al.[38,39]

MW: molecular weight. This property and all the other ones described here refer to the largest fragment/molecule only if more than one is present. Typically, salts/clathrates/hydrates lose their counterions/solvents/water molecules before the calculation is performed. This behavior results from the Concord processing.

HD: number of hydrogen bonds donors. This is a built-in function of Sybyl.

HA: number of hydrogen bonds acceptors. This is a built-in function of Sybyl.

ROT: number of rotatable bonds in molecule. This is a built-in function of Sybyl.

clogP: Pomona College logP (water/octanol partition coefficient). With Sybyl 6.6 Tripos distributed version 4.0 of the clogP algorithm and version 18 of its associated fragment database as provided by BioByte Corp.[40]

VOL: Total molecular volume (Å^3) after regularizing the molecular geometry (in 3-D) via the 4.0.4 version of the Concord program[41] based on SAVol 3.7[42] using Allinger vdw radii.

SURF: Total molecular surface area (square Ångström), based on SAVol 3.7 using Allinger vdw radii.

PSA: Polar surface area (square Ångström), based on SAVol 3.7 using Allinger vdw radii. Polar atoms are oxygens, nitrogens, plus hydrogens attached to O and N.

Classification. The method of Nearest-Neighbor classification was applied using custom-made m. files for MatLab[43] version 6.5 installed with PLS-Toolbox[44] version 2.0.0b. The Mahalanobis distance based on the pooled covariance matrix across the two groups was used to determine proximity between compounds. Full leave-one-out cross-validation was applied to the calibration data set to find the optimal number of neighbors. The optimal number of neighbors was chosen by evaluation of plots depicting the number of false positives,[45] false negatives,[45] Cohen's Kappa values,[46] and nonclassified compounds, respectively, against the number of neighbors. In the selection the percentage of false positives, false negatives and nonclassified compounds were minimized, and the Cohen's Kappa value was maximized. Cohen's Kappa is a measure related to the false positive and false negative rate and expresses the improvement over chance of classification.

The applied classification method provided a set of posterior probabilities of class memberships when predicting new candidates. This allowed us to distinguish between cases with obvious class memberships and cases with doubtful class memberships when predictions of new compounds were carried out. In the present study the class posterior (Pr) was chosen to be ≥ 0.60 for classification. In this way compounds with low Pr were not classified. Outlier detection was carried out by finding the Mahalanobis distance for an individual compound to the kth nearest neighbor in the calibration data, where k is the optimal choice.

Variable Selection. In in silico studies, variable selection is an important element and the selection of a preferred set of descriptors is crucial to obtain a predictive and robust model. In the present study 511 subsets covering all possible combinations of the nine descriptors were generated and a model based on Nearest-Neighbor classification was made for each subset. Full cross-validation was applied to find the optimal subset of descriptors. The Pr was set to 0.50, and five nearest neighbors were used in all 511 models. The Cohen's Kappa value was used as a measure of model performance, and subsets with descriptors giving higher Cohen's Kappa values were favored.

Acknowledgment. We acknowledge the excellent experimental work of Lisette Gammelgaard Nielsen.

References

(1) Boobis, A.; Gundert-Remy, U.; Kremers, P.; Macheras, P.; Pelkonen, O. In Silico prediction of ADME and pharmacokinetics. Report of an expert meeting organised by COST B15. *Eur. J. Pharm. Sci.* **2002**, *17*, 183−193.
(2) Ekins, S.; Rose, J. In silico ADME/Tox: the state of the art. *J. Mol. Graph. Model.* **2002**, *20*, 305−309.
(3) Walters, W. P.; Murcko, M. A. Prediction of 'drug-likeness'. *Adv. Drug Delivery Rev.* **2002**, *54*, 255−271.
(4) Waterbeemd, H. van de; Gifford, E. ADMET in silico Modelling: Towards prediction paradise? *Nature* **2003**, *2*, 192−204.
(5) Lombardo, F.; Gifford, E.; Shalaeva, M. Y. In silico ADME prediction: Data, Models, Facts and Myths. *Mini Rev. Med. Chem.* **2003**, *3*, 861−875.
(6) Artursson, P.; Karlsson, J. Correlation between oral drug absorption in humans and apparent drug permeability coefficients in human intestinal epithelial (Caco-2) cells. *Biochem. Biophys. Res. Commun.* **1991**, *175*, 880−885.
(7) Artursson, P.; Palm, K.; Luthman, K. Caco-2 monolayers in experimental and theoretical predictions of drug transport. *Adv. Drug Delivery Rev.* **1996**, *22*, 67−84.
(8) Hidalgo, I. J. Cultured intestinal epithelial cell models. In *Models for Assessing Drug Absorption and Metabolism*; Borchardt, R. T., Smith, P. L., Wilson, G., Eds.; Plenum: New York, 1996; pp 35−50.
(9) Irvine, J. D.; Takahashi, L.; Lockhart, K.; Cheong, J.; Tolan, J. W.; Selick, H. E.; Grove, J. R. MDCK (Madin-Darby canine kidney) cells: a tool for membrane permeability screening. *J. Pharm. Sci.* **1999**, *88*, 28−33.
(10) Lipinski, C. A.; Lombardo, F.; Dominy B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.* **1997**, *23*, 3−25.
(11) Winiwarter, S.; Bonham, N. M.; Ax, F.; Hallberg, A.; Lennernäs, H.; Karlén, A. Correlation of human jejunal permeability (in vivo) of drugs with experimentally and theoretically derived parameters. A multivariate data analysis approach. *J. Med. Chem.* **1998**, *41*, 4939−4949.
(12) Raevsky, O. A.; Fetisov, V. I.; Trepalina, E. P.; McFarland, J. W.; Schaper, K.-J. Quantitative estimation of drug absorption in humans for passively transported compounds on the basis of their physicochemical parameters. *Quant. Struct.−Act. Relat.* **2000**, *19*, 366−374.
(13) Agatonovic-Kustrin, S.; Beresford, R.; Pauzi, A.; Yusof, M. Theoretically derived molecular descriptors important in human intestinal absorption. *J. Pharm. Biomed. Anal.* **2001**, *25*, 227−237.
(14) Faassen, F.; Vogel, G.; Spanings, H.; Vromans, H. Caco-2 permeability, P-glycoprotein transport ratios and brain penetration of heterocyclic drugs. *Int. J. Pharm.* **2003**, *263*, 113−122.
(15) Kelder, J.; Grootenhuis, P. D.; Bayada, D. M.; Delbressine, L. P.; Ploemen, J. P. Polar molecular surface as a dominating determinant for oral absorption and brain penetration of drugs. *Pharm. Res.* **1999**, *16*, 1514−1519.
(16) Stenberg, P.; Norinder, U.; Luthman, K.; Artursson, P. Experimental and computational screening models for the prediction of intestinal drug absorption. *J. Med. Chem.* **2001**, *44*, 1927−1937.
(17) Yazdanian, M.; Briggs, K.; Jankovsky, C.; Hawi, A. The "High Solubility" Definition of the Current FDA Guidance on Biopharmaceutical Classification System May Be Too Strict for Acidic Drugs. *Pharm. Res.* **2004**, *21*, 293−299.
(18) Jack. D. B. *Handbook of Clinical Pharmacokinetic Data*; Macmillan Publishers Ltd., New York, 1992.

(19) *Physician's Desk Reference*, 55th ed.; Medical Economic Co., Montvale, NJ, 2001.
(20) Hammarlund-Udenaes, M.; Benet L. Z. Furosemide pharmacokinetics and pharmacodynamics in health and disease -an update. *J. Pharmcokin. Biopharm.* **1989**, *17*, 1−46.
(21) Guidance for Industry, Waiver of In Vivo Bioavailability and Bioequivalence studies for Immediate Release Solid Oral Dosage Forms Based on a Biopharmaceutics Classification System, August 2000, FDA/CDER.
(22) Winiwarter, S.; Ax, F.; Lennernäs, H.; Hallberg, A.; Pettersson, C.; Karlén, A. Hydrogen bonding descriptos in the prediction of human in vivo intestinal permeability. *J. Mol. Graphics Modell.* **2003**, *21*, 273−287.
(23) Bakken, G. A.; Jurs, P. C. Classification of Multidrug-Resistance Reversal Agents Using Structure-Based Descriptors and Linear Discriminant Analysis. *J. Med. Chem*. **2000**, *43*, 4534−4541.
(24) Penzotti, J. E.; Lamb, M. L.; Evensen, E.; Grootenhuis, P. D. J. A Computational Ensemble Pharmacophore Model Identifying Substrates of P-Glycoprotein. *J. Med. Chem*. **2002**, *45*, 1737−1740.
(25) Yoshida, F.; Topliss, J. G. QSAR Model for Drug Human Oral Bioavailability. *J. Med. Chem*. **2000**, *43*, 2575−2585.
(26) Lee, K.; Ng, C.; Brouwer, K. L. R., Thakker, D. R. Secretory transport of ranitidine and famotidine across caco-2 cell monolayers. *J. Pharmacol. Exp. Ther.* **2002**, *303*, 574−580.
(27) Artursson, P.; Karlsson, J. Correlation between oral drug absorption in humans and apparent drug permeability coefficients in human intestinal epithelial (Caco-2) cells. *Biochem Biophys Res Commun.* **1991**, *175*, 880−885.
(28) Stewart, B. H.; Chan, O. H.; Lu, R. H.; Reyner, E. L.; Schmid, H. L.; Hamilton, H. W.; Steinbaugh, B. A.; Taylor, M. D. Comparison of intestinal permeabilities determined in multiple in vitro and in situ models: relationship to absorption in humans. *Pharm Res*. **1995**, *12*, 693−699.
(29) Veber, D. F.; Johanson, S. R.; Cheng, H.-Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D. Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* **2002**, *45*, 2615−26.
(30) Taub, M. E.; Shen, W.-C. Polarity in the trancytotic processing of apical and basal membrane-bound peroxidase-polylysine conjugates in MDCK cells. *J. Cell Physiol.* **1992**, *150*, 283−290
(31) Taub, M. E.; Kristensen, L.; Frøkjær, S. Optimized conditions for MDCK permeability and turbidimetric solubility studies using compounds representative of BCS classes I−IV. *Eur. J. Pharm. Sci.* **2002**, *15*, 331−340.
(32) MOE (Molecular Operating Environment) software available from Chemical Computing Group (www.chemcomp.com).
(33) Xu, Y.; Johnson, M. Algorithm for naming molecular equivalence classes represented by labeled pseudographs *J. Chem. Inf. Comput. Sci.* **2001**, *41* (1), 181−185.
(34) Definitions taken from UNITY4.2 available from Tripos, Inc. 1699 South Hanley Rd., St. Louis, MO 63144-2319.
(35) Ash, S.; Cline, M. A.; Homer, R. W.; Hurst, T.; Smith, G. B. SYBYL line notation (SLN): A versatile language for chemical structure representation. *J. Chem. Inf. Comput. Sci.* **1997**, *37* (1), 71−79.
(36) Cheshire is a part of MDL Chemistry Rules Interface available from MDL Information Systems Inc., San Leandro, CA, 94577.
(37) Sybyl 6.6 Tripos Associates Inc., 1699 South Hanley Rd., St. Louis, MO 63144.
(38) Moriguchi, I.; Hirono, S.; Liu, Q.; Nakagome, I.; Matsushita, Y. Simple Method of Calculating Octanol/Water Partition Coefficient. *Chem. Pharm. Bull*. **1992**, *40*, 127−139.
(39) Moriguchi, I.; Hirono, S.; Liu, Q.; Nakagome, I.; Hirano, H. Comparison of Reliability of logP Values for Drugs Calculated by Several Methods. *Chem. Pharm. Bull.* **1994**, *42*, 976−978.
(40) BioByte Corp. 201 W. Fourth St., Suite #204, Claremont, CA 91711-4707.
(41) Concord program. Balducci, R.; McGarity, C.; Rusinko, A., III; Skell, J. M.; Smith, K.; Pearlman, R. S. Laboratory for Molecular Graphics and Theoretical Modeling, College of Pharmacy, University of Texas, Austin, TX 78712.
(42) SAVol 3.7. Pearlman, R. S.; Skell; J. M.; Deanda, F. Laboratory for Molecular Graphics and Theoretical Modeling, College of Pharmacy, University of Texas, Austin, TX 78712.
(43) MatLab, Math Works. Ins., Natick, MA 01760.
(44) PLS-Toolbox, Wise & Gallagher, Eigenvector Technologies, Manson, WA 98831.
(45) Fielding, A. H.; Bell, J. F. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Enviromen. Conserv.* **1997**, *24*, 38−49.
(46) Cohen, J. A. A coefficient of agreement for nominal scales. *Educ. Psych. Meas*. **1960**, *20*, 37−46.

JM049661N