Experimental and Computational Screening Models for the Prediction of Intestinal Drug Absorption

Patric Stenberg,[†] Ulf Norinder,[‡] Kristina Luthman,*[,§] and Per Artursson[†]

Department of Pharmaceutics, Uppsala Biomedical Center, Uppsala University, P.O. Box 580, SE-751 23 Uppsala, Sweden; Department of Medicinal Chemistry, AstraZeneca Research and Development, SE-151 85 Södertälje, Sweden; and Department of Medicinal Chemistry, Institute of Pharmacy, University of Tromsø, N-9037 Tromsø, Norway

Received October 19, 2000

The aim of this study was to devise experimental protocols and computational models for the prediction of intestinal drug permeability. Both the required experimental and computational effort and the accuracy and quality of the resulting predictions were considered. In vitro intestinal Caco-2 cell monolayer permeabilities were determined both in a highly accurate experimental setting (P_c) and in a faster, but less accurate, mode (P_{app}). Computational models were built using four different principles for generation of molecular descriptors (atom counts, molecular mechanics calculations, fragmental, and quantum mechanics approaches) and were evaluated for their ability to predict intestinal membrane permeability. A theoretical deconvolution of the polar molecular surface area (PSA) was also performed to facilitate the interpretation of this composite descriptor and allow the calculation of PSA in a simplified and fast mode. The results indicate that it is possible to predict intestinal drug permeability from rather simple models with little or no loss of accuracy. A new, fast computational model, based on partitioned molecular surface areas, that predicts intestinal drug permeability with an accuracy comparable to that of time-consuming quantum mechanics calculations is presented.

Introduction

The search for pharmacologically active compounds often involves the screening of large compound libraries. Although this method has proven to be successful in terms of generating leads with high affinity for the selected receptor, failure to comply with biopharmaceutical demands, such as permeation of the intestinal mucosa, is likely to terminate the development of many drug candidates.[1,2] Recently, several experimental and computational models of varying complexities have been devised for the prediction of biopharmaceutical properties such as intestinal drug permeability.[3]

Among the experimental models, drug transport across intestinal epithelial Caco-2 cell monolayers grown on permeable filter supports has received particular attention. In its simplest form, an apparent permeability coefficient (P_{app}) may be obtained by measuring the rate of drug transport across the cell monolayer by a single point determination. Alternatively, a more demanding experimental procedure involving multiple data points may be used.[4] By using the latter method, it is possible to avoid experimental artifacts introduced by unstirred water layers or by the filter on which the cells are cultured, thus obtaining the true cell monolayer permeability coefficient (P_c). The choice of methodology will therefore depend on the accuracy needed for predictions of human intestinal permeability in vivo.

The varying complexities of the computational models, which range from simple (and rapid) counting of atoms or fragments (e.g., rule of five, see ref 5) to time-demanding approaches involving quantum mechanics calculations (see, e.g., ref 6), also call for cost−benefit considerations. So far, the more computationally intensive methods have proven to be more accurate than the faster methods for the prediction of intestinal permeability.[7,8] However, to our knowledge a cost−benefit analysis of these methods has not yet been performed to assess what the tradeoffs are in terms of accuracy when a faster method is applied as a virtual permeability screen. Among the computational methods, molecular surface properties, in particular the polar surface area (PSA),[9] have found application as a computational filter for membrane permeability.[10,11] Although it is generally assumed that this composite descriptor is closely related to hydrogen bonding, the nature of PSA has yet to be explored in quantitative terms.

The aim of this study was to devise optimal experimental protocols and computational models for the prediction of intestinal drug permeability. Consequently, two themes have been explored. First, we compared Caco-2 cell monolayer permeabilities in vitro using both a simple (P_{app}) and an experimentally demanding, and presumably more accurate (P_c), measure of permeability for a set of structurally diverse model drugs with known human intestinal absorption. P_{app} and P_c were investigated for their abilities to describe in vivo permeability. Second, we performed a comparative analysis of computational models of varying complexities for the prediction of intestinal drug permeability. The models investigated were selected to represent four different principles for the generation of computational descriptors, namely, atom counts, molecular mechanics calculations, fragmental, and quantum mechanics approaches. The analysis included a

* Address correspondence to this author at the Department of Chemistry, Göteborg University, SE-412 96 Göteborg, Sweden [telephone +46-31 772 2894; fax +46-31 772 3840; e-mail luthman@mc.gu.se].

[†] Uppsala University.

[‡] AstraZeneca.

[§] University of Tromsø.

10.1021/jm001101a CCC: $20.00
Published on Web 05/16/2001

deconvolution of the composite descriptor PSA into several readily interpreted physicochemical properties. Finally, a new computational model based on partitioned molecular surface areas for the prediction of intestinal membrane permeability was devised.

Methods

Cell Culture. Caco-2 cells obtained from American Tissue Collection, Rockville, MD, were maintained in an atmosphere of 90% air and 10% CO_2 as described previously.[12,13] For transport experiments, 5×10^5 cells of passage numbers 96−105 were seeded on polycarbonate filter inserts (12 mm diameter; pore size = 0.4 μm; Costar) and allowed to grow and differentiate for 21−34 days before the cell monolayers were used for transport experiments.

Model Drugs and Radiolabeled Markers. Amiloride, diazepam, hydrocortisone, lactulose, mannitol, metolazone, nordazepam, oxazepam, phenazone, phosphonoformic acid (foscarnet), D-raffinose, sulfasalazine, sulpiride, and terbutaline were purchased from Sigma (St. Louis, MO). Olsalazine was a gift from Pharmacia and Upjohn (Uppsala, Sweden). SB209670 and SB217242 were gifts from SmithKline Beecham Pharmaceuticals. Sildenafil was a gift from Pfizer Limited. Ciprofloxacin and [^{14}C]ciprofloxacin were gifts from Bayer AG (Wuppertal, Germany). [^{14}C]Phosphonoformic acid was purchased from Moravec Biochemicals (Brea, CA), [^{3}H]lactulose was from American Radiolabeled Chemicals, Inc. (St. Louis, MO), and [^{14}C]mannitol and D-[^{3}H]raffinose were obtained from NEN Life Sciences (Boston, MA).

The solvents of the radiolabeled markers were allowed to evaporate in a LAF hood. Tritiated markers were subject to an additional evaporation step whereby 100 μL of purified water was added and evaporated in an LAF hood overnight.

Transport Studies. Drug transport rates across Caco-2 cell monolayers were determined for the compounds in data set 1 as described below or were obtained from previous publications from our laboratory (see refs in Table 3).[14] Drug transport experiments were performed as described previously.[15] In general, the markers were dissolved in Hank's balanced salt solution containing 25 mM Hepes at pH 7.4 (HBSS pH 7.4) to a final concentration of 0.1−5 mM. The amount used of each particular compound depended upon its solubility, the detection limits of the HPLC analysis, the presence of saturable active transport mechanisms, and effects on monolayer integrity as described below. Transport experiments were initiated by incubating the monolayers in HBSS pH 7.4 at 37 °C for 20 min in a humidified atmosphere. Drug solution was added to the donor side of the cell monolayers, and the inserts were placed on a calibrated shaker (IKA-Schüttler MTS4) and kept at 37 °C throughout the experiment. By using two different stirring rates (100 and 500 rpm, respectively) it was possible to obtain the cellular permeability, P_c, which is unaffected by the aqueous boundary layer as described below.[4] Monolayer permeability to the paracellular marker [^{14}C]mannitol was routinely used to investigate the integrity of the monolayers under the experimental conditions. To detect the possible influence of active transport or efflux mechanisms, the permeabilities for all compounds but SB217242 and SB209670 were measured in both apical to basolateral (a−b) and basolateral to apical (b−a) directions. Under "sink" conditions, the apparent permeability coefficients (P_{app}) were calculated from

$$P_{app} = \frac{\Delta Q}{\Delta t}\frac{1}{AC_0} \quad (1)$$

where $\Delta Q/\Delta t$ is the steady-state flux (mol/s), C_0 is the initial concentration in the donor chamber at each time interval (mol/mL), and A is the surface area of filter insert (cm^2). For rapidly transported compounds, where sink conditions could not be maintained for the full duration of the experiments, P_{app} was calculated as described previously[16] from

$$C_R(t) = \frac{M}{V_D + V_R} + \left(C_{R,0} - \frac{M}{V_D + V_R}\right) e^{-P_{app}A(1/V_D + 1/V_R)t} \quad (2)$$

where $C_R(t)$ is the time-dependent drug concentration in the receiver compartment, M is the total amount of drug in the system, V_D and V_R are the volumes of the donor and receiver compartments, respectively, $C_{R,0}$ is the drug concentration in the receiver compartment at the beginning of the interval, and t is the time from the start of the interval. P_{app} was obtained from nonlinear regression, minimizing the sum of squared residuals ($\sum(C_{R,i,obsd} - C_{R,i,calcd})^2$), where $C_{R,i,obsd}$ is the observed receiver concentration at the end of the interval and $C_{R,i,calcd}$ is the corresponding concentration calculated according to eq 2.[16]

P_c was calculated from the slope of the relationship between V/P_{app} and V

$$\frac{V}{P_{app}} = \frac{1}{K} + \left(\frac{1}{P_c} + \frac{1}{P_f}\right)V \quad (3)$$

where P_{app} is the apparent permeability obtained at 100 or 500 rpm, V is the stirring rate, and P_f is the calculated permeability of the filter support.[4]

Analytical Methods. Radioactive samples were analyzed using a liquid scintillation counter (Packard Instruments 1900CA TRI-CARB; Canberra Packard Instruments, Downers Grove, IL). SB209670 and SB217242 were analyzed using LC-MS/MS. The LC-MS/MS system incorporated an analytical Hypersil BDS column (50 × 2 mm) and negative APCI with detection of the 417 Da fragments. All other model compounds were analyzed using reversed-phase HPLC. The HPLC system consisted of a Perkin-Elmer isocratic LC pump 250, a Perkin-Elmer advanced LC sample processor ISS-200, a Spectra Physics UV100 detector, and an integration software (Chromatography Station for Windows). An analytical Hichrom Partisil ODS3 column (100 × 4.6 mm) with a mean particle size of 5 μm was used. The mobile phase was composed of phosphate buffer (pH 3.0−6.5; 60 mM KH_2PO_4, 8 mM H_3PO_4) and acetonitrile; 10−80% acetonitrile was used to obtain retention times of 6−15 min at flow rates of 0.5−1.5 mL/min. Injection volumes were 10−200 μL.

Calculation of Electrotopological State Indices. The Molconn-Z program[17] was used to calculate electrotopological state indices related to hydrogen, nitrogen, and oxygen atoms as described recently.[18] Briefly, the electrotopological state index for a particular atom is a value resulting from its topological and electronic environment. The index will encode the electronegativity as well as the local topology of each atom by considering perturbation effects from its neighbors. For each molecule, the indices of hydrogen bonding hydrogen, all nitrogen, and all oxygen atoms were added together into three groups termed SumH, SumN, and SumO, respectively. An additional term related to the electrotopological state indices of non-hydrogen-bonding hydrogen atoms (SHother) was also included. In addition, the numbers of hydrogen bond acceptors and donors were derived.

Calculation of Hydrogen Bond Strength. Hydrogen bond donor and acceptor strengths were calculated using Hybot.[19] The program Hybot, which is based on experimental data from ~12000 hydrogen bond complexes,[20] was used as a benchmark for the calculated hydrogen bond strengths for the other methods investigated in this study.

Lipophilicity and Refractivity. Drug lipophilicity (Clog*P*) and molecular refractivity (CMR) were calculated using BioByte software[21] accessed via the Sybyl interface.[22]

Conformational Analysis. For the compounds in data set 1 (Table 3), a 10−120000-step Monte Carlo conformational search[23] was performed using the MMFF force field in MacroModel v6.5[24] on a Silicon Graphics workstation. Energy minimizations were made with and without simulated water. By using this procedure, the entire conformational space of low-energy conformations ($\Delta E_s \leq 2.5$ kcal/mol) was expected

to be covered. For the flexible compounds, the conformational search was performed by two or more Monte Carlo simulations. The global minimum of each simulation was then used as starting conformation for a subsequent conformational search. The resulting conformers from the searches were then combined, and duplicate conformers were removed. For the compounds in data sets 2 and 3, a single 3D conformation was generated using Concord.[25] The generated conformers were then geometry optimized using the MMFF force field in MacroModel without the presence of simulated solvent.

Surface Area Calculations. The in-house computer program MAREA[26] was used to calculate the surface area of each conformer as described previously.[27] The program calculates the free surface area of each atom as well as the molecular volume (*V*). The atomic van der Waals radii used were the following: sp^2 carbons, 1.94 Å; sp^3 carbon, 1.90 Å; oxygen, 1.74 Å; nitrogen, 1.82 Å; sulfur, 2.11 Å; electroneutral hydrogen, 1.50 Å; hydrogen bonded to oxygen, 1.10 Å; and hydrogen bonded to nitrogen, 1.125 Å (obtained from PCMODELv4.0, see ref 28).

The polar surface area (PSA) was defined as the area occupied by nitrogen and oxygen atoms plus the area of the hydrogen atoms attached to these heteroatoms. The nonpolar part of the surface area (NPSA) was defined as the total surface area (SA) minus PSA. We also calculated the fraction of the surface area occupied by polar groups (%PSA).

Dynamic surface properties were calculated for the compounds in data set 1. The dynamic surface area (SA_d) is a statistical average obtained by weighting the surface area of each low-energy conformation ($\Delta E_s \leq 2.5$ kcal/mol) by its probability of existence according to a Boltzmann distribution.[9,29] The dynamic polar surface area (PSA_d), dynamic nonpolar surface area ($NPSA_d$), percentage of the total dynamic surface area that is polar ($\%PSA_d$), and dynamic volume (V_d) were determined in the same manner.

Semiempirical Calculations. The global minima from the conformational searches in MacroModel were used as starting geometries for semiempirical geometry optimizations using the AM1 method in Spartan.[30]

Ab Initio Calculations. An ab initio single-point calculation using the 3-21G* basis set in Spartan was performed on each optimized geometry to obtain a wave function suitable for input to the MolSurf program.

MolSurf Calculations. The wave function of each compound was used as input in the MolSurf program.[31] MolSurf calculates various chemical and physicochemical properties. The molecular descriptors for each compound derived from MolSurf were as follows: the strength of hydrogen bond donors of each compound; the number of hydrogen bond donors; the strength of hydrogen bond acceptor oxygens (HBAO); the number of hydrogen bond acceptor oxygens; the strength of hydrogen bond acceptor nitrogens (HBAN); the number of hydrogen bond acceptor nitrogens; the sum of HBAO and HBAN; the molecular electrostatic potential surface area; the polarizability; the polarity; the Lewis acid strength; the Lewis base strength; the acid dissociation constants; octanol−water partition and distribution coefficients; heptane−water partition and distribution coefficients; and the amount of electrostatic potential surface area being strongly positive, weakly positive, strongly negative, or weakly negative.[31,32]

Data Analysis. The molecular descriptors were examined by principal component analysis (PCA).[33] Descriptors showing a skewness in excess of 1.5 were cubic root transformed before being subjected to further analysis. Correlations between multiple calculated descriptors or between calculated descriptors and permeability were established by the method of partial least squares projections to latent structures (PLS).[34] The number of computed PLS components was determined by Q^2, the leave-one-out cross-validated R^2 (where R^2 is the fraction of the sum of squared permeability coefficients explained by the model). Only PLS components resulting in a positive Q^2 were computed, and the number of components was never allowed to exceed one-third of the number of descriptors used for the model. The model was further developed by a stepwise selection of variables. Initially, all descriptors were included in the PLS model. After the first round, the descriptor with least influence on the prediction was deleted and the PLS repeated. If exclusion of a descriptor resulted in a more predictive model (as assessed by Q^2), the descriptor was permanently left out of the model. The variable selection was repeated until no further improvement of Q^2 was achieved. PCA and PLS analyses were performed using Simca.[35] The external predictivity of the models was assessed by the root-mean-square error of the test set (RMSEte). The strength of linear relationships between single variables was assessed by the correlation coefficient, *r*.

Table 1. Physicochemical Characteristics of the Compounds of the Three Data Sets in This Study[a]

	mean				standard deviation			
descriptor[b]	all	1	2	3	all	1	2	3
P_c (cm/s)	8.7×10^{-5}				1.6×10^{-4}			
Clog*P*	2.5	0.7	2.3	2.6	2.1	3.0	2.3	1.9
M_w	309	338	287	315	105	125	122	97
#HBAo	2.7	4.2	2.4	2.7	2.0	3.5	2.0	1.8
#HBAn	2.1	2.4	1.8	2.1	1.6	2.3	1.8	1.5
#HBD	1.7	3.6	1.5	1.7	1.6	3.2	1.7	1.4
PSA (Å^2)	66	103	57	67	37	55	43	32
SA (Å^2)	371	398	332	383	119	131	132	111

[a] Mean values and standard deviations are given for all compounds used in this study (all) and separately for data sets 1, 2, and 3. See Supporting Information for histograms of these properties. Permeabilities (P_c) were obtained only for the compounds in data set 1 (Table 2). [b] Permeability (P_c), lipophilicity (Clog*P*), number of hydrogen bond accepting oxygens (#HBAo) and nitrogens (#HBAn), number of hydrogen bond donating hydrogens (#HBD), polar molecular surface area (PSA), and total molecular surface area (SA) were calculated as described under Methods.

Training Set Selection. To validate the derived PLS models, the data set was further divided to obtain a training and a test set for each respective model. Unless otherwise stated, the training set was selected from two-thirds of the data set used to develop the model. The training sets were also selected to cover a maximum range in descriptor space, which was achieved by selecting the extreme values from the first three components of the PCA.

Results and Discussion

Data Set. A total of 674 drugs covering a wide range in physicochemical properties and permeability were investigated in this study (Table 1; Supporting Information). The data set was divided into three subsets based on the level of experimental and computational effort spent on each compound (Table 2). Most information was obtained for the 27 compounds in data set 1, for which experimental in vitro and in vivo data as well as seven sets of computationally derived descriptors were generated (Table 3). These compounds were carefully selected from published in vivo data so that only compounds that are predominantly absorbed over the intestinal wall by passive diffusion were included in the study. The selection was also performed to minimize the contribution from complicating factors such as low solubility, instability in the gasintrointestinal lumen, or presystemic metabolism. Although the structural diversity among these compounds was large enough to span a significant space of conventional drugs, the structural diversity of data set 1 was increased by the inclusion of compounds representing newer drug classes such as opioid peptides (TAPP and lef553, which represent large, hydrophilic tetrapeptides) and peptide receptor antagonists (SB217242 and SB209670, which represent large, lipophilic compounds obtained from

Table 2. Overview of Generated Experimental and Calculated Parameters of the Three Data Sets in This Study

data set	*n*	P_c[a]	P_{app}[a]	FA[a]	Hybot[a]	MolSurf[a]	Clog*P*[a]	#HB[a]	CMR[a]	surface area[a]	electro-topological[a]
1[b]	27	x	x	x[c]	x	x	x	x	x	x[d]	x
2	170				x	x	x[e]	x	x	x	x
3	477						x[f]	x	x[g]	x	x

[a] Experimentally determined cell monolayer permeabilities (P_c), apparent permeabilities (P_{app}), and the fraction of drugs absorbed after oral administration to humans (FA) as well as molecular descriptors determined by the programs Hybot and MolSurf, lipophilicity (Clog*P*), molecular refractivity (CMR), surface areas, and fragmental electrotopological state indices were obtained as described under Methods. [b] The compounds of data set 1 are detailed in Table 3. [c] FA was obtained for 21 of the compounds in data set 1. [d] Both dynamic and static surface areas were calculated. [e] Due to missing fragments, Clog*P* was calculated for 168 compounds. [f] Due to missing fragments, Clog*P* was calculated for 427 compounds. [g] Due to missing fragments, CMR was calculated for 475 compounds.

Table 3. Intestinal Permeability to the Compounds in Data Set 1

compound[a]	P_c a−b (cm/s × 10^6)[b]	P_{app} a−b (cm/s × 10^6)[b]	a−b/b−a[b]	FA[c]
alfentanil*	310[d] ± 16.8	83 ± 7.7	1.1[d]	n/a[e]
alprenolol	242[f] ± 14.0	88 ± 7.2	≈1[f]	96
amiloride	0.78 ± 0.064	0.56 ± 0.041	0.83	58 ± 5.0
antipyrin	215 ± 11	77 ± 7.9	1.1	97 ± 7
atenolol*	1.0[g] ± 0.10	0.59 ± 0.056	≈1[g]	54 ± 17
cimetidine*	1.2[d] ± 0.087	1.23 ± 0.082	1.8[d]	79
ciprofloxacin	1.9 ± 0.076	1.7 ± 0.096	0.44	69 ± 7
diazepam	756 ± 73	119 ± 5.4	1.7	97
foscarnet	0.050 ± 0.0035	0.045 ± 0.0064	0.65	17 ± 3.5
hydrocortisone	41.6 ± 1.25	27 ± 1.8	0.56	n/a[e]
lactulose	0.27 ± 0.064	0.30 ± 0.012	1.4	0.6 ± 0.3
lef553[h]	0.023 ± 0.0061	0.037 ± 0.015	0.41	n/a[e]
mannitol*	0.19 ± 0.014	0.11 ± 0.0084	0.89	26 (1−89)
metolazone	6.1 ± 0.31	6.2 ± 0.24	0.64	64 ± 23
metoprolol	92[f] ± 4.0	64 ± 4.1	≈1[f]	100 ± 5
nordazepam*	307 ± 8.0	96 ± 1.0	1.0	99 ± 19
oxazepam*	246 ± 4.4	91 ± 1.6	1.1	97 ± 11
oxprenolol*	120[f] ± 6.7	61 ± 10	≈1[f]	97 ± 13
pindolol*	55[f] ± 0.60	36 ± 1.8	≈1[f]	92 ± 11
practolol*	3.5[f] ± 0.53	1.8 ± 0.08	≈1[f]	95 ± 3
raffinose	0.047 ± 0.0040	0.051 ± 0.0041	1.2	0.3 (0.1−0.9)
SB209670[i]	8.8 ± 0.86	9.8 ± 0.32	n/a[e]	n/a[e]
SB217242[i]	70 ± 22	54 ± 11	n/a[e]	n/a[e]
sildenafil	87 ± 1.6	48 ± 0.90	0.83	92
sulfasalazine	0.16 ± 0.021	0.16 ± 0.014	0.087	12 ± 5
sulpiride*	0.39 ± 0.054	0.23 ± 0.0045	0.87	36 ± 20
TAPP[h]	0.022 ± 0.0024	0.040 ± 0.022	*j*	n/a[e]

[a] Compounds designated with an asterisk were assigned to the test set in the PLS analyses of structure−permeability relationships unless otherwise stated. [b] Cell monolayer permeability coefficients (P_c) and apparent permeability coefficients (P_{app}) are presented for transport studies in the apical to basolateral direction and are given as means ± standard deviation. $n = 3-12$. The ratios of transport in the absorptive (a−b) to the secretory (b−a) directions are also shown.[51] [c] Values of the fraction absorbed after oral administration to humans (FA) were obtained from the literature (amiloride,[52] sildenafil,[53] cimetidine,[54] and the references to the remaining compounds are collated in ref 27). Values represent mean ± standard deviation or (range) where available. [d] Obtained from ref 16. [e] Not available. [f] Obtained from ref 9. [g] Obtained from ref 15. [h] For lef553 (Tyr-Arg-Phe-Phe-NH_2) and TAPP (Tyr-D-Ala-Phe-Phe-NH_2), P_c values were calculated by assuming that the effect of the unstirred water layer was negligible.[55] [i] Structures are given in ref 38. [j] Due to lack of substance, permeability was not determined in the b−a direction. However, as TAPP was used at a high concentration (5 mM), active transport or efflux was probably not a problem.

high-throughput pharmacological screening) for which human in vivo data are not yet available. The variability in lipophilicity, molecular weight, and charge of the compounds in data set 1 is comparable to those of registered drugs intended for oral administration.[5,36]

An in-house database of wave functions of publicly available compounds at AstraZeneca was used to obtain structures for the compounds in data set 2 (Table 2). The database mainly consisted of conventional drugs (≈85%) but also some hydrocarbons and a few other model compounds such as alcohols and sugars. First, compounds that overlapped with data set 1 were removed. Thereafter, compounds bearing permanent charges (i.e., quaternary ammonium ions) were also removed as the charges are not well parametrized when molecular mechanics calculations are applied. As the number of permanently charged drugs is negligible,[36] this exclusion criterion is not likely to limit the applicability of any derived model for drug screening.

Data set 3 was compiled from the CMC database[37] by extracting compounds belonging to the therapeutic classes consisting of analgesic and antihypertensive drugs and removing those compounds that were already present in data set 1 or 2. A large structural variability of the compounds in data set 3 was secured because analgesic and antihypertensive drugs are represented by chemically quite different classes of compounds.

Experimental Predictions of Intestinal Permeability. The experimental and theoretical methods employed in the present study primarily aim at predicting passive membrane transport. However, recent research indicates that an increasing number of drugs are not only passively but also actively transported or secreted by the intestinal epithelium.[3] Consequently, it was important to investigate if the transport mechanism of the compounds in data set 1 in Caco-2 cell monolayers was mainly passive diffusion across the monolayers (as suggested by the in vivo data). Bidirectional perme-

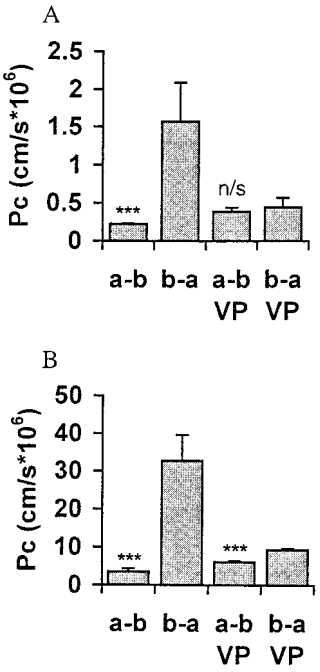

Figure 1. Effect of 200 μM verapamil (VP) on the bidirectional transport of sulpiride (A) and metolazone (B). Caco-2 cell monolayer permeabilities (P_c) were determined in the apical to basolateral (a−b) or the basolateral to apical (b−a) direction. The error bars represent standard deviations (n = 3−4). n/s = no significant difference between a−b and b−a transports.

ability coefficients were therefore determined for all compounds except SB217242 and SB209670, which have already been shown to permeate Caco-2 cell monolayers by a passive mechanism.[38] Most compounds had a <2-fold difference in P_c in the apical to basolateral (a−b) direction as compared with the basolateral to apical direction (b−a) and were therefore considered to be passively transported (Table 3). It was not possible to determine P_c in the b−a direction for olsalazine as this compound affected the membrane integrity when administered to the basolateral side (results not shown). Consequently, olsalazine was excluded from further analyses.

Sulpiride and metolazone had 6- and 9-fold higher P_c values in the b−a than in the a−b direction, respectively, indicating involvement of an efflux mechanism. However, after saturation of the efflux mechanism with the *p*-glycoprotein substrate verapamil, a−b and b−a transports were equalized, and the passive permeability could be determined (Figure 1). Ciprofloxacin had a b−a/a−b ratio of 4 at 1 μM (Figure 2). Because *p*-glycoprotein substrates such as verapamil are not good inhibitors of ciprofloxacin efflux,[39] the efflux was partly inhibited by increasing the ciprofloxacin concentration 1000-fold. This caused P_c b−a to decrease while P_c a−b remained constant, indicating that active efflux contributes only quantitatively to P_c in the b−a direction of ciprofloxacin. Similarly, sulfasalazine has a b−a/a−b ratio of 12, but it has been demonstrated that inhibition of the efflux system affects transport only in the b−a direction.[40] Thus, we conclude that the compounds used for in vitro−in silico and in vitro−in vivo analyses are transported across the epithelial membrane in the a−b direction mainly by passive mechanisms. Caco-2 cell monolayer permeabilities (P_c and P_{app}) and the fraction

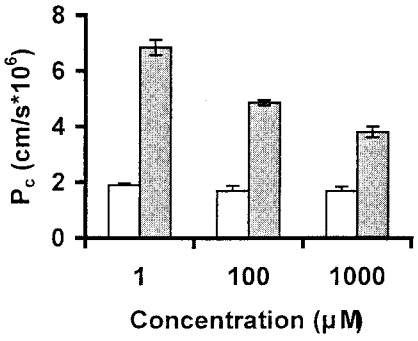

Figure 2. Concentration dependence of Caco-2 cell permeability to ciprofloxacin (P_c) in the basolateral to apical (shaded bars) and apical to basolateral (open bars) direction. The error bars represent standard deviations (n = 3−4).

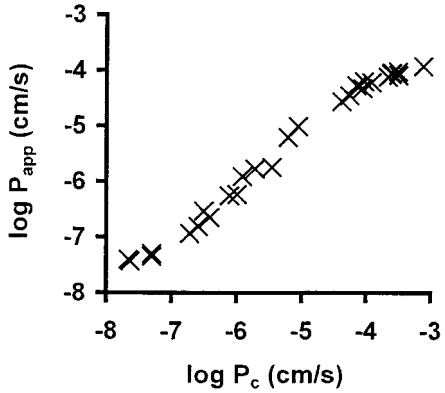

Figure 3. Relationship between Caco-2 cell monolayer permeability (P_c) and apparent permeability (P_{app}) to the 27 compounds in data set 1.

of drug absorbed after oral administration to humans (FA) for the model drugs are shown in Table 3.

Experimental Prediction of Intestinal Drug Absorption in Vivo. Determination of P_{app} using multiple sampling at two different stirring rates allows the calculation of P_c, which is unaffected by extracellular artifacts, such as the resistance of the unstirred water layer at the cell surfaces and the filter support itself.[4] Therefore, the experimentally demanding parameter P_c would hypothetically provide stronger relationships to predictors of membrane permeability and also be expected to better predict drug absorption in vivo than P_{app} values obtained at a single stirring rate. However, a strong correlation with a slope of 0.87 was observed between log P_c and log P_{app} (r^2 = 0.99; Figure 3). A slope of less than unity was expected because the relative effect of the unstirred water layer and filter support on P_{app} is larger for rapidly transported compounds. This is also apparent from the plateau observed as P_{app} reaches 10^{-4} cm/s (Figure 3). Fitting the data to the sigmoidal equation FA = (0−100)/[1 + (log P_c/log $P_{c50\%}$)^k] + 100, where log $P_{c50\%}$ is the in vitro permeability at 50% FA and k is a slope factor,[41] shows that 90% FA is expected at log P_c of −5.3 or 5.0 × 10^{-6} cm/s, a value far below 10^{-4} cm/s. A log P_c below −6.9 or 1.3 × 10^{-7} cm/s would result in <10% FA (R^2 = 0.90; Figure 4). The corresponding 90 and 10% values for P_{app} were 3.5 × 10^{-6} and 7.4 × 10^{-8} cm/s, respectively (R^2 = 0.88; data not shown). We conclude that P_{app} can substitute for P_c in predictions of FA.

Computational Methods for the Prediction of Intestinal Drug Permeability. A. Polar Surface

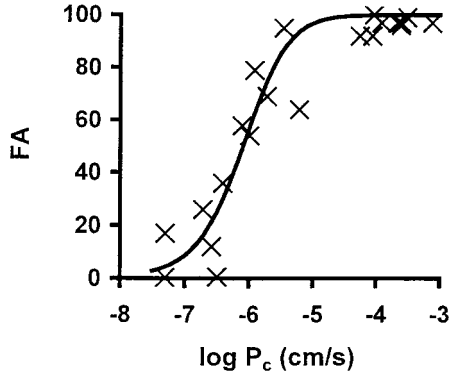

Figure 4. Relationship between Caco-2 cell monolayer permeability (P_c) and fraction absorbed after oral administration to humans (FA) of the 21 compounds in data set 1 for which FA was available. The line was fitted to the data as described in the text ($R^2 = 0.90$).

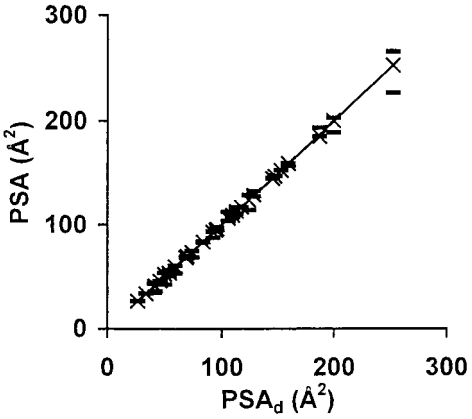

Figure 5. Variability of polar surface area (PSA) among low-energy conformations of the 27 compounds in data set 1. PSA of the global minimum conformation (×) is plotted against the dynamic polar surface area (PSA_d). The horizontal bars indicate the largest and smallest PSA of each respective compound. The solid line has a slope of 1 and an intercept of 0.

Area. 1. Dynamic and Static Surface Areas. Calculation of the dynamic polar surface area (PSA_d) takes all low-energy conformations of the drug molecule into account, which allows a more accurate analysis of how steric effects, for example, influence the preferred conformation. However, although this may be important for the prediction of membrane permeability to more flexible molecules,[15] our own results as well as those of others suggest that PSA from single low-energy conformations may be sufficient to predict drug permeability.[8,27] Indeed, a narrow range of PSAs of the low-energy conformations of data set 1 was observed also in this study, indicating that PSA of any low-energy conformation can be used to describe membrane permeability (Figure 5). However, it is also apparent from Figure 5 that the span of PSA among low-energy conformations increases with increasing PSA, suggesting that flexible compounds with larger PSA are more sensitive to the choice of conformation for the calculation of PSA.[42]

2. Effect of Simulated Solvent. The effect of simulated solvent (which slows the calculations) on calculated surface properties was also investigated using data set 1. Again, the surface areas calculated from geometry optimizations performed in simulated water and in a vacuum were strongly correlated ($r^2 = 0.99$; Supporting Information), thus strengthening earlier results obtained with small analogous data sets.[7,15,43]

We conclude that PSA calculated from single conformations gives results comparable to those obtained after a more computationally demanding conformational search. Similarly, we conclude that PSA calculated from conformations generated in a vacuum can be used instead of the more time-consuming calculations in a simulated water environment.

3. Deconvolution of PSA. The surprisingly high accuracy and the increasing application of PSA as a predictor of intestinal membrane permeability in drug discovery settings call for an analysis of what this composite descriptor actually represents. We reasoned that the deconvolution of PSA into various well-known, and presumably readily interpreted, physicochemical properties will make results from PSA calculations easier to translate into drug design and structure optimization efforts. Various measures of hydrogen bonding, lipophilicity, and charge were calculated using quantum mechanics calculations as detailed under Methods. A multivariate PLS projection to latent structures was used to describe PSA in terms of the parameters derived by quantum mechanics. Because no experimental permeability data were needed for this analysis, an extended training set comprising 128 of the compounds from data sets 1 and 2 was investigated. The remaining 69 compounds of data set 2 were assigned as a test set. Five principal components were used to build the model of PSA ($R^2 = 0.98$, $Q^2 = 0.96$, model 1 in Table 4). The importance of each variable for the prediction of PSA (VIP) is shown in Figure 6. It is apparent that the number rather than the strength of hydrogen bonding atoms best describes PSA in this diverse data set. Despite large variations in polarity and size-related descriptors,[44] these descriptors seem to be of less importance. This was confirmed by a PLS model based on all compounds in data sets 1 and 2, which described PSA using #HBAo, #HBAn, and #HBD only. The model based on these approximate descriptors of hydrogen bonding accounted for 93% of the variance in PSA (Figure 7). The predictivity of the derived model, as assessed by the 477 compounds in data set 3, was also good (model 2, Table 4). It should be noted that these models were built from PSA calculated from single conformations where the effect of intramolecular hydrogen bonding might be less pronounced because conformations stabilized by internal hydrogen bonds may not be accounted for. The extensive computations required to determine PSA_d prevented us from performing a full conformational search on all 674 compounds. Furthermore, PSA of molecules larger and more flexible than those described herein is not expected to be accurately predicted by atom or fragment counts,[42] as intramolecular hydrogen bonding will be more important for these compounds.

The strong relationship between PSA and hydrogen bond count led us to investigate correlations between various hydrogen bonding measures obtained from computational methods based on hydrogen bond counts, fragment descriptors (Hybot and electrotopological state indices), and quantum mechanics calculations (MolSurf) (Table 5). Fair to good correlations were observed among

Table 4. Statistics of Derived PLS Models

model[a]	descriptor set[a]	response[a]	R^2 [b]	Q^2 [b]	A[c]	Ntr[d]	RMSEtr[e]	Nte[d]	RMSEte[e]
1	MolSurf	PSA	0.98	0.96	5	128	7.47	69	13.2
2	#HB	PSA	0.93	0.93	1	197	12.4	477	10.9
3a	surface	P_c	0.95	0.86	2	17	0.33	10 (9)[f]	1.15 (0.85)[f]
3b	surface	P_c	0.85	0.79	2	27	0.56		
4a	fragmental	P_c	0.71	0.67	1	17	0.84	10	0.86
4b	fragmental	P_c	0.68	0.67	1	27	0.84		
5a	MolSurf	P_c	0.87	0.78	1	17	0.56	10	0.83
5b	MolSurf	P_c	0.79	0.74	2	27	0.66		
5c	MolSurf	P_c	0.84	0.79	1	25	0.56		

[a] Models 1 and 2 describe polar surface area (PSA) using descriptors derived by quantum mechanics (MolSurf) and hydrogen bond counts (#HB), respectively. Model 3a describes Caco-2 cell permeability (P_c) in terms of surface areas (surface). Models 4a and 5a describe P_c in terms of fragmental electrotopological state indices (including CMP and Clog*P*) and descriptors derived by quantum mechanics (MolSurf), respectively. Models 3b, 4b, and 5b differ from models 3a, 4a, and 5a by including all 27 compounds in the training set. Model 5c differs from model 5a by including all compounds but raffinose and lactulose in the training set. [b] R^2 is the coefficient of determination, and Q^2 is the leave-one-out cross-validated R^2. [c] A is the number of principal components used to construct the model. [d] Ntr and Nte represent the number of compounds used for training the model and testing the model, respectively. [e] RMSEtr and RMSEte are the root-mean-square errors of the training and test sets, respectively. [f] Numbers within parentheses indicate the model statistics after exclusion of cimetidine from the test set, as discussed in the text.

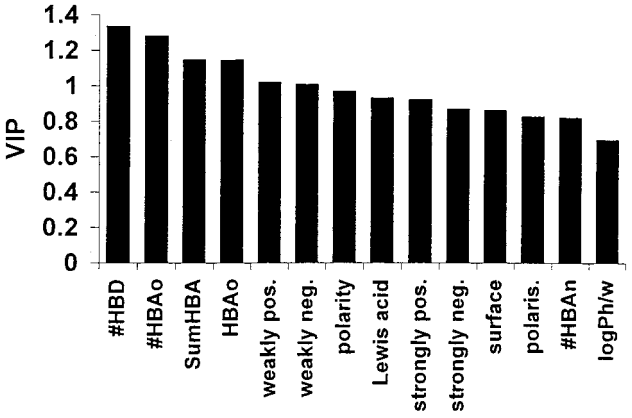

Figure 6. VIP plot for the description of PSA by parameters calculated by a quantum mechanics methodology. The plot indicates a dominating contribution of hydrogen bond counts. The PLS model was built using five principal components. The descriptors used in the final model were number of hydrogen bond donating hydrogens (#HBD), number of oxygens (#HBAo), total strength of hydrogen bond acceptors (SumHBA), strength of hydrogen bond accepting oxygens (HBAo), amount of the isoelectric molecular surface area being weakly positive (weakly pos.) or negative (weakly neg.),[32] polarity, Lewis acid strength, amount of isoelectric molecular surface area being strongly positive (strongly pos.) or negative (strongly neg.), total isoelectric surface area (surface), polarizability (polaris.), number of nitrogens (#HBAn), and heptane−water partition coefficient (log*P* h/w).

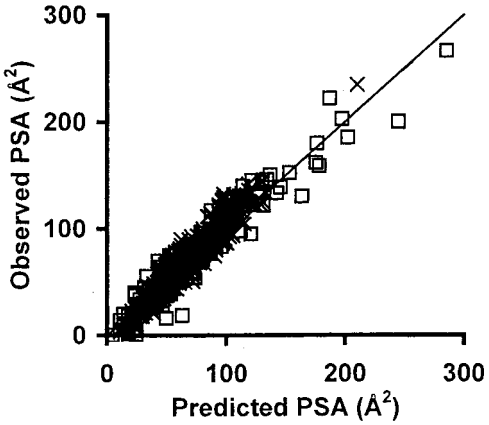

Figure 7. Prediction of PSA from hydrogen bond count for 674 compounds (model 2, Table 4). Compounds in the training and test sets are denoted □ and ×, respectively.

all measures of hydrogen bond acceptors. However, hydrogen bond donor strengths calculated by MolSurf did not correlate well with calculations using the other three methods. The reason for this is most likely that MolSurf assigns hydrogen bond strengths to all polarized hydrogen atoms, not just the ones connected to oxygen and nitrogen atoms. In conclusion, deconvolution of PSA revealed a major role of the number of hydrogen bond moieties rather than the strength of the hydrogen bond interaction. The intercorrelations among various hydrogen bond properties calculated by hydrogen bond counts, fragmental methods, or quantum mechanics was generally weaker. This was particularly apparent for hydrogen bond donating properties. Therefore, it does not seem likely that hydrogen bond count could successfully substitute for, for example, quantum mechanics calculations, in predictions of membrane permeability.

B. Computational Prediction of Intestinal Drug Permeability. We wished to compare the rather simple computational protocols based on hydrogen bond counts and surface areas discussed above with fragmental methods and computationally demanding quantum mechanics calculations in order to suggest an efficient computational protocol for predictions of intestinal membrane permeability. Because the computational requirements of these methods differ quite markedly (Table 6), we set out to determine if the computationally demanding methods provided higher accuracy in the predictions of intestinal membrane permeability. Thus, molecular descriptors derived by computational methods representing four different principles were used to construct models of intestinal permeability (P_c)[45] and FA. We also note that the experimental variations and complex nature of FA make it difficult to quantitatively compare computational models based on FA only. However, because it is FA rather than P_c that represents the end point in the drug development process, the use of FA as an indicator of model quality is still motivated. The principles (methods) investigated were atom count (hydrogen bond count), molecular mechanics calculations (molecular surface areas), fragmental count (electrotopological state indices, Clog*P*, and CMR), and quantum mechanics calculations (MolSurf).

1. Hydrogen Bond Count. A sigmoidal relationship has previously been used to relate various molecular

Table 5. Correlation Matrix (r) of Hydrogen Bond Properties Calculated with Four Different Methods[a]

	electrotopological[b]	#HB[b]	MolSurf[b]	Hybot[b]
electrotopological		0.97/0.96/0.89	0.94/0.89/0.69	0.94/0.83/−0.92
#HB			0.91/0.94/0.49	0.94/0.86/−0.93
MolSurf				0.93/0.91/−0.51
Hybot				

[a] Correlations between hydrogen bond accepting oxygen atoms/hydrogen bond accepting nitrogen atoms/hydrogen bond donating hydrogen atoms are shown in the table for the 197 compounds from data sets 1 and 2. [b] Electrotopological state indices, hydrogen bond counts (#HB), and hydrogen bond strengths from MolSurf and Hybot were determined as described under Methods.

Table 6. Scoring of the Investigated Computational Models Based on Time Consumption, Ease of Interpretation of Results, and Accuracy in Predicting Intestinal Permeability

	time for generation of descriptors[a]	interpretability of results[b]	accuracy of prediction[c]
hydrogen bond count	3 (ms)	3	1
static surface areas, composite	3 (ms)	2	2
static surface areas, partitioned	3 (ms)	3	3
electrotopological state indices	3 (s)	1	3
quantum mechanics	1 (hours−days)	1	3
dynamic surface areas, composite	1 (minutes−weeks)	2	2
dynamic surface areas, partitioned	1 (minutes−weeks)	3	3

[a] Indicates the approximate time needed for generation of descriptors from a 2D structure. May in certain cases involve the generation of a 3D structure, a conformational search, or a wave function as detailed in the text. A score of 1−3, where 1 is slow and 3 is fast, is assigned to the methods. ms = milliseconds. [b] The interpretability indicates the ease by which predictions generated by the model are transferred to structural information for use in drug design and structural optimization. The methods are arbitrarily scored from 1 to 3, where 3 means that the results are readily interpreted. [c] Accuracy of prediction of intestinal drug permeability is based on the performance of the methods assessed by the external test set (Table 4). The methods are arbitrarily scored from 1 to 3, where 3 means that the method provides accurate predictions.

descriptors to FA.[27] Fitting hydrogen bond count to the sigmoidal relationship $FA = 100/[1 + (x/x_{50})^{\gamma}]$, where x is the hydrogen bond count, x_{50} is the count at 50% FA, and γ is a slope factor, resulted in weak relationships ($R^2 = 0.64$ for the total number of hydrogen bonding atoms, $R^2 = 0.53$ for the number of hydrogen bond acceptors, and $R^2 = 0.58$ for the number of hydrogen bond donors). Highly scattered correlations were obtained also when log P_c was correlated to hydrogen bond counts ($r^2 = 0.57$ for the total number of hydrogen bonding atoms, $r^2 = 0.38$ for the number of hydrogen bond acceptors, and $r^2 = 0.62$ for the number of hydrogen bond donors). Thus, it was not possible to describe either in vivo or in vitro intestinal membrane permeability using hydrogen bond counts only.

2. Molecular Surface Properties. Fairly strong sigmoidal relationships between FA and PSA_d calculated without solvent and with simulated water were observed ($R^2 = 0.81$ and 0.82, respectively; Figure 8a). These relationships were only marginally weaker than those observed between FA and in vitro permeability coefficients (see above). However, it is important to note that there were no FA data available for some of the compounds in data set 1 (Table 3). Surprisingly, the log P_c values of the compounds in data set 1 were only poorly described by PSA_d calculated without solvent and with simulated water ($r^2 = 0.63$ in both cases, Figure 8b). PSA, as predicted by the hydrogen bond count (Table 4, model 2), also resulted in a weak correlation with log P_c ($r^2 = 0.58$). In agreement with the strong correlation between PSA and PSA_d, dynamic surface areas did not provide stronger correlations to log P_c than surface areas calculated from single conformations ($r^2 = 0.63$ 0.64, respectively, for conformations generated without or with simulated water). This correlation is weaker than those observed previously when PSA_d was used to describe Caco-2 cell monolayer permeability to various homologous series of compounds,[7,9,15] suggesting that the reason for the weaker correlation was related to the high diversity of data set 1. Intestinal membrane permeability to a structurally diverse set of conventional drugs has previously been successfully correlated to PSA, but the present investigation also includes peptide

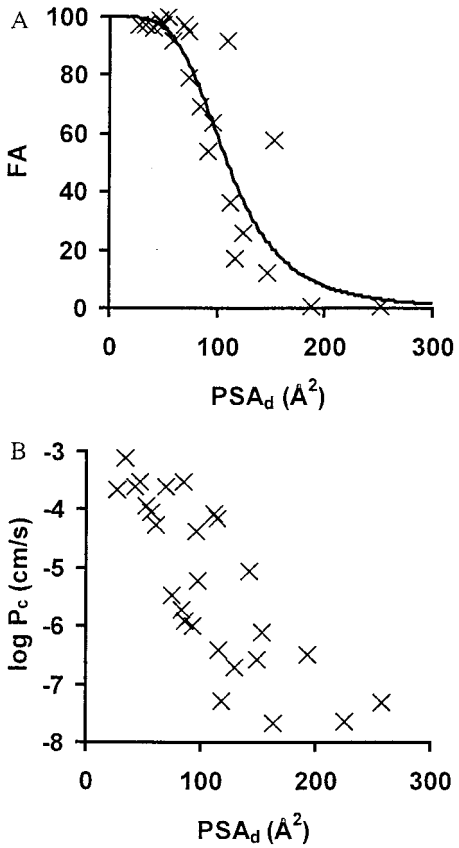

Figure 8. Relationship between dynamic polar surface areas (PSA_d) and (A) fraction of drug absorbed after oral administration to humans (FA) or (B) cell monolayer permeability coefficients (P_c) of the compounds in data set 1. PSA_d was calculated from conformational searches performed in a vacuum.

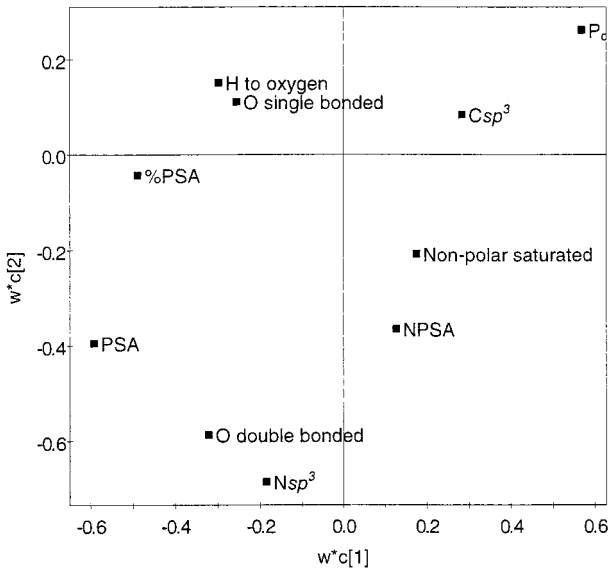

Figure 9. Loading plot showing the inter-relationship between molecular surface areas in the description of Caco-2 cell monolayer permeability (P_c), PLS model 3a (Table 4). The most important variable for prediction of P_c is the polar surface area (PSA), followed by (in decreasing order of importance) the fraction of the total surface area being polar (%PSA), the surface area of double-bonded oxygen atoms (O double bonded), the surface area of sp^3-hybridized nitrogen atoms (Nsp^3), the surface area of hydrogen atoms bonded to oxygen atoms (H to oxygen), the surface area of sp^3-hybridized carbon atoms (Csp^3), the surface area of single-bonded oxygen atoms (O single bonded), the total nonpolar surface area (NPSA), and the saturated part of the nonpolar surface area (Nonpolar saturated).

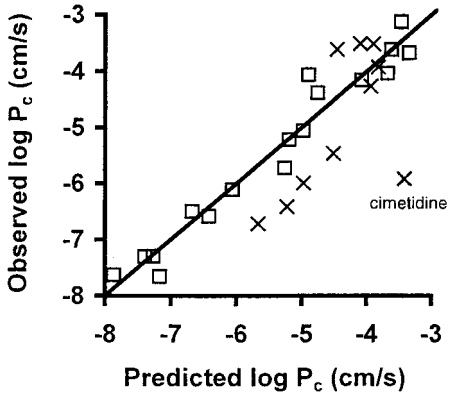

Figure 10. Relationship between observed Caco-2 cell monolayer permeabilities (P_c) and permeabilities calculated from the partitioned molecular surface properties detailed in Figure 9 (PLS model 3a; Table 4). Compounds in the training and test sets are denoted □ and ×, respectively.

derivatives as well as some newer, more lipophilic drugs as discussed above. No outliers are immediately apparent from the correlation between log P_c and PSA_d, so we speculated that the diversity of the data set introduces a general scatter, possibly because of a lack of scaling of the atomic surface areas making up the PSA or because the nonpolar part of the surface area (NPSA) strongly influences the permeability to the investigated compounds. The latter phenomenon was previously observed for peptides and endothelin receptor antagonists.[7,43] However, an optimal linear combination of PSA and NPSA provided only a marginally stronger correlation to log P_c than PSA_d alone ($r^2 = 0.68$).

We therefore investigated whether it was possible to fragment the total molecular surface area further. A multivariate methodology was used to correlate log P_c to surface areas of single atom types as well as collated measures by means of PLS. Before the analyses were performed, nonabundant atom types that showed a highly bimodal distribution (here defined as a skewness >4) were removed. Thus, 19 molecular surface properties[46] calculated from single conformations generated without simulated solvent were used for variable selection of a PLS model describing log P_c. Nine surface properties[47] remained after the variable selection in the final two-component model (Table 4, model 3). The interrelationships among these nine properties in making up the model are presented in the loading plot (Figure 9). Despite the presence of oxygen, nitrogen, and polar hydrogen surfaces in the model, the most important descriptor for the prediction of P_c was PSA. Interestingly, both NPSA and the nonpolar saturated surface areas are situated opposite to P_c in the second principal component. This might indicate that the hydrophobic effect of these properties is somewhat balanced by the size increase which is associated with them, especially because an sp^3-hybridized carbon is located in the same quadrant as P_c (Figure 9). Although the explanatory power and internal predictivity of this partitioned total surface area (PTSA) model are excellent ($R^2 = 0.95$, $Q^2 = 0.86$), its predictivity in the external test set is impaired by its failure to account for the permeability of cimetidine (Figure 10). We note that sp- and sp^2-hybridized nitrogens, which are abundant in cimetidine, were removed after variable selection of the training set (Figure 9). Thus, these atom types are accounted for only by PSA and %PSA in the model, which does not seem to be sufficient in the case of cimetidine.

In conclusion, the PTSA model provides a much stronger correlation with intestinal drug permeability than PSA. This also indicates that the PTSA model can accommodate data sets of a larger structural diversity than the PSA model. However, further studies including more compounds are probably required before PTSA can be advanced as a global model for the prediction of intestinal drug permeability.

3. Fragmental Methods. Electrotopological state indices calculated by the Molconn-Z program,[17] Clog*P*, and CMR have previously been correlated to intestinal membrane permeability.[18] These rapidly calculated fragmental descriptors are related to hydrogen bonding, lipophilicity, and size, respectively. In the present investigation, a PLS methodology was used to build models of log P_c in terms of these fragmental descriptors (Table 4, models 4a and 4b). Only three of the original nine[48] descriptors remained after variable selection of the one-component PLS model. These were the sum of electrotopological state indices related to hydrogen bond donors, the number of hydrogen bond donors, and the sum of electrotopological state indices related to nitrogen atoms (given in the order of importance for the prediction). Surprisingly, descriptors associated with hydrogen bonding oxygens and lipophilicity were excluded from the model by the variable selection process. Despite this, the RMSE for the external test set was only 0.86 log P_c unit. We conclude that a predictive model for the prediction of intestinal membrane perme-

ability could be constructed from rapidly calculated fragmental descriptors.

4. Quantum Mechanics. Molecular descriptors derived from quantum mechanics calculations have previously been correlated to in vitro[7,49] and in vivo[6,50] intestinal membrane permeability with some success. In this study, quantum mechanics derived descriptors calculated by the program MolSurf[31] were used to build PLS models of log P_c (Table 4, model 5). After variable selection, the one-component PLS model contained three descriptors of the original 25. These were (in the order of importance for the prediction) the number of hydrogen bond donors, the polarity, and the number of nitrogen atoms. These descriptors are similar to those remaining after variable selection of the fragmental methods (see above). Apparently, both models account for hydrogen bonding oxygen atoms by giving the attached hydrogen atoms high importance, whereas the hydrogen bonding nitrogen atoms are accounted for by a separate descriptor. Given the large number of hydroxyl groups in the oligosaccharides raffinose and lactulose, the hydrogen bond count was probably assigned a too important role in this particular model at the expense of descriptors related to lipophilicity and size. Therefore, a modified model (5c, Table 4) based on all compounds except for raffinose and lactulose was generated. Given the unusual structures of the oligosaccharides as compared with drug-like molecules, this exclusion would probably not affect the general applicability of the derived model in a drug discovery setting. The resulting model contained one principal component, which was made up from the number of hydrogen bond donors, the polarity, the strength of hydrogen bond accepting oxygen atoms, the hexane-water partitioning coefficient and the number of nitrogen atoms. Thus, we verified our theory that the absence of oxygen atoms in the model was probably due to the presence of the oligosaccharides. We also confirmed that variable selection of the fragmental descriptors after exclusion of the oligosaccharides resulted in a model comprising (in the order of importance for the prediction) the number of hydrogen bond donors, the sum of electrotopological state indices related to hydrogen atoms, ClogP and the sum of electrotopological state indices related to oxygens and nitrogens, respectively.

In conclusion, these results indicate that it is possible to speed up calculations of molecular surface areas by performing geometry optimizations without simulated solvent and by considering single conformations only without a significant loss of accuracy in the predictions. In contrast, simple counting of hydrogen bond acceptors and donors results in loss of valuable information, especially when surface areas of individual atoms are considered. Models based on PTSA and fragmental descriptors can be used for rapid predictions of intestinal membrane permeability with an accuracy comparable to those obtained with descriptors derived by computationally demanding quantum mechanics calculations. However, the model based on PTSA has the advantage over the fragmental methods in terms of interpretability (Table 6). For example, information about which particular atoms that are detrimental or beneficial to membrane permeability to a given molecule is readily derived from the former model.

Conclusions

In conclusion, our results indicate that experimental and computational methods for the prediction of intestinal drug permeability may be applied in a highly simplified manner without compromising the accuracy of the predictions (Table 6). Deconvolution of the composite descriptor PSA indicated a major contribution of the number of hydrogen bonding atoms rather than the hydrogen bond strengths. Interestingly, despite the strong correlation between PSA and hydrogen bond count, the latter descriptor was inferior to PSA in terms of predicting P_c. We also demonstrated that partitioning of the total molecular surface area beyond the level of PSA and NPSA resulted in quite an accurate description of intestinal membrane permeability. This improved model, which we name the partitioned total surface area model (PTSA), has an accuracy comparable to that of computationally demanding quantum mechanics calculations and is therefore adaptable to the virtual screening format.

Acknowledgment. This work was supported by The Swedish Medical Research Council (9478), Centrala försöksdjursnämnden (97-46) and The Swedish Natural Science Research Council (K 11163-301). We thank Dr. Katrin Palm for providing P_{app} data for the β-receptor antagonists, alfentanil and cimetidine. We gratefully acknowledge Vladimir Sherbukhin, AstraZeneca R&D Mölndal, for help with the Hybot calculations. We greatly appreciate experimental assistance from Mats Brodin, Christel Johansson, and Eva Ragnarsson.

Supporting Information Available: Histograms showing the distribution of calculated physicochemical properties of the three data sets in this study, a scatter plot showing the relationship between dynamic polar surface areas for compounds generated in a vacuum and in a simulated water environment, and permeability coefficients predicted from selected PLS models in Table 4. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Kennedy, T. Managing the drug discovery/development interface. *Drug Discov. Today* **1997**, *2*, 436−444.
(2) Prentis, R. A.; Lis, Y.; Walker, S. R. Pharmaceutical innovation by the seven UK-owned pharmaceutical companies (1964−1985). *Br. J. Clin. Pharmacol.* **1988**, *25*, 387−96.
(3) Stenberg, P.; Luthman, K.; Artursson, P. Virtual screening of intestinal drug permeability. *J. Controlled Release* **2000**, *65*, 231−243.
(4) Karlsson, J.; Artursson, P. A method for the determination of cellular permeability coefficients and aqueous boundary layer thickness in monolayers of intestinal epithelial (Caco-2) cells grown in permeable filter chambers. *Int. J. Pharm.* **1991**, *71*, 55−64.
(5) 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.
(6) Norinder, U.; Österberg, T.; Artursson, P. Theoretical calculation and prediction of intestinal absorption of drugs in humans using MolSurf parametrization and PLS statistics. *Eur. J. Pharm. Sci.* **1999**, *8*, 49−56.
(7) Stenberg, P.; Luthman, K.; Ellens, H.; Lee, C.-P.; Smith, P. L.; Lago, A.; Elliott, J. D.; Artursson, P. Prediction of the intestinal absorption of endothelin receptor antagonists using three theoretical methods of increasing complexity. *Pharm. Res.* **1999**, *16*, 1520−1526.
(8) Clark, D. E. Rapid calculation of polar molecular surface area and its application to the prediction of transport phenomena. 1. Prediction of intestinal absorption. *J. Pharm. Sci.* **1999**, *88*, 807−814.
(9) Palm, K.; Luthman, K.; Ungell, A.-L.; Strandlund, G.; Artursson, P. Correlation of drug absorption with molecular surface properties. *J. Pharm. Sci.* **1996**, *85*, 32−39.

(10) 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.
(11) Pickett, S. D.; McLay, I. M.; Clark, D. E. Enhancing the hit-to-lead properties of lead optimization libraries. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 263−272.
(12) Artursson, P. Epithelial transport of drugs in cell culture. I: A model for studying the passive diffusion of drugs over intestinal absorbtive (Caco-2) cells. *J. Pharm. Sci.* **1990**, *79*, 476−482.
(13) Artursson, P.; Karlsson, J.; Ocklind, G.; Schipper, N. Studying transport processes in absorptive epithelia. In *Cell Culture Models of Epithelial Tissues—A Practical Approach*; Shaw, A. J., Ed.; Oxford University Press: New York, 1996; pp 111−133.
(14) The inclusion of historical data was motivated by the low variability of the permeability in Caco-2 cell monolayers in our laboratory.
(15) Palm, K.; Luthman, K.; Ungell, A.-L.; Strandlund, G.; Beigi, F.; Lundahl, P.; Artursson, P. Evaluation of dynamic polar molecular surface area as predictor of drug absorption: comparison with other computational and experimental predictors. *J. Med. Chem.* **1998**, *41*, 5382−5392.
(16) Palm, K.; Ros, J.; Gråsjö, J.; Luthman, K.; Artursson, P. Effect of molecular charge on intestinal epithelial drug transport. *J. Pharmacol. Exp. Ther.* **1999**, *291*, 435−443.
(17) Molconn-Z v3.15S, Hall Associates Consulting, Quincy, MA.
(18) Norinder, U.; Österberg, T. Theoretical calculation and prediction of drug transport processes using simple parameters and PLS statistics. The use of electrotopological state indices. *J. Pharm. Sci.* **2001**, submitted for publication.
(19) HYBOT 98, *pION* Inc., 5 Constitution Way, Woburn, MA 01801.
(20) Raevsky, O. A. Hydrogen bond strength estimation by means of the HYBOT program package. In *Computer-Assisted Lead Finding and Optimization*; van de Waterbeemd, H. B. T., Folkers, G., Eds.; Verlag Helvetica Chimica Acta: Basel, Switzerland, 1997; pp 367−378.
(21) BioByte v4.0, BioByte Corp., 201 W. Fourth St., Suite 204, Claremont, CA 91711-4707.
(22) Sybyl v6.5 Update 3, Tripos Inc., 1699 S. Hanley Rd., Suite 303, St. Louis, MO 63144-2913.
(23) Chang, G.; Guida, W. C.; Still, W. C. An internal coordinate Monte Carlo method for searching conformational space. *J. Am. Chem. Soc.* **1989**, *111*, 4379−4386.
(24) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. MacroModel—An integrated software system for modeling organic and bioorganic molecules using molecular mechanics. *J. Comput. Chem.* **1990**, *11*, 440−467.
(25) Concord v3.2.4, Tripos Inc., 1699 S. Hanley Rd., Suite 303, St. Louis, MO 63144-2913.
(26) MAREA v1.4. The program MAREA is available upon request from the authors. The program is provided free of charge for academic users. Contact Johan Gråsjö (e-mail johan.grasjo@galenik.uu.se).
(27) Palm, K.; Stenberg, P.; Luthman, K.; Artursson, P. Polar molecular surface properties predict the intestinal absorption of drugs in humans. *Pharm. Res.* **1997**, *14*, 568−571.
(28) Gajewski, J. J.; Gilbert, K. E.; McKelvey, J. MMX an enhanced version of MM2. *Adv. Mol. Model.* **1990**, *2*, 65−92.
(29) Lipkowitz, K. B.; Baker, B.; Larter, R. Dynamic molecular surface areas. *J. Am. Chem. Soc.* **1989**, *111*, 7750−7753.
(30) Spartan v5.0, Wavefunction Inc., 18401 Von Karman Ave., Suite 370, Irvine, CA 92612.
(31) MolSurf v98, Qemist AB, Hertig Carls allé 29, SE-691 41 Karlskoga, Sweden. http://members.nbci.com/Qemist.
(32) An electrostatic surface potential below −40 kcal/mol is considered strongly negative, −40 to −15 kcal/mol is weakly negative, 15−40 kcal/mol is weakly positive, and >40 kcal/mol is strongly positive.
(33) Jackson, E. J. *A Users Guide to Principal Components*; Wiley: New York, 1991.
(34) Höskuldsson, A. PLS regression methods. *J. Chemom.* **1988**, *2*, 211−228.
(35) Simca-P v8.0, Umetri AB, Box 7960, SE-907 19, Umeå, Sweden.
(36) Wells, J. I.; Aulton, M. E. Preformulation. In *Pharmaceutics. The Science of Dosage Form Design*; Aulton, M. E., Ed.; Churchill Livingstone: Edinburgh, Scotland, 1988; pp 223−253.
(37) CMC 99.1, MDL Information Systems, San Leandro, CA.
(38) Ellens, H.; Eddy, E. P.; Lee, C.-P.; Dougherty, P.; Largo, A.; Xiang, J.-N.; Elliott, J. D.; Cheng, H.-Y.; Ohlstein, E.; Smith, P. L. In vitro permeability screening for identification of orally bioavailable endothelin receptor antagonists. *Adv. Drug Delivery Rev.* **1997**, *23*, 99−109.
(39) Cavet, M. E.; West, M.; Simmons, N. L. Fluoroquinolone (ciprofloxacin) secretion by human intestinal epithelial (Caco-2) cells. *Br. J. Pharmacol.* **1997**, *121*, 1567−1578.
(40) Powers, M. R.; Lockhart, K.; Irvine, J.; Cheong, J.; Selick, B. Efflux transport of sulfasalazine across Caco-2 cells. *Pharm. Res.* **1996**, *13*, S237.
(41) 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.
(42) Ertl, P.; Rohde, B.; Selzer, P. Fast Calculation of Molecular Polar Surface Area as a Sum of Fragment-Based Contributions and Its Application to the Prediction of Drug Transport Properties. *J. Med. Chem.* **2000**, *43*, 3714−3717.
(43) Stenberg, P.; Luthman, K.; Artursson, P. Prediction of membrane permeability to peptides from calculated dynamic molecular surface properties. *Pharm. Res.* **1999**, *16*, 205−212.
(44) For the training and test sets, the polarity ranges from 9 to 85 (±13) and from 21 to 63 (±8), whereas the isoelectric surface area ranges from 60 to 666 Å^2 (±118 Å^2) and from 191 to 652 Å^2 (±95 Å^2), respectively. Standard deviations are given in parentheses.
(45) P_{app} and P_c are strongly correlated for permeability values $<10^{-4}$ cm/s (which allows any of these parameters to be used for predictions of *in vivo* drug transport as discussed above) but at permeabilities $>10^{-4}$ cm/s experimental artifacts associated with P_{app} make the correlatation weaker. As we wanted to develop computational models that described permeability also for rapidly transported compounds, the use of P_c in favor of P_{app} was motivated, especially since we already had determined P_c for the first part of the study.
(46) The surface areas used as input for variable selection were PSA, %PSA, saturated, unsaturated and total NPSA, SA, Csp^2, Csp^3, single- and double-bonded O, Nsp^2, Nsp^3, electroneutral H, H bonded to O, and N, S, P, F, and Cl.
(47) The surface areas remaining after variable selection were (in decreasing order of importance for the prediction of P_c) PSA, %PSA, double-bonded O, Nsp^3, H bonded to O, Csp^3, single-bonded O, total NPSA, and saturated NPSA.
(48) The fragmental descriptors used as input for variable selection of the PLS model were SumH, SHother, SumN, SumO, #HBAn, #HBAo, #HBD, Clog*P*, and CMR.
(49) Norinder, U.; Österberg, T.; Artursson, P. Theoretical calculation and prediction of Caco-2 cell permeability using MolSurf parametrization and PLS statistics. *Pharm. Res.* **1997**, *14*, 1785−1790.
(50) Winiwarter, S.; Bonham, N. M.; Ax, F.; Hallberg, A.; Lennernäs, H.; Karlen, 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.
(51) Ratios are given for $P_{app,a-b}/P_{app,b-a}$. For metolazone, sulpiride, and ciprofloxacin, which required more extensive analyses of bidirectional transport (as discussed in the text), the ratios for $P_{c,a-b}/P_{c,b-a}$ after saturation of the efflux mechanisms are given.
(52) Weiss, P.; Hersey, R. M.; Dujovne, C. A.; Bianchine, J. R. The metabolism of amiloride hydrochloride in man. *Clin. Pharmacol. Ther.* **1969**, *10*, 401−406.
(53) Walker, D. K.; Ackland, M. J.; James, G. C.; Muirhead, G. J.; Rance, D. J.; Wastall, P.; Wright, P. A. Pharmacokinetics and metabolism of sildenafil in mouse, rat, rabbit, dog and man. *Xenobiotica* **1999**, *29*, 297−310.
(54) Burland, W. L.; Duncan, W. A.; Hesselbo, T.; Mills, J. G.; Sharpe, P. C.; Haggie, S. J.; Wyllie, J. H. Pharmacological evaluation of cimetidine, a new histamine H2-receptor antagonist, in healthy man. *Br. J. Clin. Pharmacol.* **1975**, *2*, 481−486.
(55) Karlsson, J.; Artursson, P. A new diffusion chamber system for the determination of drug permeability coefficients across the human intestinal epithelium that are independent of the unstirred water layer. *Biochim. Biophys. Acta* **1992**, *1111*, 204−210.

JM001101A