ELSEVIER

Journal of Molecular Graphics and Modelling 21 (2003) 273–287

Journal of Molecular Graphics and Modelling

www.elsevier.com/locate/JMGM

Hydrogen bonding descriptors in the prediction of human in vivo intestinal permeability

Susanne Winiwarter [a], Fredrik Ax [a], Hans Lennernäs [b], Anders Hallberg [a], Curt Pettersson [a], Anders Karlén [a,*]

[a] *Department of Medicinal Chemistry, BMC, Uppsala University, SE-751 23 Uppsala, Sweden*
[b] *Department of Pharmacy, BMC, Uppsala University, SE-751 23 Uppsala, Sweden*

Received 24 April 2002; accepted 5 August 2002

Abstract

Hydrogen bonding has been identified as an important parameter for describing drug permeability. Recently, we derived models for predicting intestinal permeability using the hydrogen bonding descriptors polar surface area (PSA) and number of hydrogen bond donors (HBD), and a lipophilicity descriptor [J. Med. Chem. 41 (1998) 4939]. We have now explored other types of hydrogen bonding descriptors to see if these improve the models. Both an experimental hydrogen bonding descriptor, $\Delta\log P$, and calculated descriptors, based either on semiempirical calculations or on experimentally derived hydrogen bond strength values of small molecules, were used. Principal component analyses (PCA) were performed in order to characterize the different parameters, using both a drug data set and a data set of small non-drug-like molecules for which $\Delta\log P$-values had been published.

For a set of diverse drug molecules, for which human intestinal permeability data was available, a PLS-analysis was performed to study the correlation of permeability to the different hydrogen bonding parameters. No correlation could be identified between $\Delta\log P$ and human intestinal permeability in this data set. However, the combination of a hydrogen bond donor descriptor, a general hydrogen bonding descriptor and a lipophilicity descriptor enabled the prediction of human intestinal permeability, whereas hydrogen bond acceptor descriptors were found to be less important. The obtained models successfully predicted the intestinal permeability values of two external data sets.

Keywords: Human intestinal permeability; Hydrogen bonding descriptors; PSA; $\Delta\log P$; Lipophilicity; $\log P_{oct}$; $\log P_{hep}$; QSAR

1. Introduction

Early prediction of drug absorption is of considerable importance in drug development [1–4]. Two key determinants of intestinal drug absorption are dissolution rate of the drug in the gastrointestinal fluids and passage across gut wall (permeability). The measurement of effective intestinal permeability in humans in vivo has become possible through the development of the regional jejunal perfusion system [5]. Using this method, the local absorption rate coefficient P_{eff} (cm/s) across the intestinal barrier can be estimated directly. We recently correlated the human jejunal P_{eff}-values of a set of 22 structurally diverse compounds to a number of physico-chemical descriptors using multivariate data analysis [6]. Three models, with good statistics, were derived for predicting passive intestinal absorption. The obtained equations included terms for the hydrogen bonding descriptors number of hydrogen bond donor atoms (HBD) and polar surface area (PSA) either alone or combined with a lipophilicity descriptor ($C\log P$, the calculated partition coefficient in octanol/water, or $\log D_{5.5}$, the distribution coefficient in octanol/water at pH 5.5). We now sought to explore other types of hydrogen bonding descriptors to see if these improved the above models or added any new information.

Simple theoretical descriptors encoding hydrogen bonding information include a count of the number of potential HBD and acceptor atoms (HBA) or the sum of both (HB) [4,6–8]. These descriptors do not describe the strength of the hydrogen bond nor do they account for the possibility of internal hydrogen bonds. However, more sophisticated descriptors may be able to include these effects. For example, Abraham's α- and β-values are based on experimentally-derived hydrogen bond donor acidities or hydrogen bond acceptor basicities

* Corresponding author. Tel.: +46-18-471-4293; fax: +46-18-471-4474.
E-mail address: anders.karlen@orgfarm.uu.se (A. Karlén).

1093-3263/02/$ – see front matter
PII: S1093-3263(02)00163-8

[9–11], and Raevsky's $\sum C_d$- and $\sum C_a$-descriptors are estimated from thermodynamic data of hydrogen bonding [12,13]. Theoretical descriptors encoding hydrogen bonding information can be derived from programs like Volsurf [14] or Molsurf [15]. The Volsurf descriptors correspond to interaction energies calculated between the compound and different probes (e.g. water-, carbonyl- or lipophilic-probe) [16], whereas Molsurf uses the wave function obtained by ab initio calculations for deriving its parameters.

The PSA of a molecule also encodes hydrogen bonding information [13,17,18]. PSA has been defined in numerous ways in the literature [6,17,19–21], depending on the kind of molecular surface area used (e.g. Connolly, van der Waals, solvent accessible surface area, etc.) or on which atoms are considered to be polar. The definition using nitrogen-, oxygen- and attached hydrogen-atoms seems the most common [17,20], but additionally sulfur [6] or other electronegative atoms may be included. Recently, Feher et al. defined PSA by considering only those surface atoms with an absolute charge above a given limit [21]. Another difference in definition can be with respect to the number of conformations used for calculations. Either one conformation is used to give the 'static' PSA or a group of low energy conformations is used to obtain a more computationally expensive 'dynamic' PSA [20,22,23]. It has been shown that the simpler 'static' PSA gives similar results to the 'dynamic' PSA [13,24,25]. Recently, an even more rapid approach to calculate PSA has been proposed, which does not consider the 3D-structure [26].

Another theoretical approach to describe hydrogen bonding is to combine partial atomic charges with other molecular properties. For example, the combination of the partial atomic charge of the most positive atom with E_{LUMO}, the energy of the lowest unoccupied molecular orbital was found to give a good estimate for hydrogen bond donor acidity [27,28]. It is also possible to combine atomic charges with partial molecular surface areas [29,30]. Many of the aforementioned parameters have been used previously for the modeling of drug permeability or oral bioavailability [6,13,17,20–24,31–40].

An interesting and entirely experimental descriptor for modeling hydrogen bonding capacity is $\Delta\log P$ which corresponds to the difference between the partition coefficient determined in octanol/water ($\log P_{oct}$) and the partition coefficient in alkane/water ($\log P_{alk}$) [41]. $\Delta\log P$ has been found to correlate both to blood brain barrier, skin and corneal permeability [42–45]. $\Delta\log P$ was also found to be useful in a study of a congeneric series of compounds when considering oral absorption [46], whereas $\Delta\log D$, the difference of the distribution coefficients measured in octanol/water and alkane/water at pH 7.4, gave only a poor correlation to Caco-2 cell permeability of a large and structurally diverse drug data set [47]. According to El Tayar et al. $\Delta\log P$ expresses mainly the hydrogen bond donor acidity of a substance [48]. Others have suggested that $\Delta\log P$, besides encoding hydrogen bonding information, describes the polarity or polarizability of a given compound [27,49].

The aim of the present study was to derive and compare different hydrogen bonding descriptors and to see if any of these could improve our previously presented models for predicting human intestinal permeability. We determined $\Delta\log P$-values using heptane as alkane for eight of the passively absorbed drugs for which human jejunal permeability (P_{eff}) had been measured previously, and correlated these to $\log P_{eff}$. In addition, we compiled different hydrogen bonding descriptors for all drugs listed in the Swedish list of drugs (FASS 2000) [50] and compared them using principal component analysis (PCA) in order to study what information they encode. Finally, we used the hydrogen bonding descriptors and PLS analysis to derive new equations for the prediction of intestinal permeability of the human jejunal data set. We both tested our equations on a test set of four molecules with known $\log P_{eff}$ data, which were not used in the PLS analysis, and against an external data set of 67 compounds with reported FA% data (Section 2).

2. Methods

2.1. Compound data sets

Data set 1 consists of 673 pharmaceutical products taken from the Swedish list of drugs (FASS 2000) [50]. All structures were used, except inorganic compounds, metallo-organic compounds, polymer structures and drugs with a molecular weight (MW) above 900. *Data set* 1a consists of 14 structurally diverse drugs taken from data set 1 for which human permeability data (P_{eff}-values) have been determined as part of the Biopharmaceutical classification system (BCS): [6,51] hydrochlorothiazide, furosemide, propranolol, atenolol, metoprolol, piroxicam, verapamil, enalaprilat, ketoprofen, naproxen, phenazone (antipyrine), carbamazepine, terbutaline and fluvastatin. All of these drugs are supposed to be passively, transcellularily absorbed. Only verapamil is a known substrate of the P-glycoprotein efflux system. However, the effective permeability of this drug was measured at a high concentration, where the efflux transport was saturated [52]. Desipramine, another passively absorbed drug, for which P_{eff} had been determined [6], was not included in data set 1a since it is not listed in FASS 2000. However, this compound was used in the test set together with losartan, cimetidine and ranitidine, three drugs from data set 1, for which P_{eff}-values had been reported recently [53,54]. Data set 1b consists of 67 drugs taken from data set 1 for which fraction absorbed (% human intestinal absorption, FA%) values were compiled by Wessel et al. [38]. Data set 2 contains 70 small, non-drug like molecules, for which Abraham's hydrogen bonding descriptors (α- and β-), $\log P_{oct}$-, $\log P_{hep}$- (and thus $\Delta\log P$-) values have been reported in literature [48].

2.2. Experimental determination of $\Delta \log P$

$\Delta \log P$, the difference of the partition coefficients measured in the octanol/water system ($\log P_{oct}$) and in the heptane/water system ($\log P_{hep}$), was determined for data set 1a. $\log P_{oct}$-, $\log P_{hep}$- and $\Delta \log P$-values for data set 2 were taken from El Tayar et al. [48].

$\log P_{oct}$- and pK_a-values for data set 1a were taken from our previous study [6] in which we used potentiometric titration [55,56]. $\log P_{oct}$- and pK_a-values of losartan, cimetidine, ranitidine and enalaprilat were determined in the present study with the same method.

Heptane–water partition coefficients ($\log P_{hep}$-values) were determined using the conventional shake flask method. Desipramine HCl was dissolved directly in heptane. A small volume of 1 M NaOH was added to remove the HCl from the lipophilic phase. Propranolol HCl and verapamil HCl were first dissolved in a small amount of 1 M NaOH to obtain the free base, which was then extracted quantitatively to heptane. Antipyrine, naproxen, ketoprofen, carbamazepine and metoprolol tartrate were not lipophilic enough and had to be dissolved in the aqueous phase (using an appropriate buffer). To obtain high precision in the measurements the pH-value of the aqueous phase was chosen to give a distribution D close to 1 ($\log D = 0$) for the investigated compound. The optimal pH was estimated from previous experiments with dodecane [57] and the pK_a-value of each compound. The value was validated and, if necessary, adjusted by a preliminary experiment. Five pH-values were chosen within a range of ± 0.5 pH units from the calculated center point. The pH-value was regulated by appropriate buffer solutions with defined ionic strength (see later). From these five values the $\log P_{hep}$-value was calculated from a rectilinear plot [58]. No indication of any secondary side-reaction, e.g. dimerization or ion-pair extraction, was observed. The partition experiments were performed in centrifuge tubes. The tubes were shaken in a water bath for at least 8 h at 25.0 ± 0.1 °C, and after centrifugation for 10 min at 4000 rpm the organic phase was removed by a capillary siphon. Except for carbamazepine the total concentration of the analyte was determined in both phases, using photometric measurements at the absorbance maximum. The measurement in the aqueous phase was determined at such a pH-value that the analyte was completely protonated or deprotonated. If the extinction coefficient ε could not be determined in heptane due to too low solubility (metoprolol, ketoprofen and naproxen) the organic phase was extracted with an aqueous solution at a pH where quantitative extraction of the analyte to the aqueous phase was achieved before measurement. All experiments were made in duplicate.

2.2.1. Chemicals

Antipyrine, atenolol, carbamazepine, desipramine hydrochloride, furosemide, hydrochlorothiazide, ketoprofen, (+/−) metoprolol hemitartrate, naproxen, piroxicam, D/L-propranolol hydrochloride, ranitidine, terbutaline hemisulphate, and verapamil hydrochloride were all obtained from Sigma and were more than 99% pure. Naproxen was of USP-grade. Enalaprilat was kindly provided by MSD. Sodium fluvastatin, potassium losartan and cimetidine were provided by Dr. Hans Lennernäs. The substances were used without further purification.

Citric acid monohydrate, *ortho*-phosphoric acid (85%), sodium hydrogen carbonate, anhydrous disodium carbonate, disodium hydrogen phosphate dodecahydrate, sodium dihydrogen-phosphate and potassium chloride were all p.a. from Merck. Sodium hydroxide was from Eka Chemicals, Bohus. Trisodium phosphate dodecahydrate 99% was from Merck. α,α,α-tris(Hydroxymethyl)-methylamin 99% and α,α,α-tris(hydroxymethyl)-methylamin hydrochloride 99% were obtained from Aldrich.

Buffers used were citric acid/NaOH (pH 3.0–5.9), sodium dihydrogen phosphate/disodium hydrogen phosphate (pH 5.8–8.0), sodium hydrogen carbonate/disodium carbonate (pH 9.1–11.1) disodium hydrogen phosphate/trisodium phosphate and tris(hydroxymethyl)-methylamin (pH 7.0–9.0). Buffers were all of 0.1 ionic strength. KOH was 0.4992 N volumetric standard from Aldrich. HCl for titrations was 0.5 mol Fixanal® from Riedel de Haene.

Heptane was analytical reagent (Mallinckrodt Chemical Works, St. Louis, USA) or pro analysi (Merck, Darmstadt, Germany) and used without further purification. 1-Octanol used was >99% HPLC-grade (Sigma–Aldrich) pre-equilibrated with water at room temperature.

2.2.2. Apparatus

The photometric determinations were made with Zeiss PMQ II Spektralphotometer in the range 261–303 nm using 10.0 mm quartz cells. Slit width was 0.1 at all times. All spectra were recorded on a Cecil 3000 scanning spectrophotometer (Cecil Instruments, Cambridge, UK). The pH measurements were made with a Metrohm 632 pH-meter. pK_a and $\log P_{oct}$ determinations were made using a Sirius PCA-101 automatic titrator (Sirius Analytical Instruments Ltd., Forrest Row, UK).

2.3. Calculation of molecular descriptors

All structures were drawn using ChemDraw or Sybylsketch, saved in SMILES format [59] and as such read into Sybyl [60]. 3D starting structures were generated using Concord [61]. If Concord could not generate a 3D-structure, e.g. because the molecule was too large, the structure was built with the help of Sybyl's Build-module and roughly optimized using the Tripos Force Field [62]. Acids and bases were drawn in their neutral form, only quarternary ammonium compounds, e.g. pyridostigmine, were calculated as charged species. Some structures were given as zwitterions in FASS and were drawn as such. Chirality was used as given in FASS. If no chirality was specified either the more active enantiomer was used, if such information was available (e.g. in the case of the NSAIDs), or one chiral

Table 1
List of descriptors used in this study

General descriptors			
MW	Molecular weight	H	Hardness $= (E_{LUMO} - E_{HOMO})/2$
V	Molecular volume	DM	Dipole moment
S	Molecular surface area	NPSA	Non-polar surface area $= S - PSA$
O	Ovality $= S/[4\pi(3V/4\pi)^{2/3}]$	$\log P_{Cr}$	Partition coefficient between octanol and water (calculated according to Ghose–Crippen)
E_{LUMO}	Energy of the lowest unoccupied molecular orbital	$\log P_{oct}$	Partition coefficient between octanol and water (calculated via ABSOLV (data set 1) or experimental values (data set 2))
E_{HOMO}	Energy of the highest occupied molecular orbital	$\log P_{hep}$	Partition coefficient between heptane and water
General hydrogen bonding descriptors			
HB	Number of hydrogen bonding atoms (= HBA + HBD)	PSA/NPSA	–
PSA	Surface area of all N-, O-, S- and attached H-atoms	PSA_{CW}	Sum of (absolute) charge weighted surface area of all N-, O-, S-, F-, Cl-, Br-, I- and attached H-atoms
PSA_{NO}	Surface area of all N-, O- and attached H-atoms	$\Delta\log P$	$\log P_{oct} - \log P_{hep}$
PSA_{rel}	Relative PSA = PSA/S × 100	–	–
Hydrogen bond donor descriptors		*Hydrogen bond acceptor descriptors*	
α	Hydrogen bond donor acidity	β	Hydrogen bond acceptor basicity
Q_H	Highest partial charge of a hydrogen atom (electrostatic component of hydrogen bond donor ability)	Q_{MN}	Absolute value of lowest partial charge of a non-hydrogen atom (electrostatic component of hydrogen bond acceptor ability)
HBD	Number of hydrogen bond donor atoms	HBA	Number of hydrogen bond acceptor atoms
$\sum Q_H$	Sum of the partial charges of all H-atoms attached to an O-, N-, or S-atom	$\sum Q_{MN}$	Sum of the absolute partial charges of all negatively charged O-, N-, S- and F-atoms
$CWPSA_{HBD}$	Sum of charge weighted surface area of all HBD-atoms	$CWPSA_{HBA}$	Sum of absolute charge weighted surface area of all negatively charged N-, O-, S-, F-, Cl-, Br- and I-atoms

form was selected randomly. All molecular structures were optimised by the semi-empirical AM1 method [63] of the MOPAC-program [64] within Sybyl, adding the keywords PRECISE, NOMM and XYZ. For the quarternary ammonium compounds, the keyword CHARGE was also used. Dipole moment (DM), energy of highest occupied molecular orbital (E_{HOMO}), energy of lowest unoccupied molecular orbital (E_{LUMO}) and the partial charges of all atoms, which were used for computing other descriptors, were taken from these calculations. Savol [65,66], a program that calculates the solvent accessible area for each atom, was used to derive molecular surface and volume properties (Table 1). The probe radius used in these calculations was defined as 0.0 in order to obtain a van der Waals surface that resulted in a polar surface area (PSA) that was comparable to that derived from our previous study [6].[1]

Simple hydrogen bonding descriptors correspond to the count of the number of hydrogen bond donor (HBD) or acceptor atoms (HBA) or the sum of both (HB) [6]. Other descriptors were based on partial atomic charges, either alone or in combinations. Q_H is the charge of the most positively charged hydrogen atom, and $\sum Q_H$ the sum of the charges of all hydrogen atoms connected to a heteroatom. Q_{MN} is the absolute value of the charge of the atom with the most negative charge and $\sum Q_{MN}$ the sum of the absolute values of the charges of all negatively charged heteroatoms [27,67].[2] PSA_{CW}, $CWPSA_{HBD}$ and $CWPSA_{HBA}$ combine partial atomic charges and partial surface areas by summing up the charge weighted surface areas of defined atoms (Table 1 for exact definition). The charge weighted surface area of a given atom is the surface area of that atom multiplied with the absolute value of its atomic charge. Polar surface area was either defined as the surface area of the oxygen- and nitrogen-atom(s) and attached hydrogen(s) (PSA_{NO}) [17,20] or as the surface area of the oxygen-, nitrogen- and sulfur- and attached hydrogen-atom(s) (PSA) [6]. Based on the PSA descriptor we also derived the non-polar surface area (NPSA), the relative polar surface area (PSA_{rel}) and the ratio between PSA and NPSA (PSA/NPSA). Fig. 1 shows PSA_{NO}, PSA, NPSA and the parts of the surface area considered for calculating PSA_{CW}, $CWPSA_{HBA}$ and $CWPSA_{HBD}$ using furosemide as an example (this figure was generated using Molcad in Sybyl). Note that the values of the charge weighted surface areas are smaller due to the influence of the atomic charges. Sybyl–SPL-scripts automated the descriptor calculation whenever possible.

[1] Molcad, the program used previously for calculating surface and volume properties [6], was primarily created for visualizing surfaces. When calculating PSA with this program sometimes ambiguous results were generated, that had to be worked around manually. Comparing the PSA values for a set of 270 drugs calculated with either Molcad or Savol we obtained a correlation coefficient (R^2) of 0.97.

[2] $\sum Q_H$ and $\sum Q_{MN}$ were first presented by us on a poster at the log P 2000 Conference in Lausanne in March 2000, log P2000.

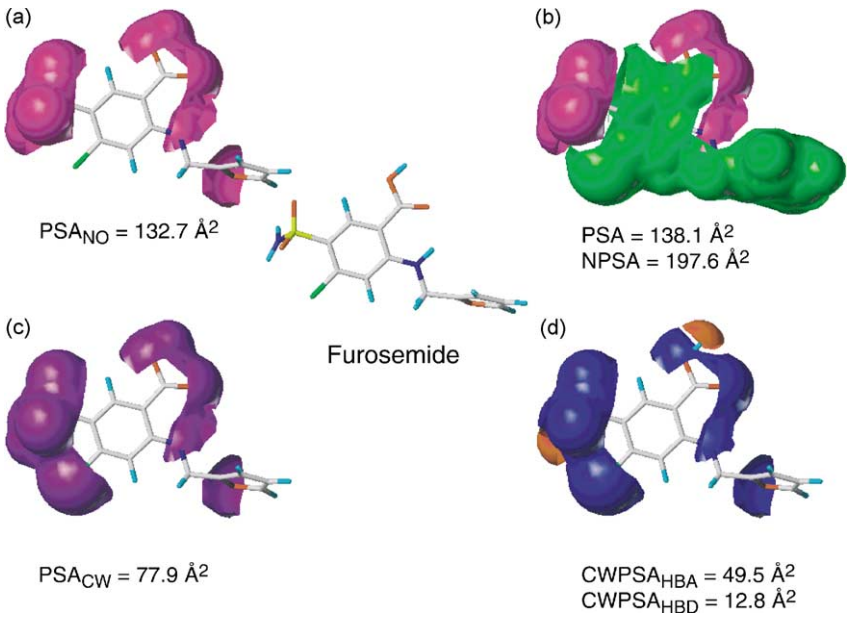

Fig. 1. Illustration of some different molecular surface areas used in the study: (a) PSA_{NO} (magenta); (b) PSA (cyan) and NPSA (green); (c) PSA_{CW} (purple); and (d) $CWPSA_{HBA}$ (blue) and $CWPSA_{HBD}$ (red).

For data set 1 (and thus also data sets 1a and b) Abraham's hydrogen bond donor acidity (α) and hydrogen bond acceptor basicity (β) values [9] and $\log P_{oct}$-values were calculated using the ABSOLV-program [68]. The measured $\log P_{oct}$-values for data set 1a were only used for the determination of $\Delta\log P$ (see earlier), whereas the calculated $\log P_{oct}$-values were used in the multivariate studies. For data set 2 the α-, β-, and $\log P_{oct}$-values were taken from literature [48]. An additional lipophilicity descriptor, $\log P_{Cr}$, was calculated for all molecules according to the atomic fragment method developed by Ghose and Crippen [69] with the help of a Sybyl–SPL-script. We also included the size descriptors molecular weight (MW), surface area (S) and volume (V), the shape descriptor ovality (O) and an additional electronic descriptor, hardness (H). All descriptors used in this study are listed and explained in Table 1. The descriptors were tentatively classified into general descriptors, general hydrogen bonding descriptors, hydrogen bond donor and hydrogen bond acceptor descriptors.

2.4. Data analysis

Correlation coefficients (R^2) between the parameters within each of the data sets used were calculated within Sybyl (results only shown for data set 1, Table 3). Multivariate data analyses were performed with Simca-P [70] (version 8.0), using default settings. Principal component analyses were performed to investigate the relation between the parameters. PLS-analysis (Projection to Latent Structures Partial Least Squares analysis) was employed to derive models for predicting intestinal permeability. Cross validation was used to quantify the predictive power of the PLS models. Variable Importance in Projection (VIP-) analysis showed which of the parameters were important for the model. The less important parameters were excluded immediately. Further reduction of the variable set was done iteratively: In a number of PLS-analyses each of the remaining descriptors was excluded once at a time. The PLS analysis with the highest cross-validated R^2 (Q^2) was used in the next step. The procedure was continued, until no further improvement, either increase of Q^2 or reduction of number of significant components with no or only a slight decrease of Q^2, could be obtained. $\log P_{eff}$-values for the molecules of data set 1b were calculated from the most interesting model and plotted against known experimental FA%-values.

3. Results and discussion

3.1. Experimental determination of $\Delta\log P$ (data set 1a)

$\Delta\log P$-values could be measured for 7 out of the 14 compounds of data set 1a (antipyrine, carbamazepine, ketoprofen, metoprolol, naproxen, propranolol, and verapamil) and for desipramine (not included in data set 1). For the remaining compounds $\log P_{hep}$ was too low to be measured reliably. The obtained $\Delta\log P$-values range between 1.1 for desipramine and 3.5 for carbamazepine (Table 2). In theory, a high $\Delta\log P$-value indicates high hydrogen bonding capacity and thus poor absorption. However, when $\Delta\log P$ is plotted against $\log P_{eff}$ (Fig. 2) no correlation is seen ($R^2 = 0.08$, $n = 8$). One reason may be that the data set was biased since $\Delta\log P$-values could only be obtained for the more lipophilic compounds ($\log P_{oct} = 0.56$ for

Table 2
P_{eff}-, pK_a- and log P-values for the compounds of data set 1a and desipramine, cimetidine, losartan and ranitidine

Substance	P_{eff} (10^{-4} cm/s)	pK_a	Octanol		Heptane	$\Delta\log P$
			log P_{oct}	log P_{ion}	log P_{hep}	
Antipyrine	4.5 ± 2.5[a]	1.44[a]	0.56 ± 0.04[a]	–	−2.5 ± 0.01	3.1
Atenolol	0.2 ± 0.2[a]	9.60[a]	0.18 ± 0.02[a]	–	–	–
Carbamazepine	4.3 ± 2.7[a]	[b]	2.45[a]	–	−1.05 ± 0.05	3.50
Enalaprilat	0.2 ± 0.3[a]	1.60 ± 0.11	−0.24 ± 0.02	−1.0 ± 0.05	–	–
	–	3.09 ± 0.02	–	−0.59 ± 0.25	–	–
	–	7.68 ± 0.01	–	–	–	–
Fluvastatin	2.4 ± 1.8[a]	4.31[a]	4.17 ± 0.01[a]	1.12 ± 0.03[a]	–	–
Furosemide	0.05 ± 0.04[a]	3.34	2.53 ± 0.01[a]	−1.03 ± 0.03[a]	–	–
	–	10.46[a]	–	–	–	–
Hydrochlorothiazide	0.04 ± 0.05[a]	8.84	−0.17 ± 0.01[a]	–	–	–
	–	10.07[a]	–	–	–	–
Ketoprofen	8.4 ± 3.3[a]	3.89[a]	3.37 ± 0.01[a]	−0.34 ± 0.03[a]	−0.14 ± 0.02	3.51
Metoprolol	1.3 ± 1.0[a]	9.60[a]	2.07 ± 0.01[a]	−0.63 ± 0.06[a]	−0.11 ± 0.02	2.18
Naproxen	8.3 ± 4.8[a]	4.01[a]	3.58 ± 0.01[a]	−0.22 ± 0.02[a]	0.40 ± 0.02	3.18
Piroxicam	7.8 ± 7.5[a]	NM[c]	–	–	–	–
Propranolol	2.9 ± 2.2[a]	9.52[a]	3.43 ± 0.02[a]	0.70 ± 0.05[a]	1.45 ± 0.01	1.98
Terbutaline	0.3 ± 0.3[a]	8.72	−0.09 ± 0.01[a]	−0.96 ± 0.06	–	–
	–	10.00	–	−0.81 ± 0.05	–	–
	–	11.10[a]	–	−1.35 ± 0.04[a]	–	–
Verapamil	6.7 ± 2.9[a]	8.66[a]	3.96 ± 0.01[a]	0.96 ± 0.03[a]	2.56 ± 0.04	1.40
Desipramine	4.4 ± 1.8[a]	10.65[a]	4.54 ± 0.01[a]	1.05 ± 0.04[a]	3.48 ± 0.01	1.06
Cimetidine	0.3 ± 0.11[d]	12.34 ± 0.03	0.49 ± 0.01	−1.80 ± 0.28	–	–
	–	6.96 ± 0.00	–	–	–	–
Losartan	1.14 ± 1.1[e]	4.43 ± 0.01	3.52 ± 0.01	2.305 ± 0.18	–	–
	–	2.49 ± 0.01	–	0.91 ± 0.02	–	–
Ranitidine	0.27 ± 0.17[d]	13.11 ± 0.11	0.15 ± 0.01	−1.18 ± 0.04	–	–
	–	8.40 ± 0.00	–	–	–	–

[a] Values taken from Ref. [6].
[b] Aprotic.
[c] Not measured.
[d] Values taken from Ref. [53].
[e] Values taken from Ref. [54].

antipyrine, and above two for the seven other compounds). Another reason may be that $\Delta\log P$ as defined in this study (log P_{oct} − log P_{hep}) does not contain information important for modelling intestinal permeability.

Table 2 also lists the pK_a- and log P_{oct}-values for cimetidine, losartan and ranitidine. Unfortunately, log P_{hep}, and thus $\Delta\log P$, could not determined for these compounds due to the low distribution to heptane.

3.2. Comparison of hydrogen bonding descriptors

All drugs in data set 1 were compiled and the parameters listed in Table 1 were calculated as described in Section 2. To detect which variables are correlated to each other we have both used principal component analysis and calculated the correlation coefficients between two variables at a time. In the principal component analysis of data set 1 (PCA1) we included all 27 theoretical descriptors available for the 673 molecules. Ten components were significant according to Simca standard criteria. The first two components explained almost two-thirds (65.8%) of the data variance, whereas the third component describes only 8.1%. In the loadings plot showing the first two principal components (Fig. 3) it can be seen that the size descriptors, especially molecular volume (V) and molecular surface area (S), are very similar. Cross correlation coefficients within data set 1, which are listed in Table 3, confirm this.

All hydrogen bonding descriptors (depicted in black, red and blue in the plot) can be found in one region of the loadings plots (Fig. 3). PSA_{rel} and PSA/NPSA are found somewhat outside the general cluster, indicating that they also contain other information. In fact, the loadings plot shows that both descriptors are opposite to log P_{oct} (and log P_{Cr}). Therefore, it seems likely that both PSA_{rel} and PSA/NPSA contain lipophilicity information. Q_{MN} and Q_H also fall out of the general cluster of hydrogen bonding descriptors. This

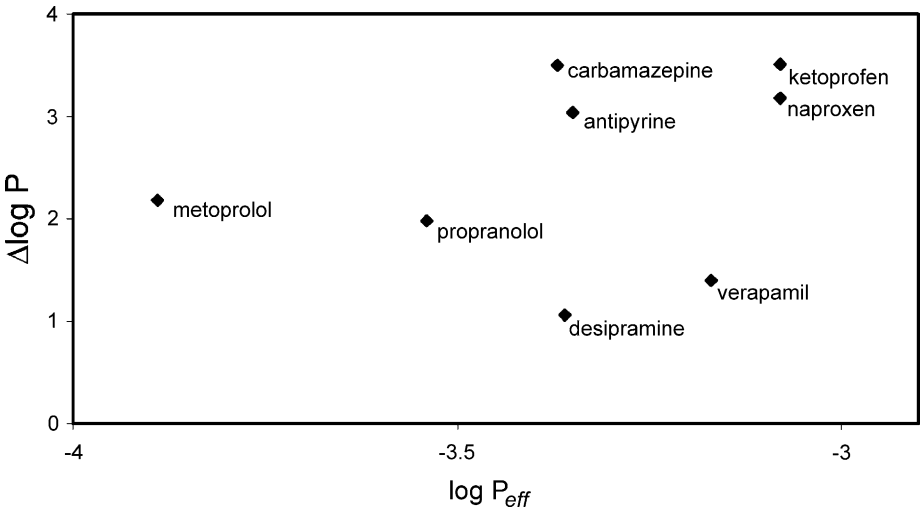

Fig. 2. $\Delta \log P$ vs. intestinal permeability ($\log P_{\text{eff}}$).

is probably due to the fact that they only include information from one single atom, while there is usually more than one hydrogen bond donor or acceptor group in a drug. The remaining hydrogen bond donor descriptors, $\sum Q_{\text{H}}$, HBD, $\text{CWPSA}_{\text{HBD}}$ and α, cluster tightly, whereas, the remaining hydrogen bond acceptor descriptors are more widely distributed. The correlation coefficients between the hydrogen bond donor descriptors and between the hydrogen bond acceptor descriptors, respectively, confirm that the descriptors in each group contain similar information (Table 3). PSA and PSA_{NO} were found to contain almost identical information with a cross correlation coefficient (R^2) of 0.96.

Since information concerning $\Delta \log P$ is lacking for most of the above molecules we searched for a data set, which included this parameter. El Tayar et al. [48] have published both $\log P_{\text{oct}}$-, $\log P_{\text{hep}}$-, $\Delta \log P$- and Abraham's α- and β-values for a set of small molecules (data set 2). A principle component analysis of this data set (PCA2) was performed and compared to the PCA of the combination of both data set 1 and 2 (PCA3). With both analyses, ten components were found to be significant according to the standard criteria in Simca. The first two components explained approximately two-thirds of the data variance in both analyses (65.8 and 66.8%, respectively), whereas the third component gave more information in PCA2 than in PCA3 (10.1% versus 7.7%). The loadings plot based on data set 2 (PCA2, Fig. 4) shows $\Delta \log P$ to be close to hydrogen bond donor descriptors. This is in accordance with the previous suggestion, based on the same data set, that $\Delta \log P$ is mainly a measure of hydrogen bond donor acidity [48]. The correlation coefficient of $\Delta \log P$ to α is greatest ($R^2 = 0.84$), followed by the correlation coefficient to $\log P_{\text{hep}}$ and Q_{H} (R^2 is 0.66 and 0.65, respectively). $\Delta \log P$ is also correlated to

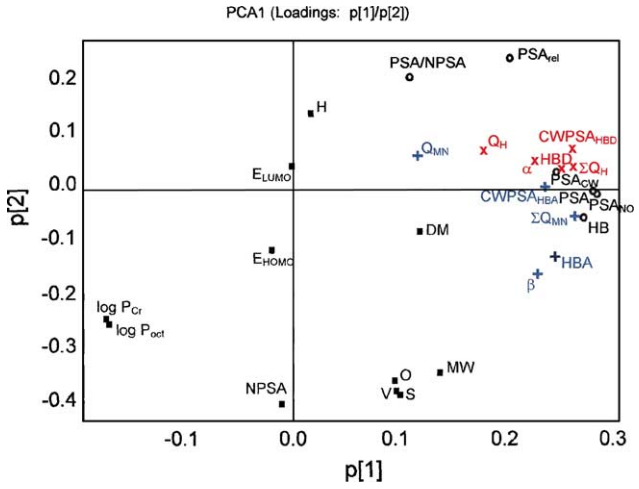

Fig. 3. Loadings plot of PCA1 (general descriptors in green, general hydrogen bonding descriptors in black, hydrogen bond donor descriptors in red and hydrogen bond acceptor descriptors in blue).

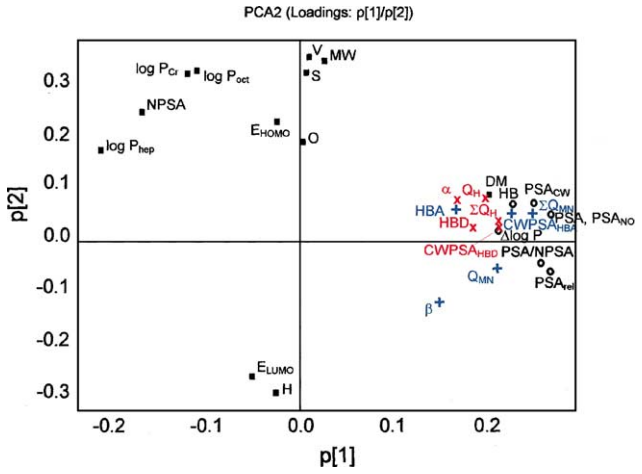

Fig. 4. Loadings plot of PCA2 (general descriptors in green, general hydrogen bonding descriptors in black, hydrogen bond donor descriptors in red and hydrogen bond acceptor descriptors in blue).

Table 3
Correlation matrix (R^2) of all variables within data set 1 ($n = 673$)

	MW	V	S	O	E_{LUMO}	E_{HOMO}	H	DM	NPSA	$\log P_{Cr}$	$\log P_{oct}$	HB	PSA	PSA_{NO}	PSA_{rel}	PSA/NPSA	PSA_{CW}
MW	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
V	0.867	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
S	0.851	0.980	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–
O	0.693	0.793	0.889	1	–	–	–	–	–	–	–	–	–	–	–	–	–
E_{LUMO}	0.017	0.003	0.004	0.007	1	–	–	–	–	–	–	–	–	–	–	–	–
E_{HOMO}	0.019	0.027	0.031	0.043	0.146	1	–	–	–	–	–	–	–	–	–	–	–
H	0.058	0.038	0.044	0.066	0.366	0.255	1	–	–	–	–	–	–	–	–	–	–
DM	0.100	0.067	0.073	0.074	0.009	0.026	0.051	1	–	–	–	–	–	–	–	–	–
NPSA	0.610	0.836	0.843	0.733	0.004	0.037	0.050	0.017	1	–	–	–	–	–	–	–	–
$\log P_{Cr}$	0.051	0.110	0.108	0.087	0.000	0.039	0.033	0.020	0.354	1	–	–	–	–	–	–	–
$\log P_{oct}$	0.047	0.119	0.119	0.108	0.003	0.062	0.026	0.018	0.358	0.713	1	–	–	–	–	–	–
HB	0.297	0.186	0.182	0.144	0.007	0.001	0.010	0.110	0.005	0.285	0.233	1	–	–	–	–	–
PSA	0.214	0.100	0.110	0.113	0.000	0.000	0.000	0.140	0.005	0.348	0.308	0.834	1	–	–	–	–
PSA_{NO}	0.223	0.110	0.118	0.114	0.001	0.002	0.004	0.133	0.002	0.340	0.292	0.883	0.963	1	–	–	–
PSA_{rel}	0.026	0.101	0.095	0.083	0.002	0.024	0.031	0.036	0.391	0.515	0.518	0.304	0.509	0.476	1	–	–
PSA/NPSA	0.041	0.078	0.078	0.083	0.007	0.032	0.054	0.002	0.193	0.162	0.190	0.059	0.117	0.112	0.524	1	–
PSA_{CW}	0.121	0.028	0.036	0.051	0.014	0.013	0.000	0.125	0.016	0.214	0.228	0.457	0.612	0.565	0.393	0.105	1
α	0.050	0.014	0.014	0.009	0.006	0.002	0.011	0.055	0.027	0.141	0.189	0.432	0.464	0.503	0.350	0.115	0.413
Q_H	0.014	0.000	0.000	0.001	0.017	0.004	0.005	0.108	0.048	0.126	0.131	0.199	0.315	0.323	0.323	0.081	0.298
HBD	0.094	0.044	0.042	0.029	0.031	0.000	0.024	0.040	0.018	0.342	0.282	0.779	0.695	0.744	0.378	0.104	0.337
$\sum Q_H$	0.100	0.041	0.039	0.027	0.018	0.000	0.016	0.055	0.023	0.343	0.303	0.777	0.742	0.788	0.436	0.126	0.410
$CWPSA_{HBD}$	0.064	0.017	0.018	0.015	0.015	0.000	0.016	0.048	0.048	0.366	0.337	0.721	0.739	0.782	0.510	0.161	0.436
β	0.462	0.407	0.393	0.300	0.010	0.001	0.005	0.103	0.126	0.144	0.244	0.628	0.540	0.541	0.081	0.009	0.269
Q_{MN}	0.000	0.005	0.002	0.000	0.021	0.013	0.001	0.072	0.023	0.043	0.085	0.024	0.065	0.049	0.117	0.055	0.503
HBA	0.415	0.286	0.284	0.233	0.000	0.003	0.002	0.145	0.047	0.164	0.133	0.859	0.680	0.712	0.167	0.019	0.408
$\sum Q_{MN}$	0.248	0.142	0.154	0.153	0.002	0.005	0.000	0.182	0.006	0.193	0.224	0.610	0.648	0.644	0.280	0.067	0.842
$CWPSA_{HBA}$	0.120	0.042	0.052	0.066	0.014	0.028	0.001	0.156	0.004	0.168	0.194	0.408	0.519	0.501	0.318	0.085	0.901

	α	Q_H	HBD	$\sum Q_H$	$CWPSA_{HBD}$	β	Q_{MN}	HBA	$\sum Q_{MN}$	$CWPSA_{HBA}$
α	1	–	–	–	–	–	–	–	–	–
Q_H	0.359	1	–	–	–	–	–	–	–	–
HBD	0.531	0.213	1	–	–	–	–	–	–	–
$\sum Q_H$	0.629	0.288	0.972	1	–	–	–	–	–	–
$CWPSA_{HBD}$	0.662	0.314	0.931	0.976	1	–	–	–	–	–
β	0.211	0.081	0.414	0.421	0.365	1	–	–	–	–
Q_{MN}	0.140	0.135	0.012	0.028	0.041	0.021	1	–	–	–
HBA	0.240	0.129	0.411	0.424	0.378	0.607	0.027	1	–	–
$\sum Q_{MN}$	0.408	0.252	0.385	0.451	0.445	0.507	0.418	0.606	1	–
$CWPSA_{HBA}$	0.371	0.269	0.224	0.291	0.309	0.263	0.559	0.441	0.893	1

some polar surface area and hydrogen bond donor descriptors in the small molecule data set (PSA_{rel}, $CWPSA_{HBD}$, $\sum Q_H$, PSA_{NO} and PSA all have an R^2 of more than 0.4). The most significant correlation coefficient for α is found to $\Delta \log P$ ($R^2 = 0.84$) followed by Q_H ($R^2 = 0.69$). $\log P_{hep}$, the other descriptor which is new with this analysis, can be found near the lipophilicity descriptors $\log P_{oct}$ and $\log P_{Cr}$, with correlation coefficients of 0.53 and 0.47, respectively.

The PCA of both data sets ($n = 743$) using all 29 descriptors (PCA3) should be considered with caution, since $\Delta \log P$ and $\log P_{hep}$ are available for only about 10% of the investigated compounds (70 compounds of data set 2 plus 7 compounds of data set 1a). The results of this analysis are very similar to those of PCA1. However, it is interesting to consider the score plot, which is shown in Fig. 5. Data set 2, depicted with blue triangles, is only found in one region of the plot, overlapping just slightly with the molecules of data set 1. This illustrates clearly the difference between the two data sets, a drug data set and a non-drug data set. One important reason for this difference is the molecular weight, which ranges between 60 and 855 for data set 1 but only between 30 and 205 for data set 2. However, other properties may also differ significantly between the two data sets. For example, it can be noticed that the correlations between the hydrogen bond donor descriptors are different within the two datasets: $\sum Q_H$, a parameter that is well correlated to α in the drug data set ($R^2 = 0.63$, Table 3), is only weakly correlated ($R^2 = 0.31$) in the small molecule data set, whereas the situation is contrary for Q_H. Hydrogen bond donor acidity is obviously better explained by a descriptor based on only one atom (Q_H) in the small molecule data set whereas the larger drug molecules need a descriptor based on more atoms to describe the whole functionality ($\sum Q_H$). Also β correlates better to the parameter based on many atoms, $\sum Q_{MN}$, in the drug data set (Table 3), whereas in the small molecular data set the correlation coefficient between β and Q_{MN} is higher ($R^2 = 0.52$ versus 0.18 for the correlation between β and Q_{MN} and β and $\sum Q_{MN}$, respectively). Thus, the observation that $\Delta \log P$ clusters among the hydrogen bond donor descriptors in the small molecule data set (PCA2) does not necessarily imply that this is also valid for data set 1.

The score plot of PCA3 can also be used to investigate the diversity of data set 1a within the drug data set (data set 1). The molecules of data set 1a, depicted with red triangles, can be found reasonably well distributed amongst the drug data set. This supports our previous findings, which were based on a smaller reference data set, that data set 1a is representative for drugs [6].

3.3. Correlation of human jejunal permeability (in vivo) to different hydrogen bonding descriptors

Data set 1a consists of drug compounds for which human intestinal permeability data was available. All 29 descriptors derived in this study were used in a first PLS analysis which resulted in a two component model with $Q^2 = 0.67$ and $R^2 = 0.87$ (Table 4, PLS1). In the scores plot (Fig. 6a) it can be seen that piroxicam is an outlier. We have previously observed that this compound, which has a rather high $\log P_{eff}$-value, is predicted incorrectly, due to its low lipophilicity ($\log P_{oct} = -0.75$) [68] and its numerous

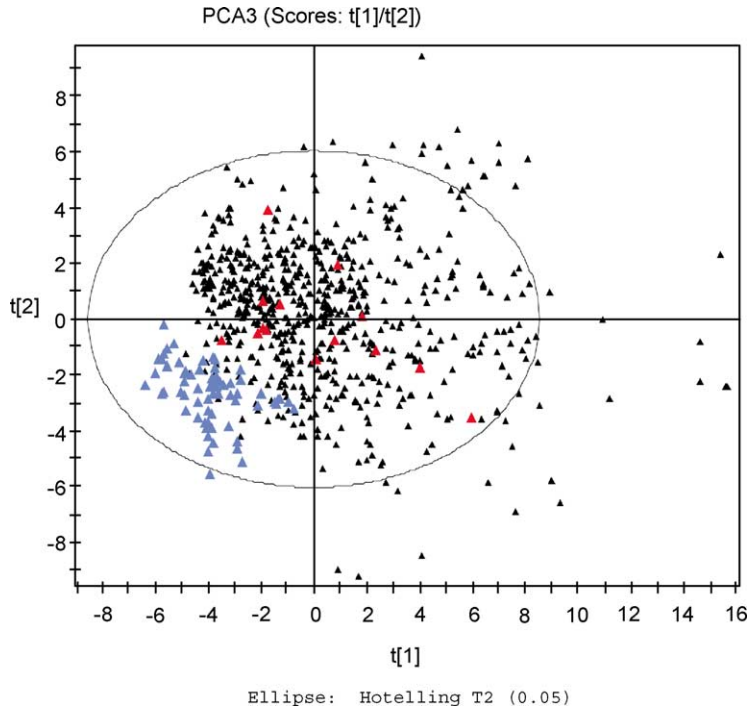

Fig. 5. Scores plot of PCA3 (data set 1a depicted with red triangles, data set 2 with blue triangles and data set 1 without data set 1a with black triangles).

Table 4
List of most interesting PLS models

Model	R^2	Q^2	Number of compounds	Number of molecules	Descriptors used
PLS1	0.872	0.670	2	14	All 29 descriptors
PLS2	0.960	0.873	3	13	All 29 descriptors
PLS3	0.953	0.907	2	13	15 Descriptors[a]
PLS4	0.952	0.942	1	13	HB, $\sum Q_H$ and $\log P_{oct}$
PLS5	0.945	0.935	1	13	PSA, HBD, $\log P_{oct}$
PLS6	0.864	0.815	1	13	PSA, HBD
PLS7	0.863	0.824	1	13	PSA, CWPSAHBD
PLS8	0.855	0.846	1	13	HBD, PSA/NPSA
PLS9	0.947	0.935	1	13	HB, $\sum Q_H$ and $\log P_{Cr}$
PLS10	0.945	0.932	1	13	PSA, HBD, $\log P_{Cr}$

[a] HB, PSA, PSA_{NO}, PSA_{rel}, PSA/NPSA, PSA_{CW}, α, $\sum Q_H$, HBD, $CWPSA_{HBD}$, Q_{MN}, $\sum Q_{MN}$, $CWPSA_{HBA}$, $\log P_{Cr}$, $\log P_{oct}$.

hydrogen bonding groups (properties that in general give low permeability) [6]. One reason for this inconsistency may be the problem to unambigously define the structure and thus calculate the properties correctly, since several tautomers can exist [71–73]. Furthermore, it is also known that piroxicam is not very soluble in water, despite its low lipophilicity. Thus, we omitted piroxicam and performed a new PLS analysis. PLS2 resulted in a three component model with much higher Q^2- and R^2-values (Table 4), whereas the scores plot, considering only the first component, did not look very different (compare Fig. 6a and b).

Analyzing the Variable Importance in the Projection (VIP-) plot (see Fig. 7) we identified 15 descriptors that had a VIP-value greater than 1, which indicates that they have an above average importance for explaining $\log P_{eff}$ [74]. In PLS3 only these 15 descriptors were taken into account. The analysis resulted in a two-component model with $Q^2 = 0.91$ and $R^2 = 0.95$. However, many of the variables were highly correlated, e.g. $CWPSA_{HBD}$, $\sum Q_H$ and HBD or $\log P_{oct}$ and $\log P_{Cr}$ (compare the correlation coefficients for these parameters within data set 1 in Table 3). We aimed to reduce the number of variables even further to obtain a simpler model. By eliminating the variables one by one as described in Section 2 we found the combination of $\sum Q_H$, HB and $\log P_{oct}$ to give a high Q^2-value in a one component model (PLS4).

Table 4 additionally lists models PLS5 and PLS6, containing the same descriptors that were used in our previous study [6]. However, PSA was now calculated using Savol instead of Molcad and $\log P$ using ABSOLV instead of the $C\log P$ program. Roughly the same statistics and equations were obtained. When $\log P_{Cr}$ was used as lipophilicity descriptor (PLS10) the resulting model was very similar to PLS5.

To summarize, it seems that combinations of a general hydrogen bonding descriptor, a hydrogen bond donor descriptor and a lipophilicity descriptor give the best models. We have tested several such combinations (PSA, PSA_{NO} or HB as general hydrogen bonding descriptor, $\sum Q_H$, HBD or $CWPSA_{HBD}$ as hydrogen bond donor descriptor and $\log P_{oct}$ or $\log P_{Cr}$ as lipophilicity descriptor) and found Q^2-values between 0.91 and 0.94 in all cases (e.g. PLS4, PLS5, PLS9 and PLS10 in Table 4). If the lipophilicity descriptor was omitted, Q^2 decreased and values between 0.72

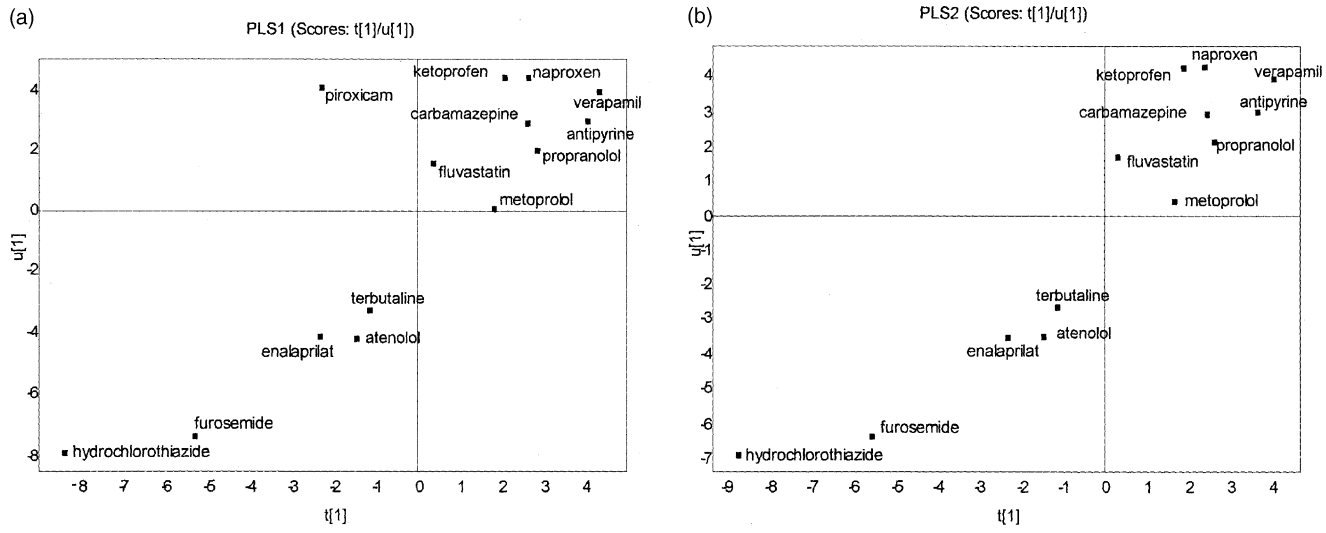

Fig. 6. Scores plot of (a) PLS1, (b) PLS2.

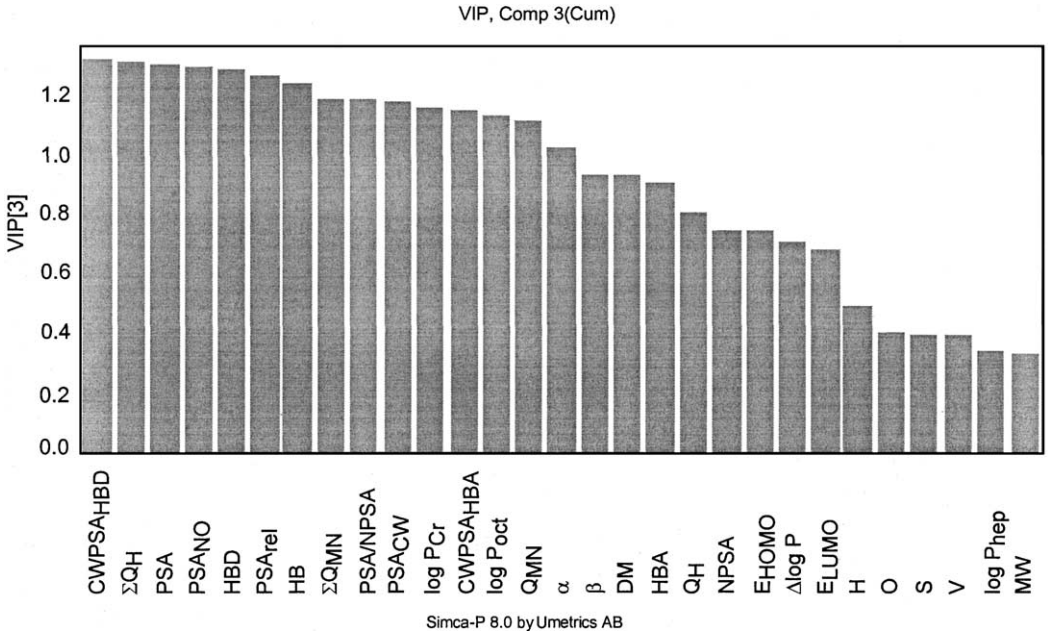

Fig. 7. Variable importance in projection plot (PLS2).

Table 5
Coefficients for the model equations

Model	Const	HB	PSA	HBD	$\sum Q_H$	$CWPSA_{HBD}$	$\log P_{oct}$	$\log P_{Cr}$	PSA/NPSA
PLS4	−3.102	−0.115	–	–	−0.993	–	0.140	–	–
PLS5	−3.092	–	−0.0089	−0.217	–	–	0.137	–	–
PLS6	−2.434	–	−0.0111	−0.271	–	–	–	–	–
PLS7	−2.499	–	−0.0106	–	–	−0.090	–	–	–
PLS8	−2.834	–	–	−0.295	–	–	–	–	−1.304
PLS9	−3.142	−0.113	–	–	−0.980	–	–	0.175	–
PLS10	−3.128	–	−0.0088	−0.215	–	–	–	0.172	–

(combination of HBD and HB alone) and 0.82 (PLS7, Table 4) were found. If NPSA/PSA, a descriptor that gives information about the lipophilic–hydrophilic properties of the molecule, was included as the lipophilicity descriptor Q^2-values up to 0.86 could be obtained. If either NPSA/PSA or PSA_{rel} was used instead of both a general hydrogen bonding descriptor and a lipophilicity descriptor Q^2-values between 0.82 and 0.85 (PLS8, Table 4) were found. The coefficients for the resulting equations are listed in Table 5. It should be noted that the coefficients for PLS5 (or PLS10) and PLS6 deviate slightly from the equations given previously [6], which is mainly due to the changes in how the descriptors were calculated (polar surface area and lipophilicity were calculated using different programs).

To test the seven models suggested above (Tables 4 and 5) we used them to predict the intestinal permeability of desipramine,[3] losartan [54], cimetidine and ranitidine [53], which had not been used for deriving the models. All but four of the 28 predicted $\log P_{eff}$-values were within 0.5 units of the respective measured values (three were within 0.6 and the worst value was still within 0.7 units). For example, PLS model PLS4 gave good predictions for cimetidine and losartan (residues 0.12 and 0.27, respectively) and slightly worse predictions for desipramine (residue 0.45) and ranitidine (residue 0.52).

Due to the high variability of human in vivo permeability data (see S.D. values for $\log P_{eff}$ in Table 2) one has to be prepared that data predicted by any derived model will show at least as much variability as the experimental data. Therefore, we cannot expect to obtain better predictions than those described above.

3.4. Comparison of predicted log P_{eff}-values to experimental fraction absorbed (FA%-) values

We have further tested the models derived above by plotting predicted permeability values against reported FA%-data of 67 compounds (data set 1b [38]) and found them all to give reasonable estimates. Fig. 8 shows such

[3] $\log P_{eff}$ is −3.36 [6], HB = 2, PSA = 17.87, HBD = 1, $\sum Q_H$ = 0.15, CWPSAHBD = 1.33, $\log P_{Cr}$ = 3.64 and PSA/NPSA = 0.06 were calculated as described in Section 2 and the $C\log P$-value (= 4.09, [6]) was used as $\log P_{oct}$.

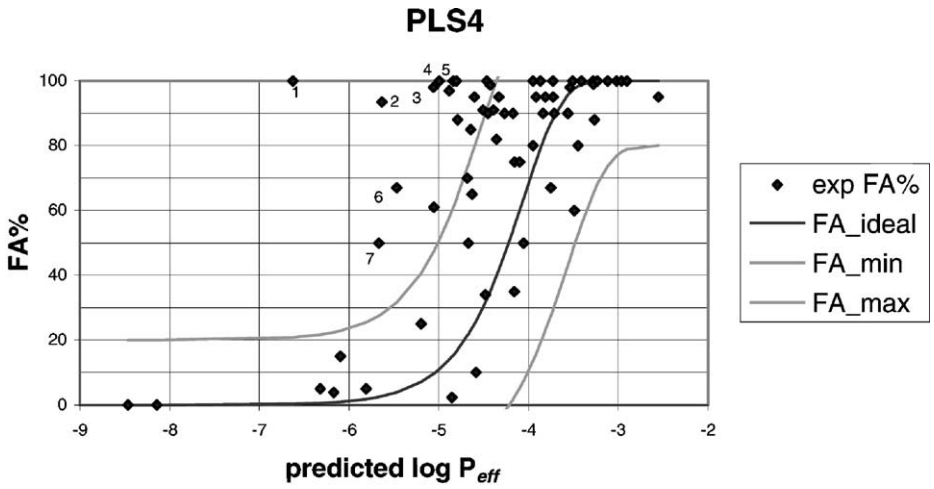

Fig. 8. Predicted log P_{eff} according to PLS4 vs. FA% (methotrexate (1), amoxicillin (2), cephalexin (3), loracarbef (4), zidovudine (5), hydrochlorothiazide (6) and etoposide (7)).

a plot for PLS4. We added a sigmoidal curve to model the ideal relationship between log P_{eff} and FA%, using the equation FA% $= (1 - e^{(2P_{\text{eff}}(t_{\text{res}}/r)f)}) \times 100$ ($t_{\text{res}} = 3$ h, $r = 1.75$ cm and $f = 2.8$) [75]. We also drew curves to indicate upper and lower limits for acceptable prediction. A variability of log P_{eff} by ± 0.5 units and FA% by $\pm 20\%$ was tolerated. Fig. 8 shows that the model predicts reasonable log P_{eff}-values for most drugs, i.e. most compounds fall into the region between the curves indicating minimum and maximum FA%-values for a given permeability. However, some compounds can be found that are more extensively absorbed than would be expected from their log P_{eff}-values.

The two most extreme outliers, having predicted log P_{eff}-values below -5.5, but FA%-values higher than 90%, are methotrexate and amoxicillin. Both are known to be actively absorbed [6,24]. Also cephalexin, loracarbef and zidovudine have been reported to be actively absorbed [24,76]. All three compounds show an experimental FA%-value of almost 100%, but their log P_{eff}-values were predicted to be around -5. In the medium absorption range only three compounds were found to have greater absorption than would be expected from their predicted log P_{eff}-values. Hydrochlorothiazide, a compound that was used in the model development, is reported to be absorbed to 67% but its calculated permeability is -5.47 (FA% should be below 10%). The experimental log P_{eff}-value for this compound is -5.4 [6], which shows that small changes in the medium range can give large differences due to the steepness of the curve in this region. The other outlier is etoposide, again showing a low log P_{eff}-value (-5.67) compared to its FA% (50%). A recent study indicates that this compound may be actively transported across the blood brain barrier [24,77]. Active transport may thus be a reason for its higher than predicted intestinal permeability. However, it has also been suggested that etoposide is subject to efflux mediated transport [78].

4. Conclusions

In this study we did not identify a correlation between $\Delta \log P$ and human intestinal permeability, although this parameter has previously been found useful for modelling structure permeability relationships. Furthermore, $\Delta \log P$ was not found to be an important parameter in the multivariate analysis for modelling intestinal permeability, which is probably due to the fact that $\Delta \log$ P could only be determined for the seven most lipophilic compounds of our thirteen training set molecules.

In order to compare the different hydrogen bonding descriptors we used two datasets: a drug data set, data set 1, for which the experimental $\Delta \log P$ value had not been determined and a data set consisting of small molecules, data set 2, for which both $\Delta \log P$ and α and β values were available. It could be confirmed that $\Delta \log P$ is highly correlated to α in data set 2. However, this is not necessarily true for the drug data set, since many differences between the two datasets were found. This underlines the necessity of using an appropriate data set. One should be cautious when using non-drug data sets for deriving information about drugs.

Using thirteen diverse drugs for which log P_{eff}-values describing the passive transcellular absorption were known several models were developed combining hydrogen bonding and lipophilicity descriptors. Interestingly, we found that independent of which of the hydrogen bond donor descriptors we used (except α and $\Delta \log$ P) statistically good equations could always be obtained. The inclusion of hydrogen bond acceptor descriptors, which has been reported to be

valuable for describing permeability across the blood brain barrier [21], did not seem to be equally important for modeling intestinal permeability.

All derived models were able to predict $\log P_{\text{eff}}$ within 0.7 units from the experimental values for the test set of four compounds. These models gave also reasonable estimates of FA% for an external data set of 67 compounds. The statistics of model PLS4, which was used as an example, are not necessarily superior as compared to models PLS5, PLS6 or PLS10, which correspond to the models previously presented [6]. However, the variables $\sum Q_{\text{H}}$, HB and $\log P_{\text{oct}}$ may be more straightforward to calculate.

Acknowledgements

Sirius Analytical Instruments Ltd. is gratefully acknowledged for calculating the α-, β- and $\log P_{\text{oct}}$-values of data set 1 using the ABSOLV program. We also thank the Swedish Foundation for Strategic Research (SFF) for financial support.

References

[1] C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, Adv. Drug Delivery Rev. 23 (1997) 3–25.
[2] D.A. Smith, B.C. Jones, D.K. Walker, Design of drugs involving the concepts and theories of drug metabolism and pharmacokinetics, Med. Res. Rev. 16 (1996) 243–266.
[3] S. Ren, A. Das, E.J. Lien, QSAR analysis of membrane permeability to organic compounds, J. Drug Target. 4 (1996) 103–107.
[4] H.W. Hamilton, B.A. Steinbaugh, B.H. Stewart, O.H. Chan, H.L. Schmid, R. Schroeder, M.J. Ryan, J. Keiser, M.D. Taylor, C.J. Blankley, J.S. Kaltenbronn, J. Wright, J. Hicks, Evaluation of physicochemical parameters important to the oral bioavailability of peptide-like compounds: implications for the synthesis of renin inhibitors, J. Med. Chem. 38 (1995) 1446–1455.
[5] H. Lennernäs, Ö. Ahrenstedt, R. Hällgren, L. Knutson, M. Ryde, L.K. Paalzow, Regional Jejunal perfusion, a new in vivo approach to study oral drug absorption in man, Pharm. Res. 9 (1992) 1243–1251.
[6] S. Winiwarter, N.M. Bonham, F. Ax, A. Hallberg, H. Lennernäs, A. Karlén, Correlation of human jejunal permeability (in vivo) of drugs with experimentally and theoretically derived parameters. A multivariate data analysis approach, J. Med. Chem. 41 (1998) 4939–4949.
[7] R.A. Conradi, A.R. Hilgers, N.F.H. Ho, P.S. Burton, The influence of peptide structure on transport across Caco-2 cells, Pharm. Res. 8 (1991) 1453–1460.
[8] W.D. Stein, The molecular basis of diffusion across cell membranes, in: The Movement of Molecules Across Cell Membranes, Academic Press, New York, 1967, pp. 65–125.
[9] M.H. Abraham, Scales of solute hydrogen-bonding: their construction and application to physicochemical and biochemical processes, Chem. Soc. Rev. 22 (1993) 73–83.
[10] M.H. Abraham, H.S. Chadha, G.S. Whiting, R.C. Mitchell, Hydrogen bonding 32. An analysis of water–octanol and water–alkane partitioning and the $\Delta\log P$ parameter of Seiler, J. Pharm. Sci. 83 (1994) 1085–1100.
[11] J.A. Platts, D. Butina, M.H. Abraham, A. Hersey, Estimation of molecular linear free energy relation descriptors using a group contribution approach, J. Chem. Inf. Comput. Sci. 39 (1999) 835–845.
[12] O.A. Raevsky, V.Y. Grigorév, D.B. Kireev, N.S. Zefirov, Complete thermodynamic description of H-bonding in the framework of multiplicative approach, Quant. Struct. Act. Relat. 11 (1992) 49–63.
[13] H. van de Waterbeemd, G. Camenisch, G. Folkers, J.R. Chretien, O.A. Raevsky, Estimation of blood-brain barrier crossing of drugs using molecular size and shape, and H-bonding descriptors, J. Drug Target. 6 (1998) 151–165.
[14] G. Cruciani, M. Pastor, W. Guba, VolSurf: a new tool for the pharmacokinetic optimization of lead compounds, Eur. J. Pharm. Sci. 11 (2000) S29–39.
[15] P. Sjöberg, MolSurf—a generator of chemical descriptors for QSAR, in: H. van de Waterbeemd, B. Testa, G. Folkers (Eds.), Computer-Assisted Lead Finding and Optimization: Current Tools for Medicinal Chemistry, VHCA, Basel, 1997, pp. 83–92.
[16] P. Goodford, A computational program for determining energetically favorable binding sites on biologically important macromolecules, J. Med. Chem. 28 (1985) 849–857.
[17] H. van de Waterbeemd, M. Kansy, Hydrogen-bonding capacity and brain penetration, Chimia 46 (1992) 299–303.
[18] T. Österberg, U. Norinder, Prediction of polar surface area and drug transport processes using simple parameters and PLS statistics, J. Chem. Inf. Comput. Sci. 40 (2000) 1408–1411.
[19] C. Frömmel, The apolar surface area of amino acids and its empirical correlation with hydrophobic free energy, J. Theor. Biol. 111 (1984) 247–260.
[20] K. Palm, K. Luthman, A.-L. Ungell, G. Strandlund, P. Artursson, Correlation of drug absorption with molecular surface properties, J. Pharm. Sci. 85 (1996) 32–39.
[21] M. Feher, E. Sourial, J.M. Schmidt, A simple model for the prediction of blood-brain partitioning, Int. J. Pharm. 201 (2000) 239–247.
[22] K. Palm, P. Stenberg, K. Luthman, P. Artursson, Polar molecular surface properties predict the intestinal absorption of drugs in humans, Pharm. Res. 14 (1997) 568–571.
[23] K. Palm, K. Luthman, A.-L. Ungell, G. Strandlund, F. Beigi, P. Lundahl, P. Artursson, Evaluation of dynamic molecular surface area as predictor of drug absorption: comparison with other computational and experimental predictors, J. Med. Chem. 41 (1998) 5382–5392.
[24] D.E. Clark, Rapid calculation of polar molecular surface area and its application to the prediction of transport phenomena. Part 1. Prediction of intestinal absorption, J. Pharm. Sci. 88 (1999) 807–814.
[25] P. Stenberg, K. Luthman, P. Artursson, Virtual screening of intestinal drug permeability, J. Contr. Rel. 65 (2000) 231–243.
[26] P. Ertl, B. Rohde, P. Selzer, 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. 43 (2000) 3714–3717.
[27] J.C. Dearden, T. Ghafourian, Hydrogen bonding parameters for QSAR: comparison of indicator variables, hydrogen bond counts, molecular orbital and other parameters, J. Chem. Inf. Comput. Sci. 39 (1999) 231–235.
[28] L.Y. Wilson, G.R. Famini, Using theoretical descriptors in quantitative structure-activity relationships: some toxicological indices, J. Med. Chem. 34 (1991) 1668–1674.
[29] D.T. Stanton, P.C. Jurs, Development and use of charged partial surface area structural descriptors in computer-assisted quantitative structure-property relationship studies, Anal. Chem. 62 (1990) 2323–2329.
[30] A.R. Katritzky, L. Mu, V.S. Lobanov, M. Karelson, Correlation of boiling points with molecular structure. Part 1. A training set of 298 diverse organics and a test set of 9 simple inorganics, J. Phys. Chem. 100 (1996) 10400–10407.
[31] P.S. Burton, R.A. Conradi, N.F.H. Ho, A.R. Hilgers, R.T. Borchardt, How structural features influence the biomembrane permeability of peptides, J. Pharm. Sci. 85 (1996) 1336–1340.

[32] L. Hjorth Krarup, I.T. Christensen, L. Hovgaard, S. Frokjaer, Predicting drug absorption from molecular surface properties based on molecular dynamics simulations, Pharm. Res. 15 (1998) 972–978.
[33] D.E. Clark, Rapid calculation of polar molecular surface area and its application to the prediction of transport phenomena. Part 2. Prediction of blood–brain barrier penetration, J. Pharm. Sci. 88 (1999) 815–821.
[34] J. Kelder, P.D.J. Grootenhuis, D.M. Bayada, L.P.C. Delbressine, J.-P. Ploemen, Polar molecular surface as a dominating determinant for oral absorption and brain penetration of drugs, Pharm. Res. 16 (1999) 1514–1519.
[35] U. Norinder, T. Österberg, P. Artursson, Theoretical calculation and prediction of intestinal absorption of drugs in humans using MolSurf parametrization and PLS statistics, Eur. J. Pharm. Sci. 8 (1999) 49–56.
[36] G. Cruciani, P. Crivori, P.-A. Carrupt, B. Testa, Molecular fields in quantitative structure-permeation relationships: the VolSurf approach, J. Mol. Struct. (THEOCHEM) 503 (2000) 17–30.
[37] L. Hjorth Alifrangis, I.T. Christensen, A. Berglund, M. Sandberg, L. Hovgaard, S. Frokjaer, Structure-property model for membrane partitioning of oligopeptides, J. Med. Chem. 43 (2000) 103–113.
[38] M.D. Wessel, P.C. Jurs, J.W. Tolan, S.M. Muskal, Prediction of human intestinal absorption of drug compounds from molecular structure, J. Chem. Inf. Comput. Sci. 38 (1998) 726–735.
[39] H. van de Waterbeemd, G. Camenisch, G. Folkers, O.A. Raevsky, Estimation of Caco-2 cell permeability using calculated molecular descriptors, Quant. Struct. Act. Relat. 15 (1996) 480–490.
[40] T.I. Oprea, J. Gottfries, Toward minimalistic modeling of oral drug absorption, J. Mol. Graphics Mod. 17 (1999) 261–274.
[41] P. Seiler, Interconversion of lipophilicities from hydrocarbon/water systems into the octanol/water system, Eur. J. Med. Chem. 9 (1974) 473–479.
[42] R.C. Young, R.C. Mitchell, T.H. Brown, C.R. Ganellin, R. Griffiths, M. Jones, K.K. Rana, D. Saunders, I.R. Smith, N.E. Sore, T.J. Wilks, Development of a new physicochemical model for brain penetration and its application to the design of centrally acting H2 receptor histamin antagonists, J. Med. Chem. 31 (1988) 656–671.
[43] N. El Tayar, R.-S. Tsai, B. Testa, P.-A. Carrupt, C. Hansch, A. Leo, Percutaneous penetration of drugs: a quantitative structure-permeability relationship study, J. Pharm. Sci. 80 (1991) 744–749.
[44] E.G. Chikhale, K.-Y. Ng, P.S. Burton, R.T. Borchardt, Hydrogen bonding potential as a determinant of the in vitro and in situ blood-brain barrier permeability of peptides, Pharm. Res. 11 (1994) 412–419.
[45] F. Yoshida, J.G. Topliss, Unified model for the corneal permeability of related and diverse compounds with respect to their physicochemical properties, J. Pharm. Sci. 85 (1996) 819–823.
[46] T.W. von Geldern, D.J. Hoffman, J.A. Kester, H.N. Nellans, B.D. Dayton, S.V. Calzadilla, K.C. Marsh, L. Hernandez, W. Chiou, D.B. Dixon, J.R. Wu-Wong, T.J. Opgenorth, Azole Endothelin Antagonists. Part 3. Using $\Delta\log P$ as a tool to improve absorption, J. Med. Chem. 39 (1996) 982–991.
[47] M. Yazdanian, S.L. Glynn, J.L. Wright, A. Hawi, Correlating partitioning and Caco-2 cell permeability of structurally diverse small molecular weight compounds, Pharm. Res. 15 (1998) 1490–1494.
[48] N. El Tayar, R.-S. Tsai, B. Testa, P.-A. Carrupt, A. Leo, Partitioning of solutes in different solvent systems: the contribution of hydrogen-bonding capacity and polarity, J. Pharm. Sci. 80 (1991) 590–598.
[49] A. Tsantili-Kakoulidou, A. Varvaresou, T. Siatra-Papastaikoudi, O.A. Raevsky, A comprehensive investigation of the partitioning and hydrogen bonding behavior of indole containing derivatives of 1,3,4-thiadiazole and 1,2,4-triazole by means of experimental and calculative approaches, Quant. Struct. Act. Relat. 18 (1999) 482–489.
[50] FASS 2000, Elanders, Stockholm (2000).
[51] G.L. Amidon, H. Lennernäs, V.P. Shah, J.R. Crison, A theoretical basis for a biopharmaceutical drug classification: the correlation of in vitro drug product dissolution and in vivo bioavailability, Pharm. Res. 12 (1995) 413–420.
[52] R. Sandström, A. Karlsson, L. Knutson, H. Lennernäs, Jejunal absorption and metabolism of R/S-Verapamil in humans, Pharm. Res. 15 (1998) 856–862.
[53] N. Takamatsu, O.-N. Kim, L.S. Welage, N.M. Idkaidek, Y. Hayashi, J. Barnett, R. Yamamoto, E. Lipka, H. Lennernäs, A. Hussain, L. Lesko, G.L. Amidon, Human jejunal permeability of two polar drugs: cimetidine and ranitidine, Pharm. Res. 18 (2001) 742–744.
[54] H. Lennernäs, unpublished data.
[55] D. Dyrssen, Studies on the extraction of metal complexes. Part IV. The dissociation constants and partition coefficients of 8-quinolinol (oxine) and *N*-nitroso-*N*-phenylhydroxylamine (Cupferron), Svensk kem. tidsk. 64 (1952) 213–224.
[56] A. Avdeef, pH-metric log *P*. Part 1. Difference plots for determining ion-pair octanol–water partition coefficients of multiprotic substances, Quant. Struct. Act. Relat. 11 (1992) 510–517.
[57] N.M. Bonham, unpublished data.
[58] G. Schill, H. Ehrsson, J. Vessman, D. Westerlund, Separation methods for drugs and related organic compounds, 2nd ed., Swedish Pharmaceutical Press, Stockholm, 1978.
[59] D. Weininger, SMILES, a chemical language and information system. Part 1. Introduction to methodology and encoding rules, J. Chem. Inf. Comput. Sci. 28 (1988) 31–36.
[60] SYBYL: Molecular Modeling Software, Tripos Inc., 1699 South Hanley Rd., St. Louis, MO 63144, USA.
[61] R.S. Pearlman, CONCORD: rapid generation of high quality approximate 3D molecular structures, Chem. Design Automation News 2 (1987) 1.
[62] M. Clark, R.D. Cramer III, N. Van Opdenbosch, Validation of the general purpose Tripos 5.2 force field, J. Comput. Chem. 10 (1989) 982–1012.
[63] M.J.S. Dewar, E.G. Zoebisch, E.F. Healy, J.J.P. Stewart, AM1: a new general purpose quantum mechanical molecular model, J. Am. Chem. Soc. 107 (1985) 3902–3909.
[64] J.J.P. Stewart, MOPAC: a semiempirical molecular orbital program, J. Comput. Aided Mol. Des. 4 (1990) 1–105.
[65] R.S. Pearlman, in: W.J. Dunn, J.H. Block, R.S Pearlman (Eds.), Partition coefficient: determination and estimation, Pergamon Press, New York, NY, 1986, pp. 165.
[66] R.S. Pearlman, Sarea, QCPE # 413.
[67] F. Ax, S. Winiwarter, C. Pettersson, H. Lennernäs, A. Karlén. Correlation of Human in vivo Jejunal Permeability with Hydrogen Bonding Descriptors, Lipophilicity in Drug Disposition, University of Lausanne (Switzerland), VHCA Zürich Wiley-VCH, Weinheim, New York, 2000.
[68] ABSOLV Software for predicting solute properties from molecular structure, Sirius Analytical Instruments Ltd., Riverside, East Sussex, RH18 5DW, UK, 2000.
[69] A.K. Ghose, G.M. Crippen, Atomic physicochemical parameters for three-dimensional structure-directed quantitative structure-activity relationships. Part I. Partition coefficients as a measure of hydrophobicity, J. Comput. Chem. 7 (1986) 565–577.
[70] SIMCA-P, 1999, Umetri AB: SE-90719 Umeå.
[71] E. Bernhard, F. Zimmermann, Contribution to the understanding of oxicam ionization constants, Arzneim. Forsch./Drug Res. 34 (1984) 647–648.
[72] G. Reck, G. Dietz, G. Laban, W. Günther, G. Bannier, E. Höhne, X-ray studies on piroxicam modifications, Pharmazie 43 (1988) 477–481.
[73] K. Takács-Novák, J. Kökösi, B. Podányi, B. Noszál, R.-S. Tsai, G. Lisa, P.-A. Carrupt, B. Testa, Microscopic protonation/deprotonation equilibria of the anti-inflammatory agent piroxicam, Helv. Chim. Acta 78 (1995) 553–562.
[74] S. Wold, Validation of QSAR's, Quant. Struct. Act. Relat. 10 (1991) 191–193.

[75] U. Fagerholm, M. Johansson, H. Lennernäs, Comparison between permeability coefficients in rat and human jejunum, Pharm. Res. 13 (1996) 1336–1342.
[76] J.D. Irvine, L. Takahashi, K. Lockhart, J. Cheong, J.W. Tolan, H.E. Selick, J.R. Grove, MDCK (Madin–Darby Canine kidney) cells: a tool for membrane permeability screening, J. Pharm. Sci. 88 (1999) 28–33.
[77] D.E. Burgio, M.P. Gosland, P.J. McNamara, Effects of P-glycoprotein modulators on etoposide elimination and central nervous system distribution, J. Pharmacol. Exp. Ther. 287 (1998) 911–917.
[78] W.M. Kan, Y.T. Liu, C.L. Hsiao, C.Y. Shieh, J.H. Kuo, J.D. Huang, S.F. Su, Effect of hydroxyzine on the transport of etoposide in rat small intestine, Anticancer Drugs 12 (2001) 267–273.