Eur. J. Med. Chem. 33 (1998) 799–807

Quantitative estimation of hydrogen bond contribution to permeability and absorption processes of some chemicals and drugs

Oleg A. Raevsky[a], Klaus-Jürgen Schaper[b]*

[a]*Laboratory of Computer-Aided Molecular Design, Institute of Physiologically Active Compounds, Russian Academy of Sciences, 142432, Chernogolovka, Moscow region, Russia*
[b]*Research Center Borstel, Center for Medicine and Biosciences, Parkallee 1–40, D-23845 Borstel, Germany*

(Received 26 January 1998; accepted 13 May 1998)

Abstract – The H-bond donor and acceptor descriptors ΣC_d and ΣC_a, which are estimated directly from thermodynamic data of hydrogen bonding, were successfully used for the correlation with permeability and absorption data for some chemicals and drugs. The evaluation of different types of permeability test systems and of different classes of compounds showed that in addition to steric bulk effects both the H-bond donor and acceptor strength play an important role in explaining differences in permeability and absorption of neutral chemical compounds and drugs. However, because of the frequently observed intercorrelation between ΣC_d and ΣC_a, often only the more significant of them leads to a significant regression coefficient in multiple linear regression equations. In comparison with ΣC_d and/or ΣC_a less significant correlations are obtained with the experimental parameter $\Delta\log P$ (the difference between the octanol/water partition coefficient $\log P_{Oct}$ and $\log P$ for the system alkane or cyclohexane/water) which has to be considered as a composed descriptor containing H-bond donor as well as H-bond acceptor effects. © Elsevier, Paris

H-bond donor / H-bond acceptor / descriptors / membrane / permeability / absorption / QSAR

1. Introduction

The recognition of factors governing permeability and absorption processes of chemicals as well as the development of possibilities for their calculation is very important for drug design. $\text{Log}P_{Oct}$, the octanol/water partition coefficient, is the most popular parameter for the description of passive transport processes of chemicals in organisms [1]. In 1974 Seiler proposed the experimental $\Delta\log P$ parameter as the difference between $\log P_{Oct}$ and $\log P$ for the system alkane or cyclohexane/water to account for hydrogen bonding contribution to transport processes [2]. Young and coworkers [3] applied the concept of $\Delta\log P$ to design brain-penetrating H_2 histamine receptor antagonists, while Scott et al. [4] correlated the percutaneous absorption of a limited number of compounds to $\Delta\log P$. It has been shown by El Tayar et al. [5] that $\Delta\log P$ reflects mainly the hydrogen bond donor ability of chemicals. Measurements of $\log P_{Oct}$ and $\Delta\log P$ can be done only after the synthesis of the compounds. That is why different procedures for the determination of $\log P_{Oct}$ and $\Delta\log P$ on an empirical or theoretical basis have been developed during the last years. Based on the so-called solvation equation several applications of hydrogen bond parameters (acidity α and basicity β) have been presented to quantitatively describe drug transport properties [6–8]. The solvation equation contains a rather large number of descriptors including the excess molar refraction, the solute dipolarity/polarizability parameter, the solute hydrogen bond acidity, the hydrogen bond basicity and the characteristic volume of McGowan. The authors emphasize that hydrogen bond acidity and basicity values "must be amended where necessary in order to reflect possible multiple hydrogen bonding in solution".

Raevsky et al. [9] proposed the following equation for the quantitative description of the free energy of hydrogen bonds by means of the product of H-bond donor and acceptor factors:

$$\Delta G = k' C_d \, C_a + k_0 \qquad (1)$$

*Correspondence and reprints

where C_d and C_a are free energy H-bond donor and H-bond acceptor factors which characterize the relative donor and acceptor strength of compounds.

Recently those factors were successfully used together with polarizability (Pol, which quantitatively describes bulk or volume-related effects) for the description of the octanol/water partition coefficient [10–12] of neutral organic compounds:

$$\log P_{Oct} = 0.25(\pm 0.01)\text{Pol} - 0.68(\pm 0.02)\Sigma C_a - 0.01(\pm 0.08); \quad (2)$$
$$n = 234,\ r = 0.961,\ s = 0.34,\ F = 1383,\ \text{SDEP} = 0.35.$$

In equation (2), ΣC_a is the sum of the hydrogen bond acceptor factor values of all substructures in a molecule accepting H-bonds, n is the number of compounds, r is the correlation coefficient, s is the standard deviation between observed and fitted log$P$, F is the ratio of (explained variance)/(unexplained variance), SDEP is the standard deviation of the error of prediction, and the values in parentheses are the 95% confidence intervals of the regression coefficients.

Here it should be noted that the inclusion of the H-bond donor descriptor ΣC_d in equation (2) did not improve the relationship despite the presence of eighty H-bond donor molecules in the data set. So it is possible to conclude on the basis of equation (2) that only the H-bond acceptor ability plays an essential role for the lipophilicity of organic compounds.

The H-bond descriptors C_a and C_d (and also C_{ad}, the sum of the absolute values of the free energy H-bond factors, characterizing the total H-bond ability of a compound) have recently been used for the first time by van de Waterbeemd et al. [13] to describe the Caco-2 cell permeability of a few compounds.

The present report is devoted to the application of a similar approach to the quantitative description of permeability and absorption processes of chemicals and drugs using only a few significant descriptors.

2. Results

Before a correlation can be derived between physicochemical or other descriptors and permeability or absorption the quality of the available permeability and absorption data has to be considered. Unfortunately in most cases there are not enough data obtained under the same experimental conditions and in the same laboratory. The combination of such data of different authors in a common training set is problematic and may distort the physicochemical meaning of observed regression equations. That is why in the present investigation the authors decided to derive correlations between permeability or absorption and physicochemical descriptors for separate data sets of compounds.

2.1. Human red cell basal permeability (BP) of alcohols, water, urea and thiourea

The set contains ten rather simple compounds. The cell membrane permeability data (log[BP]) and lipophilicity values in *table I* are taken from Stein [14]. The permeability has the dimension of cm/sec and is the speed with which a diffusant moves across the medium considered. The *basal* permeability (BP) is the permeability in the absence of any specific transport pathway (i.e. observed after inhibition of active transport mechanisms). *Table I* also contains polarizability and H-bond donor and acceptor data calculated for those compounds. The correlation of the permeability data with lipophilicity is rather poor:

$$\log[\text{BP}] = -3.31(\pm 1.75) + 0.79(\pm 0.81)\log P_{Oct}; \quad (3)$$
$$n = 10,\ r = 0.619,\ s = 1.86,\ r_{cv}^2 = -0.056.$$

In this equation r_{cv} is the 'leave-one-out' cross-validated correlation coefficient (observed vs. predicted log[BP], $r_{cv}^2 = 1 - \text{PRESS/MS}$, PRESS = Prediction Error Sum of Squares, MS = Sum of Mean Squares).

Table I. Human red cell basal permeability (BP, cm/s), lipophilicity (log$P$), polarizability (Pol), and H-bond acceptor/donor ability (ΣC_a, ΣC_d) of alcohols, water and (thio)urea.

No.	Compound	logP_{Oct}	Pol	ΣC_a	ΣC_d	log[BP]$_{obs.}$	log[BP]$_{calc.}$ (eq. (5))
1	Erythritol	–2.92	10.66	6.48	–7.06	–8.17	–8.30
2	Ethanediol	–1.92	5.72	3.14	–3.32	–4.54	–4.28
3	Ethanol	–0.32	5.08	1.82	–1.42	–2.68	–2.23
4	Glycerol	–2.55	8.19	4.81	–5.19	–6.80	–6.29
5	n-Hexanol	2.04	12.42	1.79	–1.59	–2.06	–2.41
6	Methanol	–0.74	3.25	1.76	–1.50	–2.43	–2.32
7	n-Propanol	0.34	6.92	1.60	–1.27	–2.19	–2.07
8	Thiourea	–1.14	8.08	5.61	–5.00	–5.96	–6.08
9	Urea	–2.66	5.72	3.44	–5.00	–6.11	–6.08
10	Water	–3.80	1.41	1.44	–2.88	–2.92	–3.80

According to Stein [14], Anderson et al. [15], and Lien et al. [16] the molecular bulk effect descriptor log[Molecular Weight] should be considered as a factor influencing permeability processes. And indeed a significant equation was obtained using log$P$ and log[MW] (intercorrelation: $r = 0.276$), whereas $(\log P)^2$ was not significant:

$$\log[BP] = 9.49(\pm 3.60) + 1.06(\pm 0.28)\log P - 7.01(\pm 1.95)\log[MW]; \quad (4)$$
$$n = 10,\ r = 0.972,\ s = 0.59,\ r_{cv} = 0.939.$$

However, for this set of compounds an excellent one-parameter correlation was obtained using only the H-bond donor strength. This correlation could not be improved by adding log[MW].

$$\log[BP] = -0.70(\pm 0.64) + 1.08(\pm 0.16)\Sigma C_d; \quad (5)$$
$$n = 10,\ r = 0.983,\ s = 0.43,\ r_{cv} = 0.976.$$

The H-bond acceptor strength (ΣC_a) is also significantly correlated with log[BP], but the statistical criteria indicate a lower significance as compared with equation (5):

$$\log[BP] = -0.76(\pm 1.19) - 1.14(\pm 0.33)\Sigma C_a; \quad (6)$$
$$n = 10,\ r = 0.943,\ s = 0.78,\ r_{cv} = 0.913.$$

A strong intercorrelation between the two descriptors ΣC_d and ΣC_a ($r = 0.928$) does not permit to use them simultaneously (see also Discussion). Nevertheless, it is obvious from this first data set that the H-bond ability plays an essential role in the human red cell basal permeability. In this data set as well as in all others discussed in this paper the authors have also systematically investigated the use of the hydrogen bond descriptor ΣC_{ad} mentioned in the introduction (ΣC_{ad}, the sum of the absolute values of the free energy H-bond factors, characterizing the total H-bond ability of a compound). However, in no case a correlation of permeability with this parameter was found with a significantly better predictive power. The reason for this result is possibly the contribution of both $|\Sigma C_a|$ and $|\Sigma C_d|$ to ΣC_{ad} with a fixed weight of one.

2.2. *Permeation of non-electrolytes through cells of the alga Chara ceratophylla*

This set of compounds has been studied by Lien et al. [16]. *Table II* contains the data of permeability (log[Per]), ether/water partition coefficients ($\log P_{e/w}$), polarizability, and H-bond strength of the compounds. The cited authors

Table II. Permeation (log[Per]) of various non-electrolytes through *Chara ceratophylla* cells, ether/water partition coefficient ($\log P_{e/w}$), polarizability, and H-bond strength.

No.	Compound	log[Per]	$\log P_{e/w}$	Pol	ΣC_a	ΣC_d
11	Trimethylcitrate	–0.6198	–0.367	19.21	9.00	–1.92
12	Cyanamide	–0.6778	–0.959	3.98	4.96	–2.50
13	i-Valeramide	–0.7212	–0.770	11.58	4.41	–4.16
14	Butyramide	–0.7696	–1.237	9.75	4.41	–4.16
15	Propionamide	–0.8861	–1.886	7.91	4.41	–4.16
16	Monochlorhydrin	–1.0458	–1.097	9.48	3.76	–3.53
17	Propyleneglycol	–1.0605	–1.620	7.55	3.24	–3.53
18	Glycerinethylether	–1.1135	–1.585	11.86	4.94	–3.55
19	Formamide	–1.1135	–2.854	4.24	4.36	–3.38
20	Succinimide	–1.2291	–1.509	9.13	6.68	–2.10
21	Diethylurea	–1.2291	–1.721	13.06	5.27	–3.86
22	Acetamide	–1.2757	–2.602	6.08	4.28	–4.16
23	Glycerinemethylether	–1.3665	–1.721	10.30	4.99	–3.53
2	Ethyleneglycol	–1.3665	–2.167	5.72	3.14	–3.32
24	Dimethylurea	–1.4685	–2.337	9.39	5.27	–3.86
25	Monacetin	–1.7959	–1.387	11.65	5.68	–3.53
26	Ethylurea	–1.9208	–2.387	9.39	5.27	–4.43
8	Thiourea	–2.1135	–2.201	8.08	5.61	–5.00
27	Methylurea	–2.1675	–2.921	7.55	5.27	–4.43
28	Lactamide	–2.2518	–2.745	8.55	6.08	–6.03
9	Urea	–2.3979	–3.328	5.72	3.44	–5.00
29	Dicyanodiamide	–2.9586	–2.538	8.29	11.24	–6.48
4	Glycerine	–3.1308	–3.180	8.19	4.81	–5.19
30	Mucic acid diethylester	–3.2757	–2.060	22.53	11.10	–7.48
31	Malonamide	–3.8539	–3.523	9.54	8.82	–8.32
1	Erythritol	–4.3372	–3.959	10.66	6.48	–7.06
32	Arabinose	–4.5086	–4.301	12.36	8.14	–8.29

obtained the best one-parameter correlation with $\log P_{e/w}$. For those 27 compounds for which we could estimate H-bond factor values a correlation with $r = 0.830$ was obtained with $\log P_{e/w}$.

Using H-bond donor strength values instead of log*P* a statistically significant improvement of the correlation was achieved:

$$\log[\text{Per}] = 0.83(\pm 0.57) + 0.59(\pm 0.12)\Sigma C_d; \quad (7)$$
$$n = 27,\ r = 0.903,\ s = 0.49,\ r_{cv} = 0.885.$$

In this set a significant intercorrelation of ΣC_d and $\log P_{e/w}$ is found ($r = 0.741$). So it is not possible to use them together. The addition of other descriptors (ΣC_a, Pol, MW or log[MW]) to equation (7) did not improve the result. It can be concluded that also in the case of the permeation of various non-electrolytes through Chara cells the H-bond *donor* strength plays a very important role. And again the permeability decreases with increasing H-bond donor strength.

2.3. Absorption and intraduodenal (id) bioavailability of azole endothelin antagonists

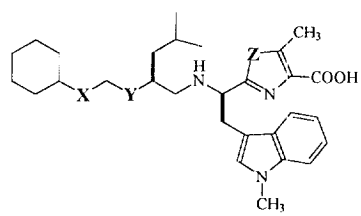
In the remarkable article of von Geldern et al. [17] it has been attempted to use $\Delta\log P$ ($\Delta\log P = \log P_{OctOH} - \log P_{cy\text{-}hexane}$) to improve the design of azole endothelin antagonists. By modification of the lead compound **33** (changing strong H-bond donor groups to H-bond acceptor or non-H-bonding groups) the bioavailability (intraduodenal absorption) in the rat as measured by the 'area under the serum level vs. time curve' in the first 60 min after intraduodenal injection (AUC_{id}) was increased by a factor of 400. In addition to this very important practical result the authors found a dependence of the bioavailability on $\Delta\log P$.

Von Geldern et al. assume their AUC_{id} results to represent mainly intestinal drug absorption. In principle AUC_{id} is a bioavailability parameter and not necessarily a descriptor of pure absorption processes. Therefore, AUC_{id} is certainly also influenced by metabolism or by distribution into lipid compartments. On the basis of the available data there is no way to separate the contribution of these effects to AUC_{id} from absorption. However, the assumption of von Geldern et al. is supported by the fact that the data have been collected only during the first hour after the intraduodenal administration.

Table III contains the data for $\log[AUC_{id}]$ (AUC in M × min/mL), $\Delta\log P$, polarizability and H-bond strength for compounds **33–46**. The correlation for nine compounds where quantitative data for AUC_{id} and $\Delta\log P$ was available is presented in the following equation:

$$\log[AUC_{id}] = -2.93(\pm 2.25) - 0.98(\pm 0.49)\Delta\log P; \quad (8)$$
$$n = 9,\ r = 0.871,\ s = 0.44,\ r_{cv} = 0.661.$$

Table III. Absorption data (AUC_{id} [μg × min/mL], and $\log[AUC_{id}]$ [M × min/mL]), $\Delta\log P_{Oct\text{-}cyHex}$, polarizability, and H-bond factor values for azole derivatives.



No.	X	Y	Z	AUC_{id}	$\log[AUC]_{id}$	$\Delta\log P$	Pol	ΣC_a	ΣC_d
33	NH	NH	NH	0.26	–9.303	5.92	60.20	19.80	–11.34
34	NCH_3	NH	NH	6.2	–7.937	5.40	62.04	18.30	–9.01
35	O	NH	NH	10.1	–7.715	> 6.2	59.00	18.17	–9.01
36	NH	NCH_3	NH	14.8	–7.559	> 6.0	62.04	18.30	–9.41
37	NH	O	NH	20.8	–7.401	> 6.1	59.00	18.17	–9.41
38	NH	CH_2	NH	4.9	–8.027	> 6.2	60.56	17.44	–9.41
39	NH	NH	O	11.6	–7.655	> 6.1	59.00	16.40	–9.28
40	NCH_3	NH	O	50.0	–7.032	4.05	60.83	16.40	–6.95
41	O	NH	O	81.5	–6.809	4.66	57.80	16.27	–6.95
42	NH	NCH_3	O	59.0	–6.96	4.59	60.83	16.40	–7.35
43	NH	O	O	48.9	–7.031	4.04	57.80	16.27	–7.35
44	NH	CH_2	O	17.5	–7.475	4.16	59.36	15.54	–7.35
45	NCH_3	O	O	110.3	–6.688	3.82	59.63	16.34	–5.02
46	O	O	O	105.8	–6.696	3.71	56.59	15.17	–5.02

It is obvious that the variation in $\log[AUC_{id}]$ is much better described and also predicted (compare r_{cv} values!) by the H-bond donor strength than by $\Delta\log P$:

$$\log[AUC_{id}] = -4.32(\pm 1.05) + 0.41(\pm 0.14)\Sigma C_d; \quad (9)$$
$$n = 9,\ r = 0.935,\ s = 0.32,\ r_{cv} = 0.818.$$

The corresponding equation for the whole set of compounds is:

$$\log[AUC_{id}] = -4.70(\pm 0.95) + 0.34(\pm 0.11)\Sigma C_d; \quad (10)$$
$$n = 14,\ r = 0.882,\ s = 0.34,\ r_{cv} = 0.787.$$

Again the H-bond donor and H-bond acceptor strengths are intercorrelated ($r = 0.845$, $n = 14$; plot $\log[AUC_{id}]$ vs. ΣC_a: $r = -0.807$, $n = 14$). The reason for the frequently observed $\Sigma C_d/\Sigma C_a$ intercorrelation is that many heteroatom groups not only contain H-bond acceptor atoms (O, N, S) but also H-bond donor groups. However, in the case of this set of compounds the importance of the H-bond donor strength as the most influencing factor for the biological effect is obvious. The change of a –NH– group in compound **39** to $-NCH_3-$ (compound **40**) does not change the C_a value but changes essentially C_d and as a result changes the activity value. A comparison of the data for compounds **39** and **45** leads to the same conclusion. Compounds **39–43** and **45** have approximately the same ΣC_a values but differ essentially in ΣC_d values. The consequence is that for this small group of compounds the activity is changed by a factor of 10.

In the discussed set of compounds there are also correlations of $\Delta\log P$ with ΣC_d, ΣC_a (ΣC_d vs. ΣC_a: $r = -0.888$) and with ΣC_{ad}:

$$\Delta\log P = 1.84(\pm 0.99) - 0.36(\pm 0.13)\Sigma C_d; \quad (11)$$
$$n = 9,\ r = 0.926,\ s = 0.30\ r_{cv} = 0.898;$$

$$\Delta\log P = -3.54(\pm 3.02) + 0.48(\pm 0.18)\Sigma C_a; \quad (12)$$
$$n = 9,\ r = 0.922,\ s = 0.31,\ r_{cv} = 0.893;$$

$$\Delta P = -0.75(\pm 1.53) + 0.22(\pm 0.06)\Sigma C_{ad}; \quad (13)$$
$$n = 9,\ r = 0.951,\ s = 0.25,\ r_{cv} = 0.924.$$

Although these equations have been derived from only nine measurements this result shows that $\Delta\log P$ seems to reflect not only the H-bond donor strength of compounds [5], but also the acceptor strength and possibly also other factors [18]. However, if the easily estimated HYBOT–H-bond factor values can explain a well defined part of the experimental parameter $\Delta\log P$ and give better correlations with biological properties the advantage of the first is obvious.

2.4. Human skin permeability coefficients (log k_p) of phenols

Data on $\log k_p$ (permeation rate k_p in cm/s) were taken from El Tayar et al. [5] and have been measured by Roberts et al. [19]. Those data together with polarizability, H-bond strength, and $\Delta\log P_{Oct-hept}$ values are listed in *table IV*. The correlation of permeability with $\Delta\log P$ is rather poor:

$$\log k_p = -4.44(\pm 0.54) - 0.38(\pm 0.21)\Delta\log P_{Oct-hept}; \quad (14)$$
$$n = 17,\ r = 0.704,\ s = 0.43,\ r_{cv} = 0.342.$$

At the same time for this data set a rather good correlation between permeability and H-bond donor strength was derived, whereas the consideration of Pol, ΣC_a, and ΣC_{ad} did not improve significantly the statistical criteria:

$$\log k_p = -3.39(\pm 0.59) + 0.71(\pm 0.21)\Sigma C_d; \quad (15)$$
$$n = 17,\ r = 0.883,\ s = 0.28,\ r_{cv} = 0.814.$$

Table IV. Human skin permeability coefficients ($\log k_p$, k_p in cm/s), polarizability, $\Delta\log P_{Oct-heptane}$, and H-bond strength of phenols.

No.	Compound	Pol	$\Delta\log P$	ΣC_a	ΣC_d	$\log k_p$	$\log k_p$ (eq. (16))
47	Phenol	10.71	2.16	1.37	–2.49	–5.64	–5.52
48	Resorcinol	10.98	4.85	2.36	–5.00	–7.18	–7.04
49	p-Nitrophenol	12.55	3.91	2.43	–3.65	–5.81	–5.88
50	m-Nitrophenol	12.55	3.40	3.26	–3.49	–5.81	–5.78
51	p-MeOOC-phenol	14.81	2.09	3.54	–2.73	–5.60	–5.29
52	m-Cresol	12.54	2.31	1.47	–2.44	–5.37	–5.34
53	o-Cresol	12.54	1.70	1.61	–2.27	–5.36	–5.28
54	p-Cresol	12.54	2.29	1.31	–2.44	–5.31	–5.34
55	2-Naphthol	17.98	1.07	1.99	–2.64	–5.11	–5.17
56	o-Chlorophenol	12.64	0.10	1.92	–2.18	–5.04	–4.98
57	p-Ethylphenol	14.38	2.14	1.32	–2.44	–5.01	–5.21
58	3,4-Xylenol	14.38	1.95	1.52	–2.28	–5.00	–5.10
59	p-Bromophenol	13.33	2.68	1.55	–2.83	–5.00	–5.10
60	p-Chlorophenol	12.64	2.50	1.61	–2.91	–5.00	–5.47
61	Thymol	18.05	1.68	1.60	–2.01	–4.83	–4.70
62	2,4,6-Trichlorophenol	16.49	2.79	2.61	–2.21	–4.78	–4.54
63	2,4-Dichlorophenol	14.56	1.69	2.27	–2.48	–4.78	–4.93

In contrast to Pol, however, log[MW] (intercorrelation $r = 0.640$) in this case significantly improved the description of $\log k_p$:

$$\log k_p = -8.72(\pm 2.79) + 0.67(\pm 0.15)\Sigma C_d + 2.47(\pm 1.28)\log[MW]; \quad (16)$$
$$n = 17,\ r = 0.949,\ s = 0.20,\ r_{cv} = 0.915.$$

For this set of phenols a significant correlation is found between $\Delta\log P$ and ΣC_d whereas the statistical criteria of correlations with ΣC_a, ΣC_{ad} or with ΣC_a and ΣC_d are much poorer:

$$\Delta\log P = -1.02(\pm 1.28) - 1.22(\pm 0.45)\Sigma C_d; \quad (17)$$
$$n = 17,\ r = 0.828,\ s = 0.62,\ r_{cv} = 0.776.$$

2.5. Human skin permeability coefficients (log k_p) of alcohols and steroids

Data on $\log k_p$ were also taken from El Tayar et al. [5] and have been measured by Scheuplein et al. [20, 21]. Like in the preceding section the data reflect the permeation of compounds through the human skin. However, as the phenol data and the alcohol/steroid data have been observed in different laboratories, we (also) prefer to consider these data sets separately. The permeability data of alcohols and steroids together with polarizability, $\Delta\log P_{Oct-hept}$, and H-bond strength values are presented in *table V*.

The correlation of permeability with $\Delta\log P_{Oct-hept}$ is rather poor:

$$\log k_p = -3.76(\pm 1.16) - 1.25(\pm 0.46)\Delta\log P_{Oct-hept}; \quad (18)$$
$$n = 22,\ r = 0.789,\ s = 0.73,\ r_{cv} = 0.738.$$

Interestingly for this data set a significant regression equation was obtained on the basis of both H-bond parameters. The low correlation between ΣC_a and ΣC_d ($r = 0.527$) permits to use the two parameters simultaneously:

$$\log k_p = -4.88(\pm 0.48) - 0.31(\pm 0.10)\Sigma C_a + 0.23(\pm 0.17)\Sigma C_d; \quad (19)$$
$$n = 22,\ r = 0.912,\ s = 0.50,\ r_{cv} = 0.889.$$

It is interesting to note that for this group of compounds there is no significant correlation between $\Delta\log P$ and ΣC_d ($r = -0.583$). However, there is a moderate correlation between $\Delta\log P$ and ΣC_a (which is not improved by the addition of ΣC_d) and a slightly better correlation between $\Delta\log P$ and ΣC_{ad}:

$$\Delta\log P = 1.42(\pm 0.38) + 0.22(\pm 0.07)\Sigma C_a; \quad (20)$$
$$n = 22,\ r = 0.827,\ s = 0.42,\ r_{cv} = 0.782.$$

$$\Delta\log P = 1.32(\pm 0.39) + 0.16(\pm 0.05)\Sigma C_{ad}; \quad (21)$$
$$n = 22,\ r = 0.834,\ s = 0.41,\ r_{cv} = 0.789.$$

Table V. Human skin permeability ($\log k_p$, k_p in cm/s), polarizability, $\Delta\log P_{Oct-Heptane}$, H-bond ability, length L, and width W (in Å) of alcohols and steroids.

No.	Compound	Pol	$\Delta\log P$	ΣC_a	ΣC_d	L	W	$\log k_p$ (obs.)	$\log k_p$ (eq. (24))	$\log k_p$ (eq. (25))
6	Methanol	3.25	2.03	1.76	−1.50	2.91	1.809	−6.56	−6.20	−6.82
3	Ethanol	5.08	1.79	1.82	−1.42	4.18	1.806	−6.56	−6.14	−6.56
7	1-Propanol	6.92	1.77	1.60	−1.27	5.46	1.806	−6.41	−5.93	−6.15
64	1-Butanol	8.75	1.58	1.83	−1.32	6.72	1.806	−6.16	−5.98	−5.99
65	1-Pentanol	10.5	1.96	1.79	−1.31	8.00	1.806	−5.78	−5.89	−5.70
5	1-Hexanol	12.4	1.58	1.79	−1.59	9.26	1.806	−5.44	−5.88	−5.50
66	1-Heptanol	14.2	1.74	1.79	−1.37	10.5	1.806	−5.05	−5.76	−5.17
67	1-Octanol	16.0	1.35	1.79	−1.35	11.7	1.806	−4.84	−5.69	−4.90
68	Progesterone	36.3	2.47	4.57	0.00	12.2	6.079	−6.38	−5.91	−6.16
69	Pregnenolone	36.8	2.51	3.89	−1.48	13.0	6.103	−6.38	−5.92	−6.03
70	HO-Progesterone	36.9	2.34	5.84	−2.13	13.0	6.068	−6.78	−6.98	−7.01
71	HO-Pregnenolone	37.4	2.80	5.20	−3.61	13.8	6.096	−6.78	−7.01	−6.92
72	Deoxycorticosterone	36.9	2.40	6.14	−1.66	12.9	6.064	−6.90	−7.01	−7.06
73	Testosterone	33.1	2.90	4.68	−2.13	10.8	6.064	−6.95	−6.58	−6.96
74	Cortexolone	37.1	3.52	8.32	−1.66	14.1	6.062	−7.68	−8.03	−7.71
75	Corticosterone	37.6	3.56	7.67	−3.04	12.8	6.089	−7.78	−8.03	−8.04
76	Cortisone	37.7	1.97	7.89	−3.79	13.4	6.072	−8.56	−8.31	−8.19
77	Hydrocortisone	38.2	3.57	8.98	−5.19	12.9	6.098	−9.08	−9.13	−9.09
78	Aldosterone	39.3	3.78	9.28	−3.79	12.7	6.425	−9.08	−8.90	−8.97
79	Estrone	30.5	2.28	3.61	−2.40	11.7	6.071	−6.00	−6.24	−6.41
80	Estradiol	31.0	2.89	3.49	−4.53	11.7	6.052	−7.08	−6.66	−6.85
81	Estriol	31.6	3.11	5.34	−6.66	12.4	6.066	−7.95	−8.01	−7.97

Also for the subgroup of the steroids in *table V* the statistical criteria of the two parameter equation for $\log k_p$ are very good:

$$\log k_p = -4.36(\pm 0.61) - 0.38(\pm 0.09)\Sigma C_a + 0.24(\pm 0.11)\Sigma C_d; \quad (22)$$
$$n = 14,\ r = 0.961,\ s = 0.30,\ r_{cv} = 0.939.$$

Because of systematic deviations between observed $\log k_p$ values and those calculated for the alcohols by equation (19), the steroids and alcohols were investigated separately in more details. For the alcohols the best correlation was obtained with their polarizability:

$$\log k_p = -7.29(\pm 0.33) + 0.15(\pm 0.031)\text{Pol}; \quad (23)$$
$$n = 8,\ r = 0.979,\ s = 0.15,\ r_{cv} = 0.954.$$

The virtually constant ΣC_a and ΣC_d values of the alcohols were non-significant for the description of $\log k_p$.

In contrast to the alcohols in the subgroup of the steroids the polarizability was not significant if added to equation (22). This is not unexpected as for these compounds Pol is almost constant. By combining steroids and alcohols and using Pol as an additional descriptor the following highly significant correlation was obtained (in contrast to [5] without exclusion of cortisone **76**):

$$\log k_p = -5.14(\pm 0.45) - 0.47(\pm 0.14)\Sigma C_a + 0.23(\pm 0.14)\Sigma C_d + 0.038(\pm 0.027)\text{Pol}; \quad (24)$$
$$n = 22,\ r = 0.941,\ s = 0.42,\ r_{cv} = 0.914.$$

Interestingly in this equation the regression coefficient of Pol deviates significantly from that of the preceding correlation, whereas the other slopes are similar to those of equation (22). Furthermore, there is still a systematic deviation in the calculated $\log k_p$ values of the alcohols (see *table V*). Possibly the bulk descriptor Pol is here (in equation (23)/(24)) mainly a descriptor of the bulk in one dimension (i.e. molecular length). Obviously the almost constant molecular shape of the steroids differs from the shape of the n-alkanols. If the indicator variable $I_{steroid}$ is added to indicate the deviating molecular width of the steroids (i.e. $I_{steroid} = 1$ for steroids, $I_{steroid} = 0$ for alcohols) a correlation is obtained ($r = 0.968$, $s = 0.32$, $r_{cv} = 0.944$) that shows no systematic deviation for the calculated $\log k_p$ values of the alcohols. Unfortunately, there are significant cross-correlations between $I_{steroid}$, Pol, and ΣC_a (maximum: $r = 0.966$, $I_{steroid}$ vs. Pol).

To investigate the assumption that the molecular length L and width W (in Å) is important for this data set we determined both by the ALCHEMY 2000 program package (see Methods, *table V*). Interestingly a highly significant equation is obtained if Pol and $I_{steroid}$ are replaced by L and W:

$$\log k_p = -6.14(\pm 0.39) - 0.42(\pm 0.07)\Sigma C_a + 0.23(\pm 0.08)\Sigma C_d + 0.21(\pm 0.06)L - 0.11(\pm 0.10)W; \quad (25)$$
$$n = 22,\ r = 0.984,\ s = 0.23,\ r_{cv} = 0.974.$$

The highest descriptor intercorrelation for this equation (W vs. ΣC_a: $r^2 = 0.656$) is at an acceptable level. The observation that the permeability increases with length L and decreases with width W seems to be reasonable. As the highly structured phospholipid bilayers in cell membranes consist mainly of parallel, long alkyl chains attached to polar headgroups it may be expected that molecules with similar properties (i.e. long and narrow shape) will pass these domains with more ease than sterically demanding solutes [22, 23].

3. Discussion

Detailed knowledge of the influence of molecular properties on solute permeabilities is helpful in understanding the mechanisms by which various solutes permeate through the skin [15, 23] or through lipophilic barriers separating aqueous phases (e.g. blood brain barrier, red cell membrane, cell walls of microorganisms, etc.). For a long time molecular descriptors like $\log P$, molecular volume or molecular weight, etc., have been used in structure/property–permeability relationship analyses [1, 14–16, 24, 25]. Later also $\Delta\log P_{OctOH-alkane}$ values have been considered in several papers [3–5, 17] to describe mainly H-bond donating effects of permeants. Recently the influence of hydrogen bonds on permeability processes has been investigated in more details by differentiating between H-bond acidity and H-bond basicity [6–8, 23]. In this paper we analyzed the possibility to describe H-bond effects on solute permeation using HYBOT acceptor and donor factor values (ΣC_a, ΣC_d) [9–13, 26] that are easily accessible for a large number of compound classes.

Five literature data sets providing data on the permeation or absorption behavior of different classes of compounds in different test systems have been analyzed. In no case the permeability of membranes was found to be *improved* by solute hydrogen bonds. On the contrary the permeation through lipophilic phases was in all cases negatively dependent on the strength of the overall hydrogen bonding effect, irrespective of the H-bond donor or acceptor effect of permeants. Qualitatively, this may be somewhat obvious. However, though results of five data sets were available we could not obtain a consistent model of the real situation. Only in one data set (alcohols and steroids, *table V*) correlations (equation (24)/(25)) were found that contained *both* ΣC_a and ΣC_d as significant descriptors among others. On the other hand, the other four data sets could be best described by using only ΣC_d values. Nevertheless, two of these data sets (*tables I* and *III*) also showed slightly less significant relationships with ΣC_a that indicate reduced permeability with increasing ΣC_a values. Two explanations are conceivable for this result: First, there is simply a chance (?) intercorrelation between ΣC_a and ΣC_d leading to significant relationships with one or the other; or second, both H-bond donor and acceptor effect are important physicochemical properties for permeation processes. However,

because of the intercorrelation, only the more significant of them leads to an unbiased, significant regression coefficient in the dual parameter equation. To enable the recognition of the correct explanation we performed principal component analyses with subsequent VARIMAX rotation of the components [27, 28]. By this procedure orthogonal principal components (PCs) were obtained from the intercorrelated ΣC_a and ΣC_d variables. According to the loadings of the original descriptors on the PCs the latter represent mainly one or the other of the former descriptors. For instance *table VI* shows that the first principal component (PC1) derived from the ΣC_a and ΣC_d values of *table I* contains mainly the information of ΣC_a, whereas PC2 corresponds to ΣC_d (loadings are identical with correlation coefficients between original variables and PCs!).

Table VI. Loadings of ΣC_a and ΣC_d (*table I*) on Varimax-rotated principal components.

	PC1	PC2
ΣC_a	*0.829*	–0.560
ΣC_d	–0.560	*0.829*

Using the largely separated H-bond acceptor and donor effects as expressed by the principal component scores the basal permeability data of *table I* could be described by the following highly significant equation:

$$\log[BP] = -4.39(\pm 0.31) - 1.38(\pm 0.32)PC1(\cong\Sigma C_a) + 1.71(\pm 0.32)PC2(\cong\Sigma C_d); \quad n = 10,\ r = 0.987,\ s = 0.41. \tag{26}$$

Similar results were also obtained for the algal cell permeability data and for the bioavailability of the azole endothelin antagonists. The only exception was the phenol skin permeability data set of *table IV* that showed a significant correlation only with the PC corresponding to ΣC_d. Possibly the reason for this observation is the rather small range of ΣC_a values (minimum 1.32, maximum 3.54), whereas the range of ΣC_d values (min. – 5.0, max. – 2.01) is more extended.

The overall result of this analysis is our conviction that – in addition to steric bulk descriptors – both H-bond donor *and* acceptor effect are important physicochemical properties for permeation processes (see also [13]). However, because of their frequently observed intercorrelation, only the more significant of them leads to an unbiased, significant regression coefficient in multiple linear regression equations.

4. Conclusion

Instead of the experimental parameters log*P* or Δlog*P* the H-bond donor and acceptor descriptors (ΣC_d and ΣC_a) are proposed as parameters for modeling the relationships between the physicochemical properties of chemicals or drugs and their permeability and absorption behavior. It is not necessary to synthesize a compound for an a priori estimation of ΣC_d and ΣC_a values as these can be easily calculated by the program HYBOT. The often used experimental parameter Δlog*P* should be avoided as it includes not only the H-bond donor property of a solute but also other possible interactions.

5. Methods

5.1. Determination of descriptor values and statistical data analysis

All information about the H-bonding strength of chemical compounds and drugs was obtained by means of the commercially available program package HYBOT-96 (HYdrogen BOnd Thermodynamics) [11, 12, 26]. The program package contains databases for the free energy, enthalpy and H-bond descriptor values for 414 H-bond donors and 1298 H-bond acceptors that have been calculated on the basis of the multiplicative approach [9]. By knowing C_d values for 414 H-bond donors and C_a values for 1298 H-bond acceptors, it is possible to estimate free energy values for 537 372 H-bond complexes (in [8] H-bond acidity data for 150 compounds and basicity data for 500 H-bond acceptors are given, providing the possibility to estimate free energy values for only 75 000 H-bond complexes). The prediction procedure for H-bond factors in HYBOT is based on the search of the nearest neighbor in the data base of factor values and is carried out by fragmentation of molecules with subsequent analysis of the structural environment of H-bonding atoms according to [29].

The polarizability values for the compounds were calculated on the basis of an additive scheme [30] by our special program. The molecular length *L* and width *W* (in Å) for the alkanol/steroid data set (*table V*) was determined by the ALCHEMY 2000 package (Tripos Assoc. Inc., St. Louis, MO). The flexible alkanols were constructed in the fully extended conformation. After molecular mechanics geometry optimization the length *L* was calculated as the distance between the OH hydrogen atom and the most distant H atom of the terminal methyl group, whereas the width *W* corresponds to the distance between two methylene group H atoms. In the geometry optimized steroids the length *L* is the distance between the carbonyl oxygen atom at the steroid 3-position (or corresponding 3-OH hydrogen) and the most distant atom at the opposite side (an atom within a substituent in 17- or 16-position). The steroid width W corresponds to the distance between the equatorial H atoms in positions 7 and 12 (however, for corticosterone, hydrocortisone: 7-H_{eq}/11-OH hydrogens; aldosterone: 7-H_{eq}/OH hydrogen atom of the 13-CH–OH group).

Multiple linear regression and principal component analyses were performed by means of in-house programs (KJS).

Acknowledgements

O.A.R. is very grateful to the Forschungszentrum Borstel for a two-month stipendium.

References

[1] Hansch C., Leo A., Exploring QSAR, Vol. 1: Fundamentals and Applications in Chemistry and Biology, ACS, Washington, 1995.

[2] Seiler P., Eur. J. Med.Chem. 9 (1974) 473–479.

[3] Young R.C., Mitchell R.C., Brown T.H., Ganellin C.R., Griffiths R., Jones M., Rana K.K., Saunders D., Smith I.R., Sore N.E., Wilks T.J., J. Med. Chem. 31 (1988) 656–671.

[4] Scott R.C., Leahy D.E., Evans J.R., Bole E., J. Pharm. Pharmacol. 38 (1986) 66P.

[5] El Tayar N., Tsai R.S., Testa B., Carrupt P.A., Hansch C., Leo A., J. Pharm. Sci. 80 (1991) 744–749.

[6] Abraham M.H., Chadha H.S., Mitchell R.C., J. Pharm. Sci. 83 (1994) 1257–1268.

[7] Abraham M.H., Chadha H.S., In: Pliska V., Testa B., van de Waterbeemd H. (Eds.), Lipophilicity in Drug Action and Toxicology, VCH, Weinheim, 1996, pp. 311–337.

[8] Abraham M.H., In: Politzer P., Murray J.S. (Eds.), Quantitative Treatments of Solute/Solvent Interactions. Theoretical and Computational Chemistry, Vol. 1, Elsevier, Amsterdam, 1994, pp. 83–134.

[9] Raevsky O.A., Grigor'ev V.J., Kireev D.B., Zefirov N.S., Quant. Struct. Act. Relat. 11 (1992) 49–63.

[10] Raevsky O.A., Schaper K.-J., Seydel J.K., Quant. Struct. Act. Relat. 14 (1995) 433–436.

[11] Raevsky O.A., In: van de Waterbeemd H., Testa B., Folkers G. (Eds.), Computer-Assisted Lead Finding and Optimization; Current Tools for Medicinal Chemistry, Wiley-VCH, Weinheim, 1997, pp. 367–378.

[12] Raevsky O.A., J. Phys. Org. Chem. 10 (1997) 405–413.

[13] van de Waterbeemd H., Camenish G., Folkers G., Raevsky O., Quant. Struct. Act. Relat. 15 (1996) 480–490.

[14] Stein W.D., Lieb W.R., Transport and Diffusion Across Cell Membranes, Academic Press, Orlando, 1986, pp. 69–112.

[15] Anderson B.D., Raykar P.V., J. Invest. Dermatol. 93 (1989) 280–286.

[16] Lien E.J., Lien L.L., Gao H., In: Sanz F., Giraldo J., Manaut F. (Eds.), QSAR and Molecular Modelling: Concepts, Computational Tools and Biological Applications, Prous Science, Barcelona, 1995, pp. 94–100.

[17] von Geldern T.W., Hoffman D.J., Kester J.A., Nellans H.N., Dayton B.D., Calzadilla S.V., Marsh K.C., Hernandez L., Chiou W., Dixon D.B., Wu-Wong J.R., Opgenoth T.J., J. Med. Chem. 39 (1996) 982–991.

[18] Abraham M.H., Chadha H.S., Whiting G.S., Mitchell R.C., J. Pharm. Sci. 83 (1994) 1085–1100.

[19] Roberts M.S., Anderson R.A., Swarbrick J., J. Pharm. Pharmacol. 29 (1977) 677–683.

[20] Scheuplein R.J., J. Invest. Dermatol. 45 (1964) 334–346.

[21] Scheuplein R.J., Blank I.H., Brauner G.J., MacFarlane D.J., J. Invest. Dermatol. 52 (1969) 63–70.

[22] Xiang T.X., Anderson B.D., J. Membrane Biol. 140 (1994) 111–122.

[23] Potts R.O., Guy R.H., Pharm. Res. 12 (1995) 1628–1634.

[24] Ho N.F.H., Higuchi W.I., J. Pharm. Sci. 60 (1972) 537–541.

[25] Schaper K.-J., Quant. Struct. Act. Relat. 1 (1982) 13–27.

[26] Commercially available software HYBOT (Raevsky O., Grigor'ev V., Institute of Physiologically Active Compounds of the Russian Academy of Sciences, 142432, Chernogolovka, Moscow region, Russia, Fax + 7-095-302-9159, E-mail raevsky@ipac3.sherna.msksu, Internet E-mail raevsky@ipac.ac.ru).

[27] Schaper K.-J., Kaliszan R., In: Mutschler E., Winterfeldt E. (Eds.), Trends in Medicinal Chemistry, VCH, Weinheim, 1987, pp. 125–139.

[28] Cooley W.W., Lohnes P.R., Multivariate Data Analysis, Wiley, New York, 1971.

[29] Trepalin S., Yarkov A., Dolmatova L., Zefirov N.S., J. Chem. Inform. Comput. Sci. 35 (1995) 405–411.

[30] Miller K.J., J. Am. Chem. Soc. 112 (1990) 8533–8538.