Physicochemical Descriptors in Property-Based Drug Design

Oleg A. Raevsky*

Department of Computer-Aided Molecular Design, Institute of Physiologically Active Compounds of Russian Academy of Sciences, 142432, Chernogolovka, Moscow region, Russia

Abstract: The contribution of physicochemical descriptors to lipophilicity, water solubility, and intestinal absorption and oral bioavailability in humans is considered.

Partitioning in the octanol/water system is presented as a competition between two opposing effects: volume and hydrogen bond acceptor ability. Water solubilities of liquid compounds are roughly equal to their reciprocal logP values. However, there is also a detectable contribution of H-bond donor ability to water solubility. The main problem in predicting the solubilities of solid chemicals and drugs is the estimation of their crystal lattice energies. QSAR approaches that add terms such as melting point, and the product of H-bond donor and acceptor parameters are not sufficient to make these predictions practical.

Human intestinal absorption for passively transported drugs is almost completely correlated with hydration processes that are determined by H-bond acceptor and donor abilities.

It is emphasized that structural features of drug molecules have significant influences on their properties. Classic QSAR approaches are not enough to create stable, predictive models for diverse drugs. A combination of Similarity and QSAR approaches is one possibility to take all physicochemical properties in addition to structural features into account.

Keywords: ADMET, Descriptors, Lipophilicity, Solubility, Absorption, QSAR, H-bond, Similarity.

INTRODUCTION

Although physicochemical properties as they relate to pharmacokinetic and biopharmaceutical properties were almost completely neglected in earlier efforts at drugs design, it is now obvious that they are important for the discovery of prospective new drug candidates. As a result a new tool, *Property-Based Design* in Medicinal Chemistry, is currently being intensively developed [1].

The concepts of quantitative structure-property and structure-activity relationships (QSPR and QSAR), are the main platforms for optimizing a drug candidate's properties. They include two main elements: the description of structure by means of independent parameters (descriptors) and chemometric tools. The latter includes many different regression/correlation approaches, methods of pattern recognition, classification, and experimental design [2].

This short review deals with physicochemical descriptors that contribute to transport and distribution properties of drugs in the human organism, and discusses possibilities to create stable, predictive models based on them.

LIPOPHILICITY

Lipophilicity is the affinity of drug molecules for a lipophilic environment, and is often considered as a key property in the transport processes of drugs in human beings. These include intestinal absorption, membrane permeability, protein binding, and distribution among different tissues [3]. It is usually defined as the partition coefficient (P) of a compound distributed between octanol and water phases, and is commonly expressed as logP, its logarithmic form.

Many approaches to calculate lipophilicity have been proposed since the 1960s, beginning with the π-system developed by Hansch and Fujita [4], and the initial fragment approach of Rekker [5]. Since then, modern approaches based on neural network interpretations of structural and quantum chemical descriptors have come to the fore [6-9]. Current commercially available programs to predict lipophilicity are described here in [10].

There are a lot of different calculation procedures to estimate octanol-water partition coefficients that have been used with varying degrees of success and applicability [10,11]. For the most part, these procedures are based on fragment or atom contributions and are outside the scope of this review. As to QSAR models that are based on physicochemical descriptors it is necessary to mention [11] the application of molecular weight, molecular volume, solvatochromic parameters, molecular surface area, solvent-accessible surface area, atomic surface tensions, molecular surface properties, molecular polarizability, different molecular orbital descriptors (charge density, electrostatic potentials, highest occupied and lowest unoccupied molecular orbital energies) and other properties (ionization potentials, positive and negative electrostatic potentials, and dipole moments) [12-20].

Currently, thermodynamic approaches for describing molecular properties (including lipophilicity) are being developed by three groups: M.Abraham [21], P.Ruelle [22] and O.Raevsky [23]. All of these methods are directly connected by their application of thermodynamic properties

*Address correspondence to this author at the Department of Computer-Aided Molecular Design, Institute of Physiologically Active Compounds of Russian Academy of Sciences, 142432, Chernogolovka, Moscow region, Russia; Tel/Fax: 007-095-785-70-24; E-mail: raevsky@ipac.ac.ru

1389-5575/04 $45.00+.00

to various intermolecular interactions. The differences in the approaches relate to the different number of descriptors used in the training sets, and to the type of QSAR models.

Five molecular descriptors are used in the solvation equation of Kamlet, Taft and Abraham: (1) E – excess molar refraction, which models dispersion force interactions arising from the polarizability of pi- and n-electrons; (2) S- solute polarity/polarizability (due to solute-solvent interactions between bond dipoles and induced dipoles); (3) A – solute H-bond acidity; (4) B – solute H-bond basicity; and (5) V – McGowan characteristic molar volume [24]. The H-bonding parameters are summation terms relevant to the behavior of solutes in solvents. The acidity "A" relates to the strength and number of H-bonds formed by donor groups in the solute when they interact with lone pairs of acceptor groups in solvent molecules. The basicity "B" relates to the strength and number of H-bonds formed by the lone pairs of acceptor groups in the solute when they interact with donor solvents. These descriptors are based on the H-bond scales that refer to the overall or summation H-bond acidity and basicity effects.

Distribution in the octanol-water system was characterized in framework of this approach by equation (1) [25]:

$$\log P_{ow} = 0.088 + 0.562\,E - 1.054\,S - 0.032\,A - 3.460\,B + 3.814\,V \quad (1)$$

n=613, R=0.9974, sd=0.116, F=23161.6

where n is number of compounds, R is the correlation coefficient, sd is the standard deviation, and F is the Fisher criterion.

Some words about statistical criteria that are too good: the standard deviation for this training set (which was created on the basis of different measurements by various authors) is less than even the experimental error of determination (± 0.5 logP unit [26]). One reason such a situation (as indicated in [1]) could arise is the mutual intercorrelation of descriptors "leading to over-optimistic statistics". Another reason for this good result may be the lack of diversity in the training set. The application of the solvation approach to the data extracted from the MedChem97 database gave much more modest result: n=8844, r=0.909, RMS (root-mean-square error) = 0.674, F = 8416 [27].

In accordance with Mobil Order and Disorder Theory five components at the most contribute to the Gibbs free energy of partitioning of a solute in a biphasic system of two essentially immiscible solvents [22]:

$$\log P = \Delta B + \Delta D + \Delta F + \Delta O + \Delta OH \quad (2)$$

where the entropy of mixing term, ΔB, gives information about differences between the two phases in the entropy of the solute/solvent exchange; the hydrophobic effect- term, ΔF, accounts for differences in the propensities of the solvent phases to squeeze the solute out of the solution; the two H-bond interaction-related terms, ΔO and ΔOH express differences in the strengths of the H-bonds that bind the solute and solvent molecules in each phase; the term ΔD is similar to the two previous ones, but accounts for non-specific forces only. The application of eq. (2) to a set of 482 compounds resulted in an adequate correlation between experimental and calculated values; the standard deviation of the computed logP values was at the level of 0.50.

The thermodynamic approach followed by Raevsky's group considers the property P to be based on contributions from three main intermolecular interactions: steric, electrostatic and hydrogen bonding [28,29]:

$$P = f(\alpha, \sum q, \sum C) \quad (3)$$

where α is molecular polarizability (a volume-related term), ∑q is a sum of partial atomic charges (an electrostatics-related term) in a molecule, ∑C comprises free energy H-bonding factors [30]. All these descriptors are calculated by program package HYBOT (HYdrogen BOnd Thermodynamics) [31]. Besides polarizability and partial atomic charges, the current version of this program has the possibility to calculate enthalpy and free energy as well as overall H-bond factors by searching for nearest neighbors in a data bases containing 250 000 structural fragments. Recently, original hydrogen bond potentials based on hydrogen bond factors were proposed [32] as new 3-D H-bond descriptors [33].

Physicochemical models of lipophilicity based on volume-related terms and polarities or hydrogen bonding capacities of solutes were first described in the 1980s [34-36]. In 1995, an equation was published [37] that used molecular volume (MV) and hydrogen bonding to describe the octanol/water partition coefficients for 38 neutral carbonyl and hydroxyl compounds. For the hydrogen bonding part, the authors used free energy H-bond acceptor (∑Ca) and donor factors (∑Cd). Later they changed from using molecular volume to using molecular polarizability (α), and obtained a good correlation for logP on about three thousand simple compounds [38,39]:

$$\log P = 0.267\,\alpha - 1.00 \sum Ca \quad (4)$$

n=2850, r=0.970, s=0.23

Eq. (4) was used for logP calculations in the program SLIPPER-98 [40].

There are two important features of eq. (4) to note:

- In spite of being an essential part of the training set, H-bond donor factors (∑Cd) made no contribution in estimating logP.
- The regression equation has a zero intercept. Hence, in the case of a compound where the magnitudes of the terms are equal, the compound will be distributed equally between both phases. Because, in this case, the terms have opposite signs they cancel each other, and thus logP equals zero (0), and therefore P equals one (1). Polarizability can have only positive values; hence, it can only make a positive contribution to logP. Thus, increasing polarizability leads to higher concentrations of the solute in the octanol phase. H-bond acceptor factors also can only be positive, but because a negative sign proceeds its term in the equation, H-bond acceptors can only increase the solute's distribution in water phase.

Although the application of eq. (4) in the program SLIPPER-98 demonstrated adequate predictive power for logP values, the predictions of logP for some drug

molecules containing several functional groups deviated significantly from those observed. Those deviations were approximately the same for related chemical structures. So it was reasonable to suppose that structural features of molecules also influence a drug's distribution in octanol-water system.

There are many approaches for selecting related chemical structures. These include a chemist's intuition to select compounds with similar structural frameworks, and special QSAR procedures such as SIMCA/PLS [41], DIREM [42], and different types of cluster analysis [2,43]. The present state of Similarity concepts permit one to formalize quantitative measures of relatedness among chemical structures [44].

An original combination of Similarity and QSAR for creating stable, predictive models of lipophilicity was recently proposed [45]. In this work four approaches were considered:

1. The application of eq. (4).
2. The logP value of the nearest neighbor in a large training set was used as the calculated value for the compound-of-interest [46].
3. The mean logP value of three nearest neighbors was used as the calculated value for compound-of-interest [47].
4. LogP values of the nearest neighbors were used only in the first step. In addition, the contribution to lipophilicity arising from differences in polarizabilities and H-bond acceptor factors between the compound-of-interest and its nearest neighbors were also taken into consideration. In this case the eq. (5) was used employing coefficient values from eq. (4):

$$\log P_i = \sum_{j=1}^{N} [((\log P_j + 0.267(\alpha_i - \alpha_j) - 1.00\ (Ca_i - Ca_j)]/N \quad (5)$$

where index i indicates the compound-of-interest, index j indicates a near neighbor; and N is the number of closely related structures used.

For 48 drugs, the statistical criteria for the correlation between experimental and calculated logP values on the basis of eq.(5) are better than those obtained on the same drugs using the methods tested by Mannhold and Dross [48]:

$$\log P_{exp} = 1.012\ (\pm 0.019)\ \log P_{calc(eq.(5))} \quad (6)$$

N=48, r=0.964, sd=0.419, F=597

Later eq (5) was used to calculate logP values for a database containing 10 937 compounds and drugs in the program SLIPPER-2001 [49]. The correlation coefficient found between experimental and calculated logP was 0.972. The results of such calculations for 24 drugs studied in [49] are presented in Table **1**.

Some important points to remember:

- logP refers to the neutral state of molecule. In the presence of acidic and/or basic groups the distribution becomes pH dependent. The pH dependent distribution coefficient logD is related to logP through the ionization constant (pKa) [50,51]. Examples of procedures for calculating logD on the basis of physicochemical descriptors are presented in [52,53].
- logP and logD are widely used in QSAR to study permeability, absorption and drug distribution in organisms. However, it was recently found that hydrogen-bonding ability is the main contributor to passive transport [1]. (For a more detailed discussion look in the Chapter devoted to intestinal absorption). Besides, as demonstrated above, logP is a composite descriptor that includes, in hidden form, volume-related and H-bonding terms. The octanol-water system should be considered only as one model to test different descriptors and methodological approaches. Other systems (in particularly immobilized artificial membranes [54-56]) are now proposed as alternatives to the octanol-water system.

WATER SOLUBILITY

Aqueous solubility is an important property that influences a drug's release, permeability through different biologic membranes, transport and absorption in humans. Thus, there is an increasing need develop methods to estimate water solubility from molecular structure and calculable physicochemical descriptors that relate to ADMET (absorption, distribution, metabolism, excretion and toxicity).

In 1968, on the basis of aqueous solubility data for 150 liquid compounds, Hansch already logically proposed that solubility and the partition coefficient are reciprocally related [57]. Yalkovsky *et al*. developed the idea further, and proposed the following quantitative relationship between solubility and partition coefficient values for liquid compounds [58]:

$$\log S = -1.07 \log P + 0.67 \quad (7)$$

r=0.954, sd = 0.344

A few years ago, investigations [37-40] of logP for small sets of chemicals showed that the octanol-water partition coefficient is sufficiently described by ΣCa, a quantitative descriptor of hydrogen bonding (HB) acceptor strength (sum over all HB accepting substructures within a molecule), and by a bulk effect descriptor such as molar volume, MV, or polarizability (α). Interestingly, the descriptor of HB donor strength, ΣCd, was shown to be not significant. Clearly, if logP can be quantitatively described by ΣCa and α, then the same can be expected for solubility logS. This was recently confirmed by a correlation using a training set of 630 liquid compounds [59]:

$$\log S = 0.578(\pm 0.133) - 0.305(\pm 0.010)\alpha + 1.155(\pm 0.048)\ \Sigma Ca \quad (8)$$

n = 630 r = 0.936 s = 0.586 q = 0.935

However already with 45 neutral polar liquid compounds, Raevsky showed [38] that besides polarizability, H-bond acceptor factors, and hydrogen bond donor factors ($|\Sigma Cd|$) also influence water solubility. Later, the training set was increased to 142 liquids [39]. Because

Table 1. Calculations Results of Lipophilicity, Solubility and Human Intestinal Absorption Prediction for 24 Drugs by means of Program Package SLIIPER-2001[49]

NN	Name	α	ΣC_a	ΣC_d	logPcalc			logPexp	$logS_{calc}$			$logS_{exp}$	Fa_{calc}			Fa_{exp}
					1nn	2nn	3nn		1nn	2nn	3nn		1nn	2nn	3nn	
1	acetaminophen	16.0	4.0	-4.5	0.73	0.44	0.33	0.51	-0.82	–0.89	–0.86	–1.03	0.92	0.93	0.83	0.80
2	caffeine	19.4	5.4	0.0	-0.26	-0.10	0.04	-0.07	-0.42	-0.38	-0.32	-0.95	0.99	0.88	0.92	0.99
3	corticosterone	37.6	6.8	-3.3	1.48	1.46	2.05	1.94	-1.66	-3.35	-3.27	-3.24	0.99	1.00	0.99	0.99
4	dapsone	26.9	5.8	-6.5	1.21	1.34	1.39	0.97	-3.57	-3.53	-3.30	-2.80	0.99	0.97	0.95	0.93
5	digitoxin	77.1	16.5	-7.7	2.64	2.67	2.18	1.74	-5.67	-5.05	-4.76	-5.29	0.96	0.89	0.88	0.90
6	flurbiprofen	25.9	2.9	-2.8	3.58	3.68	3.75	4.16	-3.21	-3.45	-3.44	-4.50	0.99	0.99	0.97	0.92
7	imipramine	35.0	3.9	0.0	4.74	4.60	4.74	4.80	-4.76	-4.30	-4.57	-4.19	1.00	0.94	0.92	1.00
8	lidocaine	28.1	4.8	-2.5	1.64	2.38	2.34	2.26	-1.24	-0.97	-1.63	-1.75	0.52	0.44	0.43	0.35
9	mannitol	15.6	8.5	-9.4	-3.59	-3.88	-4.05	–3.10	-1.06	-2.01	-0.68	-0.03	0.02	0.03	0.22	0.26
10	metronidazole	15.7	4.3	-1.6	-1.17	-0.53	-0.49	-0.02	-2.01	-0.93	-0.41	-1.22	0.98	0.99	0.98	0.99
11	morphine	29.9	6.1	-2.9	0.58	0.44	0.67	0.76	-4.21	-3.31	-2.79	-3.28	0.27	0.17	0.28	0.24
12	oxazepam	29.9	6.2	-4.3	1.94	1.70	1.88	2.24	-3.70	-3.69	-3.79	-3.95	0.93	0.76	0.82	0.97
13	phenobarbital	24.1	5.0	-3.4	2.31	1.80	1.65	1.47	-2.41	-2.16	-1.97	-2.33	0.82	0.79	0.86	0.99
14	phenitoin	28.1	5.8	-3.7	2.29	2.31	2.27	2.47	-3.23	-4.08	-3.99	-3.99	0.90	0.89	0.81	0.90
15	prednisolone	38.0	8.1	-5.1	0.17	0.93	1.31	1.62	-2.43	-2.92	-3.11	-3.18	0.81	0.87	0.81	0.99
16	progesterone	36.3	4.2	0.0	3.76	3.68	3.73	3.87	-4.72	-4.65	-4.67	-4.42	0.99	0.77	0.84	0.91
17	quinidine	37.7	6.7	-1.5	3.39	2.66	2.84	3.44	-3.15	-2.79	-3.26	-3.12	0.74	0.78	0.84	0.80
18	salicylic acid	13.0	1.4	-4.8	1.82	1.92	2.07	2.26	-0.14	-1.17	-1.60	-1.89	0.99	0.97	0.97	0.99
19	spironolactone	44.2	5.7	0.0	4.32	2.98	2.86	2.78	-5.30	-5.26	-5.26	-4.30	0.73	0.83	0.57	0.25
20	sulfadiazine	25.5	7.2	-5.3	-0.17	-0.15	0.24	-0.09	-2.49	-2.82	-2.58	-3.40	0.92	0.91	0.91	0.98
21	sulfisoxazole	26.6	6.1	-5.3	2.65	1.15	1.02	1.01	-3.43	-3.07	-2.75	-3.02	0.97	0.98	0.98	0.96
22	testosterone	33.1	3.8	-1.4	3.42	3.17	3.11	3.32	-3.48	-3.49	-4.10	-4.08	0.84	0.90	0.91	0.98
23	theophylline	17.6	5.2	-2.1	0.42	0.31	0.30	-0.02	-2.76	-1.99	-1.78	-1.36	0.97	0.59	0.73	0.96
24	tolbutamide	28.4	5.1	-3.7	2.06	2.21	2.45	2.34	-3.64	-3.51	-3.51	-3.55	0.81	0.94	0.94	0.93

all ΣCd values obtained from HYBOT have negative values, the authors recently used absolute values $|\Sigma Cd|$ in this investigation; more positive $|\Sigma Cd|$ values indicate stronger HB donor effects. The following equation was obtained [59]:

$$\log S = 0.434(\pm 0.124) - 0.298(\pm 0.0088)\alpha + 1.090(\pm 0.046)\Sigma Ca + 0.304(\pm 0.054)|\Sigma Cd| \quad (9)$$

n = 630 r = 0.947 s = 0.536 q = 0.945

Application of available *experimental* logP values for the chemicals from this training set gave the equations:

$$\log S = 0.66(\pm 0.06) - 1.12(\pm 0.02)\log P \quad (10)$$

n = 365 r = 0.941 s = 0.50

$$\log S = 0.54(\pm 0.06) - 1.09(\pm 0.03)\log P + 0.15(\pm 0.03)|\Sigma Cd| \quad (11)$$

n = 365 r = 0.945 s = 0.48

Obviously, H-bond donor ability is a significant descriptor of logS. Thus, an important part of the solute-solvent interaction was neglected in correlations of logS with logP. According to eq. (9), solubility in water is expected to increase with increasing HB effects and to decrease with increasing α, the steric bulk effect for solutes (cavity formation in water).

In cases of solid chemicals and drugs, the significant contribution of crystal lattice energy to solubility is important. Unfortunately, there are many studies where the training sets contain both liquids and solids, and where the contribution of crystal lattice energy to solubility is not directly considered. Under such conditions the real relationships between molecular structure and solubility can be masked. Jurs et.al [60-63] published a series of papers on aqueous solubility in which the mathematical models were based on different topological, geometrical and electronic descriptors in a framework of neural network applications. In particular, this group presented multiple linear regression and computational neural network models for three training sets (containing 176 compounds having one or more nitrogen atoms with some oxygen, 223 compounds having one or more oxygen atoms with no nitrogen, and 399 compounds from the two previous sets). Many types of descriptors were considered including various topological and geometric descriptors, path counts, distance edge between different carbon atoms, sum of E-state values over

all heteroatoms, shadow area on YZ plane, partial negative surface area, sum of charges on all donable hydrogen atoms, first and third geometric moment, and cube root of the gravitation index over atom pairs [63]. In similar publications molecular topology was used to predict solubility by means of neural network modeling [64-67]. Many different physicochemical descriptors were used in linear regression analyses and non-linear relationships based on artificial neural networks to calculate water solubility. These included LSER descriptors [68], molecular volume, partial atomic charges, fractional hydrogen donor surface area [69-71], molecular weight, log P, polar surface area (PSA) [72], molecular polarizability [73], dipole moment, moments of inertia, ionization potential, heat of formation, total energy, electronic energy and other descriptors calculated by the PM3 method [74], molecular refractivity, number of hydrophilic rotatable bonds, number of H-bond donors and/or acceptors [75], molecular weight and the set of electrotopological E-state indices [76], about 100 descriptors emphasizing surface properties [77], and a set of 32 Radial Distribution Function code values [78]. A Monte Carlo simulation of water solubility was presented in [79]. A recent review devoted to solubility calculation methods is presented in [80]. Over time, the number of compounds and drugs included in training sets has increased. For example, 3351 compounds were included in the study [75]. Nevertheless, in spite of good statistical criteria for correlations between calculated solubility values and experimental ones, the chance to estimate solubilities accurately is questionable because of the large number of parameters and complex architectures used in the neural networks.

Different approaches are being developed to take the crystal lattice energy contribution to solubility into account directly. Yalkowsky *et al* proposed a "General Solubility Equation" (GSE) [81- 85]:

$$\log S_w = 0.5 - \log P_{ow} - 0.01(MP-25) \quad (12)$$

where MP is the melting point in C°. If the solute melts below 25 C°, its melting point is set equal to 25 C° so that the melting point term vanishes. The following assumptions are used in the GSE [83]: (1) the reduction in solubility due to the crystallinity of the solute is described by the van't Hoff equation; (2) the entropy of melting for most organic compounds is approximated by Walden's rule; (3) for liquid solutes, the octanol-water partition coefficient is approximately equal to the octanol-water solubility ratio; (4) most organic liquids are completely miscible with octanol; and (5) pure octanol has a molarity of 6.3. The application of GSE by the authors gave an average absolute error at the level of 0.55 log S units and a root-mean-square error at the level of 0.76 log S unit for the 380 compounds tested [85]. Next training set contained already 1026 compounds including 497 liquids and application of GSE gave even better statistic criteria [86]. However, recent verification of eq. (12) with a training set containing 752 only solid compounds and drugs in a correlation with a forced zero intercept between experimental and calculated values gave rather modest results [87]:

$$\log S_{exp} = 1.00(\pm 0.01) \log S_{eq(12)} \quad (13)$$

n=752, R=0.873, s=1.01

Based on clustering, a detailed correlation analysis of water solubility with the octanol-water partition coefficient and the melting point was presented in [88]. While eq (12) gives important insights into the physical processes needed for compounds to dissolve, it fails as a predictive tool for compounds conceived but not yet made, an important aspect for virtual libraries. The problem is that to know its $\log P_{ow}$ and MP, a compound must first be prepared before these properties can be measured. The situation for $\log P_{ow}$ is not critical because good estimates can be made as indicated in the previous section. However, there is at present no reliable way to predict MP.

An "Amended Solvation Energy Relationship" was proposed to take the crystal lattice energy contribution into account [89]. The authors incorporated in their solvation equation a term that supposedly reflects intermolecular interactions in pure liquids or solids: the product of H-bond acidity and H-bond basicity descriptors. Although the inclusion of this term led to an improved correlation, one needs to evaluate this new composite descriptor with others in the equation, in particular with those descriptors associated with H-bond acidity and basicity. For example, the correlation matrix of H-bond factors for about 2000 chemicals and drugs shows an intercorrelation between the product H-bond acceptor and H-bond donor factors and H-bond donor factors alone on the order of r = 0.90 [87].

Other approaches to calculating solubility using terms directly connected to crystal lattice energy were proposed in the framework of Mobil Order Theory (MOT) [90-93] and COSMO-RS (the conductor-like screening model for real solvents) [94]. They gave enough good results. However, as noted in [89] "the method (MOT) requires not only the entropy of fusion of solid solutes (or a MP correction term) but also a modified nonspecific solute cohesion parameter. The latter is obtained either from experimental solubilities in hydrocarbon solvents or is deduced by analogy to similar compounds". Similar situations exist in the cases of estimating free energy differences between crystal and liquid states in the framework of COSMO-RS.

Because almost all drugs are solids, the problem of calculating crystal lattice energies is important. Such direct calculations are expensive and are still far from being routine. Thus, a new approach was recently proposed to avoid this problem [49]. The method combines Similarity and QSAR concepts. It supposes that closely related compounds have similar crystal structures. So the solubilities of the nearest related structures are used as starting points. Then additional solubility increments are calculated for the compound-of-interest on the basis of HYBOT descriptors and the following equation [49]:

$$\log S_i = \sum_{j=1}^{N} [((\log S_j - 0.275(\alpha_i - \alpha_j) + 0.96\,(Ca_i - Ca_j)$$

$$-0.27(Cd_i - Cd_j)]/N \quad (14)$$

The extracted results of such solubility calculations for 24 drugs studied in [49] are presented in Table **1**.

The use of HYBOT descriptors to estimate aqueous solubility is further supported by the work of McFarland *et al*. [95,96]. They showed that quantitative H-bond

descriptors along with logP, volume and partial atomic charge terms are significantly correlated with water solubility.

In concluding this section, it should be emphasized that while development of the above mentioned approaches to take crystal lattice energy into account has improved the situation, a satisfactory method to calculate the aqueous solubility of crystalline drugs has not been completely decided.

INTESTINAL ABSORPTION

For the most part, orally administered drugs use the intestinal epithelium for transport in the human organism. There are at least four different routes of this type: (*i*) the passive transcellular, (*ii*) paracellular, (*iii*) the carrier mediated route and (*iv*) transcytosis [97,98]. A set of *in vitro* experimental methods is used for detailed study of drug transport in organisms. These include Caco-2 monolayers, systems for evaluating metabolic susceptibility (employing human liver microsomes, hepatocytes, and recombinant P450 isozymes), and artificial membranes [99-103]. However, all of these methods are expensive. Thus, the development of approaches to calculate and predict permeability and intestinal absorption (in particular) are desirable for drug design.

Three remarkable systematic investigations of the contribution of physicochemical properties to permeability levels were carried out during the middle 1990s. Based on their analysis of World Drug Index data, Lipinsky *et al*. [104] proposed that four physicochemical parameters are significantly related to permeation: molecular weight, logP, the number of H-bond donors and the number of H-bond acceptors. They suggested "the Rule of 5". This rule declares that poor drug absorption or permeation is more likely when there are more than five H-bond donors (the count of OH and NH groups), MW is over 500, logP is over 5, and there are more than 10 (2×5) H-bond acceptors (the count of nitrogen and oxygen atoms). Compounds that are substrates for biological transporters were excluded from this consideration. Waterbeemd *et al*. [105] discussed Caco-2 cell permeability using calculated molecular descriptors including volume-related parameters (molecular weight, molecular volume, polar and nonpolar parts of surface area), polarity terms (the number of H-bond donors and acceptors, and H-bond donor and acceptor factors), and a few physicochemical properties (logP, logD, and the difference between the octanol/water and the alkane/water partition coefficients). Using graphical and equation-based approaches, they demonstrated that in combinations of appropriate size H-bonding descriptors might be of potential use to estimate membrane permeation. Winiwart *et al*. [106] studied the effective permeability in the human jejunum (*in vivo*) as a function of $logD_{5.5}$, $logD_{6.5}$, the highest occupied and lowest unoccupied molecular orbital energies, dipole moments, molecular weight, number of atoms, molecular volume, molecular surface area, number of potential hydrogen bond donors and potential hydrogen bond acceptors, and polar surface area. In the framework of multivariate data analysis, the best PLS models were obtained in cases involving the number of hydrogen bond donor atoms, polar surface area and $logD_{5.5}$ or $logD_{6.5}$. Unfortunately, the two latter publications were based on small training sets, and so their real predictive powers are not likely reliable.

A non-linear six-descriptor neural network model based on molecular properties was constructed for 86 drugs [107]. The descriptors were: the number of single bonds, the normalized 2D projection of molecules on the YZ plane, the charge on donor hydrogen atoms, the surface area of hydrogen bond acceptor atoms, the charge on hydrogen bond acceptor atoms, and the gravitational index. A similar work for the same data is presented in recent publication [108]. Nitrogen with one or two hydrogen atoms, nitrogen with three single bonds attached to heavy atoms, oxygen with one hydrogen atom, oxygen with two single bonds attached to heavy atoms, oxygen with one double bond attached to a heavy atom, the number of rotatable bonds, and the number of aromatic rings were recognized as significant descriptors in this study. Quantum-chemical descriptors have also been used [109] as well as a novel numerical molecular representation called the "molecular hashkey" [110]. A non-aqueous partitioning system has been proposed for the prediction of oral peptide absorption [111].

The polar surface area (PSA) is now the most popular descriptor for permeability at the whole cell level, and for human intestinal absorption in particular [105, 112-123]. As to the physical meaning, it is supposed that the polar groups of a molecule are solvated in the intestine's aqueous environment, and the molecule's entry to the more lipophilic environment of the intestinal epithelium requires the desolvation of those groups. This process is endothermic, so absorption must be hindered. In this approach, the hydrogen bond potential is expressed as the polar part of the molecular surface area. The PSA of a molecule is defined as the area of its van der Waals surface that arises from oxygen and nitrogen H-bond acceptor atoms, and hydrogen atoms attached to such atoms. Only a single conformation of a molecule was taken into consideration in the first application of the method [105,113]. Later, "dynamic" PSA_d, which is a Boltzmann-weighted average value calculated from a set of low energy conformers, was proposed [114-118]. Because detailed conformation searches for large molecules are slow and expensive, Clark demonstrated that using a single low-energy conformer as a representative of the ensemble of conformers likely to be present *in vivo* is a reasonable approximation [120-122]. As a result of correlating absorption with polar surface area, the threshold of poor absorption was estimated to be $PSA \geq 140$ A^2 [120]. However, for some drugs there are significant deviations from that rule. For example, pyridostigmine has a low PSA value but is poorly absorbed [120]. One reason for this may be that PSA is an imperfect descriptor. As an example, in [119] to obtain a good correlation with Caco-2 permeability, timolol's calculated PSA value was "corrected" by omitting the two weak H-bond acceptor nitrogens in the thiadiazole ring. Thus, it seems that a more correct definition and determination of polar surface area is necessary.

The application of thermodynamic H-bond parameters to model human intestinal absorption is presented in [124-126]. Data for 169 drugs were included in the analysis. The descriptors in the solvation equations correlated directly with the percentage of absorption [124]. In a recent publication [124], for those drugs whose absorptions were neither 0 nor

100%, regression analysis on the effective rate of absorption (GI k_{eff}) led to the following equation:

$$\log GI\ k_{eff} = 0.544 - 0.025\ E + 0.141\ S - 0.409\ A - 0.513\ B + 0.204\ V \quad (15)$$

N=127, R=0.89, s=0.29, F=0.84

This approach was applied also to benzamidine analog inhibitors [126].

HYBOT descriptors were used to quantitatively estimate the oral drug absorption in humans for 31 passively transported drugs [127]. For the first time, the ionization states of the compounds at physiological pH were taken into account in calculating physicochemical descriptors. The authors also discussed the relationship between descriptors and published absorption data: linear [114,117,125] compared to sigmoid relationships [105,115, 117, 120]. Data presented as percentage effect or fractional effect (i.e. fraction absorbed, FA) are problematic in linear regression techniques. For example, linear models correlating FA data directly with physicochemical descriptors can predict negative FA values or FA values greater than 1 (i.e. > 100%). However, this difficulty may be overcome by transforming the FA values into logit FA values [128]:

$$\text{logit FA} = \log (FA/(1-FA)) \quad (16)$$

In spite of good correlation coefficients and standard deviations, cross-validation coefficients (Q) were not satisfactory for correlating logit FA with the independent variables (for the use of logit FA see also [123]). This caused the authors to consider empirical non-linear models for estimating relationships between FA and physicochemical parameters. For this purpose eq. (16) was rewritten as:

$$Z = \text{logit FA} = \log [FA/(1-FA)] \quad (17)$$

$$FA/(1-FA) = 10^{Z} \quad (18)$$

$$FA = 10^{Z} / (1+10^{Z}) \quad (19)$$

$$FA = 1/(1+ 10^{-Z}) \quad (20)$$

where $Z = F(X_1, X_2, \ldots\ldots X_i)$.

This non-linear approach led to $\sum C_{ad}$ as being the best descriptor for a one-parameter equation:

$$FA = 1/(1+10^{-[5.02 - 0.307 \sum C_{ad}]}) \quad (21)$$

N=31, R=0.943, s=0.12, Q=0.918

This term is the sum of absolute free energy values of H-bond acceptor and donor factors in the molecule, and characterizes the total ability of a compound to form hydrogen bonds. The equation shows that compounds with

Table 2. New 3-D Hydrogen Bond Descriptors with Definitions [33,129]

$WEASA = \sum_{n} k_a E_a$	Van der Waals'acceptor surface area in Å^2 which is proportional to H-bond enthalpy factors E_a of acceptor atoms. n is number of acceptors in the molecule of interest. $k_a = \frac{1}{5}(\frac{1}{3} S_O)$. So is a surface sphere with a radius of 1.36 Å (Osp3)
$WOFEASA = \sum_{n} k_a C_{a(o)}$	Van der Waals'acceptor surface area in Å^2 which is proportional to H-bond overall free energy factors $C_a(o)$ of acceptor atoms.
$WEDSA = \sum_{n} k_d E_d$	Van der Waals donor surface area in Å^2 which is proportional to H-bond enthalpy factors of donor atoms.n is number of donors in a molecule, $k_d = \frac{1}{5}(\frac{1}{3} S_H)$, SH is a surface sphere with a radius of 1.08 Å (H atom).
$WFEDSA = \sum_{n} k_d C_d$	Van der Waals donor surface area in Å^2 which is proportional to H-bond free energy factors of donor atoms.
$OEASAprobe = \sum_{n} k_a(H_d) E_a E_{d\,(probe)}$	Surface area around a molecule in Å^2 where interactions of acceptor atoms of a molecule with a H-bond donor probe have been optimumly placed and which is proportional to product of H-bond enthalpy factor absolute values of those atoms. $E_{d(probe)}$ is the enthalpy factor of the probe H-bond donor, $k_a \quad \frac{1}{20}(\frac{1}{3} S_{rm})$, Srm is the surface area of sphere with a radius of r_m = 2.45 Å for the strongest H-bonding
$OEASAprobe = \sum_{n} k_a(H_d) E_a E_{d\,(probe)}$	Surface area around a molecule in Å^2 where interactions of acceptor atoms of a molecule with H-bond donor probe have been optimumly placed and which is proportional to product of H-bond free energy factor absolute values of those atoms.
$OEDSAprobe = \sum_{n} k_d(H_a) E_d E_{a(probe)}$	Surface area in Å^2 around a molecule where interactions of donor atoms of a molecule with H-bond acceptor probe have been optimumly placed and which is proportional to product of its H-bond enthalpy factor absolute values.
$OFEDSAprobe = \sum_{n} k_d(H_a) C_d Ca(o)_{probe)}$	surface area in Å^2 around a molecule where interactions of donor atoms of a molecule with H-bond acceptor probe have been optimumly placed and which is proportional to product of its H-bond free energy factor absolute values.
$SIEAprobe = \phi H \cdot d(s)$	surface integral for enthalpy values (kcal/M*Å^2) of interactions between acceptor atoms of a molecule and a donor probe on the surface $OEASA$.
$SIEDprobe = \phi H \quad d(s)$	surface integral for enthalpy values (kcal/M*Å^2) for interactions between donor atoms of a molecule and an acceptor probe on the surface $OEDSA$

$\sum C_{ad}$ values < 8 are completely absorbed in humans. In cases where $\sum C_{ad} \cong 16$, the FA values are approximately equal to 0.5. Absorption is poor in the cases where $\sum C_{ad} > 22$. In the sigmoid model, descriptors that gave poorer results included molecular weight and polarizability as volume-related terms, partial charges on atoms as electrostatic terms and logD as a lipophilicity term. Separating the composite descriptor $\sum C_{ad}$ into the H-bond acceptor and donor components, $\sum C_a$ and $\sum C_d$, significantly improved the correlation:

$$FA=1/(1+10^{-[5.05 - 0.36\ \sum Ca + 0.26\ \sum Cd]}) \qquad (22)$$

N=31, r=0.974, s=0.09, Q=0.959

Addition of other non-correlated descriptors had only weak influences on the results. Eq.(22) demonstrates that hydrogen bonding is a key factor in absorption processes.

Recently, this study was significantly improved by using a training set of 154 passively transported drugs, and by employing a larger set of descriptors [33]. The descriptor set included the following 2-D properties: MW, molecular polarizability (α), number of H-bond acceptor atoms (N_a) in a molecule, number of H-bond donor atoms (N_d), H-bond acceptor (E_a) and H-bond donor (E_d) enthalpy factors, H-bond acceptor (C_a) and H-bond donor (C_d) free energy factors, and logP. The 3-D descriptors were SAREA, PSA, and 10 new descriptors based on new H-bond potentials [33, 129]; these are defined in Table **2**. The correlation matrix showed that almost all of these descriptors are significantly intercorrelated. High levels of correlation exist not only among H-bond acceptor descriptors, and among H-bond donor descriptors, but also between H-bond acceptor and donor descriptors. There is a significant correlation between molecular weight (MW) and the number of H-bond acceptor and donors. Such a correlation could be explained by tendency for the number of chemical functional groups to increase with molecular size. LogP values correlate significantly with number of H-bond donor groups in the molecule. LogP is a composite descriptor that includes, in hidden form, information about molecular size and H-bond acceptor ability. Under such conditions, the simultaneous consideration of molecular weight, the number of H-bond donors and acceptors and logP as presented in the "Rule 5" [104] possibly masks the real relationship between structure and absorption.

In the framework of one parameter sigmoid models H-bond descriptors demonstrated obvious advantages when compared to the volume-related terms (SAREA, polarizability) and lipophilicity (logP). These correlations show that H-bond acceptor ability and H-bond donor ability play significant roles in absorption. For absorption, equations with the best statistical criteria for a training set containing 154 drugs were obtained with the composite descriptors WEASA+WEDSA (n=154, r=0.89, s=0.16, F=507.4), WOFEASA+WFEDSA (n=154, r=0.91, s=0.15, F=734.7), OEASA+OEDSA (n=154, r=0.90, s=0.16, F=627.0) and OFEASA+OFEDSA (n=154, r=0.91, s=0.15, F=738.5). These descriptors quantitatively characterize the total ability of a drug to participate in hydrogen bonding.

The composite indicator descriptor N_{ad} (recently used with logP to calculate the human intestinal absorption of 124 drugs [130]) and PSA (recently used with logP in a pattern recognition model [131]) give enough good correlation with human intestinal absorption. These descriptors are obtained respectively by counting the number of atoms in a molecule that can (in principle) participate in hydrogen bonding, and by calculating the van der Waals

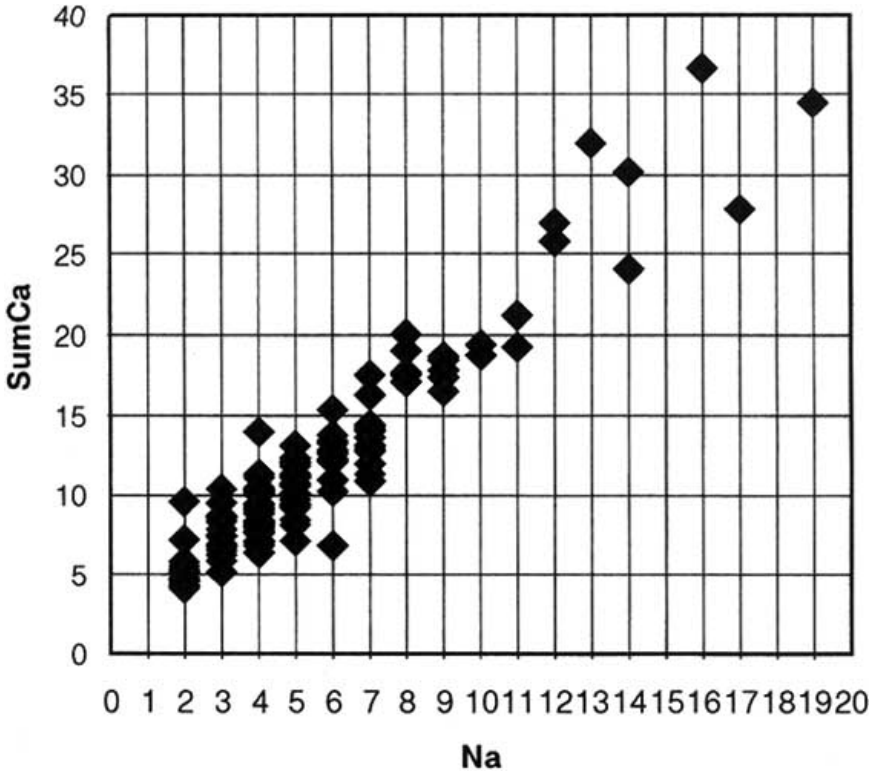

Fig. (1). Graphical comparison of number of acceptor atoms in a molecule and overall free energy factors.

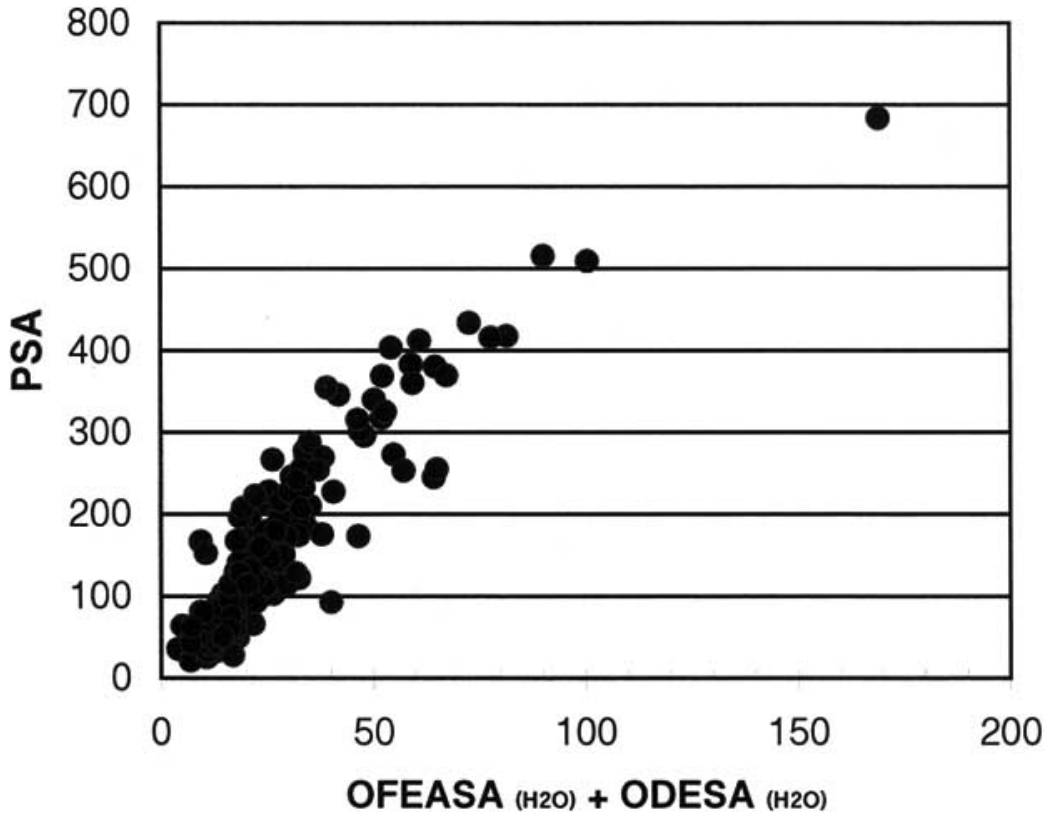

Fig. (2). Graphical comparison of 3D-HYBOT descriptors and PSA for 154 drugs.

surface for oxygen, nitrogen and hydrogens at those heteroatoms. The statistical parameters for models with those descriptors are only a little worse than those based on thermodynamic descriptors. There are strong intercorrelations among those descriptors. However, a comparison of values between these two types of descriptors shows that for any fixed number of H-bond acceptors and donors there are wide ranges of thermodynamic descriptor values (see Fig. **1**). And PSA as descriptor doesn't reflect real ability to form H-bonds (graphical comparison PSA and 3D HYBOT descriptors are presented in Fig. **2**). Hence, to estimate human intestinal absorption, N_{ad} and PSA are only crude preliminary instruments. The proper estimation of intestinal absorption is really only possible on the basis of descriptors that directly relate to thermodynamic data.

The probability of different mechanisms of absorption hinders the construction of QSAR for diverse chemicals and drugs. A novel approach based on the Similarity concept and QSAR was recently proposed to predict intestinal absorption in humans [132]. The approach is based on the assumption that nearest related structures (nrs) have the same absorption mechanisms. For each pair compared, the difference in absorption values is related to the differences in the physicochemical parameters that contribute in passive transport. As example, a set of ten β-lactams was considered (Table **3** in [132]). All of these compounds have strong H-bond acceptor and donor groups. If these drugs relied solely on passive transport, they would all be poorly absorbed. The last four compounds from this list are in fact poorly absorbed. However, cephalexin and five related structures contain the group –CH(NH3+)Ph, which is associated with active transport. Despite the strong H-bond donor strength of these compounds, each is almost completely absorbed because of this special mechanism. The Tanimoto, Euclidean and Cosine similarity indices allow one to readily select related compounds. The final absorption calculation in the framework of this approach is carried out by means of eq (23):

$$FA=1/(1+10^{[(-\log(FA_{nrs}/(1-FA_{nrs}))-\ 0.36\Delta\Sigma C_{a(drug-nrs)}+\ 0.26\ \Delta\Sigma C_{d(drug-nrs)}]}) \quad (23)$$

Eq. (23) was used to calculate the absorption of 100 drugs including 38 neutral compounds, 29 cations, 20 anions and 13 zwitterions. At pH 7.4, the ionization state of the drugs was taken into account. The correlation between the experimental and calculated absorption values had almost a zero intercept and good statistics parameters (n=100, r=0.945, s=0.11, Q=0.943). Using eq. (22) gave much worse results (n=100, r=0.740, s=0.23, Q=0.719). This approach was used in the program SLIPPER–2001 [49] to predict lipophilicity, aqueous solubility and intestinal absorption in humans. The results of calculations for 24 drugs are presented in Table **1**.

ORAL BIOAVAILABILITY

Oral administration is the preferred way to insert a drug in humans. The fraction of the drug dose absorbed intact depends not only on epithelial permeability but also on other processes. Of course, membrane permeation is recognized as a common requirement for oral bioavailability. Thus, all of the already estimated relationships between

physicochemical descriptors and intestinal absorption may be useful for studies of structure-bioavailability relationships. However, many other factors are now recognized as limiting oral bioavailability. These include energy-driven export from blood to gut, and first pass metabolism by enzymes of intestinal or liver cells. Bioavailability is a global parameter that incorporates many permeability and transport processes [123]. So contrary to simple correlations (for example, MW with bioavailability [133]) it is possible that the relationship between physicochemical descriptors and this complex phenomenon may also be complex. Structure-bioavailability relationships studies are only beginning to be developed. In this short review it is possible only to mention two recent publications. The structure-bioavailability relationships among 232 structurally diverse drugs were studied [134]. The oral bioavailabilities in human adults were each assigned to one of four ratings, and analyzed in relation to structural and physicochemical parameters by means of a multicategorical classification method using a simplex technique. The set of descriptors included $logD_{7.4}$, $logD_{6.5}$, and their difference as well as structure fragments that might participate in various metabolic reactions. Although the predictive power of the model was rather modest (only 60% of drugs were correctly classified), the development of this approach seems promising. It is possible that an approach based on combining Similarity and QSAR [49] could be also useful to model these complex phenomena.

The oral bioavailabilities in rats were included in a structure-property analysis of over 1100 compounds [135]. Molecular flexibility, as measured by the number of rotatable bonds, and polar surface area (or total hydrogen bond count, sum of donor and acceptor groups) were recognized to be important predictors of oral bioavaolability. No significant influence of molecular weight was discovered. This conclusion must be considered only as an estimated trend because of the above-indicated critique of PSA and the number of H-bond donors and acceptors as descriptors. Its application to structure-bioavailability relationships in humans is problematic.

ACKNOWLEDGEMENTS

The author appreciate valuable comments given by Dr. J. W. McFarland and Dr. K.-J.Schaper during the preparation of this manuscript.

THE LIST OF ABBREVIATIONS

ADMET	=	Absorption, distribution, metabolism, excretion and toxicity
QSAR	=	Quantitive structure-activity relationships
QSPR	=	Quantitative Structure-property relationships
n	=	Number of compounds
R	=	Correlation coefficient
sd	=	Standard deviation
RMS	=	Root-mean-square error
F	=	Fisher criterion
logP	=	Partition coefficient (P) of a compound distributed between octanol and water phases is expressed as logP, its logarithmic form
logS	=	Water solubility, its logarithmic form
FA	=	Fraction absorbed
E	=	Excess molar refraction
S	=	Solute polarity/polarizability
A	=	Solute H-bond acidity
B	=	Solute H-bond basicity
V	=	McGowan characteristic molar volume
ΔB	=	Differences between the two phases in the entropy of the solute/solvent exchange
ΔF	=	Hydrophobic effect- term, the two H-bond interaction-related terms
ΔO and ΔOH	=	Differences in the strengths of the H-bonds that bind the solute and solvent molecules in each phase
ΔD	=	Accounts for nonspecific forces only
α	=	Molecular polarizability
MW	=	Molecular volume
Σq	=	Sum of partial atomic charges in a molecule
ΣE	=	Enthalpy H-bonding factors
ΣC	=	Free energy H-bonding factors
PSA	=	Polar surface area
MP	=	Melting point
GSE	=	General solubility equation
HYBOT	=	Hydrogen bond thermadynamics

REFERENCES

[1] van de Waterbeemd, H., Smith, D. A., Beaumont, K., Walker, D. K. *J. Med. Chem*., **2001**, *44*, 1313.
[2] van de Waterbeemd, H. in *Structure-Property Correlation in Drug research;* van de Waterbeemd, H. Ed.; R. G. Landes Company and Academic Press: Georgetown, **1996**; pp. 55-109.
[3] van de Waterbeemd, H., Jones, B. C. *J. Comput. -Aid Mol. Des*., **2001**, *15*, 273.
[4] Fujita, T., Iwasa, J., Hansch, C., *J. Am. Chem. Soc*., **1964**, *86*, 5175.
[5] Rekker, R. F., *A means of characterizing membrane systems*. Pharmachem. Library Vol. 1, Elsevier: Amsterdam, 1977.
[6] Schaper, K. -J., Rosado, S. M. L. *Quant. Struct. -Act. Relat.,* **1997**, *16*, 224.
[7] Devillers, J., Domimine, D., Guillon, C. *Europ. J. Med. Chem*., **1998**,*33*,659.
[8] Livingstone, D. J., Ford, M. G., Huuskonen, J. J, Salt, D. W. *J. Comput. -Aid Mol. Des*., **2001**, *15*, 741.
[9] Beck, B., Breindl, A., Clark, T. *J. Chem. Inf. Comput. Sci*., **2000**, *40*, 1046.
[10] Mannhold, R, van de Waterbeemd, H., Smith, D.A. *J. Comput. -Aid Mol. Des*., **2001**, *15*, 337.
[11] Ruelle, P. *J. Chem. Inf. Comput. Sci*., **2000**, *40*, 681.
[12] Edward, J. T. *Can. J. Chem*., **1998**, *76*, 1294.
[13] Leahy, D. E. *J. Pharm. Sci.,* **1986**, *75*, 629.
[14] Kamlet, M. J., Abraham, M. H., Doherty, R. M., Taft, R. W. *J. Am. Chem. Soc*., **1984**, *106*, 464.
[15] Hawkins, G. D., Liotard, D. A., Cramer, C. J., Trublar, D. G. *J. Org. Chem*., **1998**, *63*, 4305.

[16] Ghuloum, A. M., Sage, C. R., Jain, A. N. *J. Med. Chem*., **1999**, *42*, 1739.
[17] de Bruijn, J. D., Hermens, J. *Quant. Struct. -Act. Relat*., **1990**, *9*, 11.
[18] Dunn, W. J., Koehler, M. G., Grigoras, S. *J. Med. Chem*., **1987**, *30*, 1121.
[19] Lewis, D. F. V. *J. Comput. Chem*., **1989**, *10*,145.
[20] Klopman, G., Namboodiri, K., Schochet, M. *J. Comput. Chem*., **1985**, *6*, 28.
[21] Abraham, M. H., Ibrahim, A., Zissimos, A. M., Zhao, Y. H., Comer, J., Reynolds, D. P. *Drug Discov. Today*, **2002**, *7*,1056.
[22] Ruelle, P. *Perspect. Drug Discovery Des*., **1999**, *17*, 61.
[23] Raevsky, O. A. *J. Phys. Org. Chem*., **1997**, *10*, 369.
[24] Abraham, M. H., Chadra, H. S., in Lipophilicity in Drug Action and Toxycology; Pliska, V., Testa, B, Van de Waterbeemd, H., Eds.; VCH:Weinheim, **1996**; pp. 311-337.
[25] Abraham, M. H., Chadra, H. S., Whiting, G. S., Mithell, R. C. *J. Pharm. Sci.,* **1994**, *83*, 1085.
[26] Dearden, J., Beesmen, G. M. *Quant. Struct. -Act. Relat*., **1988**, *7*, 133.
[27] Platts, J. A., Abraham, M. H., Butina, D., Hersey, A. *J. Chem. Inf. Comput. Sci*., **2000**, *40*, 71.
[28] Goodford, P. J. *J. Med. Chem*., **1985**, *28*, 849.
[29] Raevsky, O. A. *Russ. Chem. Rev*. **1999**, *68*, 505.
[30] Raevsky, O. A., Grigor'ev, V. Ju., Kireev, D. B., Zefirov, N. S. *Quant. Struct. -Act. Relat*., **1992**, *11*, 49.
[31] Raevsky, O. A., Grigor'ev, V. Ju., Trepalin, S. V. HYBOT program package, *Registration by Russian State Patent Agency* N 990090 of 26. 02. 99; Raevsky, O. A., Skvortsov, V. S., Grigor'ev, V. Ju., Trepalin, HYBOT in UNIX/LINUX, *Registration by Russian State Patent Agency* N 2002610496 of 05. 02. 02; On line version www//http/ibmh. msk. su/molpro/HYBOT.
[32] Raevsky, O. A., Skvortsov, V. S. *J. Comput. -Aided Mol. Des*., **2002**, *16*, 1.
[33] Raevsky, O. A., Skvortsov, V. S. *QSAR Comb. Sci*, **2003** (submitted).
[34] Dearden, J. C. *Environ. Health Perspect*., **1985**, *61*, 203.
[35] van de Waterbeemd, H., Testa, D. *Adv. Drug Res*., **1987**, *16*, 85.
[36] El Tayar, N. And Testa, B., in: *'Trends in QSAR and Molecular Modelling 92*'; Wermuth. C. Ed.; ESCOM: Leide, **1993**; pp. 101.
[37] Raevsky, O. A., Schaper, K. -J., Seydel, J. K. *Quant. Struct. -Act. Relat*., **1995**, *14*, 433.
[38] Raevsky, O. A. In *Computer-Assisted Lead Finding and Optimization;* van de Waterbeemd, H., Testa, B., Folkers, G., Eds.;Willey-VCH: Weinheim, **1997**; pp. 369-378.
[39] Raevsky, O. A., Schaper, K. -J., Waterbeemd, H., McFarland J. W. In Molecular modeling and prediction of bioactivity; Guntertofte, K., Jorgensen, F. S., Eds.; Kluwer Academic/Plenum Publishers: New York, **2000**; pp. 221-227.
[40] Raevsky, O. A., Trepalina, E. P., Trepalin, S. V. *ibid*, **2000**; pp. 489-490.
[41] Wold, S., Dunn, III, W. J., Hellberg, S. In Drug Design: Fact or fantasy, Jolles, G., Woldbridge, K. Ed.; Academic Press: London, **1984**; pp. 95-117.
[42] Raevsky, O. A., Sapegin, A, Zefirov, N. S. *Quant. Struct. -Act. Relat*., **1995**, *14*, 403.
[43] McFarland, J. W., Gans, D. J. In Chemometric Methods in Molecular Design, Waterbemd H (ed.), VCH:Weinhaim, **1995**, pp. 295-308 .
[44] Wilett, P., Barnard, J. M., Downs, G. M. *J. Chem. Inf. Comput. Sci*., **1998**, *38*, 983.
[45] Raevsky, O. A. *SAR and QSAR in Environ. Res*., **2001**, *12*, 367.
[46] Filimonov, D., Poroikov, V., Borodina, Y., Glorizova, T. *J. Chem. Inf. Comput. Sci*., **1999**, *39*, 666.
[47] Basak, S. C., Grunwald, G. D. *SAR&QSAR Environ. Res*., **1995**, *3*, 265.
[48] Mannhold, R., Dross, K. *Quant. Struct. -Act. Relat*., **1996**, *15*, 403.
[49] Raevsky, O. A., Trepalin, S. V., Trepalina, E. P., Gerasimenko, V. A., Raevskaja, O. E. *J. Chem. Inf. Comput. Sci*., **2002**, *42*, 540.
[50] Avdeef, A. *J. Pharm. Sci*., **1993**, *82*, 183.
[51] Avdeev, A., in *Lipophilicity in Drug Action and Toxicology*, (eds. Pliska, V., Testa, B., Waterbeemd, H.), Weinheim, VCH, **1996**; p. 109.
[52] Raevsky, O. A., Trepalina, E. P., Trepalin, S. V. in *Molecular Modeling and Prediction Bioactivity,* Gundertofte, K., Jorgensen F. S. eds., Kluwer/Academic/Plenum Publishers, New York, **2000**, 489.
[53] Xing, L., Glen, R. C. *J. Chem. Inf. Comput. Sci*., **2002**, *42*, 796.
[54] Reichel, A., Begley, D. *J. Pharm. Res*., **1998**, *15*, 1270.
[55] Ducarne, A., Neuwels, M., Goldstein, S., Massingham, R. *Eur. J. Med. Chem.,* 1998, 33, 215.
[56] Kansy, M., Senner, F., Gubernator, K. *J. Med. Chem*., **1998**, *33*, 215.
[57] Hansch, C. F., Quinlan, J. E., Lawrence, G. L. *J. Org. Chem*., **1968***, 33*,347.
[58] Yalkowsky, S. H., Morozowich, W., in Medicinal Chemistry, ed. Ariens, E. J. Academic Press, **1980**, *9*, 122.
[59] Schaper, K. -J., Kunz, B., Raevsky, O. A., *QSAR Comb. Sci.*, **2003**, *22*, 943.
[60] Nelson, T. M., Jurs P. *J. Chem. Inf. Comput. Sci*., **1994**, *34*, 601.
[61] Sutter, J. M., Jurs, P. *J. Chem. Inf. Comput. Sci*., **1996**, *36*, 100.
[62] Mithell, B. E., Jurs, P. *J. Chem. Inf. Comput. Sci*., **1998**, *38*, 489.
[63] McElroy, N. R., Jurs, P. *J. Chem. Inf. Comput. Sci*., **2001**, *41*, 1237.
[64] Patil, G. S. *J. Hazardous Materials*, **1994,** *36*, 35.
[65] Huuskonen, J., Salo, M., Taskinen, J. *J. Pharm. Sci.,* **1997**, *86*,450.
[66] Huuskonen, J., Salo, M., Taskinen, J. *J. Chem. Inf. Comput. Sci*., **1998**, *38*, 450.
[67] Huuskonen, J. *J. Chem. Inf. Comput. Sci*., **2000**, *40*, 773.
[68] Lee, S. B. *J. Pharm. Sci.,* **1996**, *85*,348.
[69] Huibers, P. D. T., Katritzky, A. R. *J. Chem. Inf. Comput. Sci*., **1998**, *38*, 283.
[70] Katritzky, A. R., Wang, Y, Sild, S and Tamm, T. *J. Chem. Inf. Comput. Sci*., **1998**, *38*, 720.
[71] Katritzky, A. R., Maran, U., Lobanov, V., Karelson, M. *J. Chem. Inf. Comput. Sci*., **2000**, *40*, 1.
[72] Liu, R., So, S. -S. *J. Chem. Inf. Comput. Sci*., **2001**, *41*, 1633.
[73] Chenzhong, C., Zhiliang, L. *J. Chem. Inf. Comput. Sci*., **1998**, *38*, 1.
[74] Jaffe, D., Cohen, Y., Espinosa, G., Arenas, A., Giralt, F. *J. Chem. Inf. Comput. Sci*., **2001**, *41*, 1177.
[75] Engkvist, O., Wrede, P. *J. Chem. Inf. Comput. Sci*., **2002**, *42*, 1247.
[76] Tetko, I. V., Tanchuk, V. Yu, Kasheva, T. N., Villa, A. E. P. *J. Chem. Inf. Comput. Sci*., **2001**, *41*, 1488.
[77] Bruneau, P. *J. Chem. Inf. Comput. Sci*., **2001**, *41*, 1605.
[78] Yan, A., Gasteiger, J. *J. Chem. Inf. Comput. Sci*., **2003**, *43*, 429.
[79] Jorgensen, W. L., Duffy, E. M. *Biorg. Med. Chem. Lett*., **2000**, *10*, 1155.
[80] Jorgensen, W., Duffy, E. M. *Adv. Drug Deliver. Rev*., **2002**, *54*, 355.
[81] Yalkowsky, S. H., Valvani, S. C. *J. Pharm. Sci*., **1980**, *69*, 912.
[82] Jain, N., Yalkowsky, S. H. *J. Pharm. Sci*., **2000**, *90*, 234.
[83] Ran, Y., Yalkowsky, S. H. *J. Chem. Inf. Comput. Sci*., **2000**, *40*, 773.
[84] Ran, Y., Jain, N., Yalkowsky, S. H. *J. Chem. Inf. Comput. Sci*., **2001**, *41*, 1208.
[85] Peterson, D., and Yalkowsky, S. H. *J. Chem. Inf. Comput. Sci*., **2001**, *41*, 1531.
[86] Ran, Y., He, Y., Yang, G., Johnson, J. L. H. Yalkovsky, S. H. *Chemosphere*, **2002**, *48*, 487.
[87] Raevsky, O. A., Schaper, K. -J. *unpublished data*.
[88] Nouwen, J., Hansen, B. *Quant. Struct. -Act. Relat*., **1996**, *15*, 17.
[89] Abraham, M. H., Le, J. *J. Pharm. Sci*., **1999**, *88*, 868.
[90] Ruelle, P., Buchmann, M., Ho, N. T., Kesselring, U. W. *Environ. Sci. Technol*., **1993***, 27*, 266.
[91] Ruelle, P., Kesselring, U. W. *J. Pharm. Sci*., **1997**, *96*, 179.
[92] Ruelle, P., Kesselring, U. W. *Can. J. Chem*., **1998**, *76*, 553.
[93] Ruelle, P. *J. Phys. Org. Chem*., **1999**, *12*, 769.
[94] Klamt, A., Eckert, F., Hornig, M. *J. Comput. -Aided Des*., **2001**, *15*, 355.
[95] McFarland, J. W., Avdeev, A., Berger, C. M., Raevsky O. A. *J. Chem. Inf. Comput. Sci*., **2001**, *41*, 1355.
[96] McFarland, J. W., Du, Chau M., Avdeef, A. In *Drug Bioavailability/Estimation of Solubility, Permeability and Absorption*. van de Waterbeemd, H., Lennernäs, Artursson, P., Eds., (Series: Methods and Principles in Medicinal Chemistry. Series Editors: Mannhold, R., Kubinyi, H., Folkers, G.), Wiley-VCH Verlag GmbH, Weinheim, **2003**; pp. 232-242.
[97] Artursson, P., Palm, K. , Luthman, K. *Advanced Drug Drlivery Reviews*, **1996**, *22*, 67.
[98] Stenberg, P., Luthman, K., Artursson, P. *J. Controlled Release*, **1999**, *1*, 123.
[99] Artursson, P., Borchardt, R. *Pharm. Res*., **1997**, *14*, 1655,

[100] Tarbit, M. H., Berman, *J. Curr. Opin. Chem. Biol*., **1998**, *2*, 411.
[101] Berman, J., Halm, K., Adkison, K., Shaffer, J., *J. Med. Chem*., **1997**, *40*, 827.
[102] Ito, K., Iwatsubo, T., Kanamitsu, S., Nakajima, Y, Sugiyama, *Annu. Rev. Pharmacol. Toxicol.,* **1998**, *38*, 461.
[103] Lloyd, A. W. *Pharm. Sci. Technol. Today*, **1998**, *1*, 45.
[104] Lipinski, C. A., Lombardo, F., Dominy, B. W., Feeney, P. J. *Adv. Drug. Del. Rev*., **1997**, *23*, 3.
[105] Waterbeemd, H,. Camenisch,G., Folkers, G., Raevsky, O. A. *Quant. Struct. -Act. Relat*., **1996**, *15*, 480.
[106] Winiwart, S., Bonham, N. M., Ax, F., Hallberg, A., Lennernas, H., Karlen, A. *J. Med. Chem.,* **1998**, *41*, 4939.
[107] Wessel, M. D., Jurs, P. C., Tolan, J. W., Muscal, S. M. *J. Chem. Inf. Comput. Sci*., **1998**, 38 ,726.
[108] Niwa, T. *J. Chem. Inf. Comput. Sci*., **2003**, *43*, 113.
[109] Norinder, U., Osterberg, T., Artursson, P. *Eur. J. Pharm. Sci*., **1999**, *8*, 49.
[110] Paterson, D. A., Conradi, R. A., Hilgers, A. R., Vidmar, T. J., Burton, P. S. *Quant. Struct. -Act. Relat*., **1994**,*13*, 4.
[111] Ghuloum, A. M., Sage, C. R. Jain, A. N. *J. Med. Chem*., **1999**, *42*, 1739.
[112] Hermann, R. B. *J. Phys. Chem*., **1972**, *76*, 2754.
[113] Waterbemd, H., Kansy, M. *Chimia*, **1992**, *46*, 299.
[114] Palm, K.; Luthman, K.; Ungell, A.-L., Strandlund, G., and Artursson, P. *J. Pharm. Sci.,* **1996**, *85*, 32.
[115] Palm, K., Stenberg, P., Luthman, K., and Artursson, P. *Pharm. Res*., **1997**, *14* 568.
[116] Palm, K., Artursson, P., and Luthman, K., In: van de Waterbeemd, Testa, B., Folkers, G., (eds.) , *Computer-Assisted Lead Finding and Optimization*, Willey-VCH, Weinheim, **1997**; pp. 277-289.
[117] Palm, K.; Luthman, K.; Ungell, A. -L. , Strandlund, G., Beigi, F., Lundahl, P., Artursson, P. *J. Med. Chem*., **1998**, *41*, 5382.
[118] Stenberg, P., Luthman, K., Ellens, H., Lee, C. P., Smith, P. L., Lago, A., Elliot, J. D., Artirsson, *P. Pharm. Res.,* **1999**, *16*, 1520.
[119] Krarup, L. H., Christensen, I. T., Hovgaard, L., Froljaer, S. *Pharm. Res*., **1998,** *15*, 972.
[120] Clark, D. E. *J. Pharm. Sci*., **1999**, *88*, 807.
[121] Clark, D. E. *J. Pharm. Sci*., **1999**,*88*, 815.
[122] Clark, D. E., and Pickett, S. D. *Drug Discovery Today*, **2000**, *5*, 49.
[123] Oprea, T. I., and Gottfries, J. *J. Mol. Graphics Model*., **1999**, *17*, 261.
[124] Zhao, Y. H., Le, J., Abraham, M. H., Hersey, A., Eddershaw, P. J., Luscombe, C. N., Boitina, D., Beck, G., Sherborne, B., Cooper, I., and Platts, J. A. *J. Pharm. Sci*., **2001**,*90*, 749.
[125] Abraham M. *Eur. J. Med. Chem*., **2002**, *37*, 595.
[126] Sugano, K., Yoshida, S., Takaku, M., Haramura, M., Saitoh, R., Nabuchi, Y and Ushio, H. *Biorg. Med. Chem. Lett*., **2000**, *10*, 1939.
[127] Raevsky, O. A., Fetisov, V. I., Trepalina, E. P., McFarland J. W., Schaper, K. -J., *Quant. Struct. -Act. Relat*., **2000**, *19*, 366.
[128] Schaper, K. -J., in D. Hadzy, H. Jerman-Blazic (eds*), QSAR in Drug Design and Toxicology,* p. 58-60, Elsevier, Amsterdam, **1987**.
[129] 3DHYBOT online version: http//www.ibmh.msk.su/molpro/HYBOT
[130] Egan, W. J., Merz, K. M., Baldwin, J. J. *J. Med. Chem.,* **2000***, 43*, 3867.
[131] Deretey, E., Feher, M., Schmidt, J. M. *Quant. Struct. -Act. Relat*., **2002**, *21*, 493.
[132] Raevsky, O. A., Schaper, K. -J., Artursson, P., and McFarland, J. W. *Quant. Struct. -Act. Relat*., **2002**, *21*, 402.
[133] Fecik, R. A., Frank, K. E., Gentry, E. J., Menon, S. R., Mitscher, L. A., Telikepalli, H. in Combinatorial Chemistry, John Wiley & Sons, New Jork, **1998**; pp. 149 -185.
[134] Yoshida, F., Topliss, J. G. *J. Med. Chem*., **2000**, *43*, 2575.
[135] Veber, D. F., Johnson, S. R., Cheng, H. -Y., Smith, B. R., Ward, K. W., Kopple, K. D.. *J. Med. Chem*., **2002**, *45*, 2615.