JOURNAL OF PHYSICAL ORGANIC CHEMISTRY, VOL. 10, 405–413 (1997)

QUANTIFICATION OF NON-COVALENT INTERACTIONS ON THE BASIS OF THE THERMODYNAMIC HYDROGEN BOND PARAMETERS

O. A. RAEVSKY*

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

This paper describes how empirical free energy and/or enthalpy values for hydrogen binding strength are derived from thousands of corresponding measurements between H-bond donors and acceptors, mostly in carbon tetrachloride, and how they can be used to construct common scales or factor values (increments) also for other reactions involving electron donor and acceptor ability of functions. The corresponding databases and programs (HYBOT) allow one to predict thermodynamic values for experimentally unknown equilibria, including also ionophore complexes with crown ethers or cryptands. Applications in QSAR involve the prediction of lipophilicity from any structure on the basis of only two variables, e.g. 234 systems are described this way with a correlation coefficient $r = 0{\cdot}96$. Similarly, permeabilities and some biological properties such as narcotic activities of chemicals and anti-HIV-1 activity of some porphyrins are evaluated. © 1997 by John Wiley & Sons, Ltd.

J. Phys. Org. Chem. **10**, 405–413 (1997) No. of Figures: 2 No. of Tables: 3 No. of References: 46

Keywords: non-covalent interactions; thermodynamic hydrogen bond parameter

Received 23 September 1996; revised 6 February 1997; accepted 26 February 1997

INTRODUCTION

The ability of chemical compounds to exert non-covalent interactions plays an essential role in different physical, chemical and biological properties. The obvious way to characterize these interactions quantitavely is the use of thermodynamic parameters. Thermodynamic criteria of competitive complexation in multicomponent solutions have been described.[1–3] Free energy (ΔG) or binding constant (K) and enthalpy (ΔH) are connected with each other by the well known equation

$$\Delta G = -RT\ln K = \Delta H - T\Delta S \qquad (1)$$

We note that ΔS (entropy of complexation) is needed for the complete thermodynamic description of complexation.

Four possible combinations of enthalpy and entropy values, resulting in different free energy values are possible:[4]

1. Enthalpy is positive and entropy is negative. In this case at all temperatures the equilibrium is always displaced towards to initial compounds and complexation does not occur.
2. Enthalpy and entropy are positive. In this case the final result is determined by the ratio of the absolute values of ΔH and $T\Delta S$. For example, in the case of the H-bond complex of trichloromethane with *N*-methylacetamide in CCl_4, $T\Delta S$ and ΔH are positive (these and the following thermodynamic data were taken from the database HYBOT; see below). However, the absolute value of enthalpy ($\Delta H = 1{\cdot}70$ kJ mol^{-1}) is much higher than $T\Delta S$ ($0{\cdot}03$ kJ mol^{-1}), so the free energy of complexation also has a positive sign ($\Delta G = 1{\cdot}67$ kJ mol^{-1}) and the binding constant is very small.
3. Enthalpy and entropy are negative. In this case essential complexation is present when $|\Delta H| \gg |T\Delta S|$. For example, in the case of the H-bond complex of 3-chlorophenol with *N*,*N*-diethylacetamide in CCl_4, $\Delta H = -26{\cdot}90$ kJ mol^{-1} and $T\Delta S = -6{\cdot}54$ kJ mol^{-1}. Hence $\Delta G = -20{\cdot}36$ kJ mol^{-1}, with a then rather large binding constant.
4. Enthalpy is negative and entropy is positive. In this case the complexation is larger than in cases 1–3 (of course with the same absolute values). For example, in the case of H-binding of tri-*N*-butylammonium chloride with triphenylphosphine oxide in 1,2-$Cl_2C_6H_4$, $\Delta H = -27{\cdot}40$ kJH mol^{-1} and $T\Delta S = 4{\cdot}25$ kJ mol^{-1}. So here $\Delta G = -31{\cdot}65$ kJ mol^{-1}, and the binding constant is much higher than the previous cases, although the absolute values of enthalpy and entropy are approx-

* *Correspondence to:* O. A. Raevsky

Contract grant sponsor: Volkswagen-stiftung; *Contract grant number:* I/68 951.

Contract grant sponsor: NATO; *Contract grant number:* N 931639.

Contract grant sponsor: INTAS; *Contract grant number:* 94-1914.

CCC 0894–3230/97/050405–09 $17.50

imately the same.

It should be noted that in many investigations of complexations, entropy values are either neglected completely or considered to be proportional to enthalpy. However, in a number of cases significant contributions of entropy have been established. This relates in particular to hydrophobic interactions, which play a significant role in biological systems.

An example of the essential role of entropy in complexation of crowns with calcium salts has been described.[5] The significant influence of the relationship between enthalpy and entropy on the binding of ligands with receptors has been discussed[6] and it was shown that binding of the agonists (L-isoprenalol, L-noradrenaline, L-adrenaline) with the adrenergic receptor of turkey erythrocytes takes place with a large negative entropy, while antagonists (metroprolol, practolol, pindolol) gave a relatively small enthalpy but positive entropy. For example, the complexation of L-adrenaline and practolol with the above-mentioned receptor shows approximately the same values of free energy. However, in the first case binding occurs mainly because of a negative enthalpy, whereas in the second case it occurs as a result of a positive entropy.

Quantitative estimation of non-covalent interaction contributions to the properties of many compounds were carried out in the Laboratory of Computer-Aided Molecular Design (LCAMD) of the Russian Academy of Sciences. There are three lines of this research: thermodynamics of hydrogen bonding, investigation of the structure and properties of complexes of macrocyclic compounds with different cations and quantitative structure–activity relationships (QSAR).

H-BOND STRENGTH SCALES

Interactions between H-bond donor and H-bond acceptor molecules are known to result in the formation of many molecular and ionic complexes which are of great importance in chemical and biochemical processes including enzymatic catalysis. H-bond complexes are especially important in biological systems where they play a crucial role in macromolecular structure (proteins, DNA and RNA) and in molecular recognition.

An obvious way to describe H-bonding quantitatively is to consider the thermodynamic functions of H-bond complex formation. The development of empirical scales for intermolecular interactions on the basis of thermodynamic parameters began in the middle-1960s.[7, 8] Concerning the H-bond enthalpy scales it is worth mentioning the approaches developed by Drago and co-workers.[9–11] Sherry and Purcel[12] and Iogansen.[13] Three pairs of bilinear parameters must be used to estimate the H-bond enthalpy in the case of the approach of Drago and co-workers, but only two bilinear parameters in the case of the approaches of Sherry and Purcel and Iogansen: one for the H-bond donor and one for the H-bond acceptor. Further, in this case it is possible to compare quantitatively the H-bond strengths among donor and acceptor sets. This is why we believe the H-bond scales based on a bilinear approach to be more useful than scales based on the approach of Drago and co-workers. In the 1980s Raevsky[14, 15] developed the bilinear approach of H-bond enthalpy description and proposed the following equation:

$$\Delta H = kE_dE_a \tag{2}$$

where E_d and E_a are H-bond donor and H-bond acceptor enthalpy factors which characterize the relative H-bond donor and acceptor ability of compounds. In the framework of equation (2), E_a values have to be positive and E_d values have to be negative because a positive sign was attributed to the standard H-bond acceptor hexamethylphosphoramide ($E_a = 2{\cdot}50$). In Refs 16–18 E_a values were given for 273 phosphoryl compounds.

A similar equation was proposed by Raevsky and co-workers also for free energy:[19, 20]

$$\Delta G = kC_dC_a + k_0 \tag{3}$$

where C_d and C_a are free energy H-bond donor and H-bond acceptor factors which characterize the relative H-bond donor and acceptor binding ability of compounds. A value of $C_a = 4{\cdot}00$ was selected for the standard H-bond acceptor hexamethylphosphoramide and a value of $C_d = -2{\cdot}50$ was selected for the standard H-bond donor phenol.

Taft *et al.*[8] were the first who proposed a quantitative scale of hydrogen bond strength of H-bond acceptors on the basis of log$K$ values of complexation with 4-fluorophenol in tetrachloromethane. This approach has been developed since the beginning of the 1970s,[21] and the corresponding equation can be presented in the following form,[22, 23] which differs from equation (3) only in term $2{\cdot}302\,RT$:

$$\log K = k\alpha_H\beta_{2H} + k' \tag{4}$$

where α_H and β_{2H} characterize the H-bond donor and H-bond acceptor binding ability, respectively.

The development and successful application of the above-mentioned empirical approaches are possibly only on the basis of large amounts of experimental data and the ability to calculate factor values for new compounds. Our own data and many published values on thermodynamic parameters of H-bond complexes resulted in the program package HYBOT (HYdrogen BOnd thermodynamics). HYBOT-96 contains as a complete set two modules: the program HYBOT. DataBase and the program HYBOT.Factor.

Hybot.DataBase is an interactive, menu-driven program, which runs under the Microsoft Windows 3.1. The database consists of a number of data structures, named as different files. A user can create a new file or delete an existing one by selection of the appropriate commands from the menu. Each file can contain 5000 records about the structure of H-bond partners and their names, empirical formula, molecular weights, Chemical Abstracts registry numbers, enthalpy and(or) free energy values (kJ mol^{-1}), method of determination type of H-bond complex, number of H-bonds, solvent, temperature, references and comments. Records

may be copied from one file to another or duplicated within a file via the clipboard. The user can create, delete and modify records. At present there are 12 004 entries in the database, which contains information about enthalpy and/or free energy of H-bonding between 650 H-bond donors and 2250 H-bond acceptors in 64 solvents and the gas phase.

The above-mentioned information can be used for the estimation of free energy and enthalpy factor values of different compounds or functions in the framework of common scales. The construction of such scales has been discussed in detail. One needs to emphasize the following: equation (3) differs from equation (2) by the intercept value. As already mentioned, a positive sign was attributed to bond acceptor factors and a negative sign to H-bond donors. Hence the positive values of enthalpy and free energy may be obtained only by the use of positive intercepts in the corresponding equations. The appearance of an intercept is obviously necessary and the only question is how large it should be. There are only two H-bond complexes with positive enthalpy amongst 12 004 in the HYBOT.DataBase. At the same time there are 1474 complexes (12%) in the HYBOT.DataBase where the free energy is positive. That is why the presence of the intercept in equation (3) is indispensable. We decided to fix the intercept values at the level of 5·70 kJ mol^{-1} because it corresponds to a constant value $K=0{\cdot}1$ that is close to the limit of experimental detectability, and because in the case of $K=0{\cdot}1$ it is very difficult to consider such a weak interaction as a 'true' hydrogen bond. We have therefore truncated our scale by the value of 5·7 kJ mol^{-1}, and accepted the assumption that the free energy of complex formation of compounds with zero *C*-factor will have a constant value of 5·70 kJ mol^{-1}. It must also be mentioned that the intercept value 5·7 kJ mol^{-1} in equation (3) corresponds to $\log K=-1{\cdot}1$ [intercept values in equation (4)[22]].

The above-mentioned difference of equations (2) and (3) make obvious the necessity for separate enthalpy and free energy H-bond scales. Nevertheless, correlation of the estimated *C* and *E* factor values have been carried out.[20]

The next correlation was obtained for H-bond donors:

$$C_d=-0{\cdot}52(\pm0{\cdot}13)+0{\cdot}75(\pm0{\cdot}06)E_d \qquad (5)$$
$$n=163,\ r=0{\cdot}89,\ s=0{\cdot}38,\ F=620$$

The equation for H-bond acceptors has much worse statistical criteria:

$$C_a=-0{\cdot}36(\pm0{\cdot}23)+0{\cdot}61(\pm0{\cdot}10)E_a \qquad (6)$$
$$n=195,\ r=0{\cdot}66,\ s=0{\cdot}53,\ F=152$$

Such reltionships are poor for any class of compounds. For example, for compounds containing N-H donor groups:

$$C_d=-0{\cdot}24(\pm0{\cdot}67)+0{\cdot}92(\pm0{\cdot}36)E_d \qquad (7)$$
$$n=27,\ r=0{\cdot}73,\ s=0{\cdot}55,\ F=28$$

and for compounds containing nitrogen acceptor atoms:

$$C_a=-0{\cdot}97(\pm0{\cdot}39)+0{\cdot}65(\pm0{\cdot}18)E_a \qquad (8)$$
$$n=60,\ r=0{\cdot}70,\ s=0{\cdot}62,\ F=54$$

We also mention the possibility of the construction of a scale of 'basicity-dependent properties' depending on related values for relevant compounds.[24] We estimated[20] that for the few such systems the calculated values give positive deviations from experimental values. This means that some classes of complexes including some biologically significant systems should be considered carefully in the framework of a single parameter model.

The HYBOT.Factor program represents a database of E_d, C_d and α_H values for 414 H-bond donors, and E_a, C_a and β_{2H} for 1298 H-bond acceptors that have been calculated from 6386 equations. HYBOT.Factor also contains a module for prediction of factor values for new compounds.

The majority of the measurements (7793 in the current version of HYBOT.DataBase) are based on measurements in tetrachloromethane (CCl_4). This solvent is non-polar, and is thus suitable for the estimation of H-bond parameters; 414 H-bond donors and 1298 H-bond acceptors that had data from the interaction of more than one partner were selected for the estimation of C_a and C_d values. Matrices containing 414 columns and 1298 rows were prepared. The computation algorithm includes the following steps:

1. selection of H-bond acceptors which form complexes with phenol;
2. selection of H-bond donors which form complexes with these acceptors; and
3. constructing and solving equation systems for each pair of complexes, formed by phenol and a given H-bond donor with the same acceptor.

The acceptor factors were calculated on the basis of the same algorithm using hexamethylphosphoramide as the standard H-bond acceptor.

The free energy H-bonding factor scale together with factor values for few known chemicals is presented in Figure 1. It is very important to emphasize that intervals of free energy H-bond parameter values of different functional groups are overlapping (see Table 1). This means that the role of substituents is more important in the formation of H-bond donor and acceptor properties compared with the type of atom. Comparison of experimental values of enthalpy and free energy with calculated values on the basis of equations (2) and (3) demonstrates the adequacy of the multiplicative approach for quantitative description of H-bond thermodynamics:

$$\Delta H_{calc}=-0{\cdot}49(\pm0{\cdot}29)+0{\cdot}99(\pm0{\cdot}08)\Delta H_{exp} \qquad (9)$$
$$n=2787,\ r=0{\cdot}970,\ s=2{\cdot}40,\ F=44\ 350$$

$$\Delta G_{calc}=-0{\cdot}07(\pm0{\cdot}12)+1{\cdot}04(\pm0{\cdot}05)\Delta G_{exp} \qquad (10)$$
$$n=3301,\ r=0{\cdot}991,\ s=1{\cdot}12,\ F=175\ 000$$

Knowing C_d values for 414 H-bond donors and C_a values for 1298 H-bond acceptors, one can now estimate ΔG values for 537 372 H-bond complexes in tetrachlormethane (in Ref. 25 α_H values for 150 H-bond donors are given, as

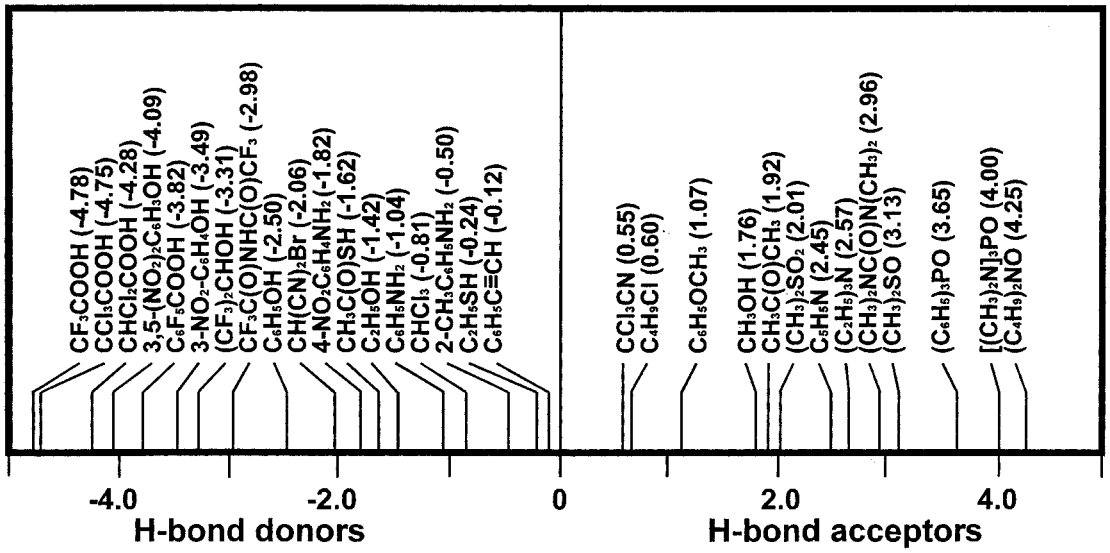

Figure 1. Free energy H-bond factor scale

well as β_H values for 500 H-bond acceptors, allowing one to estimate log K values for 75 000 H-bond complexes).

HYBOT.Factor also has a special module for factor value calculations for new compounds by searching the nearest neighbour in the above-mentioned factor value database. The algorithm for such calculations is described in Ref. 26.

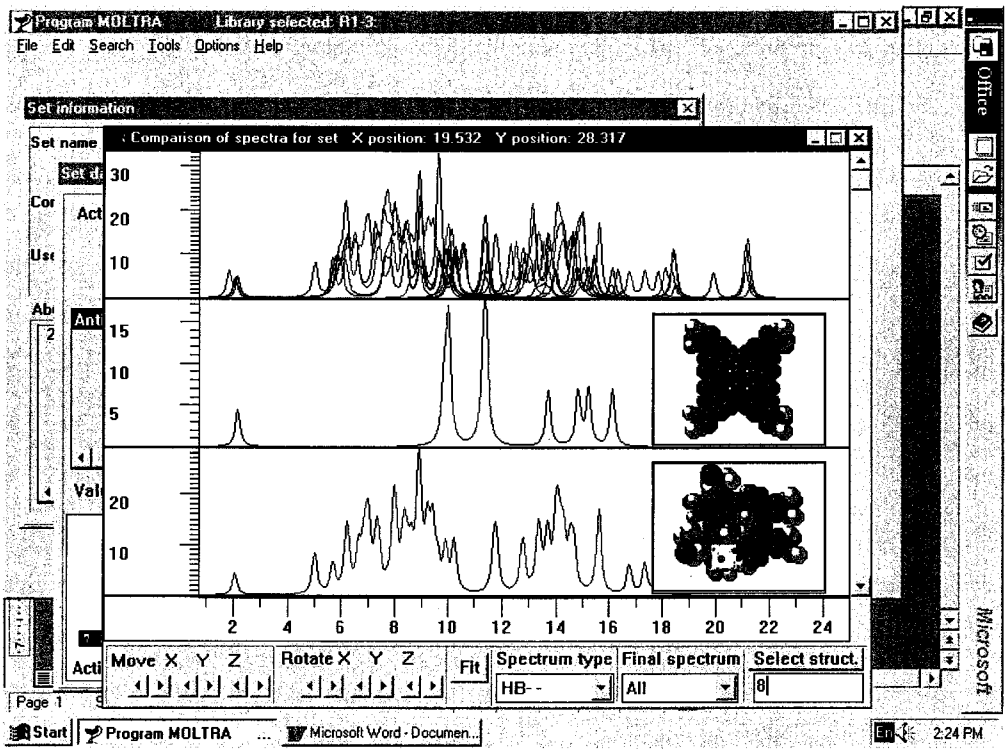

Figure 2. Spectra of interatomic distances of H-bond donors for porphyrins

Table 1. Factor value intervals of some H-bond donor and acceptor groups

H-bond donor	C_d		H-bond acceptor	C_a	
	Min.	Max.		Min.	Max.
$R_1R_2R_3C–H$	−2·06	−0·12	R–O–R	0·78	2·01
RS–H	−1·92	−0·24	R–S–R	0·57	1·51
$R_1R_2N–H$	−3·32	−0·50	$R_1R_2R_3N$	0·58	3·65
PhO–H	−4·24	−0·76	RC≡N	0·55	2·77
AlkO–H	−3·50	−1·06	$R_1R_2R_3P=O$	2·40	4·02
RC(O)O—H	−4·78	−2·50	$R_1R_2S=O$	1·16	3·25

USE OF H-BOND DESCRIPTORS FOR THE QUANTIFICATION OF COMPLEXATION STRENGTH OF MACROCYCLIC COMPOUNDS WITH DIFFERENT IONS

The second field of use of H-bond parameters in our laboratory is in the quantitative description of the thermodynamics of the complexation of macrocyclic compounds with different ions. Many experimental measurements of complexation by means of calorimetry, potentiometry and IR and x-ray spectroscopy are being carried out (see reviews[3, 5, 27]). The possibility of the use of bilinear parameters for the quantitative description of interactions of different ions as with neutral molecules and with each other was estimation for the first time in 1984.[28] In this case we used the term electron donor parameter to characterize the ability of anions and neutral compounds containing heteroatoms to interact with cations and the term electron acceptor parameter to characterize the ability of cations or neutral molecules containing X-H groups to interact with anions or neutral molecules containing heteroatoms. So for neutral molecules which contain heteroatoms there are two terms: electron donor parameter, which is used in the case of its interactions with cations, and H-bond acceptor parameter, which is used in the case of its interactions with neutral H-bond donors.

In previous work,[29, 30] the sum of H-bond acceptor parameters values ((ΣC_a)) of macrocycle fragments which were estimated in the non-polar solvent CCl_4 by the program HYBOT.Factor were used for the quantitative description of the free energy of complexation of macrocyclic compounds with different cations in polar solvent including water (ΣC_a values were calculated as the sum of C_a values of all H-bond acceptor atoms of macrocycles and C_a values for each H-bond acceptor atom were calculated by means of program HYBOT on the basis of the nearest neighbours in the database of HYBOT.Factor). It was shown that H-bond acceptor factors may be used successfully for the estimation of the free energy of the above-mentioned polar interactions as long as there is a sufficient geometric fit between hosts and guests.

The corresponding equations are presented below: for the complexation of KCl with substituted (R)-[18-crown-6] and (R)-[2.2.2]cryptands in methanol [for equations (11)–(16) all available data for corresponding ligands and solvents were used]:

$$\Delta G = 1{\cdot}30 + 3{\cdot}25\Sigma C_a \qquad n = 20,\ r = 0{\cdot}993,\ s = 1{\cdot}37,\ F = 1192 \tag{11}$$

Complexation of KX with (*R*)-[18-crown-6] and (*R*)-[2,2,2]cryptands in water (all available data for this ligand in this solvent):

$$\Delta G = -11{\cdot}02 + 2{\cdot}43\Sigma C_a \qquad n = 6,\ r = 0{\cdot}994,\ s = 1{\cdot}04,\ F = 319 \tag{12}$$

Complexation of NaX with (*R*)-[15-crown-5] and (*R*)-[2.2.1]cryptands in methanol:

$$\Delta G = -17{\cdot}43 + 4{\cdot}64\Sigma C_a \qquad n = 9,\ r = 0{\cdot}993,\ s = 1{\cdot}61,\ F = 494 \tag{13}$$

Complexation of NH_4^+ with different macrocycles in $CDCl_3$:

$$\Delta G = -17{\cdot}83 + 6{\cdot}87\Sigma C_a \qquad n = 7,\ r = 0{\cdot}910,\ sd = 4{\cdot}35,\ F = 24{\cdot}0 \tag{14}$$

Complexation of NH_4^+ with different macrocycles in D_2O:

$$\Delta G = 3{\cdot}80 + 4{\cdot}33\Sigma C_a \qquad n = 8,\ r = 0{\cdot}910,\ s = 8{\cdot}00,\ F = 27{\cdot}8 \tag{15}$$

Complexation of *t*-$BuNH_3^+$ with different macrocycles in $CHCl_3$:

$$\Delta G = -3{\cdot}92 + 3{\cdot}78\Sigma C_a \qquad n = 8,\ r = 0{\cdot}943,\ s = 2{\cdot}54,\ F = 48{\cdot}4 \tag{16}$$

Finally the question of the influence of cations on complexation Gibbs energy was addressed. From the free energy correlation with the sum of H-bond acceptor parameter values one can derive electron acceptor parameter values characterizing the electron acceptor strength of the different cations on the basis of equation (3). The thus

calculated electron acceptor parameter values for sodium, potassium, rubidium and cesium cations showed a good correlation with its hydration energy.[29]

The above-mentioned equations and their good statistical criteria demonstrate that the H-bond acceptor parameter and electron donor parameter have the same physical meaning and may be used for a quantitative description of interactions of neutral molecules between each other, interactions of neutral molecules with ions and interactions of ions between each other.

APPLICATION OF BILINEAR H-BOND PARAMETERS AS QSAR DESCRIPTORS

Since the beginning of the 1990s different H-bond descriptors based on experimental data or theoretical procedures have been proposed for the use in QSAR. We mention the overall H-bond descriptors of Abraham[25] and similar descriptors of Zhang *et al*.,[31] which are based on experimental data, and also theoretical descriptors of Famini and co-workers.[32–35] In the latter approach H-bond acidity is represented by a covalent term, e_a, the magnitude of the difference between the lowest unoccupied molecular orbital (LUMO) of the solute and the highest occupied molecular orbital (HOMO) of water, and an electrostatic term, q^+ representing the largest positive formal charge on a hydrogen atom in the solute. The H-bond basicity is defined by a covalent term, e_b (the difference between the HOMO of the solute and the LUMO of water), and an electrostatic term, q^- (the most negative formal charge in the solute molecule).

Twelve H-bond descriptors based on enthalpy and free energy parameters were proposed earlier for use in QSAR:[36] free energy and enthalpy parameter values for the strongest H-bond donor atom in the molecule ($C_{d\,max}$ and $E_{a\,max}$), the sum of C_d and E_d values for all H-bond donor atoms in the molecule ($\sum C_d$ and $\sum_d$), the sum of C_a and E_a values for all H-bond acceptor atoms in the molecule (($\sum C_a$ and $\sum E_a$) and also $\sum C_d$/MW (MW = molecular weight), $\sum E_d$/MW, $\sum C_a$/MW and $\sum E_a$/MW. Factor values are estimated from HYBOT.Factor database. The factor value of the nearest neighbour is used in the case of the absence of the studied compound in this database. The successful application of these H-bond descriptors has been demonstrated.[36–39] New examples of its applications are presented below.

Quantitative description of lipophilicity of compounds

It has been proposed[40, 41] that partition coefficients encode two major structural contributions: a volume-related term (describing steric bulk effects) and a term which reflects interactions such as dipole–dipole and hydrogen bonding. The results of our first investigation of the contribution of H-bond ability in octanol–water partition coefficient (log*P*) of carbonyl and hydroxyl compounds were published recently.[42] Molecular volume (*MV*) was used in that work to account for the volume-related term. Further work with other classes of compounds showed that in the case of the use of *MV* and $\sum C_a$ there are non-zero values for the intercepts in equations correlating log*P*. For this reason, it was decided to test different volume-related terms together with H-bond descriptors for the training set containing different types of compounds (nitriles, amines, carbonyl compounds, ethers, esters, alcohols, phenols, compounds with phenyl, nitro and halogen groups, acids). Five volume-related terms [molecular weight (*MW*), surface area (*SA*), molar volume (*MV*), molecular refractivity (*MR*) and polarizability (*Pol*)] were used together with $C_{a\,max}$, ΣC_a, ΣC_a/MW, $C_{d\,max}$ $\sum C_d$ and $\sum C_d/MW$. Only one volume-related term among those mentioned above could be used because of intercorrelations amongst the others. The best result was obtained using polarizability and H-bond acceptor strength:

$$\begin{aligned} \log P = &-0{\cdot}01(\pm 0{\cdot}08) + 0{\cdot}249(\pm 0{\cdot}005)Pol \\ &- 0{\cdot}68(\pm 0{\cdot}02)\Sigma C_a \\ &n = 234,\ r = 0{\cdot}961, \\ &s = 0{\cdot}34,\ F = 1383,\ SDEP = 0{\cdot}35 \end{aligned} \tag{17}$$

The intercept is non-significant. We note that including an H-bond donor descriptor in equation (19) did not improve the relationship in spite of the presence of 80 H-bond donors in the training set.

Abraham and Chadna[43] published results of the application of the solvation equation for log*P* in the following form:

$$\begin{aligned} \log P = &(0{\cdot}09 \pm 0{\cdot}02) + (0{\cdot}56 \pm 0{\cdot}01)r_2 - (1{\cdot}05 \pm 0{\cdot}02)\pi_{2H} \\ &+ (0{\cdot}03 \pm 0{\cdot}02)\sum\alpha_{2H} - (3{\cdot}46 \pm 0{\cdot}03)\sum\beta_{2H} \\ &+ (3{\cdot}81 \pm 0{\cdot}01)V_x \\ &n = 613,\ r^2 = 0{\cdot}995,\ s = 0{\cdot}116, \\ &F = 23\,162 \end{aligned} \tag{18}$$

where r_2 is the excess molar refraction, π_{2H} is the solute dipolarity/polarizability, $\sum\alpha_{2H}$ is the solute overall or effective hydrogen bond acidity, $\sum\beta_{2H}$ is the solute overall or effective hydrogen bond basicity and V_x is the McGowan characteristic volume. In view of the many terms used, the statistical parameters of equation (18) are expected to be better compared with equation (17). However, the H-bond parameters of equation (18) were calculated from the same equation.[43] Also for that reason it is no surprise that the standard deviation of equation (18) has a smaller value compared even with the usual error of experimental determination of log*P*. Equation (17) in contrast to equation (18) has only two parameters of clear physical significance. Also, the standard deviation of equation (17) is at the level of the experimental error, demonstrating that the use of two parameters (volume-related term and H-bond acceptor term) is sufficient for the quantitative description of log*P*.

It is obvious also from equation (17) that the polarizability and H-bond acceptor strength have opposite contributions to lipophilicity. The values for polarizability

Table 2. Estimation of log*P* from polarizability and H-bonding factor

			Contribution to log*P*		Log*P*	
Compound	ΣC_a	*Pol*	ΣC_a	*Pol*	Calc.	Exp.
Chloroacetamide	3·80	8·00	−2·57	1·99	−0·59	−0·53
Acetaldoxime	2·64	5·59	−1·79	1·39	−0·41	−0·12
Ethanolamine	4·16	6·43	−2·82	1·60	−1·23	−1·30
Nicotinamide	5·88	13·19	−3·98	3·28	−0·71	−0·37
Hydroquinone	3·00	11·32	−2·04	2·82	0·77	0·59
Dimethyl disulphide	1·50	10·44	−1·02	2·60	1·57	1·77

and H-bond acceptor ability in log *P* for some polyfunctional compounds calculated with equation (17) are presented in Table 2. As a consequence of the predominant H-bond acceptor strength compared with polarizability, the first four compounds in Table 2 have negative log*P* values. In the other two compounds the contributions of polarizability are larger than the H-bond acceptor strength influence. In consequence, these compounds have positive log*P* values.

Solubility of compounds in water

A similar approach was used for the quantitative description of the solubility of liquid neutral polar organic compounds in water. The results are presented in the following equation:

$$\log 1/S = -0{\cdot}42(\pm 0{\cdot}20) + 0{\cdot}17(\pm 0{\cdot}11)Pol - 0{\cdot}13(\pm 0{\cdot}04)\Sigma C_a + 0{\cdot}08(\pm 0{\cdot}06)\Sigma C_d \quad (19)$$

$$n = 45,\ r = 0{\cdot}925,\ s = 0{\cdot}42,\ F = 81{\cdot}4,\ SDEP = 0{\cdot}46$$

QUANTITATIVE H-BOND ABILITY–PERMEABILITY RELATIONSHIP STUDY

A number of physicochemical descriptors, such as log *P*, the difference between log*P* for octanol–water and cyclohexane–water system (ΔlogP), *MW* and water-accessible surface area (*ASA*), are commonly used for the estimation of permeability properties,[45] assuming that ΔlogP assesses hydrogen bond capacity. We used directly H-bond descriptors for QSAR analyses of permeability. Table 3 contains experimental human skin permeation data (log K_p) taken from Ref. 44, polarizability and H-bond descriptors and also calculated logK_p values (joint research with Dr K.-J. Schaper, Borstel Research Institute, Germany) for 17 phenols. The only H-bond donor factor values give a fairly stable correlation with permeability for those compounds:

$$\log K_p = -3{\cdot}39(\pm 0{\cdot}59) + 0{\cdot}71(\pm 0{\cdot}21)\Sigma C_d \quad (20)$$

Table 3. Permeability data, polarizability and H-bond ability of phenols

Compound	*Pol*	ΣC_a	ΣC_d	LogK_p*	LogK_p [calc., equation (21)]
Phenol	10·71	1·37	−2·49	−5·64	−5·36
Resorcinol	10·98	2·36	−5·00	−7·18	−6·92
p-Nitrophenol	12·55	2·43	−3·65	−5·81	−5·98
m-Nitrophenol	12·55	3·26	−3·49	−5·81	−5·88
Methyl hydroxybenzoate	14·81	3·54	−2·73	−5·60	−5·26
m-Cresol	12·54	1·47	−2·44	−5·37	−5·22
o-Cresol	12·54	1·61	−2·27	−5·36	−5·11
p-Cresol	12·54	1·31	−2·44	−5·31	−5·22
2-Naphthol	17·98	1·99	−2·64	−5·11	−5·02
o-Chlorophenol	12·64	1·92	−2·18	−5·04	−5·05
p-Ethylphenol	14·38	1·32	−2·44	−5·01	−5·11
3,4-Xylenol	14·38	1·52	−2·28	−5·00	−5·01
p-Bromophenol	13·33	1·55	−2·83	−5·00	−5·42
p-Chlorophenol	12·64	1·61	−2·91	−5·00	−5·51
Thymol	18·05	1·60	−2·01	−4·83	−4·62
2,4,6-Trichlorophenol	16·49	2·61	−2·21	−4·78	−4·84
2,4-Dichlorophenol	14·56	2·27	−2·48	−4·78	−5·12

$$n=17,\ r=0{\cdot}883,\ s=0{\cdot}28,\ F=53{\cdot}1,\ Rcv=0{\cdot}814$$

Addition of polarizability parameters *Pol* slightly improves the correlation:

$$\log K_p = -4{\cdot}44((\pm 1{\cdot}43)+0{\cdot}63(\pm 0{\cdot}22)\sum C_d + 0{\cdot}06(\pm 0{\cdot}08)Pol \quad (21)$$

$$n=17,\ r=0{\cdot}904,\ s=0{\cdot}27,\ F=31{\cdot}3,\ Rcv=0{\cdot}817$$

QUANTITATIVE H-BOND ABILITY–BIOLOGICAL ACTIVITY RELATIONSHIP STUDIES

Because H-bond descriptors reflect an important part of the ability of compounds to form intermolecular complexes, it is possible to use them in almost any investigation of structure–biological activity relationships. The high statistical significance of equation (19) allows one to estimate quantitative relationships between activity and H-bond strength in the cases where the important role of lipophilicity is already established. For example, the well known dependence of tadpole narcosis (log C)[45] on log P for simple chemicals may be presented by an equation containing two parameters (polarizability and H-bond acceptor ability):

$$\log 1/C = 0{\cdot}49(\pm 0{\cdot}20)+0{\cdot}23(\pm 0{\cdot}02)Pol - 0{\cdot}42(\pm 0{\cdot}05)\sum C_a \quad (22)$$

$$n=85,\ r=0{\cdot}954,\ s=0{\cdot}33,\ F=413{\cdot}1,\ Rcv=0{\cdot}950$$

In the description of tadpole narcosis on the basis of Wilson and Famini's approach[34] the equation contains four parameters:

$$\log C = 7{\cdot}46(\pm 4{\cdot}65) - 2{\cdot}16(\pm 12{\cdot}00)V_{mc}/100 - 42{\cdot}00(\pm 5{\cdot}76)\pi_1 - 25{\cdot}2(\pm 4{\cdot}36)e_b + 4{\cdot}11(\pm 6{\cdot}27)q \quad (23)$$

$$n=41,\ r=0{\cdot}970,\ s=0{\cdot}29,\ F=141$$

However, the interaction of molecules with their biological receptors requires in almost all cases that the three-dimensional (3D) structure is taken into account. That is why 3D H-bond descriptors for QSAR and molecular modelling are necessary. We have proposed[46] a 3D QSAR description of structure by means of spectra of interatomic distances. The spectra of interatomic distances of H-bond donor atoms in some porphyrins which have essential anti-HIV-1 activity (K_i) are presented in Figure 2. The spectra were obtained by means of the program MOLTRA (MOLecular Transform Analysis) which uses HYBOT for the calculation of 2D H-bond descriptors; and spectra of interactions between donor atoms, spectra of interactions between acceptor atoms and spectra of interactions of donor and acceptor atoms. In principle, each point of such a spectrum can be used as a 3D H-bond descriptor. In the discussed training set the distances between external H-bond donor substituents were essentially different. Thus the 3D H-bond descriptor connected with the interaction of such H-bond donor groups (HB- − 11·4A) together with the quantum-chemical descriptor (LUMO, energy of the lowest unoccupied molecular orbital) essentially contributed in activity of the compounds (joint investigation with Dr A. N. Razdolsky, LCAMD):

$$\log K_i = -6{\cdot}56(\pm 1{\cdot}44) - 2{\cdot}24(\pm 0{\cdot}11)\text{LUMO} + 0{\cdot}05(\pm 0{\cdot}03)\text{HB-} - 11{\cdot}4A \quad (24)$$

$$n=17,\ r=0{\cdot}904,\ s=0{\cdot}27,\ F=31{\cdot}3,\ Rcv=0{\cdot}817$$

ACKNOWLEDGMENTS

A great deal of the work presented here was carried out in our own laboratory. It is a pleasure to acknowledge the contributions made by Dr V. Grigorev (H-bond and factor databases), Dr V. P. Solovev (calorimetry), Dr S. Trepalin (shell of HYBOT), I. Kitova and Drs A. Razdolsky and N. Lukojanov (QSAR application of 3D descriptors). I am also very grateful to Professor Dr H.-J. Schneider (Universität des Saarlandes) for cooperation and to the Volkswagen-Stiftung (I/68 951), NATO (N 931639) and INTAS (94-1914) for financial support.

REFERENCES

1. Z. Simon, *Angew. Chem. Int. Ed. Engl.* **13**, 719–732 (1974).
2. R. C. Hayward, *Chem. Soc. Rev.* **121**, 285–308 (1985).
3. O. A. Raevsky. *Russ. Chem. Res.* (Engl. Ed.) **59**, 219–233 (1990).
4. O. A. Raevsky, *Teor. Eksp. Khim.* **22**, 450–463 (1986); *Chem. Abstr.* **105**, 221095*k* (1986).
5. O. A. Raevsky, *Koord. Khim.* **16**, 723–748 (1990).
6. G. A. Weiland, K. P. Minneman and P. B. Molinoff, *Nature* (*London*) **281**, 114–119 (1979).
7. V. Gutman, *Coordination Chemistry in Non-Aqueous Solutions*. New York, Plenum Press (1968).
8. R. W. Taft, D. Gurka, L. Loris, P. R. Schleyer and J. W. Rakshys, *J. Am. Chem. Soc.* **91**, 4801–4806 (1969).
9. R. S. Drago and B. B. Wayland, *J. Am. Chem. Soc.* **87**, 3571–3582 (1965).
10. R. S. Drago, *Structure and Bonding*. Springer, Heidelberg (1973).
11. M. K. Kroger and R. S. Drago, *J. Am. Chem. Soc.* **103**, 3250–3263 (1981).
12. A. Sherry and K. Purcel, *J. Am. Chem. Soc.* **94**, 1848–1853 (1972).
13. A. Iogansen, *Teor. Eksp. Khim.* **7**, 302–311 (1971); *Chem. Abstr.* **75**, 101848m (1971).
14. O. A. Raevsky and V. P. Novikov. *Khim. Farm. Zh.* **16**, 583–586 (1982); *Chem. Abstr.* (1982) **97**, 33128d.
15. O. A. Raevsky, V. V. Avidon and V. P. Novikov. *Khim. Farm. Zh.* **16**, 968–971 (1982), *Chem. Abstr.* **97**, 181323d (1982).
16. O. A. Raevsky, M. M. Giljazov and Ja. A. Levin, *Z. Obshch. Khim.* **53**, 563–567 (1983); *Chem. Abstr.* **99**, 5712n (1983).
17. O. A. Raevsky, M. M. Gilajazov and Ja. A. Levin, *Zh. Obshch. Khim.* **53**, 1720–1723 (1983); *Chem. Abstr.* **100**, 6676e (1988).
18. O. A. Raevsky, M. M. Giljazov and Ja. A. Levin, *Zh. Obshch. Khim.* **53**, 1724–1728; *Chem. Abstr.* **100**, 6675d (1984).
19. O. A. Raevsky, V. Ju. Grigor'ev and V. P. Solov'ev, *Zh. Oshch.*

Khim. **23**, 1294–1300 (1989); *Chem*. *Abstr*. **112**, 115266g (1990).
20. O. A. Raevsky, V. Ju. Grigor'ev, D. B. Kireev and N. S. Zefirov, *Quant*. *Struct*.–*Act*. *Relat*. **11**, 49–63 (1992).
21. J.-L. M. Abboud and L. Bellon, *Ann*. *Chim*. **5**, 63–74 (1970); J.-L. M. Abboud et al., in *Quantitative Treatments of Solute/Solvent Interactions*. *Theoretical and Computational Chemistry*, *1*, edited by P. Politzer and J. S. Murray, pp. 135–182. Elsevier, Amsterdam (1994).
22. M. H. Abraham, P. L. Grellier, D. V. Prior, R. W. Taft, J. J. Morris, P. J. Taylor, C. Lourence, M. Berthelot, R. M. Doherty, M. J. Kamlet, J.-L. M. Abboud, K. Sraidi and G. Guiheneuf, *J. Am. Chem. Soc*. **110**, 8534–8536 (1988).
23. M. H. Abraham, *Chem. Soc*. *Rev.* **22**, 73–82 (1993).
24. P. Maria, J. Gal, J. Franceschi and E. Fargin, *J. Am. Chem. Soc.* **109**, 483–492 (1987).
25. M. H. Abraham, in *Quantitative Treatments of Solute/Solvent Interactions*. *Theoretical and Computational Chemistry*, **1**, edited by P. Politzer and J. S. Murray, pp. 83–134. Elsevier, Amsterdam (1994).
26. S. Trepalin, A. Yarkov, L. Dolmatova and N. S. Zefirov, *J*. *Chem*. *Inf*. *Comput*. *Sci*. **35**, 405–410 (1995).
27. O. A. Raevsky and H.-J. Schneider. *Zh*. *Org*. *Khim*., **31**, 1793–1801 (1995).
28. O. A. Raevsky, V. Ju. Grigor'ev and V. P. Solovev, *Khim*. *Farm*. *Zh*. **18**, 578–582 (1984); *Chem*. *Abstr*. **103**, 83807n (1985).
29. H.-J. Schneider, V. Rüdiger and O. Raevsky, *J*. *Org*. *Chem*. **58**, 3648–4853 (1993).
30. H.-J. Schneider, T. Blatter, A. Eliseev, V. Rudiger and O. A. Raevsky, *Pure Appl*. *Chem.* **65**, 2329–2334 (1993).
31. J. Li, Y. Zhang, A. J. Dallas and P. W. Carr, *J*. .*Chromatogr*. **550**, 101–134 (1991).
32. G. R. Famini, R. J. Kassel, J. W. King and L. Y. Wilson, *Quant. Struct.–Act. Relat.* **10**, 344–349 (1991).
33. G. R. Famini, W. P. Ashman, A. P. Miciewicz and L. Y. Wilson, *Quant. Struct.–Act. Relat.* **11**, 162–170 (1992).
34. L. Y. Wilson and G. R. Famini, *J*. *Med*. *Chem*. **34**, 1668–1674 (1991).
35. G. R. Famini, L. Y. Wilson and C. A. Penski, *J. Phys. Org. chem.* **5**, 395–408 (1992).
36. O. A. Raevsky, V. Ju. Grigor'ev, V. P. Solov'ev, in *QSAR Rational Approaches to the Design of Bioactive Compounds*, edited by C. Silipo and A. Vittoria, pp. 135–138. Elsevier, Amsterdam (1991).
37. O. A. Raevsky, V. Ju. Grigor'ev and E. Mednikova, in *Trends in QSAR and Molecular Modelling 92*, edited by C. G. Wermuth, pp. 116–119. ESCOM, Leiden (1993).
38. O. Raevsky, V. Ju. Grigor'ev, D. B. Kireev and N. S. Zefirov, *J. Chim. Phys. Phys.–Chem. Biol*. **89**, 1747–1753 (1992).
39. O. Raevsky, L. Dolmatova and V. Grigor'ev, in *QSAR and Molecular Modelling*: *Concepts*, *Computational Tools and Biological Applications*, edited by F. Sanz, J. Giralo and F. Manaut, pp. 241–245. Prous, Barcelona (1995).
40. H. van de Waterbeemd and B. Testa, *Adv. Drug Res.* **16**, 87–227 (1987).
41. N. El Tayar and B. Testa, in *Trends in QSAR and Molecular Modelling 92*, edited by C. G. Wermuth, pp. 101–108. ESCOM, Leiden (1993).
42. O. A. Raevsky, K.-J. Schaper and J. K. Seydel, *Quant. Struct.–Act. Relat.* **14**, 433–436 (1995).
43. M. H. Abraham, H. S. Chadha, in *Lipophilicity in Drug Action and Toxicology*, edited by V. Pliska, B. Testa and H. Waterbeemd, pp. 311–337. VCH, Weinheim (1996).
44. N. El Tayar, R.-S. Tsai, B. Testa, P.-A. Carrupt, C. Hansch and A. Leo, *J. Pharm. Sci*. **80**, 744–750 (1991).
45. C. Hansch and A. Leo, *Exploring QSAR*, American Chemical Society, pp. 202–204. Washington, DC (1995).
46. O. Raevsky, in *QSAR in Drug Design and Toxicology*, edited by D. Hadzy and B. Jerman-Blazic, pp. 31–36. Elsevier, Amsterdam (1987).

