

Review

## Multivariate analysis of hydrophobic descriptors

Stefan Dove\*

Institute of Pharmacy, University of Regensburg, D-93040 Regensburg, Germany

\*Corresponding Author: E-mail: [Stefan.Dove@chemie.uni-regensburg.de](mailto:Stefan.Dove@chemie.uni-regensburg.de); Tel.:+49-941-9434673; Fax:+49-941-9434820

Received: March 03, 2014; Revised: March 06, 2014; Published: April 01, 2014

---

### Abstract

Multivariate approaches like principal component analysis (PCA) are powerful tools to investigate hydrophobic descriptors and to discriminate between intrinsic hydrophobicity and polar contributions as hydrogen bonds and other electronic effects. PCA of log *P* values measured for 37 solutes in eight solvent-water systems and of hydrophobic octanol-water substituent constants  $\pi$  for 25 meta- and para-substituents from seven phenyl series were performed (re-analysis of previous work). In both cases, the descriptors are reproduced within experimental errors by two principal components, an intrinsic hydrophobic component and a second component accounting for differences between the systems due to electronic interactions. Underlying effects were identified by multiple linear regression analysis. Log *P* values depend on the water solubility of the solvents and hydrogen bonding capabilities of both the solute and the solvents. Results indicate different impacts of hydrogen bonds in nonpolar and polar solvent-water systems on log *P* and their dependence on isotropic and hydrated surface areas. In case of the  $\pi$ -values, the second component (loadings and scores) correlates with electronic substituent constants. More detailed analysis of the data as  $\pi$ -values of disubstituted benzenes XPhY has led to extended symmetric bilinear Hammett-type models relating interaction increments to cross products  $\pi_x \sigma_y$ ,  $\pi_y \sigma_x$  and  $\sigma_x \sigma_y$  which are mainly due to mutual effects on hydrogen-bonds with octanol.

### Keywords

Partition coefficients; hydrophobic substituent constant; principal component analysis.

---

### Introduction

Per IUPAC definition [1], hydrophobicity "is the association of nonpolar groups or molecules in an aqueous environment which arises from the tendency of water to exclude nonpolar molecules". In the strict sense, this operational definition specifies "hydrophobic effects" and "hydrophobic bonding" as intramolecular or intermolecular interactions in an aqueous phase due to attracting forces (van der Waals forces based on orientation, induction and dispersion) and structural reorganization of water adjacent to nonpolar groups. Ordered water molecules in hydration shells with enthalpically stronger H-bonds are transferred into bulk water with a lower degree of order, leading to an increase of both enthalpy and entropy (for review, see [2]). With a predominating entropy term (which is not necessarily the case if restriction of flexible solutes is taken into account) hydrophobic interactions are stronger than the attracting forces themselves. In this context, hydrophobic effects are closely related to the interacting nonpolar surface and volume fractions of groups or molecules.

Hydrophobicity is operationally discriminated from lipophilicity which "represents the affinity of a molecule or a moiety for a lipophilic environment", commonly "measured by its distribution behavior in a biphasic system" (IUPAC [1]). With this definition, experimental aspects come into play: Hydrophobic or lipophilic effects are quantified by measurement of partition coefficients  $P$  in solvent-water systems or retention indices in RP-HPLC or TLC (for reviews of methods, see [3] and articles therein). These quantities do not simply reflect "hydrophobicity" as defined above, but depend on the interactions of the complete solute molecules with both phases and on the phase transfer. I.e., not only the nonpolar surface fraction of the solute plays a role, but also polar effects like dipole-dipole interactions as well as formation and breaking of hydrogen bonds (see [4] and references therein). Different conformations and intramolecular interactions in both phases may also be of impact. Hydrophobic (or lipophilic) descriptors are therefore rather complex. Nevertheless, in a first approximation they may be regarded as consisting of a nonpolar and a polar component, making factorization an attractive tool to better understand their nature.

The introduction of hydrophobic effects and parameters into systematic quantitative structure-activity relationships (QSAR) analysis was an essential part of the pioneering work of Corwin Hansch, Albert Leo and their coworkers. On one hand, they collected or measured a huge number of partition coefficients in different solvent-water systems [5, 6], documented as linear free-energy related quantities  $\log P$ , derived the hydrophobic substituent constant  $\pi$  [7] in analogy to Hammett's electronic  $\sigma$  parameter, and developed a constructional fragment method of calculating partition coefficients [8]. On the other hand, the exploration of numerous QSAR on different levels of integration led to a substantial advancement of the theoretical background, namely of the role of hydrophobic effects in ADMET (especially in transport, distribution, membrane passage) and in protein-ligand interactions ("hydrophobic bonding").

Beginning with the work of Collander [9], the correlation of  $\log P$  values from different solvent-water systems and the decomposition of  $\log P$  into more fundamental molecular descriptors like solubility, surface fractions, polarizability and hydrogen-bond strengths have contributed to quantitative structure-property analysis (QSPR) of hydrophobic effects (for review, see [4, 10]). However, already in 1964 octanol-water partition coefficients were implemented as standard in QSAR analysis because of the similarity of *n*-octanol and lipophilic biophases [7]. QSPR with focus on octanol-water  $\log P$  were integral parts of the foundation of  $\log P$  calculation software as Rekker's fragment additivity method [11, 12], CLOGP (for review, see [10]), ACDLabs [13], several atomistic approaches and the recent surface-integral model using local properties from semiempirical MO-calculations and their integrals over the molecular surface [14]. Based on large  $\log P$  databases – in 1995, Hansch, Leo and Hoekman documented ca. 17,000 values [6] – these methods have become more and more predictive, but also rather intransparent for an ordinary user with respect to the underlying QSPR, i.e., to the "factors", the specific inter- and intramolecular interactions and forces which affect the  $\log P$  under consideration. Thus, in addition to available papers and manuals, detailed QSPR of hydrophobic descriptors may be helpful to interpret and validate calculated quantities.

At this point, multivariate analysis may come into play. The linear decomposition of correlated hydrophobic descriptors into uncorrelated "inner variables" (factors or principal components, PCs) by factor or principal component analysis (PCA) yields the underlying, "inner" data structure (dimensionality, common and specific components). Correlation of the PCs with physicochemical parameters identifies basic effects accounting for the multivariate QSPR. With this feedback, hydrophobic descriptors may be modeled as linear functions of nonpolar and polar components by comparative multiple regression analysis. In the following sections, previous multivariate approaches jointly investigating  $\log P$  values from diverse solvent-water systems and substituent constants  $\pi$  derived from different aromatic scaffolds will be reviewed, accompanied by recalculations based on more recent data if available.

## Experimental

The present study is based on principal component analysis and multiple linear regression analysis of hydrophobic descriptors. In brief, PCA factorizes correlated data from  $m$  variables (systems) and  $n$  objects (compounds) into uncorrelated PCs according to the model:

$$Y_{ij} = \sum_{k=1}^p a_{ik} t_{kj} + e_{ij}$$

where  $a_{ik}$  are system-specific PC loadings and  $t_{kj}$  compound-specific PC scores. The number of significant PCs  $p$  yields the dimensionality of the data, i.e., their recombination by the model except for an error  $e_{ij}$ . Scaling of the variables (original data, normalized with zero means, or standardized with zero means and standard deviations of one) determines whether the cross product, the covariance or the correlation matrix is diagonalized. The PCs 1 to  $p$  are calculated via successive extraction of the maximal (residual) "correlation", i.e., via the eigenvalues and eigenvectors of the matrix under consideration. Present recalculations were performed with in-house programs FAPCA and REGRE.

## Multivariate analysis of partition coefficients log P from different solvent-water systems

The correlation of log P-values from different solvent-water systems was first reported by Collander [9]. The Collander-equation

$$\log P_{\text{solvent 1}} = a_0 + a_1 \log P_{\text{solvent 2}} \quad (1)$$

is restricted to homologous series or purely nonpolar solutes and models the different contribution of nonpolar solute-solvent interactions in the two solvents. The more hydrophobic solvent 1, the higher the slope  $a_1$ . The intercepts  $a_0$  are positively correlated with the water solubility of solvent 1, i.e., hydrophilic solutes ( $\log P < 0$ ) result in higher log P if solvent 1 is more polar than solvent 2. PCAs of log P-values from such restricted series extract just one significant PC. The loadings increase with the hydrophobicity of the solvent, and the scores are strongly correlated with the nonpolar surface of the solutes.

With variable polar solute moieties, the situation becomes more complex. Then the phase transfer comes along with different contributions of electrostatic interactions and, in particular, of broken solute-water and newly formed solute-solvent hydrogen bonds. PCAs of such extended series commonly lead to two significant PCs, a "hydrophobic" and a "polar" one. All analyses known from the literature comply with this rule, namely PCAs of log P-values

- from 18 solutes in six solvent-water systems (n-octanol, diethylether, chloroform, benzene, toluene, cyclohexane) [15];
- from 28 solutes [16], 50 solutes [17], and 69 solutes [18], respectively, in six solvent-water systems (n-octanol, diethylether, chloroform, carbon tetrachloride, benzene, n-hexane)

Whereas Dove et al. [15] applied the standard PCA method, diagonalization of the correlation matrix, the results of the group of Bill Dunn [16, 17, 18] were based on diagonalizing the matrix of cross products implied in the SIMCA software pocket.

In the following section, a newly calculated PCA of log P-values from 37 solutes will be presented to exemplify common principles and results. The series is a subset of the solutes analyzed by Koehler et al. [18], the number of solvents was extended to eight (by toluene and cyclohexane). Data, either taken from ref. [18] or from the tables of Hansch and Leo [5, 6], are presented in Table 1.

**Table 1.** PCA of log P-values from different solvent-water systems: data and PC scores.

| Nr | Solute                     | Octanol | Diethyl ether | Chloroform | Benzene | Toluene | CCl <sub>4</sub> | n-Hexane | c-Hexane | t <sub>1</sub> | t <sub>2</sub> |
|----|----------------------------|---------|---------------|------------|---------|---------|------------------|----------|----------|----------------|----------------|
| 1  | Methanol                   | -0.77   | -1.15         | -1.26      | -1.89   | -2.15   | -2.10            | -2.80    | -2.80    | -2.65          | -0.77          |
| 2  | Ethanol                    | -0.31   | -0.57         | -0.85      | -1.62   | -1.70   | -1.40            | -2.10    | -2.10    | -2.07          | -0.66          |
| 3  | n-Propanol                 | 0.25    | -0.02         | -0.40      | -0.70   | -0.82   | -0.82            | -1.52    | -1.52    | -1.38          | -0.46          |
| 4  | n-Butanol                  | 0.88    | 0.89          | 0.45       | -0.12   | -0.30   | -0.40            | -0.70    | -0.70    | -0.63          | -0.09          |
| 5  | n-Pentanol                 | 1.56    | 1.20          | 1.05       | 0.62    | 0.51    | 0.40             | -0.40    | -0.26    | 0.02           | 0.05           |
| 6  | n-Hexanol                  | 2.03    | 1.80          | 1.69       | 1.30    | 1.29    | 0.99             | 0.46     | 0.70     | 0.78           | 0.05           |
| 7  | n-Heptanol                 | 2.41    | 2.40          | 2.41       | 1.91    | 1.84    | 1.67             | 1.01     | 1.49     | 1.45           | 0.07           |
| 8  | Propionic acid             | 0.33    | 0.27          | -0.96      | -1.35   | -1.47   | -1.60            | -2.14    | -2.54    | -1.93          | 0.50           |
| 9  | Acetone                    | -0.24   | -0.21         | 0.24       | -0.05   | -0.31   | -0.30            | -0.91    | -0.96    | -0.97          | -1.47          |
| 10 | Trimethylamine             | 0.27    | -0.26         | 0.54       | -0.29   | -0.36   | -0.09            | -0.48    | -0.44    | -0.75          | -1.47          |
| 11 | Butylamine                 | 0.74    | 0.11          | 0.99       | 0.14    | 0.30    | -0.04            | -0.62    | -0.29    | -0.42          | -1.04          |
| 12 | Diethylamine               | 0.57    | -0.07         | 0.81       | -0.05   | -0.09   | 0.03             | -0.48    | -0.34    | -0.54          | -1.26          |
| 13 | Pyridine                   | 0.65    | 0.08          | 1.43       | 0.41    | 0.29    | 0.23             | -0.21    | -0.31    | -0.23          | -1.37          |
| 14 | Aniline                    | 0.90    | 0.85          | 1.42       | 1.00    | 0.78    | 0.60             | -0.30    | 0.02     | 0.12           | -0.85          |
| 15 | Phenol                     | 1.46    | 1.64          | 0.37       | 0.36    | 0.15    | -0.36            | -0.70    | -0.81    | -0.36          | 0.81           |
| 16 | 2-Cl-Phenol                | 2.15    | 2.05          | 1.36       | 1.46    | 1.37    | 1.19             | 0.85     | 0.86     | 0.92           | 0.19           |
| 17 | 3-Cl-Phenol                | 2.50    | 2.10          | 1.02       | 1.12    | 1.05    | 0.49             | -0.07    | 0.08     | 0.49           | 1.10           |
| 18 | 4-Cl-Phenol                | 2.39    | 2.22          | 1.01       | 1.13    | 1.08    | 0.48             | -0.11    | 0.08     | 0.49           | 1.11           |
| 19 | 2-Me-Phenol                | 1.95    | 1.70          | 1.23       | 1.14    | 1.14    | 0.67             | -0.05    | 0.15     | 0.46           | 0.35           |
| 20 | 2,4-diMe-Phenol            | 2.30    | 2.40          | 1.50       | 1.34    | 1.26    | 0.78             | 0.34     | 0.34     | 0.76           | 0.83           |
| 21 | 2,5-diMe-Phenol            | 2.33    | 2.40          | 1.59       | 1.52    | 1.43    | 1.00             | 0.38     | 0.56     | 0.90           | 0.71           |
| 22 | 3,5-diMe-Phenol            | 2.35    | 2.43          | 1.60       | 1.33    | 1.29    | 0.82             | 0.32     | 0.27     | 0.79           | 0.87           |
| 23 | 2-Naphthol                 | 2.70    | 1.77          | 1.74       | 1.74    | 1.68    | 0.99             | 0.30     | 0.32     | 0.92           | 0.57           |
| 24 | 2-OH-Benzoic acid          | 2.26    | 2.37          | 0.58       | 0.45    | 0.31    | 0.00             | -0.57    | -0.50    | 0.01           | 1.61           |
| 25 | 4-OH-Benzoic acid          | 1.58    | 1.42          | -0.50      | -1.07   | -1.17   | -1.38            | -1.82    | -1.77    | -1.31          | 1.72           |
| 26 | 2-OH-Anisole               | 1.32    | 1.44          | 1.70       | 1.32    | 1.26    | 0.98             | 0.36     | 0.48     | 0.60           | -0.60          |
| 27 | 2-OMe- Benzoic acid        | 1.59    | 0.78          | 2.53       | 2.68    | 2.59    | 2.70             | 1.65     | 2.15     | 1.71           | -2.18          |
| 28 | 2-NO <sub>2</sub> -Phenol  | 1.79    | 2.18          | 2.35       | 2.32    | 2.28    | 1.91             | 1.40     | 1.45     | 1.53           | -0.65          |
| 29 | 3-NO <sub>2</sub> -Phenol  | 2.00    | 2.18          | 0.60       | 0.48    | 0.34    | -0.64            | -1.40    | -1.22    | -0.34          | 1.80           |
| 30 | 4-NO <sub>2</sub> -Phenol  | 1.91    | 2.01          | 0.20       | 0.17    | -0.06   | -0.92            | -2.00    | -1.70    | -0.72          | 2.04           |
| 31 | 2-NO <sub>2</sub> -Aniline | 1.85    | 1.95          | 2.13       | 1.78    | 1.64    | 1.08             | 0.25     | 0.36     | 0.89           | 0.02           |
| 32 | 3-NO <sub>2</sub> -Aniline | 1.37    | 1.71          | 1.61       | 1.31    | 1.19    | 0.45             | -0.62    | -0.42    | 0.28           | 0.18           |
| 33 | 4-NO <sub>2</sub> -Aniline | 1.39    | 1.48          | 1.23       | 0.93    | 0.78    | -0.14            | -1.14    | -1.00    | -0.14          | 0.53           |
| 34 | Vanillin                   | 1.21    | 0.96          | 1.42       | 0.82    | 0.62    | 0.20             | -0.72    | -0.70    | -0.09          | -0.21          |
| 35 | o-Vanillin                 | 1.37    | 1.35          | 2.30       | 1.87    | 1.73    | 1.40             | 0.53     | 0.65     | 0.94           | -0.96          |
| 36 | i-Vanillin                 | 0.97    | 0.82          | 1.18       | 0.74    | 0.46    | 0.04             | -0.85    | -0.82    | -0.26          | -0.32          |
| 37 | p-Toluidine                | 1.39    | 1.35          | 1.99       | 1.43    | 1.35    | 1.11             | 0.44     | 0.65     | 0.73           | -0.76          |

Table 2 shows PCA results with respect to the solvents as well as the means and variances of the solvent columns which had to be additionally considered since merely the correlation structure of the data was analyzed by our PCA.

**Table 2.** PCA results for the solvents and water solubility parameters used for correlations

|              |                      | Mean  | Variance | PC loadings    |                | Extracted variance | water solubility |            |
|--------------|----------------------|-------|----------|----------------|----------------|--------------------|------------------|------------|
|              |                      |       |          | a <sub>1</sub> | a <sub>2</sub> |                    | log $\gamma$     | $\delta_h$ |
| 1            | n-Octanol            | 1.39  | 0.75     | 0.775          | 0.609          | 97.1 %             | 3.01             | 5.8        |
| 2            | Diethyl ether        | 1.24  | 0.95     | 0.724          | 0.675          | 98.0 %             | 1.28             | 2.2        |
| 3            | Chloroform           | 1.04  | 0.91     | 0.951          | -0.194         | 94.3 %             | 0.75             | 2.8        |
| 4            | Benzene              | 0.69  | 1.14     | 0.983          | -0.035         | 96.8 %             | 0.65             | 1.0        |
| 5            | Toluene              | 0.58  | 1.19     | 0.989          | -0.039         | 98.0 %             | 0.56             | 1.0        |
| 6            | Carbon tetrachloride | 0.27  | 1.03     | 0.964          | -0.253         | 99.3 %             | -0.07            | 0.3        |
| 7            | n-Hexane             | -0.39 | 1.02     | 0.946          | -0.243         | 95.3 %             | -1.87            | 0.0        |
| 8            | Cyclohexane          | -0.29 | 1.19     | 0.953          | -0.240         | 96.7 %             | -0.90            | 0.1        |
| <b>Total</b> |                      |       |          | 83.8 %         | 13.1 %         | 96.9 %             |                  |            |

The order of the means and variances reflects the general rules derived above from the Collander equation. The means are highly correlated with the water solubility of the solvents expressed in terms of log  $\gamma$  [19]:

$$\text{Mean} = 0.43 (\pm 0.16) \log \gamma + 0.39 (\pm 0.23) \quad r^2 = 0.88, s = 0.23, \alpha = 0.001 \quad (2)$$

The variances are inversely related to the hydrogen bonding component of the solvent solubility in water,  $\delta_h$  [20] (data from ref. [21]):

$$\text{Variance} = -0.065 (\pm 0.042) \delta_h + 1.13 (\pm 0.10) \quad r^2 = 0.70, s = 0.08, \alpha = 0.009 \quad (3)$$

Thus, log P scales from water soluble solvents capable of forming hydrogen bonds show higher means and lower variances compared to log P scales from nonpolar solvents.

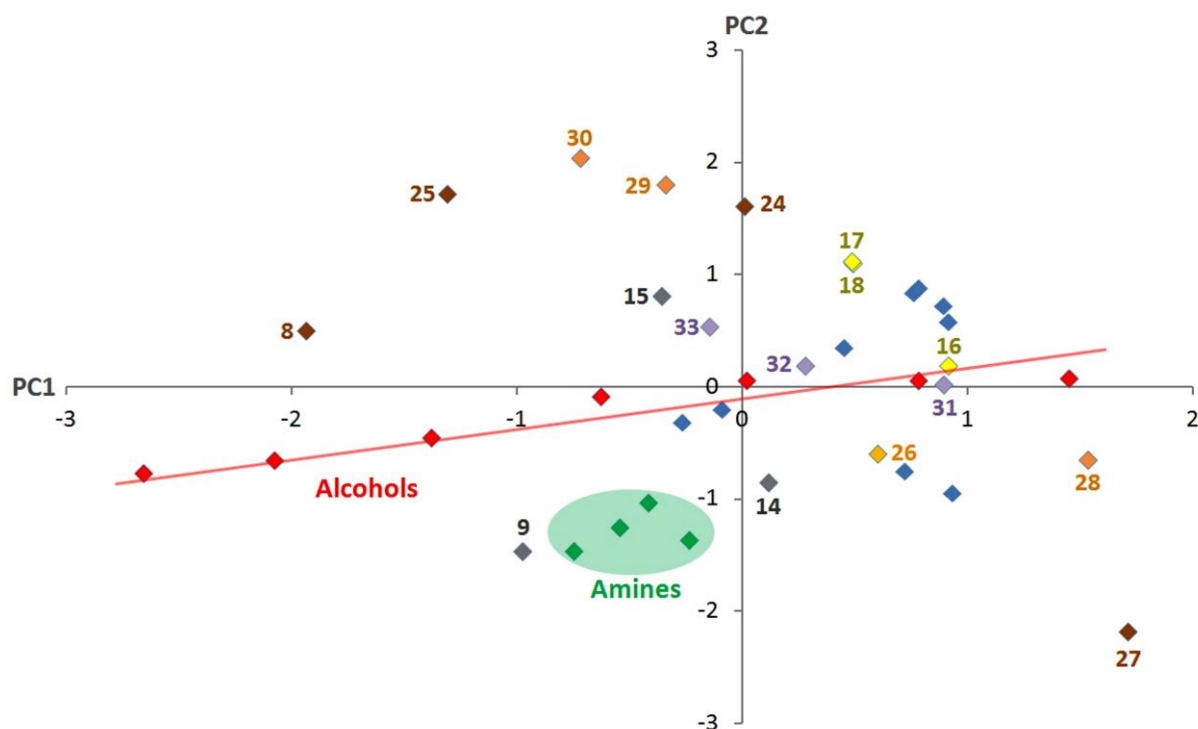
Two PCs account for 96.9 % of the data variance (first PC, 83.8 %, second PC, 13.1 %). Thus, PCA of log P-values again results in a two-component model as in the case of the previous analyses [15, 16, 17, 18]. The loadings  $a_{ik}$  (Table 2) represent correlation coefficients between log P from solvent  $i$  and scores  $t_k$  (Table 1). Obviously all nonpolar solvents, in particular benzene and toluene, are sufficiently described by the first PC which, however, extracts only ca. 55 % of the variance of the hydrogen bonding solvents n-octanol and diethylether. Therefore, the first PC represents "pure" hydrophobic effects due to the transfer of solutes from water into inert solvents. For polar solvents, a second PC accounting for hydrogen bonds and electrostatic interactions between solutes and solvents is necessary. This relationship may be modeled by correlation of  $a_2$  with the water solubility of the solvents:

$$a_2 = 0.21 (\pm 0.17) \log \gamma + 0.05 (\pm 0.24) \quad r^2 = 0.61, s = 0.24, \alpha = 0.021 \quad (4)$$

Large positive loadings of highly water-soluble solvents are in contrast to negative loadings of carbon tetrachloride, n-hexane and cyclohexane.

Inspection of the scores  $t_k$  and their correlation with suitable solute descriptors will enable more detailed insights into the QSPR. Figure 1 presents a plot of  $t_1$  vs.  $t_2$  accounting for "purely hydrophobic" effects and polar corrections, respectively, as described above. A homologous series as aliphatic alcohols is characterized by a flat line nearly parallel with the abszissa, i.e., by variation of mainly the hydrophobic component. Amines under consideration are clustered, their contribution to PC2 is significantly negative.

Aniline (14) resides close to the other amines, whereas phenol (15) has a positive  $t_2$  value, indicating strong interactions with hydrogen-bond acceptor solvents like n-octanol and diethylether.



**Figure 1.** Plot of the scores from PCA. Remarkable solutes are highlighted by colored symbols and/or numbers (cp. Table 1): red – alcohols, green – amines, brown – acids, redorange – nitrophenols, orange – 2-OH-anisole, yellow – chlorphenols, purple – nitroanilines.

Positional effects in disubstituted benzenes are evident. On one hand, *m*-nitroanilines and *m*-nitrophenols are slightly more hydrophobic than their *p*-substituted isomers (compare 29, 30 and 32, 33, respectively). On the other hand, *o*-substituted benzenes are located down to the right with respect to their *m*- and *p*-isomers as obvious from Figure 1 for hydroxybenzoic acids (24 vs. 25), chlorphenols (16 vs. 17 and 18), nitrophenols (28 vs. 29 and 30) as well as nitroanilines (31 vs. 32 and 33). The rightward shift is due to intramolecular hydrogen bonds and/or proximity effects, both increasing the nonpolar surface and reducing hydrogen bonds as well as electrostatic interactions with water. However, these effects are much more significant in nonpolar solvents.

Diethylether and n-octanol are strong hydrogen bond acceptors (n-octanol additionally a weak donor), preventing the formation of internal hydrogen bonds in solutes similarly as water (for review, see [22]). The downward shift of *o*-substituted isomers is a consequence of this phenomenon which is most pronounced in the case of *o*-methoxybenzoic acid (27). Also *o*-hydroxyanisole (26) shows a small negative "ortho-factor" in polar solvents rather due to twist than to a hydrogen bond effect [23].

Suitable descriptors for the identification of scores from PCA were provided by Dunn et al. [16, 17, 18] who defined and calculated the isotropic surface area, ISA [24], of solutes as the surface of the molecule accessible to nonspecific interactions with the solvent. The surface area of the solutes involved in specific hydrogen bonds with water, HSA, was excluded from the ISA. For calculation of ISA and HSA, Dunn et al. constructed hydrated solutes, "supermolecules", from empirical hydration rules based on crystallographic data, quantum-chemical approaches, solution modeling and experimental data from solute-gas phase equilibria (see [25, 26] and references therein).

The scores of PC1,  $t_1$ , are highly correlated with the isotropic surface area (data from [18]):

$$t_1 = 2.15 (\pm 0.25) \text{ ISA} - 4.96 (\pm 0.58) \quad r^2 = 0.90, s = 0.32, \alpha < 0.001 \quad (5)$$

Thus, the first PC accounts for the well known dependency of log P values on the nonpolar surface fraction of solutes based on entropic effects of water exclusion and nonspecific solute-solvent interactions. In the PCA approach of Koehler et al. [18], the matrix of cross products was diagonalized, leading to orthogonal but correlated scores which also depend on the means and variances of log P in six analyzed water-solvent systems. The correlation of  $t_1$  from ref. [18] with ISA was weaker than that shown in eq. 5 ( $r^2 = 0.81$ ).

The scores of PC2,  $t_2$ , are significantly related to the hydrated surface area ( $r^2 = 0.46$ ) [24], but even if four outliers (compounds 11, 12, 27, 36) are excluded from the analysis, the correlation remains rather weak:

$$t_2 = 2.67 (\pm 0.86) \text{ HSA} - 1.56 (\pm 0.60) \quad r^2 = 0.56, s = 0.63, \alpha < 0.001 \quad (6)$$

I.e., the hydrated surface area plays a role in increasing log P in polar solvents (positive PC2 loadings) and decreasing log P in nonpolar solvents (negative PC2 loadings), but HSA is not sufficient to quantitatively describe this effect. Compared to eq. 6, Koehler et al. [18] obtained a better correlation of the scores from the second PC with the hydrated fraction of the solvent accessible surface area,  $f(\text{HSA})$  (69 solutes,  $r^2 = 0.74$ ) [24]. In our PCA approach, HSA is superior to  $f(\text{HSA})$ . However, the scores  $t_2$  from ref. [18] account only for nonpolar solvent-water systems as evident from correlation coefficients between log P and  $t_2$  (in analogy to loadings  $a_2$  from our PCA): octanol, 0.23, diethylether, 0.12, chloroform, 0.84, carbon tetrachloride, 0.88, benzene, 0.78, hexane, 0.89. Thus, the correlation of  $t_2$  with  $f(\text{HSA})$  in the paper of Koehler et al. [18] mainly reflects a negative impact of the hydrated surface area fraction on the hydrophobicity of solutes in nonpolar solvents.

The discriminative effect of  $f(\text{HSA})$  on log P-values from different solvent-water systems may be explored in more detail by regression analysis of log P as function of ISA and  $f(\text{HSA})$  (see Table 3).

**Table 3.** Regression equations of log P-values (Table 1) as function of surface area descriptors [18].

|                        | Regression coefficients |                                      |                     | s    | $r^2$ | $r^2$ [18] <sup>b</sup> |
|------------------------|-------------------------|--------------------------------------|---------------------|------|-------|-------------------------|
|                        | ISA                     | $f(\text{HSA})^a$                    | intercept           |      |       |                         |
| 1 n-Octanol            | 2.16 ( $\pm$ 0.61)      | 6.20 ( $\pm$ 3.00)                   | -4.96 ( $\pm$ 1.96) | 0.54 | 0.61  | 0.57                    |
| 2 Diethyl ether        | 2.32 ( $\pm$ 0.74)      | 6.69 ( $\pm$ 3.61)                   | -5.59 ( $\pm$ 2.35) | 0.65 | 0.56  | 0.48                    |
| 3 Chloroform           | 1.94 ( $\pm$ 0.24)      | <i>-0.89 (<math>\pm</math> 1.17)</i> | -3.26 ( $\pm$ 0.76) | 0.21 | 0.95  | 0.88 <sup>c</sup>       |
| 4 Benzene              | 2.31 ( $\pm$ 0.37)      | <i>0.23 (<math>\pm</math> 1.82)</i>  | -4.68 ( $\pm$ 1.19) | 0.33 | 0.91  | 0.86                    |
| 5 Toluene              | 2.35 ( $\pm$ 0.38)      | <i>0.18 (<math>\pm</math> 1.86)</i>  | -4.88 ( $\pm$ 1.21) | 0.34 | 0.91  |                         |
| 6 Carbon tetrachloride | 1.80 ( $\pm$ 0.37)      | -2.18 ( $\pm$ 1.81)                  | -3.40 ( $\pm$ 1.18) | 0.33 | 0.90  | 0.88                    |
| 7 n-Hexane             | 1.56 ( $\pm$ 0.46)      | -3.09 ( $\pm$ 2.25)                  | -3.32 ( $\pm$ 1.47) | 0.40 | 0.84  | 0.80                    |
| 8 Cyclohexane          | 1.81 ( $\pm$ 0.49)      | -2.64 ( $\pm$ 2.40)                  | -3.89 ( $\pm$ 1.56) | 0.43 | 0.84  |                         |

<sup>a</sup> Regression coefficients in italics are not significant ( $\alpha > 0.05$ ). <sup>b</sup> Corresponding correlations of the whole dataset in ref. [18] ( $n = 69$ ). <sup>c</sup> Regression coefficient of -1.59 for  $f(\text{HSA})$  significant ( $\alpha = 0.01$ ).

All equations for nonpolar solvents provide a sufficient decomposition of log P into surface area terms. In contrast, equations for n-octanol and diethylether explain only ca. 60 % of the data variance. Whereas regression coefficients of ISA and intercepts do not significantly differ, effects of  $f(\text{HSA})$  are distinctive with respect to the solvent class: polar solvents are characterized by a positive, nonpolar solvents by a negative or no impact of the hydrated surface area fraction of solutes on log P. Solutes with a high hydration potential are poorly transferred just into carbon tetrachloride, n-hexane and cyclohexane. Koehler et al.

[18] correlated log P-values *calculated* from loadings and scores with ISA and f(HSA). Because of the nonsignificance of f(HSA) in equations for n-octanol and diethylether, they suggested that the solutes partition into the solvent as the hydrated "supermolecule" due to the water solubility of these solvents which compete with water for the hydrogen bonding sites of the solutes, leading to displacement of water from the "supermolecule". However, f(HSA) is significant if *measured* n-octanol and diethylether log P-values are correlated (Table 3), albeit both equations do not sufficiently model hydrophobicity. Therefore, additional descriptors must be taken into account for polar solvents.

The free energy of solute partition from water into n-octanol and diethylether depends on the *difference* of hydrogen bond interactions in both phases. Suitable descriptors considering these effects on log P have been derived by Taft, Kamlet and Abraham et al. [27, 28, 29, 30, 31]. Based on the solvatochromic model [32, 33, 34], log P-values are factorized into four parameters: the molecular volume V for nonpolar interactions, the solute's dipolarity/polarizability  $\pi^*$  for orientation and induction forces, as well as  $\alpha$  and  $\beta$  for the hydrogen bond donor acidity and acceptor basicity, respectively. For example, solvatochromic analysis of octanol-water partition coefficients for 103 solutes resulted in the following equation [35]:

$$\log P = 5.15 (\pm 0.16) V/100 - 1.29 (\pm 0.16) \pi^* - 3.60 (\pm 0.18) \beta + 0.45 (\pm 0.12) \quad r^2 = 0.98 \quad s = 0.16 \quad (7)$$

From the series of Koehler et al. [18], solvatochromic descriptors were available for 45 compounds [27, 30]. Also in case of this subset, the correlations of n-octanol and diethylether log P-values with ISA and HSA ( $r^2$ : 0.77 and 0.63, respectively) or with ISA and f(HSA) ( $r^2$ : 0.74 and 0.60, respectively) are not sufficient. Combining these parameters with the solvatochromic descriptors leads to the following equations for n-octanol:

$$\log P = 1.88 (\pm 0.21) ISA + 1.59 (\pm 0.34) HSA - 2.59 (\pm 0.59) \beta - 2.99 (\pm 0.70) \quad r^2 = 0.92 \quad s = 0.25 \quad (8)$$

$$\log P = 2.42 (\pm 0.31) ISA + 5.82 (\pm 1.40) f(HSA) - 2.53 (\pm 0.64) \beta - 4.32 (\pm 1.06) \quad r^2 = 0.91 \quad s = 0.27 \quad (9)$$

and for diethylether:

$$\log P = 1.80 (\pm 0.18) ISA + 1.19 (\pm 0.30) HSA - 4.19 (\pm 0.52) \beta - 1.81 (\pm 0.62) \quad r^2 = 0.95 \quad s = 0.22 \quad (10)$$

$$\log P = 2.24 (\pm 0.25) ISA + 4.55 (\pm 1.12) f(HSA) - 4.13 (\pm 0.51) \beta - 3.08 (\pm 0.85) \quad r^2 = 0.95 \quad s = 0.22 \quad (11)$$

The volume, dipolarity/polarizability and hydrogen bond acidity descriptors V,  $\pi^*$  and  $\alpha$ , respectively, are not significant. However, both hydrated surface area descriptors show a multiple collinearity with  $\alpha$  and  $\beta$ :

$$HSA = 1.05 (\pm 0.20) \alpha + 1.00 (\pm 0.39) \beta - 0.22 (\pm 0.24) \quad r^2 = 0.74 \quad s = 0.14 \quad (12)$$

$$f(HSA) = 0.37 (\pm 0.08) \alpha + 0.38 (\pm 0.16) \beta - 0.08 (\pm 0.10) \quad r^2 = 0.66 \quad s = 0.06 \quad (13)$$

Taken together, these equations represent the different impacts of hydrogen bonds in nonpolar and polar solvent-water systems. The hydrated surface area reflects hydrogen bond donor acidity and acceptor basicity in equal parts and is a suitable descriptor of the detrimental effect of solute-water hydrogen bonds on log P-values in nonpolar solvents as chloroform, n-hexane and cyclohexane. In contrast, n-octanol and diethylether are strong hydrogen bond acceptors themselves. Donor solutes are favored, i.e., log P increases if a large hydrated surface area is mainly due to a high  $\alpha$ -term. The net effect of  $\beta$  on log P is negative (compare eqs. 8-11 with 12-13). Accordingly, hydrogen bond acceptor solutes are less hydrophobic in these solvents, in particular in diethylether since n-octanol is also a hydrogen bond donor, but weak compared to water. In conclusion, the decomposition of log P-values must always consider *differences* of solute-water and solute-solvent interactions.



### Multivariate analysis of hydrophobic substituent constants $\pi$ from disubstituted benzenes

The hydrophobic substituent constant  $\pi$  was first introduced by Hansch et al. [36] as difference of the octanol-water log P of substituted and unsubstituted phenoxyacetic acid. Following this concept, numerous  $\pi$ -values were derived for *ortho*-, *meta*- and *para*-substituents X in various aromatic systems PhY as benzenes, nitrobenzenes, anilines, phenols, phenylacetic acids and phenoxyacetic acids [7]. Whereas *meta-para* positional effects on  $\pi_X$  and  $\pi_X$ -differences of inert substituents were only marginal,  $\pi_X$ -values of hydrogen-bonding, electron-attracting or -releasing substituents depend significantly on the nature of the "functional group" Y. Thus, log P values of disubstituted benzenes XPhY are not simply the sum of log P (benzene),  $\pi_X$  and  $\pi_Y$  from the benzene system, but include "interaction increments" due to electronic effects. In a first approximation, their nature was identified by correlations of the type [7]:

$$\Delta\pi = \pi_X(\text{PhY}) - \pi_X(\text{PhH}) = k \sigma_X \quad (14)$$

i.e., the difference of the  $\pi_X$ -values from a series PhY and benzene depends on the electronic properties of X, described by Hammett's  $\sigma$  constant, and on the specific impact of Y on X, reflected by k.

To further investigate these effects, principal component analyses of  $\pi_{XY}$ -values from different *meta*- and *para*-substituted aromatic series PhY [5, 7, 37] were performed by Franke et al. [38, 39]. Both approaches with separate [38] and simultaneous [39] consideration, respectively, of 27 *meta*- and *para*-substituents (PhY: benzenes, nitrobenzenes, anilines, phenols, benzoic acids, phenylacetic acids, phenoxyacetic acids, piperidinoacetanilides) resulted in two significant PCs. The first PC accounted for the "average" hydrophobicity of the substituents, and the second PC was due to electronic interactions between X and Y and correlated with  $\sigma_X$ . However, these PCAs suffered from too many unknown  $\pi_{XY}$  values (36 of 216, 17%) which had to be estimated by regression analysis in order to obtain a full data matrix. Consideration of more recent experimental log P-values [6] and withdrawal of  $\pi$ -values of piperidinoacetanilides and the CH<sub>2</sub>OH group enables a substantial reduction of calculated data (10 of 175, 6%). With this update and some substitutions by more reliable values [6, 40], the simultaneous PCA of *meta*- and *para*-disubstituted benzenes was recalculated. The data matrix is shown in Table 4.

Table 5 presents PCA results for the systems PhY. Two PCs account for 98.7 % of the data variance (first PC, 92.6 %, second PC, 6.1 %). The loadings  $a_{Yk}$  as correlation coefficients between  $\pi_{XY}$  from series Y and scores  $t_k$  (Table 4) show that  $\pi$ -values from benzenes, phenoxyacetic, phenylacetic and benzoic acids are sufficiently reproduced by the first PC, indicating only weak effects of X on Y and vice versa in these systems. Thus, PC 1 represents hydrophobicity of substituents X largely unaffected by interaction with Y. The hydrogen bonding acceptor system nitrobenzene and the donor-acceptor systems aniline and phenol bear considerable, opposed loadings in the second PC which is significantly correlated with  $\sigma_p$ -values of the "functional groups" Y:

$$a_2 = 0.49 (\pm 0.26) \sigma_{Yp} + 0.02 (\pm 0.12) \quad r^2 = 0.83, s = 0.11, \alpha = 0.005 \quad (15)$$

The correlation with  $\sigma_{Ym}$  is only weak ( $r^2 = 0.63$ ). These results must be interpreted in context with the scores  $t_1$  and  $t_2$  (Table 4). Figure 2 presents a plot of  $t_1$  vs.  $t_2$  accounting for "unaffected" hydrophobicity of substituents X and X-Y interactions, respectively. There is obviously no positional effect, corresponding *m*- and *p*-substituents overlap apart from small differences in the case of OH, OMe, COMe, CN and Br. Thus, the joint analysis of *meta*- and *para*-disubstituted benzenes is justified. The arrangement of the substituents along the abscissa (PC 1) corresponds to the common hydrophobicity scale (polar, hydrogen-bonding substituents < H < Me, F < Cl < Br < I).

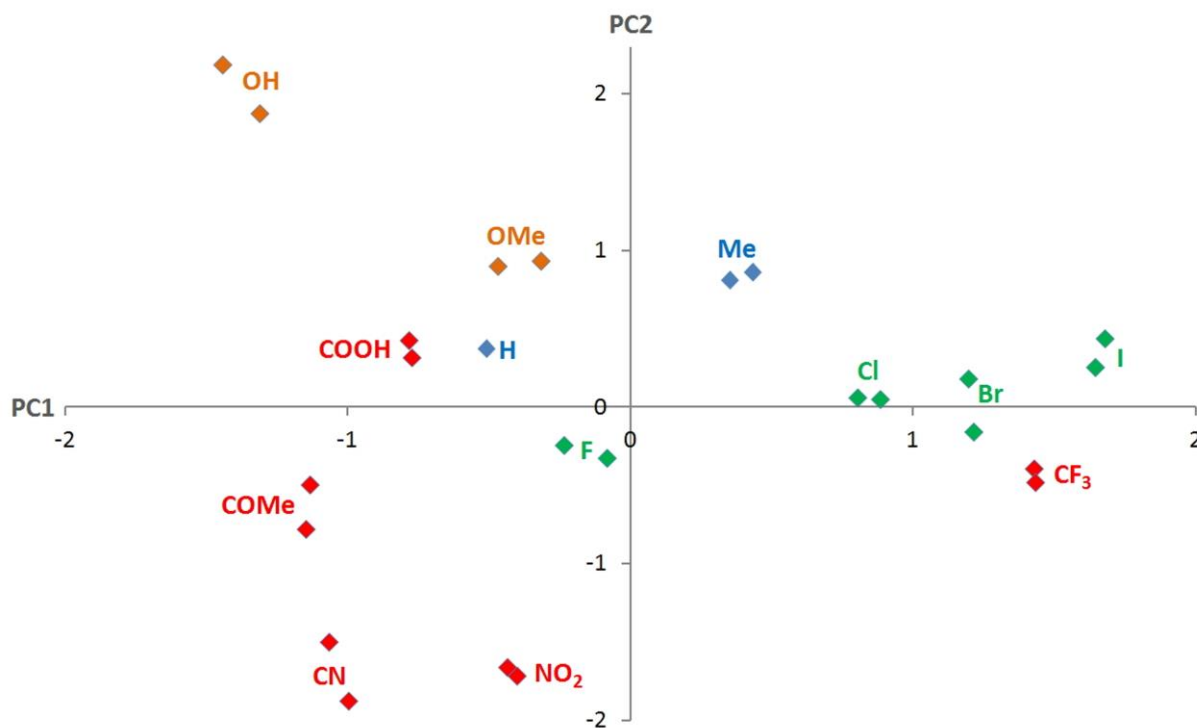
**Table 4.** PCA of  $\pi_{XY}$ -values<sup>a</sup> from different aromatic series PhY: data, PC scores and  $\sigma_X$ .

| Nr          | Substituent               | Benzene | PhOAc Acid | PhAc Acid | Benzoic Acid | Aniline | Nitro benzene | Phenol | t <sub>1</sub> | t <sub>2</sub> | $\sigma_X$ |
|-------------|---------------------------|---------|------------|-----------|--------------|---------|---------------|--------|----------------|----------------|------------|
| 1           | H                         | 0.00    | 0.00       | 0.00      | 0.00         | 0.00    | 0.00          | 0.00   | -0.51          | 0.38           | 0.00       |
| 2           | <i>m</i> -F               | 0.14    | 0.22       | 0.19      | 0.28         | 0.40    | 0.05          | 0.47   | -0.08          | -0.33          | 0.34       |
| 3           | <i>p</i> -F               | 0.14    | 0.15       | 0.14      | 0.19         | 0.25    | -0.05         | 0.31   | -0.23          | -0.25          | 0.06       |
| 4           | <i>m</i> -Cl              | 0.71    | 0.76       | 0.68      | 0.83         | 0.98    | 0.61          | 1.04   | 0.89           | 0.05           | 0.37       |
| 5           | <i>p</i> -Cl              | 0.71    | 0.70       | 0.70      | 0.78         | 0.93    | 0.54          | 0.93   | 0.81           | 0.06           | 0.23       |
| 6           | <i>m</i> -Br              | 0.86    | 0.94       | 0.91      | 0.99         | 1.20    | 0.79          | 1.17   | 1.20           | 0.18           | 0.39       |
| 7           | <i>p</i> -Br              | 0.86    | 1.02       | 0.90      | 0.98         | 1.36    | 0.70          | 1.13   | 1.22           | -0.16          | 0.23       |
| 8           | <i>m</i> -J               | 1.12    | 1.15       | 1.22      | 1.28         | 1.46    | 1.09          | 1.47   | 1.68           | 0.44           | 0.35       |
| 9           | <i>p</i> -J               | 1.12    | 1.26       | 1.23      | 1.14         | 1.44    | 1.02          | 1.45   | 1.65           | 0.25           | 0.18       |
| 10          | <i>m</i> -Me              | 0.56    | 0.51       | 0.54      | 0.52         | 0.50    | 0.57          | 0.56   | 0.43           | 0.86           | -0.07      |
| 11          | <i>p</i> -Me              | 0.56    | 0.52       | 0.45      | 0.42         | 0.49    | 0.52          | 0.48   | 0.35           | 0.81           | -0.17      |
| 12          | <i>m</i> -CF <sub>3</sub> | 0.88    | 1.07       | 1.16      | 1.07         | 1.39    | 0.77          | 1.49   | 1.43           | -0.39          | 0.43       |
| 13          | <i>p</i> -CF <sub>3</sub> | 0.88    | 1.13       | 1.04      | 1.23         | 1.49    | 0.70          | 1.36   | 1.43           | -0.48          | 0.54       |
| 14          | <i>m</i> -OMe             | -0.02   | 0.12       | 0.04      | 0.14         | 0.03    | 0.31          | 0.12   | -0.32          | 0.93           | 0.12       |
| 15          | <i>p</i> -OMe             | -0.02   | -0.04      | 0.01      | 0.08         | 0.05    | 0.18          | -0.12  | -0.47          | 0.90           | -0.27      |
| 16          | <i>m</i> -OH              | -0.67   | -0.49      | -0.52     | -0.38        | -0.73   | 0.15          | -0.66  | -1.31          | 1.87           | 0.12       |
| 17          | <i>p</i> -OH              | -0.67   | -0.61      | -0.66     | -0.30        | -0.86   | 0.11          | -0.87  | -1.44          | 2.19           | -0.37      |
| 18          | <i>m</i> -NO <sub>2</sub> | -0.28   | 0.11       | -0.01     | -0.05        | 0.47    | -0.36         | 0.54   | -0.43          | -1.66          | 0.71       |
| 19          | <i>p</i> -NO <sub>2</sub> | -0.28   | 0.24       | -0.04     | 0.02         | 0.49    | -0.39         | 0.50   | -0.40          | -1.72          | 0.78       |
| 20          | <i>m</i> -COOH            | -0.28   | -0.15      | -0.27     | -0.19        | -0.18   | -0.02         | 0.04   | -0.77          | 0.31           | 0.37       |
| 21          | <i>p</i> -COOH            | -0.28   | -0.22      | -0.49     | -0.05        | -0.22   | 0.03          | 0.12   | -0.78          | 0.42           | 0.45       |
| 22          | <i>m</i> -CN              | -0.57   | -0.30      | -0.28     | -0.37        | 0.17    | -0.68         | 0.22   | -1.00          | -1.88          | 0.56       |
| 23          | <i>p</i> -CN              | -0.57   | -0.32      | -0.35     | -0.31        | -0.06   | -0.66         | 0.14   | -1.06          | -1.50          | 0.66       |
| 24          | <i>m</i> -COMe            | -0.55   | -0.28      | -0.83     | -0.31        | -0.04   | -0.43         | -0.07  | -1.15          | -0.78          | 0.38       |
| 25          | <i>p</i> -COMe            | -0.55   | -0.37      | -0.73     | -0.26        | -0.08   | -0.36         | -0.11  | -1.13          | -0.50          | 0.50       |
| log P (PhY) |                           | 2.13    | 1.26       | 1.41      | 1.87         | 0.90    | 1.85          | 1.46   |                |                |            |

<sup>a</sup>  $\pi_{XY}$  values in italics were not available and therefore calculated as means of at least three regression equations of the type  $\pi_{XY1} = f(\pi_{XY2}, \sigma_X)$

**Table 5.** PCA results for the aromatic systems PhY and descriptors used for correlations

|              | Y                  | PC loadings           |                | Extracted variance | $\pi_{YH}$ | $\sigma_{Ym}$ | $\sigma_{Yp}$ |       |
|--------------|--------------------|-----------------------|----------------|--------------------|------------|---------------|---------------|-------|
|              |                    | a <sub>1</sub>        | a <sub>2</sub> |                    |            |               |               |       |
| 1            | Benzene            | H                     | 0.986          | 0.103              | 98.3 %     | 0.00          | 0.00          | 0.00  |
| 2            | Phenoxyacetic acid | OCH <sub>2</sub> COOH | 0.995          | -0.055             | 99.3 %     | -0.87         | 0.30          | -0.33 |
| 3            | Phenylacetic acid  | CH <sub>2</sub> COOH  | 0.985          | 0.012              | 97.1 %     | -0.72         | 0.15          | -0.07 |
| 4            | Benzoic Acid       | COOH                  | 0.992          | 0.066              | 98.8 %     | -0.28         | 0.37          | 0.45  |
| 5            | Aniline            | NH <sub>2</sub>       | 0.957          | -0.276             | 99.2 %     | -1.23         | -0.16         | -0.66 |
| 6            | Nitrobenzene       | NO <sub>2</sub>       | 0.869          | 0.489              | 99.3 %     | -0.28         | 0.71          | 0.78  |
| 7            | Phenol             | OH                    | 0.947          | -0.301             | 98.8 %     | -0.67         | 0.12          | -0.37 |
| <b>Total</b> |                    |                       | 92.6 %         | 6.1 %              | 98.7 %     |               |               |       |



**Figure 1.** Plot of the scores from PCA. Substituent types are highlighted by colored symbols: red – electron-attracting, redorange – electron-releasing, green – halogens, blue – H, Me.

The spread of PC 2 is due to electronic properties. Electron-releasing substituents (OH, OMe) are characterized by positive, electron-attracting groups (COMe, CN, NO<sub>2</sub>, CF<sub>3</sub>) by negative scores  $t_2$ . These relationships may be quantified by the following correlations with "unaffected"  $\pi$ -values from benzenes ( $\pi_{XH}$ ) and  $\sigma_{Xp}$ :

$$t_1 = 1.63 (\pm 0.12) \pi_{XH} - 0.25 (\pm 0.07) \quad r^2 = 0.97, s = 0.17, \alpha < 0.001 \quad (16)$$

$$t_2 = -2.55 (\pm 0.56) \sigma_{Xp} + 0.58 (\pm 0.23) \quad r^2 = 0.80, s = 0.46, \alpha < 0.001 \quad (17)$$

Eq. 17 is better than the correlation with position-dependent  $\sigma_{Xm}$ - and  $\sigma_{Xp}$ -values ( $r^2 = 0.68$ ). Taken together, eqs. 15 and 17 reflect an electronic X-Y interaction increment described by the product  $\sigma_{Xp} \sigma_{Yp}$ . The hydrophobicity of a substituent X increases if its electronic effects on the phenyl nucleus are counterbalanced by Y (electron-attracting X combined with electron-releasing Y and vice versa).

These findings obtained from the multivariate PCA approach may be explored in more detail by individual consideration of the series. Instead of correlating  $\Delta\pi_X$  with  $\sigma_X$  as in eq. 14 [7], the  $\pi_{XY}$ -values were directly related to  $\pi_{XH}$  and  $\sigma_X$  by multiple regression analysis (see Table 6). Calculated  $\pi_{XY}$ -values were omitted. Since correlations with  $\sigma_{Xp}$  and position-dependent  $\sigma_{Xm}$ - and  $\sigma_{Xp}$ -values led to approximately equivalent equations, the latter, "correct" descriptors were used. All regression equations except that for the nitrobenzene series result in an intercept of approximately zero and explain more than 95 % of the data variance.

As expected from the PCA (eqs. 15, 17), the regression coefficients for  $\sigma_X$  depend on electronic properties of Y. However, the coefficients for  $\pi_{XH}$  and in particular their deviation from unity seem to follow the same trend. Correlation of the regression coefficients with  $\sigma_Y$  results in:

$$c(\pi_{XH}) = -0.36 (\pm 0.11) \sigma_{Ym} + 1.01 (\pm 0.04) \quad r^2 = 0.94, s = 0.03, \alpha < 0.001 \quad (18)$$

$$c(\sigma_X) = -0.83 (\pm 0.53) \sigma_{Yp} + 0.28 (\pm 0.24) \quad r^2 = 0.76, s = 0.23, \alpha = 0.01 \quad (19)$$

**Table 6.** Regression equations of  $\pi_{XY}$ -values as function of  $\pi_{XH}$  and  $\sigma_X$  (Table 4).

|                      | Regression coefficients |                     |                                     | $r^2$ | s    | n <sup>b</sup> |
|----------------------|-------------------------|---------------------|-------------------------------------|-------|------|----------------|
|                      | $\pi_{XH}$              | $\sigma_X$          | intercept <sup>a</sup>              |       |      |                |
| 1 Benzene            | 1.00                    | 0.00                | 0.00                                | 1.00  | 0.00 | 25             |
| 2 Phenoxyacetic acid | 0.92 ( $\pm$ 0.06)      | 0.34 ( $\pm$ 0.12)  | 0.05 ( $\pm$ 0.04)                  | 0.98  | 0.07 | 23             |
| 3 Phenylacetic acid  | 0.95 ( $\pm$ 0.06)      | 0.27 ( $\pm$ 0.12)  | <i>0.03 (<math>\pm</math> 0.05)</i> | 0.98  | 0.07 | 21             |
| 4 Benzoic Acid       | 0.91 ( $\pm$ 0.07)      | 0.18 ( $\pm$ 0.14)  | 0.12 ( $\pm$ 0.06)                  | 0.97  | 0.10 | 25             |
| 5 Aniline            | 1.04 ( $\pm$ 0.12)      | 0.84 ( $\pm$ 0.23)  | <i>0.07 (<math>\pm</math> 0.09)</i> | 0.95  | 0.15 | 22             |
| 6 Nitrobenzene       | 0.72 ( $\pm$ 0.14)      | -0.44 ( $\pm$ 0.28) | 0.22 ( $\pm$ 0.12)                  | 0.86  | 0.18 | 24             |
| 7 Phenol             | 0.99 ( $\pm$ 0.06)      | 0.91 ( $\pm$ 0.13)  | -0.07 ( $\pm$ 0.06)                 | 0.98  | 0.09 | 25             |

<sup>a</sup> Intercepts in italics are not significant ( $\alpha > 0.05$ ) <sup>b</sup> Nr. of  $\pi_{XY}$ -Values used for analysis.

In both cases, the correlation with  $\sigma_m$ - and  $\sigma_p$ -values, respectively, is significantly better than with Hammett-constants of the other position ( $r^2 = 0.71, 0.53$ ). Eqs. 18 and 19 indicate that  $\pi_{XY}$ -values include two X-Y interaction increments depending on the products  $\pi_{XH} \sigma_{Ym}$  and  $\sigma_{Xp} \sigma_{Yp}$ . However, the drawback of this model is that it has been derived from separate analyses of the seven systems. A common model for all systems must be based on equivalent, "symmetric" consideration of substituents X and Y in *meta*- and *para*-disubstituted benzenes. Fujita [40] published such a model relying on bidirectional Hammett-type relationships:

$$\Delta\pi = \pi_{XY} - \pi_{XH} = \rho_Y \sigma_X + \rho_X \sigma_Y \quad (20)$$

In this equation,  $\rho_Y$  is the difference of the susceptibility constants  $\rho_Y(\text{octanol}) - \rho_Y(\text{water})$  of hydrogen bonding association between the respective solvent and the fixed substituent Y to the effect of variable substituents X, and  $\rho_X$  is the equivalent difference for the impact of Y on X. Thus, the transmission of electronic effects of substituents from X to Y is assumed to be independent of transmission from Y to X.

To make the Hammett-type relationships from eqs. 18, 19 and Table 6 bidirectional implies the introduction of an additional X-Y interaction increment  $\pi_{YH} \sigma_X$ . The following common models for *meta*- and *para*-disubstituted benzenes, respectively, were derived from the data in Table 4 (calculated data omitted, known  $\pi_{XY}$  values of X = CH<sub>2</sub>OH included):

$$\pi_{XYm} = 0.98 (\pm 0.08) \pi_{XH} - 0.25 (\pm 0.20) \pi_{XH} \sigma_{Ym} - 0.34 (\pm 0.22) \pi_{YH} \sigma_{Xm} - 0.56 (\pm 0.20) \sigma_{Xp} \sigma_{Yp} + 0.13 (\pm 0.06) \\ r^2 = 0.95 \quad s = 0.14 \quad \alpha < 0.001 \quad n = 72 \quad (21)$$

$$\pi_{XYP} = 0.93 (\pm 0.06) \pi_{XH} - 0.24 (\pm 0.12) \pi_{XH} \sigma_{Yp} - 0.42 (\pm 0.16) \pi_{YH} \sigma_{Xp} - 0.41 (\pm 0.21) \sigma_{Xp} \sigma_{Yp} + 0.11 (\pm 0.04) \\ r^2 = 0.95 \quad s = 0.15 \quad \alpha < 0.001 \quad n = 69 \quad (22)$$

Both equations are equivalent and without collinearities of descriptors. The regression coefficients of  $\pi_{XH}$  are close to unity, and the intercepts may be neglected. Small differences between the terms  $\pi_{XH} \sigma_Y$  and  $\pi_{YH} \sigma_X$  are possibly due to the imbalance of the numbers of X and Y substituents. For *meta*-disubstituted benzenes, the correlation with  $\sigma_{Xp} \sigma_{Yp}$  is better than with  $\sigma_{Xm} \sigma_{Ym}$  ( $r^2 = 0.93$ ). Eqs. 21 and 22 resemble the Hammett formalism for bidirectional inductive interaction of any two fragments [41]:

$$\Delta G_{XY} = \rho_{00} + \rho_{01} \sigma_X + \rho_{10} \sigma_Y + \rho_{11} \sigma_X \sigma_Y \quad (23)$$

i.e., the contribution of the interaction to a property as, e.g., solvation energy is a function of the electronic effects of the fragments and  $\rho$  constants depending on the skeleton between them.

Comparing these models with such bidirectional approaches [40, 41] indicates that the "susceptibility constants"  $\rho_X$  and  $\rho_Y$  depend on  $\pi_{XH}$  and  $\pi_{YH}$ , respectively. Polar substituents (negative  $\pi$ -values) increase  $\pi_{XY}$  in combination with a second, electron-attracting substituent. By this, polarity is reduced and interactions with the strong hydrogen bond acceptor octanol become slightly more favorable. In case of hydrophobic substituents (positive  $\pi$ -values), the interaction of a second, electron-releasing group with octanol is favored by the same reason. The  $\sigma_{Xp} \sigma_{Yp}$  cross term may be due to electron-releasing hydrogen bond donor substituents as OH and NH<sub>2</sub>. Combined with electron-attracting groups, their hydrogen bonds with octanol are facilitated. This effect is even underestimated by the cross term since the greatest differences between measured and calculated  $\pi_{XY}$ -values of +0.3 to +0.4 occur just in case of phenols and anilines with strongly electron-withdrawing substituents as NO<sub>2</sub>, CF<sub>3</sub> and CN. Combination of two hydrogen bond acceptors with positive  $\sigma_{Xp}$ -values reduces  $\pi_{XY}$  most. In contrast to previous suggestions [38], electronic effects on the phenyl nucleus play a minor role since otherwise *meta*-disubstituted benzenes should be correlated with a  $\sigma_{Xm} \sigma_{Ym}$  cross term. Taken together, the Hammett-type relationships represented by eqs. 21 and 22 indirectly account for mutual interactions of substituents favorable or detrimental for hydrogen bonds with octanol.

Models of this type may be used for the calculation of log P values of *meta*- and *para*-disubstituted benzenes:

$$\log P (\text{XPhY}) = \log P (\text{benzene}) + \pi_{YH} + \pi_{XY} \quad (24)$$

where  $\pi_{XY}$  is the difference between log P (XPhY) and log P (PhY) and comprises all X-Y interactions. Thus:

$$\log P (\text{XPhY}) = 2.13 + \pi_{XH} + \pi_{YH} - c_X \pi_{XH} \sigma_{Ym} - c_Y \pi_{YH} \sigma_{Xm} - c_{XY} \sigma_{Xp} \sigma_{Yp} \quad (25)$$

where  $c_X$  (ca. 0.25),  $c_Y$  (ca. 0.4) and  $c_{XY}$  (ca. 0.5) quantify the influence of the three Hammett-type increments. This is an extension of the method applied in CLOGP [10] where electronic interactions in *meta*- and *para*-disubstituted benzenes are considered by factors  $F_{XY} = \rho_Y \sigma_X$  (here,  $\sigma_X$  is no Hammett constant, but derived from log P values). For *ortho*-disubstituted benzenes, additional factors come into play representing an *ortho*-effect ( $F_{ortho}$ ), intramolecular hydrogen bonding ( $F_{HB}$ ) and alkyl-aryl interaction ( $F_{A\Phi}$ ), so that in this case

$$F_{XY} = \rho_Y \sigma_X + F_{ortho} + F_{HB} + F_{A\Phi} \quad (26)$$

## Conclusions

Multivariate, simultaneous analysis of hydrophobic descriptors by PCA may provide valuable information about the data structure (dimensionality of two, two common components). Correlated parameters from different solvent-water systems and phenyl series have been transformed into uncorrelated "inner" variables discriminating between the systems and leading to suggestions about underlying interactions. Via identification of such interactions per multiple linear regression analysis, the different impact of hydrogen bonds in nonpolar and polar solvent-water systems on log P values and their dependence on isotropic and hydrated surface areas has become obvious. The analysis of  $\pi$ -values of *meta*- and *para*-disubstituted benzenes has led to extended symmetric bilinear Hammett-type models relating interaction increments to three cross products  $\pi_X \sigma_Y$ ,  $\pi_Y \sigma_X$  and  $\sigma_X \sigma_Y$ . The resulting models from both approaches provide detailed insight into the nature of hydrophobic descriptors and fall into line with numerous other theoretical investigations on the background of hydrophobicity and lipophilicity.

## References

- [1] Glossary of terms used in computational drug design (IUPAC Recommendations 1997), [http://goldbook.iupac.org/src\\_PAC1997691137.html](http://goldbook.iupac.org/src_PAC1997691137.html) (Dec 13, 2013)
- [2] N. Muller, *Trends in Biochemical Sciences* **17** (1992) 459-463.
- [3] V. Pliška, B. Testa, H. van de Waterbeemd, *Lipophilicity in Drug Action and Toxicology*, VCH Verlagsgesellschaft mbH, Weinheim, 1996.
- [4] M.S. Tute, in: *Lipophilicity in Drug Action and Toxicology*, V. Pliška, B. Testa, H.v.d. Waterbeemd, (Eds.), VCH Verlagsgesellschaft mbH, Weinheim, 1996, pp. 7-26.
- [5] C. Hansch, A. Leo, *Substituent Constants for Correlation Analysis in Chemistry and Biology*, Wiley & Sons, New York, 1979.
- [6] C. Hansch, A. Leo, D. Hoekman, *Exploring QSAR: Hydrophobic, Electronic, and Steric Constants*, American Chemical Society, Washington, DC, 1995.
- [7] T. Fujita, J. Iwasa, C. Hansch, *Journal of the American Chemical Society* **86** (1964) 5175-5180.
- [8] A. Leo, P.Y.C. Jow, C. Silipo, C. Hansch, *Journal of Medicinal Chemistry* **18** (1975) 865-868.
- [9] R. Collander, *Acta Chemica Scandinavica* **3** (1949) 717-747.
- [10] A.J. Leo, *Chemical Reviews* **93** (1993) 1281-1306.
- [11] G.G. Nys, R.F. Rekker, *Chimie Therapeutique* **8** (1973) 521-535.
- [12] G.G. Nys, R.F. Rekker, *European Journal of Medicinal Chemistry* **9** (1974) 361-375.
- [13] A.A. Petrauskas, E.A. Kolovanov, *Perspectives in Drug Discovery and Design* **19** (2000) 99-116.
- [14] C. Kramer, B. Beck, T. Clark, *Journal of Chemical Information and Modeling* **50** (2010) 429-436.
- [15] S. Dove, R. Franke, R. Kühne, in: *Chemical Structure - Biological Activity Relationships: Quantitative Approaches*, J. Knoll, F. Darvas, (Eds.), Akademiai Kiado, Budapest 1980, pp. 247 - 255.
- [16] W.J. Dunn, S. Wold, *Acta Chemica Scandinavica Series B-Organic Chemistry and Biochemistry* **32** (1978) 536-542.
- [17] W.J. Dunn, M.G. Koehler, S. Grigoras, *Journal of Medicinal Chemistry* **30** (1987) 1121-1126.
- [18] M.G. Koehler, S. Grigoras, W.J. Dunn, *Quantitative Structure-Activity Relationships* **7** (1988) 150-159.
- [19] J. Hine, P.K. Mookerjee, *Journal of Organic Chemistry* **40** (1975) 292-298.
- [20] C.M. Hansen, K. Skaarup, *Journal of Paint Technology* **39** (1967) 511-514.
- [21] A.F.M. Barton, *Chemical Reviews* **75** (1975) 731-753.
- [22] B. Testa, P.-A. Carrupt, P. Gaillard, R.-S. Tsai, in: *Lipophilicity in Drug Action and Toxicology*, V. Pliška, B. Testa, H.v.d. Waterbeemd, (Eds.), VCH Verlagsgesellschaft mbH, Weinheim, 1996, pp. 49-71.
- [23] A. Leo, in: *Lipophilicity in Drug Action and Toxicology*, V. Pliška, B. Testa, H.v.d. Waterbeemd, (Eds.), VCH Verlagsgesellschaft mbH, Weinheim, 1996, pp. 157-172.
- [24] To get appropriate scales, ISA and HSA values from Table 1 of the paper of Koehler et al. were divided by 100; f(HSA) was calculated as HSA/(ISA+HSA).
- [25] S. Grigoras, A New Semiempirical Method for the Calculation of the Partition Coefficient, PhD Thesis, University of Illinois at Chicago, Chicago, 1985.
- [26] M.G. Koehler, A New Direction in Partition Coefficient Calculations: A Molecular Hydration Approach, PhD Thesis, University of Illinois at Chicago, Chicago, 1987.
- [27] M.H. Abraham, *Chemical Society Reviews* **22** (1993) 73-83.
- [28] M.J. Kamlet, J.L.M. Abboud, M.H. Abraham, R.W. Taft, *Journal of Organic Chemistry* **48** (1983) 2877-2887.
- [29] M.J. Kamlet, M.H. Abraham, R.M. Doherty, R.W. Taft, *Journal of the American Chemical Society* **106** (1984) 464-466.
- [30] M.J. Kamlet, R.M. Doherty, M.H. Abraham, Y. Marcus, R.W. Taft, *Journal of Physical Chemistry* **92** (1988) 5244-5255.
- [31] M.J. Kamlet, R.W. Taft, *Journal of the American Chemical Society* **98** (1976) 377-383.
- [32] N. El Tayar, B. Testa, P.A. Carrupt, *Journal of Physical Chemistry* **96** (1992) 1455-1459.
- [33] R.W. Taft, M.H. Abraham, R.M. Doherty, M.J. Kamlet, *Nature* **313** (1985) 384-386.
- [34] B. Testa, P. Seiler, *Arzneimittel-Forschung/Drug Research* **31-2** (1981) 1053-1058.
- [35] D.E. Leahy, *Journal of Pharmaceutical Sciences* **75** (1986) 629-636.

- [36] C. Hansch, P.P. Maloney, T. Fujita, *Nature* **194** (1962) 178-180.
- [37] R. Hagemann, A. Meyer, A. Barth, R. Franke, *Pharmazie* **32** (1977) 526-529.
- [38] R. Franke, S. Dove, R. Kühne, *European Journal of Medicinal Chemistry* **14** (1979) 363-372.
- [39] R. Franke, A. Gruska, in: *Chemometric Methods in Molecular Design*, H.v.d. Waterbeemd, (Ed.), VCH Verlagsgesellschaft mbH, Weinheim, 1995, pp. 113-163.
- [40] T. Fujita, in: *Progress in Physical Organic Chemistry*, R.W. Taft, (Ed.), John Wiley & Sons, New York, 1983, pp. 75-113.
- [41] V.A. Palm, *Quantitative Theory of Organic Reactions*, Khimiya, Leningrad, 1977.

©2014 by the authors; licensee IAPC, Zagreb, Croatia. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>) 