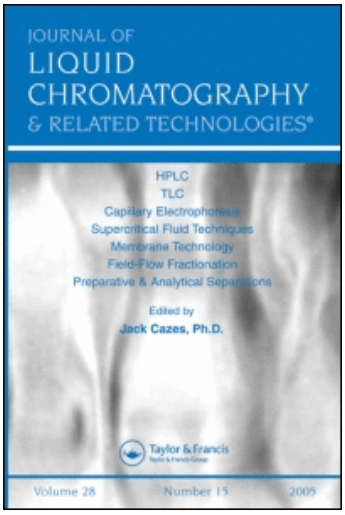


Journal of Liquid Chromatography & Related Technologies
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597273

Structure-Retention Correlation in Liquid Chromatography for Pharmaceutical Applications

Victor David[a]; Andrei Medvedovici[a]
[a] Department of Analytical Chemistry, University of Bucharest, Bucharest, Romania

To cite this Article David, Victor and Medvedovici, Andrei(2007) 'Structure-Retention Correlation in Liquid Chromatography for Pharmaceutical Applications', Journal of Liquid Chromatography & Related Technologies, 30: 5, 761 — 789
To link to this Article: DOI: 10.1080/10826070701191052
URL: http://dx.doi.org/10.1080/10826070701191052

Journal of Liquid Chromatography & Related Technologies®, 30: 761–789, 2007

ISSN 1082-6076 print/1520-572X online
DOI: 10.1080/10826070701191052

Structure-Retention Correlation in Liquid Chromatography for Pharmaceutical Applications

Victor David and Andrei Medvedovici
Department of Analytical Chemistry, University of Bucharest, Bucharest, Romania

Abstract: This review covers the most important aspects, focused on an actual topic in fundamental chromatographic research, which attempts to correlate the retentions of analytes in liquid chromatography with their structures, using various molecular descriptors/properties (hydrophobicity given by octanol/water partition coefficient, acidity/basicity constants, dipole moment, polarity parameters, molecular volume, and solubility). Some of these correlations are, however, based on retention mechanisms which are supposed to describe, thoroughly, the chromatographic process, such as the linear solvation energy relationship.

Keywords: LC retention, QSRR, Pharmaceutical compounds, Molecular descriptors, Hydrophobicity, Dissociation constants, LSER, Review

INTRODUCTION

Theoretical studies referring to Quantitative Structure-Activity (property) Relationship (QSAR, or QSPR, respectively) are recognized as fast and powerful approaches for correlation and prediction of physical, chemical, and even biological properties of the chemical compounds with respect to their structures. For instance, comparative QSAR on non-benzodiazepine compounds binding to benzodiazepine receptor has been extensively discussed by Hansch and coworkers in a recent review.[1] Various chemo-informatic techniques, such as virtual filtering and screening of combinatorial

Address correspondence to Andrei Medvedovici, Department of Analytical Chemistry, University of Bucharest, Faculty of Chemisty, Sos. Panduri, No. 90, Bucharest (5), 050663, Romania. E-mail: avmedved@yahoo.com

libraries, have been also developed relying on QSAR.[2] A large number of methods are available for modeling QSAR. A recent review examined the predictive accuracy of several methods applied to data sets of inhibitors for angiotensin converting enzyme, acetyl cholinesterase, benzodiazepine receptor, cyclooxygenase-2, dihydrofolate reductase, glycogen phosphorylase b, thermolysin, and thrombin.[3] Descriptors calculated with CoMFA, CoMSIA, EVA, HQSAR, and traditional 2D and 2.5D descriptors were used by authors for developing models with partial least squares. This topic also includes quantitative structure-retention relationship (QSRR) studies which are focused on the chromatographic process and, consequently, on the retention properties of chemical compounds.

QSRR is often related to another topic in liquid chromatography (LC), namely the retention mechanism(s), and vice versa. The correlations between the structures of the compounds and their retention behavior in liquid chromatography can offer analytical advantages, a better insight on the basic mechanisms appearing during the LC separation, and extra-analytical information about different species involved in the chromatographic process. Prediction, as the principal goal in liquid chromatography, is desirable in order to avoid time-consuming trials.[4] In the case of new stationary phases, such a theoretical approach seems to represent the unique possibility in predicting the chromatographic behavior of pharmaceutical compounds. Thus, the equations predicting LC retention of 18 dihalogeno-benzoyl-phenyl-ureas and similar compounds were established for a new type of stationary phase (polystyrene-octadecene-encapsulated zirconia). Prediction of the chromatographic retention was based on dipole moments, molar refractivities, and hydrophobicity parameters.[5]

Although the QSRR studies have been initiated several years before 1990, in this review we will discuss only some major contributions to this topic brought after this date. Taking into consideration the importance of the pharmaceutically active compounds and the necessity of their determination in more or less complex matrices by means of LC techniques, this review evaluates the recent literature reports on this subject or those having potential applications in the future. Some reviews focused on QSRR in thin-layer chromatography (TLC) are also valuable for this subject.[6,7]

GENERAL CONSIDERATIONS

The experimental parameter describing LC retention, used for a quantitative correlation with the structures of the analytes, is definitely the capacity factor (k′). The structures of the analytes should be "quantified" by means of some (i) molecular properties or descriptors (MD).

$$K' = f(MD_i) \tag{1a}$$

Additionally, the results of the chromatographic separation may be completed by means of other experimental parameters, such as the peak asymmetry (AF),

and efficiency (N). The accuracy of these experimental parameters has major implications for quantitative QSRR.[8]

$$AF = f(MD_i); \quad N = f(MD_n) \tag{1b}$$

The capacity factor (k') is related to the partition constant (K) or to the partition coefficient (D) by means of well-known formulas:

$$k' = K \cdot \frac{V_s}{V_m} \tag{1c}$$

$$k' = D \cdot \frac{V_s}{V_m} \tag{1d}$$

where V_s and V_m are the volumes of the stationary phase and mobile phase, respectively. Consequently, some practical problems arise for the accurate estimation of V_s and V_m.

As k' is usually determined in the chromatogram according to relationship,

$$k' = \frac{t_R - t_0}{t_0} \tag{1e}$$

similar difficulties in determining, exactly, the dead time (t_0) should be considered. Several proposals for estimating these parameters have been recently reviewed.[9]

Usually, QSRR models are based on the construction of predictive models using a set of known molecules and associated retention values. The mathematical models can be generated using a wide variety of methods, ranging from linear methods (e.g., linear regression and linear discriminant analysis) to nonlinear methods. Generally, the validation of a QSAR model is an important concern when extended to new compounds, and can be treated by means of different classification methods, but not so far in LC.[10]

QSRR studies are oriented especially for tested analytes, using different molecular descriptors and mathematical tools. QSRR studies with a given stationary phase are, however, characterization studies with analytical purposes in almost all cases, unless they describe the interaction with the structure of the solute during the retention process. To obtain a QSRR, the differences of the intermolecular interactions between the stationary phase and a structurally defined analyte rationalize the observed differences in terms of retention. Once the QSRR study is produced for a given stationary phase and for a set of model solutes (training set), then it can be used to predict the retentions of other compounds with a defined structure.[11] Obviously, the composition of the mobile phase has its own importance, and, consequently, should be taken into consideration in developing different solvation models for the solutes. A schematic diagram for developing a QSRR approach is shown in Fig. 1.

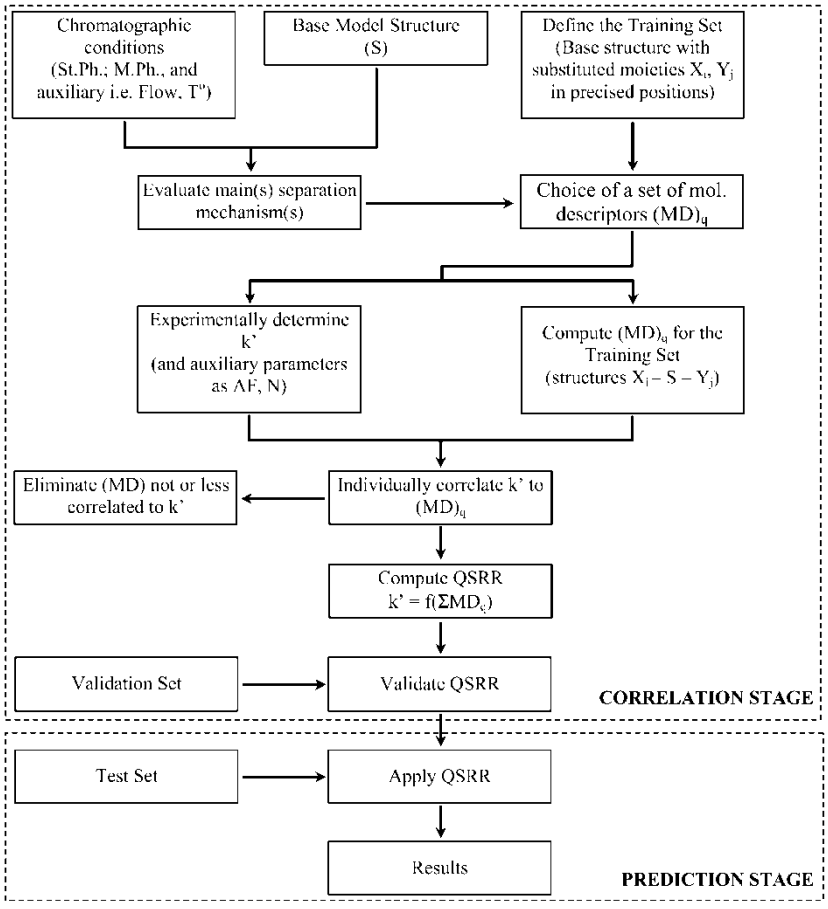


Figure 1. Schematic representation of a QSRR approach.

MOLECULAR DESCRIPTORS

Molecular descriptors are terms that characterize the chemical information contained within a symbolic representation of the molecule or the result of a standardized experimental measurement of a molecular attribute.[12]

Molecular descriptors should be unique, interpretable, independent, relevant for the expected correlation, and readily amenable for calculation.

Most molecular descriptors may be classified according to their "dimensionality" referring to the representation of the molecule from which they derive (1D, 2D, 3D).[13] Molecular descriptors should be considered as global or local ones, as they represent the expression of the whole or a part of the molecule.

The molecular descriptors most frequently correlated with the chromatographic retention are summarized in Table 1. Descriptors correlated to retention in liquid chromatography should be meaningful with respect to the separation mechanism.[14] An exhaustive approach related to molecular descriptors used in QSAR and QSPR is found in Ref. [15].

The most important problem of QSPR, and particularly QSRR. studies is the mathematical representation of the chemical structure by means of well-defined molecular descriptors and the translation of the molecular structure to a computer-readable form.[26] For the creation of the prediction model, one has to recognize and extract critical structural information that is relevant to a certain structure-property relationship. Recently, significant progress was made in the development of various topological, geometrical, electrostatic, and quantum chemical indices, to be used as molecular descriptors.[27–30] Nevertheless, this level has been attained only in a few theoretical works focused on the LC separation process. For instance, Hanai investigated the basic phenomena in RP-LC using a computational chemical method (version 5 of CAChe program, Fujitsu, Tokyo, Japan).[31] The molecular properties of the analytes were the bond stretch, the bond angle, the dihedral angle, the improper torsion, the van der Waals forces at a corresponding cut-off distance of 9 Å, the hydrogen bonding, and the electrostatic forces (denoted by MM2/MM3 bond dipoles). Thus, it has been proved that interaction energy between two test solutes (pyridine and phenol) and the surface of bonded-phase silica, under neutral pH conditions (pK_a of pyridine is 5.25, pK_a of phenol is 10.02), could explain the elution order of this pair of solutes ($k'_{phenol} > k'_{pyridine}$), although they have an opposite log P (log P for phenol is 1.54, while log P for pyridine is 0.70). By means of this method, quantitative structure-retention relationships in RP-LC were demonstrated for phenolic compounds, and for acidic and basic drugs. According to this study, the interaction energies (denoted by author as ΔFS, in kcal/mol) are correlated with the capacity factor of the undissociated form of 19 acidic tested drugs by means of the following regression:

$$\Delta FS = 6.483 \cdot \log k' + 23.145 \quad r = 0.878 \tag{2}$$

For a series of 17 basic drugs this regression equation was following:

$$\Delta FS = 7.618 \cdot \log k' + 20.924 \quad r = 0.941 \tag{3}$$

Models for the adsorption and interaction of warfarin and hematropine on a pentyl-bonded silica surface were shown.[31] Representation of covalently modified surfaces through molecular dynamics is challenging and, up to the present, only a few efforts have been undertaken.[32] Even relatively small stationary phase models contain several thousand atoms and, thus, the systems become computationally very intensive. The identification of π-π and dipole-dipole solute-stationary phase interactions and the assessment of

Table 1. Molecular descriptors more often used in QSRR studies

#	Name	Symbol	Classification	Meaning	Formula	Associated separation mechanisms	Ref.
1	Hydrophobicity	log P	1D, global	Express the affinity of a molecule for a hydrophobic environment, measured by its distribution behavior in a biphasic system (1-octanol/water)	$\log K_{ow} = \log \frac{[A]_o}{[A]_w}$	RP	
2	Hydrophobic substitution constant	π	1D, local	Express the influence of the substituent X on the K_{ow} of the base compound Y	$\pi = \log \frac{K_{ow}^{X\text{-}Y}}{K_{ow}^{Y}}$	RP	[16]
3	Equivalent carbon number	ECN	1D, global	Express the degree of unsaturation of the molecule	ECN = NC − 2 × NDB NC = no. of C atoms NDB = no. of double bonds	RP	[17]
4	Separation number	SN	1D, global	Express the degree of unsaturation of the molecule	SN = NC − NDB	NP, argentation	[18]
5	Acidity	pKa	1D, global electronic	Express the proton donating characteristics of the molecule	$pK_a = \log \frac{[A^-][H^+]}{[AH]}$	RP, IP-RP, ion exchange	
6	Hammet's constant	σ	1D, local, electronic	Express the influence of the substituent X on the K_a of the base compound Y (reflecting its electron-donating or – accepting properties)	$\log \frac{K_a^{X-Y}}{K_a^{Y}} = \sigma \times \rho$ ρ = prop. constant	IP-RP, Ion exchange Ligand exchange	[19]

7	Molar refractivity	MR	1D, global	Express the real volume of the molecules contained within a mole of substance for which the refractive index is measured at a wavelength extrapolated to infinity	$MR = (n^2 - 1)/(n^2 + 2) \times M/\rho$ M = molecular weight; n = refractive index; ρ = density	RP	[20]
8	Molar polarizability	MP	1D, global	Express the real volume of the molecules cotained within a mole of substance when instantaneus dipoles of an approaching ligand deformate the molecules	$MP = (D - 1)/(D + 2) \times M/\rho$ D = dielectric constant of the environment	RP	[20]
9	Length/breadth ratio	L/B	2D, global, shape	Express the ratio of the lenght to breadth of the maximized rectangle enclosing the molecule	$L/B = L/1$	RP, electron density transfer	[21]
10	Correlation factor	F	2D, global, topological	Express π electron density over the planar representation of the molecule	$F = NDB + NC(p,s) - 0.5 \times NCS$, NC(p,s) = no. of carbon atoms, primary and secondary; NCS = no. of cycles without aromatic character	RP, electron density transfer	
11	Connectivity index	χ	2D, topological	Express the possibility of each bond to encounter another bonds of the same molecule in a given media	$\chi = \sum (\delta_i\ \delta_j)^{-0.5}$ $\delta_{i,j}$ = degree of the vertex representing atoms i, j related through a chemical bond	RP	[22]

(*continued*)

Table 1. Continued

#	Name	Symbol	Classification	Meaning	Formula	Associated separation mechanisms	Ref.
13,14	van der Waals volume (area)	V_W A_W	3D, global, geometrical	Express the volume (surface) of a molecule considered as the addition of van der Waals volumes (surfaces) corresponding to constituing atoms taken as rigid spheres	$V_W = \frac{4\pi r_w^3}{3}$; $A_w = 4\pi r_W^2$ r_W = atomic van der Waals radius	RP, NP, steric, inclusion	[23]
15	Solvent accesible surface area	SASA	3D, global, geometrical	Express the external area of a molecule accessible to a solvent molecule considered as rigid sphere rolled over such surface	Computer assisted	RP, NP	[24]
16	Highest occupied molecular orbital energy	HOMO	3D, quantum-chemical	Express the ionization potential of a molecule and its reactivity as a nucleophile	Computer assisted from molecular orbital calculations	IP-RP	[25]
17	Lowest unoccupied molecular orbital energy	LUMO	3D, quantum-chemical	Express the electron affinity of a molecule and its reactivity as an electrofile	Computer assisted from molecular orbital calculations	Electron density transfer	[25]

their relative importance in affecting the retention process on cyano and phenyl columns in RP-LC were very recently pointed out.[33,34]

The need for fast and accurate predictors of pharmaceutically important properties has been increasing, owing to high-throughput screening, in-silico screening, and the need for rapid identification of potential pharmacokinetic issues before drugs progress through superior, more expensive clinical developmental stages. A novel method for making predictive models, based on decomposing two-dimensional structures into component structural fragments, has been proposed recently, which is based on the model of log P, water solubility, and melting point.[35]

The importance of molecular descriptors in chromatographic practice is obvious: some data on solubility, dissociation constants, and hydrophobicity (either experimental value, or computed by means of fast theoretical methods) are the primary information used in developing a liquid chromatographic method.

Hydrophobicity

Hydrophobicity is probably the most important molecular descriptor that is taken into consideration in QSRR studies. Quantitatively, this parameter gives the measure of partitioning between two immiscible phases: aqueous (denoted by the subscript index w) and 1-octanol (denoted by the subscript index o). This parameter, known as log P or log $K_{o,w}$, is strongly related to the hydrophobic interactions. It has been recently reported that protiated compounds bind to nonpolar moieties attached to silica more intensively, compared to deuterated ones.[36] The interactions responsible for binding have been characterized by studies of the effects of changes in mobile phase composition, temperature dependence of binding, and QSRR analysis, demonstrating the importance of enthalpic effects in binding and differentiation between the isotopologues.

Often, the hydrophobic character is replaced by the lipophilicity descriptor, although the lipophilic character seems to be a broader one than hydrophobicity. For instance, Avdeef proposed three indices of lipophilicity: liposome/aqueous partition system; immobilized artificial membrane/water; and the simplest partition system consisting in 1-octanol/water.[37] Recently, Katritzky et al. reported a QSPR study on an alternative approach to log P; they studied the partitioning process of a significant number of small organic compounds within a biphasic system consisting of poly(ethylene glycol) and ammonium sulfate aqueous phase.[38] Structure-lipophilicity relationship for a number of β-blockers was measured and compared by two-phase titration, centrifugal partition chromatography, and cyclic voltammetry in one or more of the following water/solvent systems: octanol, 1,2-dichloroethane, and dibutylether.[39]

Questions arising from the experimental determination of log $K_{o,w}$ of a given structure are the following: which is the lower limit of log $K_{o,w}$ for considering a compound as hydrophobic? Below this limit, can a compound be considered as hydrophilic? The answers to these questions, in connection with the chromatographic behavior, is rather relative. In RP-LC the capacity factor k' is directly proportional to $K_{o,w}$ in accordance with the Eqs. (1)c, 1d). Solutes characterized by low log $K_{o,w}$ values, or even exhibiting negative values, can have measured k' only using mobile phases with an increased content of the aqueous component. Moreover, this parameter can be estimated rather accurately by means of RP-LC; this approach supposes an accurate mathematical description of the retention behavior of solutes. The reliability of the capacity factor estimates in LC and, consequently, the reliability of the extrapolations, have been rarely discussed in the literature. Conventional protocols for estimating k' have problems that mainly arise from difficulties in the hold-up time measurements and the omission of the existence of extra-column times. Several authors analyzed this problem and proposed an approach based on the use of an external standard, leading to "relative" capacity factors.[40]

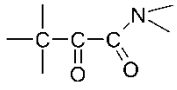
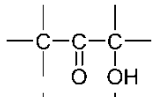
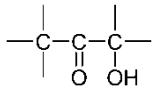
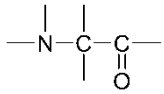
Experimental values for $K_{o,w}$ are determined for a wide variety of compounds, although there is still a certain level of uncertainty in reported $K_{o,w}$ values, as well as for the water solubility of various classes of organic compounds, a point of view emphasized by Renner.[41] The separation methods used for the indirect estimation of $K_{o,w}$ have been recently reviewed by Poole.[42] Correlation between chemical structures and $K_{o,w}$ values leads to empirical equations, which can be used in predicting this parameter for a given structure.[43] The procedure is known as the fragment methodology and several databases are already known, as well as the mathematical approximations. For instance, one of the most useful equations is written as follows:[44]

$$\log K_{o,w} = \sum_{i=1}^{n} n_i \log K_{o,w}^{(i)} + \sum_{j=1}^{m} \Phi_j + \varsigma \qquad (4)$$

where: n_i represents the number of fragments of the same type i, having log $K_{o,w}^{(i)}$; Φ_j – the factor correction for different groups, and ζ – the equation constant (usually log is taken as ten-base logarithm). The constant in Equation (4) is imposed in order to linearly correlate the experimental values of log $K_{o.w}$ with theoretical values, predicted by this methodology. Several, very common examples of hydrophobic and hydrophilic functional groups together with their contribution to log $K_{o,w}$ of a compound are given in Table 2.

An alternative for estimation of log P results from the plot of the capacity factor as function of the organic component(s) content in the mobile phase, by extrapolation to 0%. Moreover, insights on the modification of the capacity factor on the concentration of the organic solvent(s) (modifier) in the

Table 2. log $K_{o,w}$ values for several commonly used fragments by the fragment methodology

Hydrophobic fragment	log $K_{o,w}^{(i)}$	Hydrophilic fragment	log $K_{o,w}^{(i)}$	Corrections	Φ_j
-CH_3 (aliphatic)	+0.5473	-OH (to sp^3 C)	−1.4086	Two non-geminal hydroxil groups	+0.4064
-CH_2- (aliphatic)	+0.4911	-OH (to aromatic C)	−0.4802	Three vicinal hydroxil groups	+0.5944
=CH- (olefinic)	+0.3836	-NH_2 (to sp^3C)	−1.4148	Oxime structure -CH=N-OH	−1.3000
≡CH (aliphatic)	+0.3614	-NH_2 (to aromatic C)	−0.9170		+0.9755
aromatic C	+0.2940	-COOH (to sp^3 C)	−0.6895		+0.9178
-Cl (to aliphatic C)	+0.3102	-COOH (to aromatic C)	−0.1186		+0.4000
-Cl (to aromatic C)	+0.6445	-CH=O (to aliphatic C)	−0.9422		+0.7616
-Br (to aliphatic C)	+0.3997	-CH=O (to aromatic C)	−0.2828	Equation constant (ζ)	0.2290
-Br (to aromatic C)	+0.8900	-CO- (to alifatic C)	−1,5586		

mobile phase ($k' = f(C_s)$, C_s is expressed as a volume fraction) is useful for several reasons:

a) synthetically expresses the ratio between the major forces (hydrophobic/solvation) taking part in the retention process;
b) allows prediction of the retention and selectivity between target analytes without too many experiments;
c) could be used for evaluation of the variation of retention corresponding to accidental minor changes in the mobile phase composition when studying the robustness of the chromatographic separation;
d) represents an alternative for estimation of the analyte distribution constants between the organic solvent/stationary phase and aqueous

component/stationary phase (hydrophobicity), by extrapolation of the experimental function $f(C_s)$ to the limits of $C_s = 1$, or $C_s = 0$, respectively.

The function $f(C_s)$ can be derived from the theoretical models applied to the retention process.[45] However, two empirical relationships are known, from practice, to describe the dependence of the capacity factor (k') in RP mechanism on the organic solvent (modifier) content in the mobile phase:

$$k' = \sum_{i=0}^{n} \alpha_i C_s^i \tag{5}$$

$$\log k' = \sum_{j=0}^{m} \gamma_j C_s^j \tag{6}$$

Many of the experimental studies of retention and selectivity patterns on different reversed-phase columns demonstrated that the commonly used assumption of a linear relationship between $\log k'$ and the mobile phase composition is a simplification (m from Eq. (6) becomes 1), and can be valid only for a narrow interval of C_s.[46–48]

The regression coefficients α_i ($i = 0, \ldots, n$) or γ_j ($j = 0, \ldots, m$) can be estimated from the experimental dependence of k' or $\log k'$ on C_s. For a narrow interval of C_s, these dependences are reduced to a linear fit but, for a wide interval of C_s, they become polynomial functions of degree indicated in sum index from the above equations (n, or m). Extrapolation parameters for these dependences are the following:

a) For $C_s = 0$, the capacity factor corresponding to a mobile phase composed of only water, denoted by $k'_{o,w}$, is obtained. In a fair approximation agreed by almost all researchers involved in chromatography, it is taken as $k'_{o,w}$. In such a case, Equations (5) and (6) give the two extrapolated parameters, $k'_{o,w} = \alpha_0$, or $k'_{o,w} = 10^{\gamma_0}$, respectively. In its turn, the capacity factor is correlated with the octanol/water partition coefficient assigned to the analyte ($K_{o,w}$) using Eq. (1c). In accordance with the above remarks, the retention experiments should be performed in a concentration range of C_s as close to 0 as possible.
b) For $C_s = 1$, the sum of all regression parameters from the above dependences is obtained, i.e., $k'_{C_s=1} = \sum_{i=0}^{n} \alpha_i$, or $\log k'_{C_s=1} = \sum_{i=0}^{m} \gamma_i$, respectively. This is a less discussed situation, which offers an overview on the ratio of the solubilities of the analyte in 1-octanol and the organic solvent used in the mobile phase, respectively.[49]
c) For $k' = 1$, or $\log k' = 0$, the values of C_s for which $t_r = 2 \cdot t_0$ is obtained. This may indicate an upper limit of the organic modifier in the mobile phase leading to an acceptable retention of the solute.

The most delicate problem consists in the generation of an accurate functional dependence between k′ and the organic modifier content (C_s) in the mobile phase, just in the proximity of $C_s = 0$. Therefore, the fitting algorithm is of great importance in QSRR studies. According to Howkins, two subclasses of relationship problems usually arise: correlation problems and regression problems, based on a dependent variable and independent variables (denoted also as predictor variables or covariates). Interpolation (within the studied intervals) and extrapolation (outer of the studied interval) are mathematical procedures highly related to these studies.[50]

The study of the retention behavior in RP-LC and its correlation with the structures of the analytes is largely discussed in the literature. For example, the k′ values determined by RP-LC on a C_{18} column for 18 substituted indoles were correlated with the methanol content in the mobile phase. The molecular connectivity indices and quantum chemical parameters were calculated for tested compounds and used to develop QSRR.[51] Retention parameters of 45 barbituric acid derivatives were determined on an amide embedded RP silica column using non-buffered water-dioxan mobile phase.[52,53] Six retention parameters (intercept, slope, the combined retention parameter – intercept/slope, asymmetry factor, and theoretical plates according to the USP and Japanese Pharmacopoeia) were correlated with different conventional and quantum structural descriptors using QSRR. Multilinear regression analysis and principal component analysis (PCA), followed by two-dimensional nonlinear mapping and cluster analysis techniques were used to determine the retention behavior of barbituric acid derivatives. The significant effect of the hydrophobic characteristics of studied solutes estimated from retention behavior indicated that the effects of the interaction between these solutes and the residual silanol groups are negligible.

A QSRR study for RP-LC separation of amiloride, hydrochlorothiazide, and methyldopa, using the artificial neural networks (ANNs) modeling, was achieved in order to predict the separation of amiloride and methylclothiazide from different formulations.[54] The same mathematical procedure was applied to quantitative structure-gradient elution retention relationship of 18 selected amino acid derivatives.[55] The molecular structure of each amino acid was encoded with 36 calculated molecular descriptors. The application of the second most popular artificial neural networks, namely the radial basis function networks (RBFNs) has been developed by Loukas for obtaining sufficient QSRR with improved accuracy.[56] RBFNs method based on quantum chemical parameters (dipole moment, energies of the highest occupied and lowest unoccupied molecular orbitals, net charge, total energy of the molecule) was also used in predicting the LC retention of bifunctionally substituted N-benzylideneanilines.[57]

QSRR models were developed for the prediction of protein retention times in anion-exchange chromatography. Topological, subdivided surface area, and electron-density-based descriptors have been computed directly

for a set of proteins using molecular connectivity patterns and crystal structure geometries.[58] The algorithm based on support vector machine (SVM) regression may obtain predictive QSRR models using a two-step computational strategy. Seven molecular descriptors, selected by the heuristic method in CODESSA, were used as inputs for SVM in the framework of a work for quantitative prediction of log k′ of peptides by LC.[59]

Other QSRR studies reported in the literature were focused on: structurally diverse drugs separated on phospholipid modified stationary phases;[60] or monolithic ones,[61] methionine-enkephalin-related glycoconjugates as a function of the identity and position of the sugar-peptide linkage;[62] 9 corticosteroids with micellar electrokinetic chromatography (MEKC);[63] isomeric bile acids in inclusion HPLC with methyl β-cyclodextrin;[64] 33 purine nucleobases in RP-LC;[65] β-blockers with different RP stationary phases;[66] or non-steroidal anti-inflammatory drugs by using PCA based on TLC data.[67]

Acid/Base Character

Generally, an ionic form of an analyte elutes at lower retention time, compared to the undissociated form. This rule can be explained only by applying the partition model to the chromatographic process.[68,69] For this purpose, the retention behavior of an acidic compound, denoted by R-X-H, where R represents a structure with hydrophobic character, while X belongs to the ionizable functional group -X-H (for instance, X may be O, S, COO, SO_3, OSO_3, etc.) is considered; the retention process is based on the following equilibria taking place in the mobile phase (m) or at the interface between the mobile phase and the stationary phase(s) consisting in the hydrophobic coverage bound to silica gel based material:

$$(R - X - H)_m \rightleftarrows (R - X^-)_m + (H^+)_m \quad (7)$$

$$(R - X - H)_m \rightleftarrows (R - X - H)_s \quad (8)$$

The first equilibrium is characterized by acidity constant ($K_{a,i}$), and the second equilibrium is characterized by the partition constant of the indicated species i ($K_{d,i}$). A first assumption deduced from this simple representation of the retention process is the lack of partition of the dissociated form R-X^- towards the stationary phase.

Overall, the retention process is described by the partition coefficient (D_i) having the formula deduced from the above two equilibria and their constants as follows:

$$D = \frac{K_{d,i} \cdot [H^+]}{K_{a,i} + [H^+]} = \frac{K_{d,i}}{1 + 10^{pH - pK_{a,i}}} \quad (9)$$

Thus, it has been proved that at least three parameters play a major role in the retention process of dissociable compounds. Two of them are molecular descriptors ($K_{d,i}$ and $pK_{a,i}$) and the other is a parameter of the mobile phase

(pH). In the case of an ion-pairing mechanism, the partition of the species is based on equilibria (7) and (8), and two additional ones:

$$(R - X^-)_m + (R' - B^+)_m \rightarrow (R - X^{-+}B - R')_m \quad (10)$$

$$(R - X^{-+}B - R')_m \rightleftarrows (R - X^{-+}B - R')_s \quad (11)$$

where R-B^+ represents the ion-pairing agent. The mathematical formalism of the partition equilibria in reversed-phase, with or without an ion-pairing mechanism, explains, satisfactorily, the dependence of k′ on pH and log $D_{o,w}$.[70,71]

In the case of dissociable solutes, the role of $K_{o,w}$ is taken by $D_{o,w}$ (partition coefficient). The relationship between them is given by the following equation:

$$D_{o,w} = \frac{K_{o,w}}{1 + 10^{pH-pK_a}} \quad (12)$$

This equation gives the shape of the dependence of k′ on pH; the sigmoid shape is characterized by an inflexion point, which corresponds to pK_a. This property was used in a series of works to estimate the pK_a values for many drugs,[72] and requires an isocratic elution with a mobile phase containing an aqueous component with constant pH. Elution in RP-LC may also be realized based on pH gradient. On the analogy of the conventional organic modifier gradient, in the pH gradient mode, the eluting strength of the mobile phase increases due to its increasing (with acid analytes) or decreasing (with basic analytes) pH, while the content of the organic modifier is kept constant.[73] This pH gradient mode of elution was used in determining pK_as for a set of solutes in comparison with elution at constant pH, among them being pharmaceutical compounds, such as barbituric acid, warfarine, and codeine.

In the case of basic compounds (R−Y, where Y is very often an amino group), the equilibrium can be written as follows:

$$(R - Y)_m + HOH_m \rightleftarrows (R - YH^+)_m + (HO^-)_m \quad (13)$$

Such an equilibrium is characterized by means of the basicity constant ($K_{b,i}$). Usually, the molecular data refer to acidity constant, although the compound is a base. Therefore, it is important to calculate the basicity constant according to the well-known relationship:

$$pK_{b,i} = 14 - pK_{a,i} \quad (14)$$

For relatively simple structures, the methods for predicting pK_a values rely on modern computer-based molecular modeling, such as molecular orbitals.[74] Several examples are useful for this subject: Pallas 3.0; ACD; Sparc; or Discon. However, when the experimental pK_a values are available, theoretical approaches can only be verified for their accuracy.

Systematic studies of acid/base parameters in RP-LC require an accurate knowledge of the pH of the mobile phase containing organic modifiers.

The relationship between the pH value resulting in organic solvent/water mixtures and the pH measured in the aqueous component of the mobile phase is extensively discussed. Acid-base constants for dissociable solutes in organic solvent/water mixtures can be determined through calibration of the electrode system with pH standards prepared under the same conditions as the mobile phase, or with common pH standards, leading to results in one of the two generally used scales. The variation of the pK_a values for solutes with alkaline character according to the content of the organic solvent in the mobile phase, in either of the two pH scales, is different compared to pK_a variation characterizing acidic solutes. Thus, the pK_a values of acidic solutes increase with the increase of the organic solvent (i.e., acetonitrile) content, whereas the pK_a values of alkaline solutes decrease up to a minimum value and then exhibit an increasing trend.[75]

Test procedures using basic solutes as test probes provided relevant information with respect to column selection for separations of pharmaceutical compounds. The variation of pK_a of codeine, diphenhydramine, nortriptyline, quinine, and nicotine in methanol containing mobile phases was studied while using different stationary phases.[76] An RP-LC study for pK_a determination for a series of alkaline compounds related to caproctamine (a reversible inhibitor of acetyl cholinesterase) at different mobile phase compositions was also performed.[77] Experimental log $D_{o,w}$ values within a wide range (from -2 to 5) for 34 known drugs have been determined by means of liquid chromatography with mass-spectrometry.[78] The effect of ionization and the nature of the mobile phase in QSRR for ionizable and non-ionizable pharmaceutical compounds with different therapeutic effects (10 β-blockers, 7 tricyclic antidepressants, 8 steroids, and 12 sulfonamides) were studied.[79]

The acidic character for some compounds may be a consequence of tautomeric equilibria. Thus, the acid/base behavior of a series of oxicames (meloxicam, piroxicam, and tenoxicam, with tautomeric structures given in Fig. 2) was studied by means of RP-LC.[80]

Estimation of pK_a values (around 4) and correlation with their structures was possible from retention data studies (dependence of the inflexion points of the curve $k' = f(pH)$ on the content of methanol used as organic modifier in the mobile phase and extrapolation to 0). Surprisingly, from the dependences relating extrapolated values of $D_{o,w}$ for 0% methanol to the pH value of the aqueous component of the mobile phase, it results that meloxicam mimes an organic acid, while piroxicam and tenoxicam behave like organic bases. An explanation can be advanced when comparatively considering the basic characteristics of pyridine and methylthiazole rings substituted to the keto-enolic sites.

Polarity Parameters

The polarity parameter (originating from Dimroth-Reichardt polarity parameter) is a molecular descriptor that has been used to describe retention

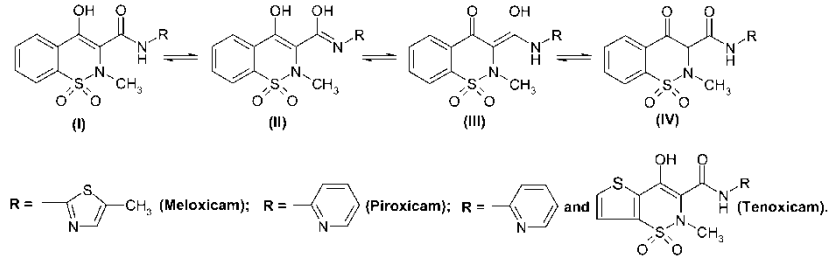


Figure 2. Tautomeric structures of oxicames (anti-inflammatory drugs).

in RP-LC. The equation relating the retention k' to polarity parameters of the solute (denoted by p), mobile phase (P_m^N) and stationary phase (P_s^N) is given by the following relationship:

$$\log k' = (\log k')_{C_s=0} + p \cdot (P_m^N - P_s^N) \tag{15}$$

Retention data reported in the literature have been used by the authors[81] to elaborate a database for p values; the polarity parameter for the mobile phase was deduced to be dependent upon the content of the organic modifier (C_s) in the mobile phase, such as for acetonitrile:

$$P_m^N = 1 - \frac{2.13 \cdot C_s}{1 + 1.42 \cdot C_s} \tag{16}$$

or for methanol:

$$P_m^N = 1 - \frac{1.33 \cdot C_s}{1 + 0.47 \cdot C_s} \tag{17}$$

The polarity parameters were demonstrated to be useful in transferring the retention data between solvent systems and between different stationary phases.[82] The authors developed a QSPR model in order to calculate the solute polarity parameter p for a set of 233 compounds of very different structures. The proposed model, derived from multiple linear regression, contains four descriptors calculated from the molecular structure and $\log K_{o,w}$. According to this study, $\log K_{o,w}$ and hydrogen bond acidities of the solutes are the most relevant descriptors to predict p values, which was embodied in a general equation to predict retention in RP-LC.

A study attempting to correlate the dipole moment of the mobile phase as a major polarity parameter with retention time values was achieved for six esters of nicotinic acid.[83] The linear solvent strength model, combined with QSRR and ANN, has been shown useful in predicting gradient elution in HPLC using three structural descriptors of tested solutes: total dipole moment, electron excess charge of the most negatively charged atom, and water-accessible molecular surface area.[84,85]

Solubility

The most important molecular interaction force in RP-LC is the interaction between the hydrophobic moiety within the analyte structure and the alkyl chains from the stationary phase. This interaction is a combination of van der Waals and London dispersion forces. Hydrogen bonding and electrostatic interactions between polar or ionic groups in the molecule of the analyte and polar centers in the stationary phase play a major role in normal phase liquid chromatographic retention processes (NP-LC). On the other hand, the solubilities of the analytes in the mobile phase represent an important feature in the chromatographic separation process. Therefore, the meaning of $\log K_{o,w}$ or $\log D_{o,w}$ should be accepted as a ratio of the solubility parameters characterizing the solute in water and the stationary phase, respectively. The simple use of these values in QSRR studies for RP-LC separation mechanism sometimes fails, as the solubility of the solute in the real mobile phase (aqueous and organic components together) is ignored.

The relationship between water solubilities of drugs (S) and their $K_{o,w}$ values can be derived from its definition as following:

$$\log \frac{1}{S} = \alpha \cdot K_{o,w} + \beta \tag{18}$$

where α and β are the regression coefficients. This equation can describe, qualitatively, the ratio between hydrophobicity and hydrophilicity; when one of the terms has a high value, the other has o low value, and conversely.

Most of the theoretical models used to describe the solute-solvent interactions are semi-empirical approaches and are based on the principle that all solution-phase processes can be modeled in terms of one or more gas-to-solution transfer processes. According to this model,[86] the free energy of each gas-to-solution transfer process is computed as the sum of the free energy of cavity formation, (where the solute molecule, modeled as a charge distribution, has to be placed), and the free energy of interaction between solute molecule and the surrounding solvent. A good correlation between predicted and experimental values for several solvents was found, among them octanol being of special interest for this purpose.

The literature reports many computational methods for solvation modeling, which can be classified into two main groups: explicit and implicit solvent-solute interaction methods. The explicit interaction methods are based on individual parameters of the molecules involved in the interaction, which supposes a large computational cost. Implicit solvent methods diminish the cost of such computations through an approximate continuum representation of solvation properties. Generally, the implicit solvent methods separate the solvation energetics into polar and nonpolar contributions. For instance, Poisson-Boltzmann and generalized Born models are two implicit solvent methods often used to approximate the polar solute-solvent interactions by representing the surrounding solvent molecules as a

simple dielectric continuum layer. Solvent-accessible surface area models are theoretical approaches to describe nonpolar solute-solvent interactions, which assume that the solvation energy is proportional to surface area.[87] Thus, a hybrid explicit/implicit solvation model has been used for calculating the solvation interaction by taking into account a first solvation shell of water molecules surrounding the alanine peptide.[88]

A fast continuum model for calculation of solvation free energies for large numbers of molecules in three biphasic systems (gas/water, gas/hexadecane and water/octanol) was proposed by Bordner et al.[89] It is based on a continuum electrostatic model with MMFF94 atomic charges combined with a nonelectrostatic term, which is a linear function of the solvent-accessible surface area. The predictive power of this model was verified by the authors using 90% of the molecule set for training and the rest as a test set. The root-mean-square errors for the investigated systems were 0.52; 0.38, and $0.58 \cdot \log P$, respectively.

Considering that the RP-LC mechanism is fundamentally a solvation process in both phases (mobile and stationary phases), Ranatunga and Carr proposed a theoretical method for estimating the free energy contributions (ΔG^0) based on solvation effects.[90] The thermodynamic cycle allowing the dissection of RP-LC retention process was simplified to the following transfer processes:

$$X_{\text{mobile phase}} \xrightarrow{\Delta G_m^0} X_{\text{gas phase}} \xrightarrow{-\Delta G_s^0} X_{\text{stationary phase}} \tag{19}$$

From this cycle, the net retention free energy ($\Delta G^0_{\text{retention}}$) results as:

$$\Delta G^0_{\text{retention}} = \Delta G^0_m - \Delta G^0_s \tag{20}$$

The authors estimated the contribution to the net retention of the free energy of CH_2 groups from a homologous series of alkyl benzenes, with measurements being achieved at different mobile phase compositions.

LINEAR SOLVATION ENERGY RELATIONSHIPS

A complex QSRR model including structural features of all partners involved in the LC separation process (solute, stationary and mobile phases) is based on linear solvation energy relationships (LSERs). During recent years, the LSERs have been applied for description and prediction of retention and selectivity in RP-LC, with or without QSRR studies. A large variety of different stationary phases have been compared and characterized by means of this model. Some publications reported the influence of the type of the organic modifier used in the mobile phase composition. According to the LSER model, the variation of the capacity factor with a property of a solute can be related to its potential against various intermolecular interactions.[91] The k' value is given by the sum of the terms from the LSER model representing various types of

molecular interactions according to the semi-empirical equation (known as Abraham's equation):

$$\log k' = \log k'_{ref} + v \cdot V_X + a \sum \alpha_2^H + b \cdot \sum \beta_2^H + s \cdot \pi_2^H + r \cdot R_2 \quad (21)$$

where V_X represents the solute molecular volume (calculated according to Abraham and McGowan's procedure), $\sum \alpha_2^H$ and $\sum \beta_2^H$ are the solute hydrogen-bond donating and accepting properties, respectively, π_2^H represents the solute polarity/polarizability, and R_2 is the excess molar refraction. The regression coefficients v, a, b, s, and r can be measured as the differences in complementary properties of the solute in the stationary and the mobile phases, respectively. An overall image of the interactions between the solute molecule and the stationary phase in RP-LC is depicted in Fig. 3.

The k'_{ref} value is defined as a contribution of the hydrophobic moiety only to the free-energy change during partition of the analyte between the mobile and the stationary phases; for instance, ethylbenzene could be a proper choice for such a reference. For the characterization of a chromatographic separation system, typically 20 to 50 compounds with known and reliable molecular descriptors are selected and eluted in the isocratic mode, using given stationary and mobile phases in order to determine their capacity factors. In the first stage, the regression coefficients from Eq. (19) are evaluated for a given combination between the stationary and the mobile phases. Then, the same procedure is applied to other sets of stationary and mobile phases. The data are treated by multiple linear regression.[92]

In a series of three publications, Snyder et al. studied, extensively, this model from a theoretical point of view and gave a useful application to the column selectivity in RP-LC.[93–95] They applied the general quantitative LSER equation to a large set of neutral, acidic, and basic solutes (many of

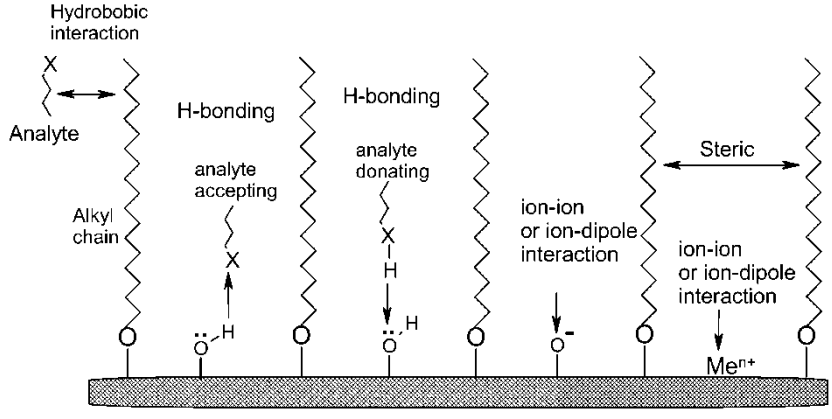


Figure 3. Representation of the main phenomena taking place during RP-LC retention.

them being pharmaceutically active compounds) of highly diverse molecular structures (size, shape, polarity, H-bonding, pK_a) on 10 different stationary phases (alkyl chain length, carbon content, pore diameter, end-capping).

Another work reported that the Abraham solute parameters resulting from the LSER equation characterizing a large set of 457 compounds of very different chemical structures, were studied and estimated by QSPR.[96] Such a method can be derived from multilinear regression analysis and computational neural network and requires several molecular descriptors. According to this model, the structure and the properties of the stationary phase, the type and the composition of the mobile phase, and the molecular properties of the solutes will influence the type of the various molecular interactions governing retention and selectivity in LC.

Phenomenological analysis of already existing hydrogen bond donor and acceptor scales and some other apparent physical considerations enabled new quantitative scales of hydrogen bond basicity and acidity by Katritzky and coworkers.[97] Chemical structures represented by molecular graphs and electronegativities of Hinze and Jaffe were used as input data. The resulting scales are well correlated with several experimental solvent polarity scales, such as $\sum\alpha_2^H$ and $\sum\beta_2^H$, pK_a and $E_T(30)$. This study has been applied to some useful partition systems: octanol-water; gas-octanol; hexadecane-water; chloroform-water; or gas-water. The previous LSER equation has been modified and used for prediction of internal standards in ion-pairing RP-LC.[98]

Last, but not least, a special direction of QSRR studies relates to enantioseparations. Due to the importance of the enantioselectivity, especially in pharmaceutical applications, attempts to correlate retention of enantiomers on different chiral stationary phases (CSPs) lead to the issue of a new topic in the field, namely QSERR (Quantitative Structure Enantioselective Retention Relationship).

Enantioseparation mechanisms are very complex. Depending on the properties of the CSPs, it is generally accepted that enantioselectivity is based mainly on inclusion and three point interaction mechanisms. As a confirmation of such a general assumption, molecular descriptors used for QSERR are 2D and especially 3D ones.

Special attention has been paid to QSERR studies on brush type CSPs (Pirkle type with π donor or π acceptor capabilities),[99,100] chemically modified cellulose or amylose derivatives,[101] and macrocyclic antibiotics.[102]

In QSERR, retention parameters functionally related to molecular descriptors are generally $\log k'_1$ ($\log$ of the capacity factor characterizing the first eluting enantiomer), $\log k'_2$, and $\log \alpha$ ($\log$ of the selectivity factor, namely $\alpha = k'_2/k'_1$).

A specificity of QSERR attempts is the molecular conformational modeling of the analytes corresponding to minimized energetic states, through dedicated software packages (e.g., SYBYL[103] or VAMP[104]). 2D or 3D molecular descriptors are, consequently, computed for the minimized energy conformations already proposed.

For QSERR studies, the following molecular descriptors are more often used: bond and torsional angles of atoms connected to the chiral center, distance from the chiral carbon atom to the first non-hydrogen atom of the functionalized moiety (DIST), polarizability parameter (α_{mop}), dipole moment, charge on carbon atoms from aromatic rings, charge on the chiral carbon atom, charge on the carbon atom attached to the chiral atom, charge on heteroatoms attached to the chiral molecular center, charge on dissociating organic functionalities, charge on the hydrogen atom bonded to the chiral center, the sum of the charges of the carbon atoms of the aromatic rings attached to the chiral center ($\Sigma q_{C_{ar}}$), the maximum and the minimum atomic charge of the molecule, HOMO (ε_{HOMO}) and LUPO (ε_{LUPO}) molecular orbital energies, the sum of electrophilic superdelocalizabilities of the aromatic carbon atoms attached to the chiral center, differences between the maximum and the minimum atomic charges, and so on. "Classical" molecular descriptors, such as log P, V_w, and M_w, conserve their predictive capabilities in QSERR, too. The perpetual quest for new molecular descriptors containing substantial information to be correlated to the enantioselective retention leads to introduction of the term of "enantiophore."[102] Such descriptors describe the molecules in terms of their ability to form favorable interactions with independent chemical groups (probes) that can be related to receptor sites. Consequently, for each analyte, it is possible to compute descriptors representing the energy contributions from all possible pairwise combinations of probes. The procedure can be simplified by considering only the most energetically favorable location and by conversion of the selected energies into alignment-independent descriptors.

Enantioselective retention data are more often related to large sets of molecular descriptors (few tens up to few hundreds may be considered) by means of combined multiple linear regression (MLR), artificial neural network (ANN), or comparative molecular field analysis (CoMFA).

CONCLUSIONS

QSRR's are mathematical procedures that try to describe, quantitatively, the retentions of compounds in accordance with their structures. For this purpose, the literature reports a large variety of more or less complex mathematical procedures using different molecular descriptors, reflecting the structural differences between analytes taking part in the chromatographic process. Many of the publications dealing with this topic are focused on the most used mechanism in liquid chromatography, i.e., the reversed-phase mechanism. Some of these publications developed QSRR studies for pharmaceutically active compounds. Not all of the cited publications should be considered as true QSRR studies. However, some of their major declared aims relate to the study of the dependence between retention and the different experimental conditions and correlation of the outcoming experimental data with a molecular property. It is our

belief that computer simulations will solve some of the debatable points of view concerning correlation between structure and retention, as well as the intimate retention mechanism in liquid chromatography.

Are QSRR studies good in the chromatographic practice? To such a question, in a last hour work focused generally on QSAR, Gedeck et al. suggested that the quality of QSAR prediction depends upon a large number of factors, including the descriptor set, the statistical method, and data sets being used.[105] They proved that not all descriptors are suitable in all data sets, and it is necessary to test each individual case and descriptor for each theoretical model.

REFERENCES

1. Hadjipavlou-Litina, D.; Garg, R.; Hansch, C. Comparative quantitative structure-activity relationship studies (QSAR) on non-benzodiazepine compounds binding to benzodiazepine receptor (BzR). Chem. Rev. **2004**, *104*, 3751–3793.
2. Dudek, A.Z.; Arodz, T.; Galvez, J. Computational methods in developing quantitative-activity relationships (QSAR): a review. Comb. Chem. & High Throughput Screening **2006**, *9*, 213–228.
3. Sutherland, J.J.; O'Brien, L.A.; Weaver, D.F. A comparison of methods for modeling quantitative structure-activity relationships. J. Med. Chem. **2004**, *47*, 5541–5554.
4. Hanai, T. Structure-retention correlation in liquid chromatography. J. Chromatogr. A **1991**, *550*, 313–324.
5. Xia, Y.; Guo, Y.; Wang, H.; Wang, Q.; Zuo, Y. Quantitative structure-retention relationships of benzoylphenylureas on polystyrene-octadecene-encapsulated zirconia stationary phase in reversed-phase high performance liquid chromatography. J. Sep. Sci. **2005**, *28*, 73–77.
6. Wang, Q.S.; Zhang, L. Review of research on quantitative structure-retention relationships in thin-layer chromatography. J. Liq. Chromatogr. Rel. Technol. **1999**, *22*, 1–14.
7. Dimov, N. Quantitative structure-retention relationship by TLC. In *Encyclopedia of Chromatography*; Cazes, J., (ed).; Taylor & Francis: New York, 2005; 1408–1411.
8. Hsieh, M.M.; Dorsey, J.G. Accurate determination of log k'_w in reversed-phase liquid chromatography. Implications for quantitative structure-retention relationship. J. Chromatogr. A **1993**, *631*, 63–78.
9. Rimmer, C.A.; Simmons, C.R.; Dorsey, J.G. The measurement and meaning of void volumes in reversed-phase liquid chromatography. J. Chromatogr. A **2002**, *965*, 219–232.
10. Guha, R.; Jurs, P.C. Determining the validity of a QSAR model–a classification approach. J. Chem. Inf. Model. **2005**, *45*, 65–73.
11. Al-Haj, M.A.; Kaliszan, R.; Buszewski, B. QSRR with model analytes as a means of an objective evaluation of chromatographic columns. J. Chromatogr. Sci. **2001**, *39*, 29–38.
12. Waterbeemd, H.; Van de Carter, R.E.; Grassy, G.; Kubinskyi, H.; Martin, Y.C.; Tute, M.S.; Willett, P. Glossary of terms used in computational drug design. Pure & Appl. Chem. **1997**, *69*, 1137–1152.

13. Xue, L.; Bajorath, J. Molecular descriptors in chemoinformatics, computational combinatorial chemistry and virtual screening. Comb. Chem. & High Throughput Screenining **2000**, *3*, 363–372.
14. Jinno, K. Study on retention in liquid chromatography. Chromatography **2001**, *22*, 1–9.
15. Kerelson, M. *Molecular Descriptors in QSAR/QSPR*; John Wiley & Sons Inc.: New York, 2000.
16. Fujita, T.; Iwasa, J.; Hanch, C. A new substituent constant π derived from partition coefficients. J. Am. Chem. Soc. **1964**, *86*, 5176–5180.
17. Medvedovici, A.; Sandra, P.; Kot, A.; David, F. *Supercritical Fluid Chromatography with Packed Columns*; Anton, K., Berger, C., Eds.; Marcel Dekker Inc.: New York, 1997; Vol. 75, Chapter 6, 161–194.
18. Sandra, P.; Medvedovici, A.; David, F. Comprehensive pSFC x pSFC – MS for the characterization of triglycerides in vegetable oils. LC-GC Europe **2003**, *12*, 2–4.
19. Hansch, C.; Leo, A.; Taft, R.W. A surwey of Hammet substituent constants and resonance and field parameters. Chem. Rev. **1991**, *91*, 165–195.
20. Carrasco Velar, R.; Padron, J.A.; Galvez, J. Definition of a novel atomic index in QSAR: the refractopological state. J. Pharm. Pharmaceut. Sci. **2004**, *7*, 19–26.
21. Wise, S.A.; Bonnet, W.J.; Guenther, F.R.; May, W.E. A relationship between RP chromatography and elution of PAHs. J. Chromatogr. Sci. **1981**, *18*, 457–465.
22. Kier, L.B.; Hall, L.H. The meaning of molecular connectivity: a bimolecular accessibility model. Croat. Chem. Acta **2002**, *75*, 371–382.
23. Radecki, A.; Lamparczyk, H.; Kaliszan, R. A relationship between the retention indices on nematic and isotropic phases and the shape of PAHs. Chromatographia **1979**, *12*, 595–600.
24. Franczkiewicz, R.; Braun, W. Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. J. Comp. Chem. **1998**, *19*, 319–333.
25. Lewis, D.F.V. Frontier orbitals in chemical and biological activity: quantitative relationships and mechanistic implications. Drug Metab. Rev. **1999**, *31*, 755–816.
26. Pompe, M.; Randic, M. Anticonnectivity": a challenge for structure-property-activity studies. J. Chem. Inf. Mod. **2006**, *46*, 2–8.
27. Katritzky, A.R.; Gordeeva, E.R. Traditional topological indexes vs electronic, geometrical and combined molecular descriptors in QSAR/QSPR research. J. Chem. Inf. Comput. Sci. **1993**, *33*, 835–857.
28. Karelson, M.; Lobanov, V.S.; Katritzky, A.R. Quantum-chemical descriptors in QSAR/QSPR studies. Chem. Rev. **1996**, *96*, 1027–1043.
29. Devillers, J.; Balaban, A.T. *Topological Indices and Related Descriptors in QSAR and QSPR*; Gordon and Breach: Amsterdam, 1999.
30. Todeschini, R.; Consonni, V. *The Handbook of Molecular Descriptors*; Mannhold, R., Kubinyi, H., Timmerman, H., (eds).; *Series of Methods and Principles in Medicinal Chemistry*; Wiley-VCH: New York, 2000; Vol. 11.
31. Hanai, T. Chromatography in silico, basic concept in reversed-phase liquid chromatography. Anal. Bioanal. Chem. **2005**, *382*, 708–717.
32. Sander, L.C.; Lippa, K.A.; Wise, S.A. Order and disorder in alkyl stationary phase. Anal. Bioanal. Chem. **2005**, *383*, 646–668.
33. Croes, K.; Steffens, A.; Marchand, D.H.; Snyder, L.R. Relevance of π-π and dipole-dipole interactions for retention on cyano and phenyl columns in reversed-phase liquid chromatography. J. Chromatogr. A **2005**, *1098*, 123–130.

34. Marchand, D.H.; Croes, K.; Dolan, J.W.; Snyder, L.R. Column selectivity in reversed-phase liquid chromatography. VII. Cyanopropyl columns. J. Chromatogr. A **2005**, *1062*, 57–64.
35. Clark, M. Generalized fragment-substructure based property prediction method. J. Chem. Inf. Model. **2005**, *45*, 30–38.
36. Turowski, M.; Yamakawa, N.; Meller, J.; Kimata, K.; Ikegami, T.; Hosoya, K.; Tanaka, N.; Thornton, E.R. Deuterium isotope effects on hydrophobic interactions: the importance of dispersion interactions in the hydrophobic phase. J. Am. Chem. Soc. **2003**, *125*, 13836–13849.
37. Avdeef, A. *Absorption and Drug Development: Solubility, Permeability and Charge State*; Wiley-Interscience: Hoboken, 2003; 67–90.
38. Katritzky, A.R.; Tamm, K.; Kuanar, M.; Fara, D.C.; Oliferenko, A.; Oliferenco, P.; Huddleston, J.G.; Rogers, R.D. Aqueous biphasic systems. Partitioning of organic molecules: a QSPR treatment. J. Chem. Inf. Comput. Sci. **2004**, *44*, 136–142.
39. Caron, G.; Steyaert, G.; Pagliara, A.; Reymond, F.; Crivori, P.; Gaillard, P.; Carrupt, P.; Avdeef, A.; Comer, J.; Box, K.J.; Girault, H.H.; Testa, B. Structure-lipophilicity relationships of neutral and protonated β-blockers. I. Intra- and intermolecular effects in isotropic solvent systems. Helv. Chim. Acta **1999**, *82*, 1211–1222.
40. Escuder-Gilabert, L.; Bermudez-Saldana, J.M.; Villanueva-Camanas, R.M.; Medina-Hernadez, M.J.; Sagrado, S. Reliability of the retention factor estimations in liquid chromatography. J. Chromatogr. A **2004**, *1033*, 247–255.
41. Renner, R. The $K_{o,w}$ controversy. Environ. Sci. Technol. **2002**, *413A*.
42. Poole, S.K.; Poole, C.F. Separation methods for estimating octanol–water partition coefficients. J. Chromatogr. B **2003**, *797*, 3–19.
43. Leo, A. Calculating log P_{oct} from structures. Chem. Rev. **1993**, *93*, 1281–1306.
44. Meylan, W.M.; Howard, P.H. Atom/fragment contribution method for estimating octanol-water partition coefficients. J. Pharm. Sci. **1995**, *84*, 83–92.
45. Nikitas, P.; Pappa-Louisi, A.; Agrafiotou, P. Effects of the organic modifier concentration on the retention in reversed-phase liquid chromatography. I. General semi-thermodynamic treatment for adsorption and partition mechanisms. J. Chromatogr. A **2002**, *946*, 9–32.
46. Nikitas, P.; Pappa-Louisi, A.; Agrafiotou, P. Effects of the organic modifier concentration on the retention in reversed-phase liquid chromatography. II. Tests using various simplified models. J. Chromatogr. A **2002**, *946*, 33–45.
47. Neue, U.D. Nonlinear retention relationships in reversed-phase chromatography. Chromatographia **2006**, *63*, S45–S53.
48. Medvedovici, A.; Albu, F.; David, V. Selection of the retention mechanism for separation of amitryptiline and the main related impurities by liquid chromatography. Rev. Roum. Chim. **2005**, *50*, 225–234.
49. David, V.; Bala, C.; Rotariu, L. Thermodynamic parameters of the reversed-phase liquid chromatography retention for some lipid-soluble vitamins. Chem. Anal. (Warsaw) **2004**, *49*, 191–199.
50. Howkins, D.M. The problem of overfitting. J. Chem. Inf. Comput. Sci. **2004**, *44*, 1–12.
51. Dai, J.; Yao, S.; Ding, Y.; Wang, L. Retention of substituted indole compounds on RP-HPLC: correlation with molecular connectivity indices and quantum chemical descriptors. J. Liq. Chromatogr. Rel. Technol. **1999**, *22*, 2271–2282.

52. Jakab, A.; Schubert, G.; Prodan, M.; Forgacs, E. Determination of the retention behavior of barbituric acid derivatives in reversed-phase HPLC by using quantitative structure-retention relationships. J. Chromatogr. B **2002**, *770*, 227–236.
53. Jakab, A.; Schubert, G.; Prodan, M.; Forgacs, E. PCA, followed by two-dimensional nonlinear mapping and cluster analysis, versus multilinear regression in QSRR. J. Liq. Chromatogr. Rel. Technol. **2002**, *25*, 1–16.
54. Agatonovic-Kustrin, A.; Zecevic, M.; Zivanovic, LJ. Use of modeling in structure-retention relationships of diuretics in RP-HPLC. J. Pharm. Biomed. Anal. **1999**, *21*, 95–103.
55. Tham, S.Y.; Agatonovic-Kustrin, A. Application of the artificial neural network in quantitative structure-gradient elution retention relationship of phenylcarbamyl amino acids derivatives. J. Pharm. Biomed. Anal. **2002**, *28*, 581–590.
56. Loukas, Y. Radial basis function networks in LC: improved structure-retention relationships compared to principal components regression (PCR) and nonlinear partial least squares regression (PLS). J. Liq. Chromatogr. Rel. Technol. **2001**, *24*, 2239–2256.
57. Xiang, Y.H.; Liu, M.C.; Zhang, X.Y.; Zhang, R.S.; Hu, Z.D.; Fan, B.T.; Doucet, J.P.; Panaye, A. Quantitative prediction of LC retention of N-benzylideneanilines based on quantum chemical parameters and radial basis function neural network. J. Chem. Inf. Comput. Sci. **2002**, *42*, 592–597.
58. Song, M.; Breneman, C.M.; Bi, J.; Sukumar, N.; Bennett, K.P.; Cramer, S.; Tugcus, N. Prediction of protein retention times in anion-exchange chromatography systems using support vector regression. J. Chem. Inf. Comput. Sci. **2002**, *42*, 1347–1357.
59. Liu, H.X.; Xue, C.X.; Zhang, R.S.; Yao, X.J.; Liu, M.C.; Hu, Z.D.; Fan, B.T. Quantitative prediction of log k of peptides in HPLC based on molecular descriptors by using the heuristic method and support vector machine. J. Chem. Inf. Comput. Sci. **2004**, *44*, 1979–1986.
60. Luco, J.M.; Salinas, A.P.; Ferriero, A.A.J.; Vazquez, R.N.; Raba, J.; Marchevsky, E. Immobilized artificial membrane chromatography: quantitative structure-retention relationships of structurally diverse drugs. J. Chem. Inf. Comput. Sci. **2003**, *43*, 2129–2136.
61. Detroyer, A.; Stokbroekx, S.; Bohets, H.; Lorreyne, W.; Timmerman, P.; Verboven, P.; Massart, D.L.; Vander-Heyden, Y. Fast monolithic micellar LC: an alternative drug permeability assessing method for high-throughput screening. Anal. Chem. **2004**, *76*, 7304–7309.
62. Varga-Defterdarovic, L.; Horvat, S.; Skuric, M.; Horvat, J. Correlation of structure and retention behavior in reversed-phase HPLC. II. Methionine-enkephalin-related glycoconjugates. J. Chromatogr. A **1994**, *687*, 107–112.
63. Yang, S.; Bumgarner, J.G.; Kruk, L.F.R.; Khaledi, M.G. Quantitative structure-activity relationships studies with micellar electrokinetic chromatography influence of surfactant type and mixed micelles on estimation of hydrophobicity and bioavailability. J. Chromatogr. A **1996**, *721*, 323–335.
64. Momose, T.; Yamaguchi, Y.; Iida, T.; Goto, J.; Nambara, T. Structure-retention correlation of isomeric bile acids in inclusion HPLC with methyl beta-cyclodextrin. Lipids **1998**, *33*, 101–108.
65. Luo, H.; Cheng, Y.K. Quantitative structure-retention relationship of nucleic-acid bases revisited. CoMFA on purine RPLC retention. QSAR & Comb. Sci. **2005**, *24*, 968–975.

66. Welerowicz, T.; Buszewski, B. The effect of stationary phase on lipophilicity determination of β-blockers using reversed-phase chromatographic systems. Biomed. Chromatogr. **2005**, *19*, 725–736.
67. Sarbu, C.; Todor, S. Determination of lipophilicity of some non-steroidal anti-inflamatory agents and their relationships by using principal component analysis based on thin-layer chromatographic retention data. J. Chromatogr. A **1998**, *822*, 263–269.
68. Espinoza, S.; Bosch, E.; Roses, M. Retention of ionizable compounds on HPLC. 5. pH scales and the retention of acids and bases with acetonitrile-water mobile phases. Anal. Chem. **2000**, *72*, 5193–5200.
69. Espinoza, S.; Bosch, E.; Roses, M. Retention of ionizable compounds in HPLC. IX. Modeling retention in reversed-phase LC as a function of pH and solvent composition with acetonitrile-water mobile phase. J. Chromatogr. A **2002**, *947*, 47–58.
70. David, V.; Albu, F.; Medvedovici, A. Retention behavior of metformin and related impurities in ion-pairing liquid chromatography. J. Liq. Chromatogr. Rel. Technol. **2005**, *28*, 83–98.
71. David, V.; Medvedovici, A. Partition model applied to the retention process of basic analytes in reversed-phase and ion-pairing liquid chromatography. Rev. Roum. Chim. **2005**, *50*, 837–843.
72. Neue, U.D.; Phoebe, C.H.; Tran, K.; Cheng, Y.F.; Lu, Z. Dependence of reversed-phase retention of ionizable analytes on pH, concentration of organic solvent and silanol activity. J. Chromatogr. A **2001**, *925*, 49–67.
73. Wiczling, P.; Markuszewski, M.J.; Kaliszan, R. Determination of pK_a by pH gradient reversed-phase HPLC. Anal. Chem. **2004**, *76*, 3069–3077.
74. Kogej, T.; Muresan, S. Database mining for pK_a predicting. Curr. Drug Disc. Technol. **2005**, *2*, 221–229.
75. Espinoza, S.; Bosch, E.; Roses, M. Acid-base constants of neutral bases in acetonitrile mixtures. Anal. Chim. Acta **2002**, *454*, 157–166.
76. Vervoort, R.J.M.; Ruyter, E.; Debets, A.J.J.; Claessens, H.A.; Cramers, C.A.; de Jong, G.J. Characterisation of reversed phase LC stationary phases for the analytes of basic pharmaceuticals: eluent properties and comparison of empirical test methods. J. Chromatogr. A **2001**, *931*, 67–79.
77. Bartolini, M.; Bertucci, C.; Gotti, R.; Tumiatti, V.; Covalli, A.; Recanatini, M.; Andrisana, V. Determination of the dissociation constants (pKa) of basic acetylcholinesterase inhibitors by RP-LC. J. Chromatogr. A **2002**, *958*, 59–67.
78. Wilson, D.M.; Wang, X.; Walsh, E.; Rourick, R.A. High throughput log D using liquid chromatography-mass spectrometry. Comb. Chem. & High Throughput Screening **2001**, *4*, 511–519.
79. Ruiz-Angel, M.J.; Carda-Broch, S.; Garcia-Alvarez-Coque, M.C.; Berthod, A. Effect of ionization and the nature of the mobile phase in QSRR studies. J. Chromatogr. A **2005**, *1063*, 25–34.
80. David, V.; Albu, F.; Medvedovici, A. Structure-retention correlation of some oxicam drugs in reversed-phase liquid chromatography. J. Liq. Chromatogr. Rel. Technol. **2004**, *27*, 965–984.
81. Torres-Lapasio, J.R.; Garcia-Alvarez-Coque, M.C.; Roses, M.; Bosch, E. Prediction of the retention in reverse-phase liquid chromatography using solute-mobile phase-stationary phase polarity parameters. J. Chromatogr. A **2002**, *955*, 19–34.
82. Bosque, R.; Sales, J.; Bosch, E.; Roses, M.; Garcia-Alvarez-Coque, M.C.; Torres-Lapasio, J.R. A QSPR study of the p solute polarity parameter to estimate retention in HPLC. J. Chem. Inf. Comput. Sci. **2003**, *43*, 1240–1247.

83. Pyka, A.; Sliwiok, J. Use of traditional structural descriptors in QSRR analysis of nicotinic acid esters. J. Liq. Chromatogr. Rel. Technol. **2004**, *27*, 785–798.
84. Bacsek, T.; Kaliszan, R. Combination of linear solvent strength model and QSRRs as a comprehensive procedure of approximate prediction of retention in gradient liquid chromatography. J. Chromatogr. A **2002**, *962*, 41–55.
85. Kaliszan, R.; Baczek, T.; Bucinski, A.; Buszewski, B.; Sztupecka, M. Prediction of gradient retention from the linear solvent strength (LSS) model, quantitative structure-retention relationships (QSRR) and artificial neural networks (ANN). J. Sep. Sci. **2003**, *26*, 271–282.
86. Deanda, F.; Smith, K.M.; Liu, J.; Pearlman, R.S. GSSI, a general model for solute-solvent interactions. I. Description of the model. Molec. Pharm. **2004**, *1*, 23–39.
87. Wagoner, J.A.; Baker, N.A. Assessing implicit models for nonpolar mean solvation forces: the importance of dispersion and volume terms. Proc. Natl. Acad. Sci. USA **2006**, *103*, 8331–8336.
88. Okur, A.; Wickstrom, L.; Layten, M.; Geney, R.; Song, K.; Hornak, V.; Simmerling, C. Improved efficiency of replica exchange simulations through use of a hybrid explicit/implicit solvation model. J. Chem. Theory Comput. **2006**, *2*, 420–433.
89. Bordner, A.J.; Cavasotto, C.N.; Abagyan, R.A. Accurate transferable model for water, n-octanol, and n-hexadecane solvation free energies. J. Phys. Chem. B **2002**, *106*, 11009–11015.
90. Ranatunga, R.P.J.; Carr, P.W. A study of the enthalpy and entropy contributions of the stationary phase in RP-LC. Anal. Chem. **2000**, *72*, 5679–5692.
91. Szepesy, L. Effect of molecular interactions on retention and selectivity in reversed-phase LC. J. Chromatogr., A **2002**, *960*, 69–83.
92. Vonk, E.C.; Lewandowska, K.; Claessens, H.A.; Kaliszan, R.; Cramers, C.A. Quantitative structure-retention relationships in reversed-phase LC using several stationary and mobile phase. J. Sep. Sci. **2003**, *26*, 777–792.
93. Wilson, N.S.; Nelson, M.D.; Dolan, J.W.; Snyder, L.R.; Wolcott, R.G.; Carr, P.W. Column selectivity in reversed-phase LC. I. A general quantitative relationship. J. Chromatogr. A **2002**, *961*, 171–193.
94. Wilson, N.S.; Nelson, M.D.; Dolan, J.W.; Snyder, L.R.; Carr, P.W. Column selectivity in reversed-phase LC. II. Effects of a change in conditions. J. Chromatogr. A **2002**, *961*, 195–215.
95. Wilson, N.S.; Dolan, J.W.; Snyder, L.R.; Carr, P.W.; Sander, L.C. Column selectivity in reversed-phase LC. III. The physico-chemical basis of selectivity. J. Chromatogr. A **2002**, *961*, 217–236.
96. Jover, J.; Bosque, R.; Sales, J. Determination of Abraham solute parameters from molecular structure. J. Chem. Inf. Comput. Sci. **2004**, *44*, 1098–1106.
97. Oliferenko, A.A.; Oliferenko, P.V.; Huddleston, J.G.; Rogers, R.D.; Palyulin, V.A.; Zefirov, N.S.; Katritzky, A.R. Theoretical scales of hydrogen bond acidity and basicity for application in QSAR/QSPR studies and drug design. Partitioning of aliphatic compounds. J. Chem. Inf. Comput. Sci. **2004**, *44*, 1042–1055.
98. Li, J. Prediction of internal standards in RP-LC. IV. Correlation and prediction of retention in reversed-phase ion-pair chromatography based on linear solvation energy relationships. Anal. Chim. Acta **2004**, *522*, 113–126.
99. Suzuki, T.; Timofei, S.; Iuroas, B.E.; Uray, G.; Verdino, P.; Fabian, W.M.F. QSERR for chromatographic separation of arylalkylcarbinols on Pirkle type stationary phases. J. Chromatogr. A **2001**, *922*, 13–23.

100. Lipkowitz, K.B. Theoretical studies of brush-type chiral stationary phases. J. Chromatogr. A **1994**, *666*, 493–503.
101. Booth, T.D.; Azzaoui, K.; Wainer, I.W. Prediction of chiral chromatographic separations using combined multivariate regression and neural networks. Anal. Chem. **1997**, *69*, 3879–3883.
102. Del Rio, A.; Piras, P.; Roussel, C. Enantiophore modeling in 3D-QSAR. A data mining application on Whelk-O1 chiral stationary phase. Chirality **2006**, *7*, 498–508.
103. *SYBYL Molecular Modeling Software*; Tripos Associated: St. Louis, MO, 1999.
104. Clark, T. *VAMP 4.4, Erlangen Vectorized Molecular Orbital Package*; Computer-Chemie-Centrum, Univ. Erlangen: Nurnberg, Germany, 1992.
105. Gedeck, P.; Rohde, B.; Bartels, C. QSAR – How good is it in practice? Comparison of descriptor sets on an unbiased cross section of corporate data sets. J. Chem. Inf. Model. **2006**, on-line 3 Aug 2006.

Received August 28, 2006
Accepted November 21, 2006
Manuscript 6980F