SHIMADZU

Excellence in Science

Small but mighty

With its compact dimensions, our Brevis GC-2050 saves valuable lab space. Thanks to Analytical Intelligence, however, it's not stingy in terms of performance and automation.

It offers you easy operation and maintenance with less consumption and is supported by LabSolutions and other leading software packages.

Space-saving
4 GC-2050 can be placed on a typical lab table

Free choice
Suitable for hydrogen, argon, helium or nitrogen as carrier gas

Cost-saving
30 % less power* and up to 90 % reduced gas consumption

Analytical Intelligence
Operator relief by automated and easier maintenance

* Compared with the conventional Shimadzu GC.

www.shimadzu.eu/small-but-mighty

ISSN 1615-9306 · JSSCCJ 46 (18) 2023 · Vol. 46 · No. 18 · September 2023

Journal of SEPARATION SCIENCE

18|2023

www.jss-journal.com

Special Issue

Dedicated to Professor František Švec on the occasion of his 80th birthday

WILEY VCH

Received: 16 May 2023 | Revised: 30 June 2023 | Accepted: 2 July 2023

DOI: 10.1002/jssc.202300346

JOURNAL OF SEPARATION SCIENCE

RESEARCH ARTICLE

Estimating the hydrophobicity extent of molecular fragments using reversed-phase liquid chromatography

Andrea Carotti[1] | **Ina Varfaj**[1] | **Ilaria Pruscini**[1] | **Ghaid W. A. Abualzulof**[1] | **Laura Mercolini**[2] | **Elisa Bianconi**[1] | **Antonio Macchiarulo**[1] | **Emidio Camaioni**[1] | **Roccaldo Sardella**[1]

[1]Department of Pharmaceutical Sciences, Via Fabretti 48, University of Perugia, Perugia, Italy

[2]Department of Pharmacy and Biotechnology (FaBiT), Alma Mater Studiorum - Via Belmeloro 6, University of Bologna, Bologna, Italy

Correspondence
Emidio Camaioni and Roccaldo Sardella, Department of Pharmaceutical Sciences, University of Perugia, Via Fabretti 48, 06123 Perugia, Italy.
Email: emidio.camaioni@unipg.it and roccaldo.sardella@unipg.it

Andrea Carotti and Ina Varfaj contributed equally to this study.

A fast HPLC method was developed to study the hydrophobicity extent of pharmaceutically relevant molecular fragments. By this strategy, the reduced amount of sample available for physico-chemical evaluations in early-phase drug discovery programs does not represent a limiting factor. The sixteen acid fragments investigated were previously synthesized also determining potentiometrically their experimental log D values. For four fragments it was not possible to determine such property since their values were outside of the instrumental working range ($2 < pK_a < 12$). An RP-HPLC method was therefore optimized. For each scrutinized method, some derived chromatographic indices were calculated, and Pearson's correlation coefficient (r) allowed to select the so-called "φ_0 index" as the best correlating with the log D. The ${}^{s}_{w}pH$ was fixed at 3.5 and a modification of some variables [organic modifier (methanol vs. ACN), stationary phase (octyl vs. octadecyl), presence/absence of the additives *n*-octanol, *n*-butylamine, and *n*-octylamine], allowed to select the best correlation conditions, producing a $r = 0.94$ ($p < 0.001$). Importantly, the φ_0 index enabled the estimation of log D values for four fragments which were unattainable by potentiometric titration. Moreover, a series of molecular descriptors were calculated to identify the chemical characteristics of the fragments explaining the obtained φ_0. The number of hydrogen bond donors and the index of cohesive interaction correlated with the experimental data.

KEYWORDS
chromatographic hydrophobicity indices, log D, linear correlation, molecular descriptors, reversed-phase chromatography

Article Related Abbreviations: ADMET, Absorption, distribution, metabolism, excretion, and toxicity; FBDD, Fragment-Based Drug Discovery; C8, Octyl Silane; C18, Octadecyl Silane; CMC, Critical Micellization Concentration; phi0, Isocratically Derived chromatographic Index; HBA×HBD1/2/SASA, Index of Cohesive Interaction; HBD, Hydrogen Bond Donors; IUPAC, International union of pure and applied chemistry; Log P, Logarithm of the Partition Coefficient; MeCN, Acetonitrile; MeOH, Methanol; RP-HPLC, Reversed-Phase High-Performance Liquid Chromatography; SP, Stationary Phase.

1 | INTRODUCTION

The lipophilicity of the drug molecules plays an important role in the absorption, distribution, metabolism, excretion, and toxicity profile of drugs [1]. Thus, the ability to early determine such properties of a new chemical entity is a central point in the drug discovery process. The pioneering works by Hansch and co-workers [2–4] first pointed out how the lipophilicity of molecules could be measured by their octanol/water partition coefficients (log P). Compared to the latter, log D is a better descriptor of the lipophilicity of an ionizable molecule. It can be determined in a similar manner to log P but instead of using a water solution, the aqueous phase must be adjusted to a specific pH.

Many methods are available for the experimental determination of octanol/water partitioning parameters, but the shake flask procedure remains the standard one [5]. However, the use of this traditional method is characterized by some drawbacks, including the sample handling, the sample sizes and it is time and labor-consuming features. The indirect analysis of the lipophilicity behavior based on RP-HPLC with *n*-alkyl chain-based stationary phases is considered a valuable alternative to the shake-flask procedure since it allows easy analysis of larger groups of compounds in a limited timeframe [6, 7]. Advantages of chromatographic measurement of the hydrophobicity of a molecule include speed, reproducibility, online detection, greater dynamic range, reduced sample handling, and sample sizes in comparison to traditional octanol/water partitioning measurements (log P) through the shake flask method. Degradation products and impurities are generally not a problem since that chromatography is inherently a method of separation [8]. In addition, chromatographic hydrophobicity measurements can be conveniently performed with automatized systems thereby allowing to obtain high throughput determination of this physicochemical parameter for hundred molecules per day with a single instrument.

The fact that retention in RP-LC systems is ruled by compound lipophilicity (or more precisely hydrophobicity) is described by the solvophobic theory [9, 10]. The expressions "hydrophobicity" and "lipophilicity" are often used interchangeably, and their frequency of usage is different in various scientific fields. However, while medicinal chemists prefer the lipophilicity term, separation scientists tend to use hydrophobicity [11]. According to the definitions suggested by IUPAC, the distribution behavior of compounds in a biphasic system, such as liquid/liquid or solid/liquid represents lipophilicity. The lipophilicity measure derived by the chromatographic method is not a simple binary solvent partition value, thus it has been called "hydrophobicity" [11]. For this reason, from now on the term hydrophobicity will be used.

The RP chromatographic retention in pure water (usually referred to as k_w, where the letter "w" stands for "water") has been reported to be one of the most useful derived chromatographic indices to estimate hydrophobicity (that is, the log P) through RP-LC analysis because the solvophobic effect is the sole hydrophobic effect [8]. The "linear solvent strength model", proposed by Snyder in the late 70s', is a simple model that can be utilized to effortlessly derive retention factors from a pure water-containing mobile phase [11]. This model is based on a linear relationship between the retention factor and the organic content in the mobile phase (solvent strength), as in Equation (1):

$$\log k = \log k_w + S\varphi \quad (1)$$

where S (which is dependent on the chromatographic system and solute) is the slope of the equation and φ is the fraction of the organic solvent in the eluent. In agreement with the above equation (Equation (1)), k values are measured for each compound with eluents containing different amounts of the organic modifier. As a result, regression analysis of the linear portion in the log k versus φ plot allows the extrapolation of the log k_w parameter. However, it should be considered that a loss of linearity is usually observed when $\varphi < 0.2$ and $\varphi > 0.8$ [8, 12]. Indeed, many of the mobile phase and stationary phase properties, that are significant for retention in an RP-LC system, do not experience any linear variation in these two regions.

It has been found that when ACN (MeCN) is used instead of methanol (MeOH) as the major organic component in the mobile phase, more pronounced curvature occurs in the correspondence of water-rich eluents, thus amplifying the uncertainty in the log k_w value determination [11–14]. Oppositely, a high correlation between these results exists when MeOH is used [14–16]. Many efforts have been made to formalize correlation equations taking into account the above loss of linearity [14]. However, the actual limitation of these equations resides in the fact that the extrapolated log k_w value for a specific compound can be different depending on the organic modifiers in use [11, 12]. Other efforts have been also made to produce materials that enable the reduction of the interference of free silanols in the whole chromatographic process. In this framework, the use of a stationary phase with some structural peculiarities as well as the incorporation of basic additives in the eluents has been proposed [7, 8, 12].

The derived chromatographic index φ_0, originally proposed by Valkò and Slégel [17] is another indirect chromatographic index to measure the hydrophobicity of

compounds [11, 14, 18]. The φ_0 value for a given chemical compound indicates the volume percentage of the organic modifier that is required to evenly distribute that compound between the two phases of the chromatographic media [10, 14, 18]. The φ_0 index can be calculated from Equation (2) as follows:

$$\varphi_0 = -\frac{\log k_w}{S} \quad (2)$$

where S has a reciprocal unit to afford a dimension less φ_0 value. The approach requires retention factor measurement (log k values), using at least three different mobile phase compositions, ideally near the concentration zone where $\log k = 0$. This is analogous to the $\log k_w$ determination. φ_0 has to be fairly regarded as a more representative hydrophobicity index because, in theory, it is exempt from the uncertainty that characterizes the most usually suggested $\log k_w$ parameter. Since φ_0 only depends on the distribution constant, it exhibits a strong link with the type and proportion of organic modifiers as well as the system temperature. Further, also the pH of the eluent has a significant impact on the φ_0 value of ionizable compounds [17].

Besides its utility for log P estimation, the φ_0 was successfully used by Natalini et al. to estimate the attitude by surface active bile acids to form aggregate, that is in the estimation of the Critical Micellization Concentration [14, 19].

In the present work, the evaluation of the hydrophilicity/hydrophobicity balance of a small series of molecular fragments (Table 1) through the φ_0 index is described.

Fragment-based drug discovery (FBDD) has become a powerful approach for discovering new biologically active hit compounds. Its success relies on screening compounds of low molecular weight and low complexity (defined fragments) during the early phase of the drug discovery program. From the initial hit fragments, lead compounds, with highly improved properties, can be efficiently obtained by growing, linking, or merging molecular fragments. Indeed, many candidates and drugs have been discovered by the FBDD technique as recently reviewed by the Walsh group [20].

All these findings prompted us to further study our recently synthesized small series of molecular fragments [21]. These acidic fragments (Table 1) were designed for a wider medicinal chemistry project aimed at discovering inhibitors capable to interact with zinc-containing enzymes [21]. In the present study, we developed a fast HPLC method enabling the evaluation of the log D values of the selected molecular fragments reported in Table 1. Very profitably, this procedure was demonstrated to require a very limited amount of sample, which represents a favorable situation when compounds coming from different synthetic routes are investigated also in terms of their physico-chemical properties. Moreover, another main advantage of the method here described is related to the capability of the φ_0 index to make possible the estimation, for some fragments, of log D values previously not accessible by potentiometric titration. Still, we also demonstrate that molecular modeling investigation can be useful to explain the experimental findings at a molecular level. All these aspects which represent a novelty in the domain of fragment-based drug discovery will be treated in detail in the following sections.

2 | MATERIALS AND METHODS

2.1 | Reagents, chemicals, and materials

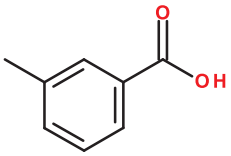
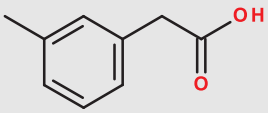
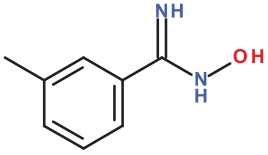
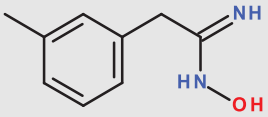
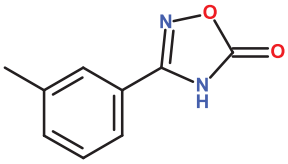
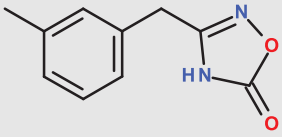
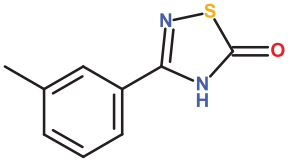
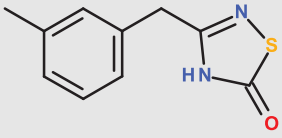
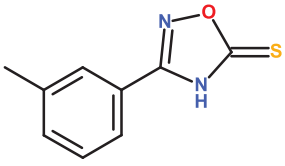
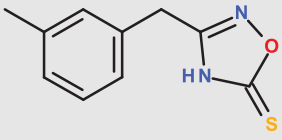
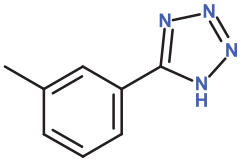
All the reagents were of analytical grade. All the test fragments were synthesized in our laboratory, according to the procedures described in [21]. ACN (MeCN), methanol (MeOH), monobasic sodium phosphate (NaH_2PO_4), dibasic sodium phosphate (Na_2HPO_4), concentrated phosphoric acid (H_3PO_4), *n*-octanol, *n*-butylamine, *n*-octylamine, and sodium nitrite ($NaNO_2$) were purchased from Merck Life Science S.r.l. Water for HPLC analysis was purified with a New Human Power I Scholar water purification system (Human Corporation) and the Milli-Q water purification system of Millipore.

2.2 | Instrumentation and chromatographic conditions

The HPLC study was performed on a Waters ALLIANCE 2695 Separations Module system (Waters Corporation), equipped with a quaternary, low-pressure mixing pump and in-line vacuum degassing, an autosampler with a maximum capacity of 120 vials, and a column heater/cooler. The system is endowed with a photodiode array detector (Waters 2996). The data management was made by Waters Millennium 32 Software. The columns Robusta C18 (250 × 4.6 mm I.D., 5 µm, 110 Å) and LiChrospher 60 RP-select B LiChroCART (250 × 4.0 mm i.d., 5 µm, 60 Å pore size) C8 were purchased from SepaChrom (Rho, Italy) and Merck KGaA, respectively. Throughout the manuscript, the employed C18 and C8, are indicated as stationary phase 1 (SP1), and stationary phase 2 (SP2), respectively.

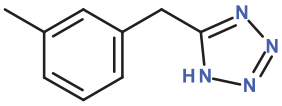
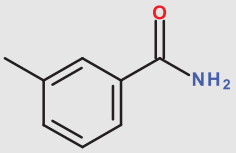
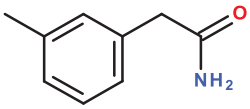
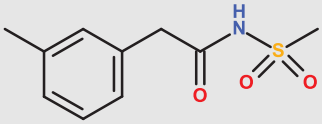
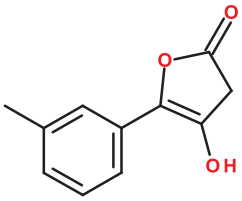
The apparent pH [${}^{s}_{w}pH$, that is the one measured in the employed hydro-organic mobile phase (s), while the calibration of the pH system was done in water (w)] of the eluents was measured with a conventional pH-meter, and then opportunely adjusted to the required value as explained below.

TABLE 1 Structure of the investigated compounds and relative properties obtained with the Sirius T3 apparatus.

Cmpd #	Structure	Name	log $D_{3.5}$	log $D_{7.4}$	pK_a [1]
1		3-Methylbenzoic acid	2.08	−1.07	3.99
2		2-(*m*-Tolyl)acetic acid	1.88	−1.3	4.13
3		*N*-Hydroxy-3-methyl-benzamidine	n.d.	n.d.	n.d.
4		*N*-Hydroxy-2-(*m*-tolyl)acetamidine	n.d.	n.d.	n.d.
5		3-(*m*-Tolyl)−4*H*-1,2,4-oxadiazol-5-one	1.93	0.47	5.12
6		3-(*m*-Tolylmethyl)−4*H*-1,2,4-oxadiazol-5-one	1.62	−0.19	5.59
7		3-(*m*-Tolyl)−4*H*-1,2,4-thiadiazol-5-one	2.22	0.79	5.99
8		3-(*m*-Tolylmethyl)−4*H*-1,2,4-thiadiazol-5-one	2.15	1.06	6.35
9		3-(*m*-Tolyl)−4*H*-1,2,4-oxadiazole-5-thione	1.74	−2.11	2.64
10		3-(*m*-Tolylmethyl)−4*H*-1,2,4-oxadiazole-5-thione	1.91	−1.82	3.19
11		5-(*m*-Tolyl)−1*H*-tetrazole	1.69	−1.33	4.07

(Continues)

TABLE 1 (Continued)

Cmpd #	Structure	Name	log $D_{3.5}$	log $D_{7.4}$	pK_a [1]
12		5-(*m*-Tolylmethyl)−1*H*-tetrazole	1.48	−1.17	4.73
13		3-Methylbenzamide	n.d.	n.d.	n.d.
14		2-(*m*-Tolyl)acetamide	n.d.	n.d.	n.d.
15		*N*-Methylsulfonyl-2-(*m*-tolyl)acetamide	0.94	−1.8	4.63
16		4-Hydroxy-5-(*m*-tolyl)−3*H*-furan-2-one	1.67	−2.13	2.95

[1]pKa data are from reference [21]. n.d.: not determined.

According to the previous determinations of p*Ka* and log D values at pH 3 and 8 [21], the Sirius T3 produced highly accurate results (standard deviations were always below 5%).

The eluent pH was fixed at 3.5 using 30 mM monobasic sodium phosphate (for ${}^{s}_{w}pH$ 3.5) and concentrated phosphoric acid.

SP1 and SP2 were conditioned with the selected mobile phase at a 0.6 ml/min flow rate for at least 40 min and at a 0.4 ml/min for at least 2 h, respectively, before use. All the analyses were carried out at 25°C column temperature. The injection volume was set at 20 µL, and the chromatographic analyses were followed at 254 nm. Before being used, all the mobile phases were vacuum filtered through a 0.45 µm membrane filter (Whatman) and then degassed with 10 min sonication. All the investigated compounds **1–16** were solubilized in MeOH and HPLC-grade water (50/50, v/v).

2.3 | Determination of log D values with the Sirius T3 apparatus

SiriusT3 apparatus (PION) measures pK_a, log P, and log D of ionizable molecules. In particular, the apparatus was used to determine experimental log D values of our series of small molecules **1–16** (Table 1) by using potentiometric titrations. Details about the procedure are fully described elsewhere [21].

2.4 | Determination of the isocratic hydrophobicity index φ0

With SP 1, the isocratic hydrophobicity index φ_0 [17], was calculated for each sample by measuring the logarithmic retention factor (log k) in the presence of different mixtures of MeOH (40, 50 e 60%, v/v) at ${}^{s}_{w}pH$ 3.5 (*Method 1*, Table 2). The retention factors data ($k = (t_R—t_0)/\ t_0$), obtained at various amounts of MeOH, were then extrapolated to 0% (v/v) MeOH and reported as k_w, using a linear procedure. The analyses were carried out at 0.5 ml/min after a proper equilibration of the chromatographic column according to the procedure reported in section 3.1, t_R represents the retention time of the target analyte, while $NaNO_2$ was used as the unretained marker to determine the dead time (t_0). By plotting the log k values with various organic solvent concentrations, the φ_0 value was obtained from the slope and the intercept of the straight line (φ_0 = —intercept/slope). For all compounds, and for

TABLE 2 Summary of the applied experimental conditions. Other details are reported in the Material and Methods sections. Codes "+" and "-" indicate, respectively, the presence and absence of the additive in the eluent.

Method#	SP#	Type of organic modifier	Flow rate (ml/min)	Type of additive		
				n-Octanol	*n*-Butylamine	*n*-Octylamine
1	1	MeOH	0.5	–	–	–
2	1	ACN	0.5	–	–	–
3	2	MeOH	0.3	–	–	–
4	2	MeOH	0.3	+	–	–
5	2	MeOH	0.3	+	+	–
6	2	MeOH	0.3	+	–	+

all the tested Methods, the retention factor values used for the determination of the φ_0 values were the average of three consecutive injections.

In the second method (*Method 2*, Table 2), MeCN was used as an organic modifier instead of MeOH, and the analyses were conducted only with the resulting better performing ${}^{s}_{w}pH$ 3.5. All other experimental parameters were kept unchanged.

To shorten the retention times with the better performing MeOH, the following analyses were isocratically conducted with the following amounts of the alcoholic modifier: 60, 65, and 70% (v/v). All the other experimental conditions were kept unchanged (*Method 3*, Table 2).

From this point of the study, SP2 was always used in place of SP1. This is a versatile, traditionally produced spherical silica gel carrier with RP properties and with a lower hydrophobicity extent than SP1, owing to its C8 chains. Since SP2 works at a pressure below 90 bar, the flow rate was reduced to 0.3 ml/min to avoid elevated backpressures. The analyses were therefore carried out with 65%, 70%, and 75% (v/v) MeOH (*Method 4*). All the other experimental conditions were kept the same as in *Method 3* as shown in Table 2.

As reported in Table 2 (*Method 5*), the next step was to use a mobile phase including *n*-octanol; the potentially profitable use deriving from the incorporation of this alcohol in such type of applications was described in detail [12, 13], and here further explained in section 3.1. A 0.25% (v/v) amount of *n*-octanol was added to MeOH to mimic the octanol hydrogen-bonding activity typical of shake-flask and other more conventional methods; *n*-octanol saturated water was used as well.

In the last two experiments (Table 2, *Methods 6* and *7*), the addition of a minor amount of either *n*-butylamine or *n*-octylamine (0.15%, v/v) to the eluent system was tested. The use of amine additives was found to be sometimes profitable in such types of studies [12, 22, 23]. 30 mM NaH_2PO_4 was dissolved in the mobile phase and 0.15% (v/v) amine modifier was added. Finally, the ${}^{s}_{w}pH$ was adjusted to 3.5 with concentrated H_3PO_4. In these experiments, the presence of *n*-octanol was maintained in the eluent.

2.5 | Statistical methods

Statistical analyses were performed with the aid of the open-source software CRAN-R version 4.2.2. [24]. To estimate and visualize correlations among all possible variables, the R packages "corrr" [25] and "corrplot" [26] were employed.

In statistical terms, the linear correlations were used to denote the association between two quantitative variables: chromatographically determined φ_0 and experimental log D. It was also assumed that the association was linear. To study the linear correlations among variables, the Pearson method was used. Indeed, Pearson's correlation coefficient is the test statistic that measures the statistical relationship, or association, between two continuous variables. It is considered among the best methods for measuring the association between variables of interest because it is based on the method of covariance. It provides information about the magnitude of the association, or correlation, as well as the direction of the relationship. p-Values < 0.001 (***) are considered statistically highly significant.

2.6 | Molecular modeling

3D Coordinates of each compound were obtained as reported in the original synthetic work [21]. Briefly, fragments were built in all the tautomeric states of interest. Next, a conformational search was performed, and the geometries of each conformer were energetically optimized at ab initio level in a simulated water environment, leading to the identification of the global minimum. The resulting conformer of each tautomeric state was used to identify the global minimum of each compound, thus, allowing to identify the preferred state of each fragment in water.

Finally, these selected tautomers were used to generate a set of nearly fifty molecular descriptors with the aid of the *QikProp* package (*QikProp*, Schrödinger, 2021) of the *Schrodinger Suite 2021*. Two of them resulted correlated with the φ_0 parameters, which were the number of Hydrogen Bond Donors (HBD) and the index of cohesive interaction (HBA×(HBD)$^{1/2}$/SASA) [27].

3 | RESULTS AND DISCUSSION

3.1 | Optimization of the chromatographic conditions to enhance the log D/φ0 correlation

Among all the calculable hydrophobic indices, the derived and isocratically measured chromatographic index φ_0 was selected [17]. φ_0 values can be estimated from within a range of values, unlike extrapolations to log k_W, which is often outside of the measurable range, and whose value is strongly dependent on the degree of curvature of the log k/φ plot (see below for details) [14, 19]. It is worth recalling that all log D values used in this work were experimentally determined at pH 3.5 [21], which explains the ${}^{s}_{w}pH$ fixed for each of the employed mobile phases.

As a general concept, correlations between two variables can be measured with the use of different indices (coefficients). The three most popular are Pearson's *r*, Spearman's *rho*, and Kendall's *tau* coefficients [28]. Herein, the Pearson method was selected as it measures the strength of the linear relationship between two normally distributed variables producing "*r*" values. This coefficient is a dimensionless measure of the covariance, with a value of −1 meaning a total negative linear correlation, 0 being no correlation, and + 1 meaning a total positive correlation.

The chromatographic conditions to improve the φ_0 versus log D correlation extent were optimized through a stepwise procedure. At first, attention was paid to the effects produced on the φ_0 chromatographic index by two different ${}^{s}_{w}pH$ values: 3.5 and 7.4. The data obtained for this first set of analysis clearly highlighted differences in performance between the two ${}^{s}_{w}pH$ values, as shown in Tables S1 and S2. As expected for ionizable compounds [29, 30], the fragments behave differently at the two ${}^{s}_{w}pH$ values, exhibiting largely different retention times.

As depicted in Figure 1A, the best correlation between experimental log D and φ_0 values were obtained at ${}^{s}_{w}pH$ 3.5 (r = 0.86; *Method 1*, Table 2), therefore this condition was maintained in all subsequent experiments.

Subsequently, the analysis of the samples at ${}^{s}_{w}pH$ 3.5 were performed using MeCN as an organic modifier (*Method 2*, Table 2) in place of MeOH. As expected, MeCN did show a higher eluotropic power over MeOH for all the investigated compounds. The chromatographic results obtained in this set of analyses attempt are listed in Table S3. The plots displayed in Figure 1B show a clear decrease in log D versus φ_0 correlation extent ($r = 0.56$).

To highlight the different results achieved with the two organic modifiers, the obtained Log k_w values are reported for all compounds in Table S4, along with the correlation coefficient (R^2) values of the φ_0 /log *k* plots. MeOH generated higher-to-comparable R^2 values than MeCN. Accordingly, the alcoholic modifier was included in all the subsequently tested mobile phases. This finding fits well with the more severe curvature already described by our group and other authors for the log k-φ_{MeCN} plots with respect to the log k-φ_{MeOH} ones [8, 31]. MeCN is well-known to generate a less extended linear portion in that plot, which enhances the uncertainty in the log k_w and *S* determination [15, 16, 31]. On the contrary, log *k* values were found to vary more linearly even for structurally unrelated compounds, when MeOH was used in the 20–80% (v/v) concentration range (namely, $0.2 \leq \varphi_{MeOH} \leq 0.8$). A more linear variation of some properties that are relevant for analyte retention was invoked to explain the better results provided by MeOH. More specifically, the eluent surface tension, which is of utmost importance for sample retention in RP systems, was demonstrated to change in an approximately linear way within the above φ_{MeOH} range [9, 31]. As a result of all the above, it is readily understandable that $\boldsymbol{\varphi_0}$ values strictly depend on the organic modifier used for the analysis [27]. Indeed, φ_0 values obtained with MeOH are however usually different from the ones deriving from the use of MeCN [31].

The following *Method 3* (Table 2) was tested with SP2 carrying C8 chains. The use of this shorter alkyl chain phase was suggested by the fact that some previous studies [12] clearly indicated distinct performance advantages from this material, over the more widely used C18-based stationary phases. The more pronounced correlations observed with C8 phases by other authors were ascribed to some profitable H-bond effects by silanol OH groups at the basis of the silica layer [12]. The chromatographic indices Log k_w, *S*, and φ_0 were calculated also under these experimental conditions and reported in Table S5. Pearson's correlation coefficient for the linear correlations among log D and φ was greatly improved ($r = 0.93$) by this experimental modification (*Method 3*, Table 2), as shown in Figure 1C.

Being inspired by the results obtained by other authors [6, 12], *n*-octanol was added as a minor alcoholic modifier in the eluent (*Method 4*, Table 2). The addition of a low amount of *n*-octanol in the mobile phase (see section 2.4 for details) turned out to be a powerful artifice for significantly increasing the correlation between the

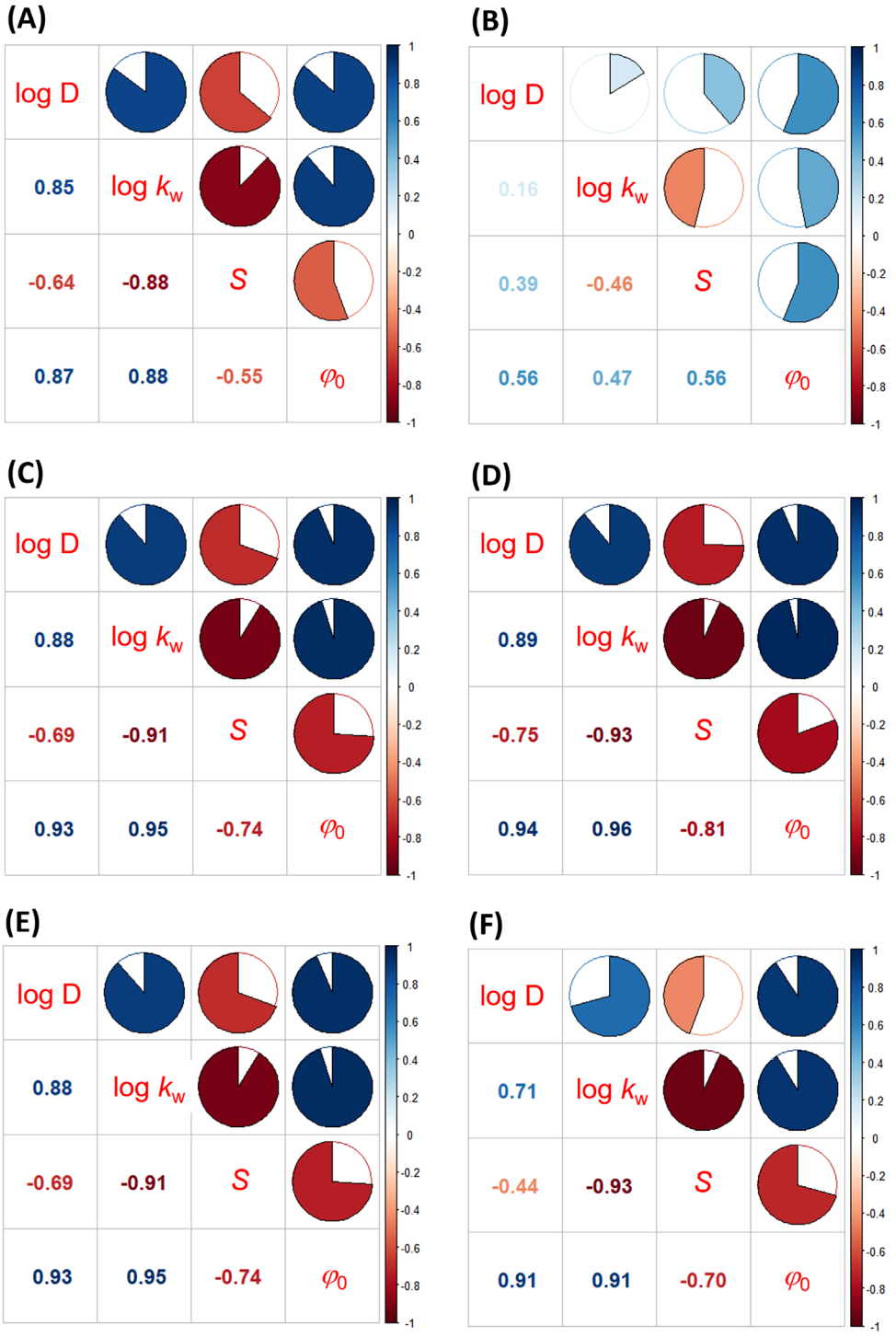


FIGURE 1 Correlation plots obtained with (A) *Method 1*; (B) *Method 2*; (C) *Method 3*; (D) *Method 4*; (E) *Method 5*; and (F) *Method 6*. Details on the scrutinized Methods are reported in Table 2. High correlation values both positive and negative are indicated in blue and red. As readily evident from the color scale of each correlation plot, the intensity of the color is related to the correlation degree. Log D values are referred to as pH 3.5.

derived log k_w parameters and the log P (either experimental or calculated) values. The benefit arising from the addition of the long-chain alcohol in the mobile phase was inferred to derive from the most faithful reproduction of the intermolecular interactions experienced by a solute in a classical *n*-octanol/water shake-flask determination. The results obtained with *Method 4* are shown in Table S6 and Figure 1D.

As shown, the Pearson's *r* value for the linear correlation among log D and φ_0 was very similar to that obtained

without the addition of *n*-octanol (Figure 1C,D), with only marginal improvement. This experimental finding further underlined the importance of introducing H-bond interactions to ameliorate the degree of correlation, implicitly indicating the minor contribution of hydrophobic interactions.

In the last two experiments, *n*-butylamine (*Method 5*, Table 2) and *n*-octylamine (*Method 6*, Table 2) were also added as amine modifiers to the eluent, and the deriving results were compared with the previous ones. The amine modifier was included in the eluent with the ultimate intent to reduce the silanophilic effect through a masking action of the residual silanols [12]. In this way, a possible different effect by the H-bonds from silanophilic activity or *n*-octanol OHs could be derived. The results of these experiments are shown in Table S7 and Figure 1E,F. Only marginal differences were found in terms of the quality of correlation. However, the results shown in Figure 1D–F lead to hypothesizing a more beneficial contribution on the log D/φ_0 correlation by the *n*-octanol OHs.

In summary the best correlation between φ_0 and experimental log D ($r = 0.94$, $p < 0.001$) was obtained with *Method 4* (Table 2, Figure 1D) when only *n*-octanol was incorporated in the MeOH/water-based eluents.

A deeper inspection of the correlation plots displayed in Figure 1D,E reveals a high correlation between the two chromatographic parameters log k_w and *S*. While the former is mostly related to solute-stationary phase interactions, the latter is generally accepted as being associated with the solute interaction with the solvent components [7, 14, 19, 32, 33]. The *S* parameter is strongly related to the number of eluent molecules in the solute solvation sphere and hence it depends on its size, shape as well as on the type, number, and orientation of polar substituents present on its backbone. Consequently, a linear relationship between *S* and log k_w values has to be expected when the solute/stationary phase interaction is governed by the same mechanism as the analyte solvation. High correlations in an *S* versus log k_w plot stand for a high uniformity in the retention mechanism for a certain set of compounds [27].

From the network plot displayed in Figure 2 where: (i) each path connecting from a variable to another variable represents a correlation value, *r*; (ii) a path with blue color represents a positive correlation between two variables, red: negative correlation; (iii) width and transparency of path explain the magnitude of a correlation between two variables, it is possible to further corroborate that log D is highly correlated with the relatively close variable φ_0. Notably, in this network, the central point is the variable log k_w showing its important role and the central connections with all other studied variables.

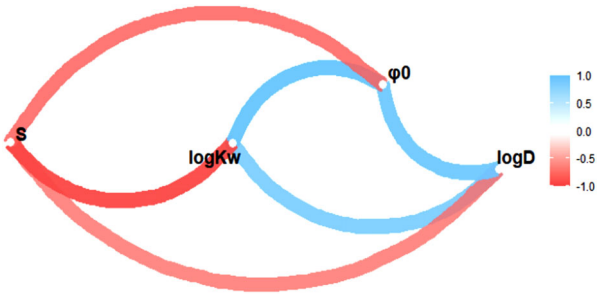


FIGURE 2 Correlation network plot; data are elaborated from Table S6. Each path connecting from one variable to another variable is representative of a Pearson correlation value. Accordingly, a blue-colored path explains a positive correlation between two interested variables, while the red color indicates a negative correlation. Both the width and transparency of each path are related to the magnitude of a correlation between two considered variables.

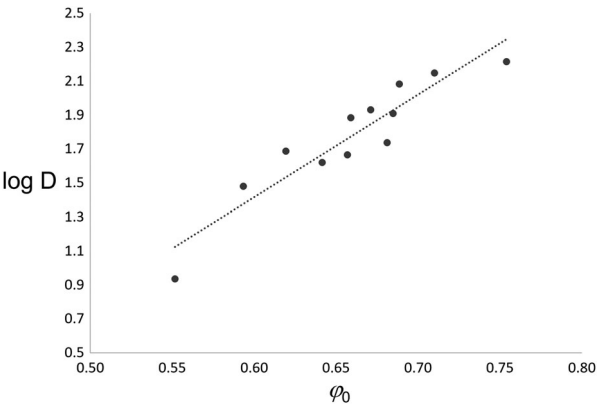


FIGURE 3 Correlation plot obtained with log D and φ_0 values applying *Method 4* (other details in Table 2 and Figure 1D). Experimental conditions: column, SP2; mobile phase, MeOH/water-based with the addition of *n*-octanol and a ${}^{s}_{w}pH$ fixed at 3.5; flow rate, 0.3 ml/min.

At this point of the study, the best correlation conditions were defined, and the corresponding regression analysis (Equation (3)) obtained by elaborating the data from Table S6, was established.

$$\log D = 6.04\,\varphi_0 - 2.21 \quad (3)$$

Log D corresponds to the experimental value obtained for each fragment by the SiriusT3 instrument and φ_0 is the corresponding derived chromatographic index. In Figure 3, the high quality of correlation is readily evident.

In Table 3 are reported the log D values for fragments **3**, **4**, **13**, and **14 are** estimated by using Equation (3).

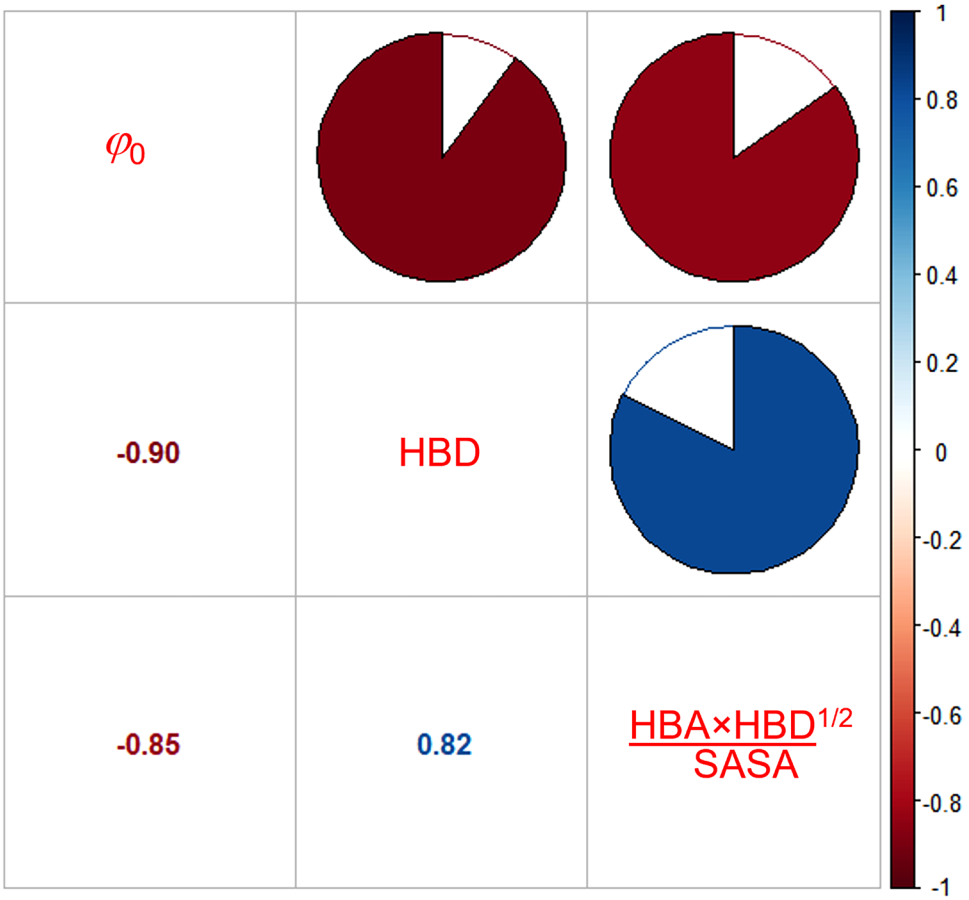


FIGURE 4 Correlation plots of φ_0 and the two selected molecular descriptors: hydrogen bond donors (HBD) and index of cohesive interaction ($HBA \times HBD^{1/2}/SASA$).

TABLE 3 Estimation of missing log D values by Equation (3).

Cmpd #	φ_0	Predicted log D
3	0.192	–1.05
4	0.389	0.14
13	0.523	0.95
14	0.547	1.1

3.2 | Explanation of the chromatographic results through selected molecular descriptors

With the aim to identify chemical characteristics of the investigated fragments useful to explain the obtained φ_0 data, a series of molecular descriptors (see section 2.6 for details) were calculated in silico. Among them, the number of HBD and the index of cohesive interaction ($HBA \times HBD^{1/2}/SASA$) descriptors particularly attracted our attention. The calculated correlation coefficients (Pearson's r), obtained for these two descriptors against the φ_0 chromatographic index, are reported in Table S8 and depicted on the correlogram (Figure 4).

The two selected descriptors are related to the ability of the analytes to interact with the eluent components. Indeed, the hydrogen bond acceptor and donor capability together with its solvent-accessible surface area determine the possibility of a molecule engaging or not interacting in the solvation shell. All these aspects define the hydrophobicity profile of a chemical entity, and this is the reason why they are found correlated with the φ_0. As expected, the negative correlation with the hydrophobicity index means that high values of HBD and index of cohesive interaction are associated with low values of φ_0.

It must be noted that the HBD descriptor, for its nature, is a discrete measure and lacks the ability to differentiate fragments that share the same HBD number but displays a different chemical structure. The index of cohesive interaction is a continuous measure that embeds different descriptors giving a more exhaustive description of the molecular features of the fragment and can fine discriminate among similar compounds, especially thanks to the presence of the SASA component that embeds three-dimensional information. Interestingly, fragments **3** and **4** that bear three HBD and display the higher indices of cohesive interaction of the series, are characterized by the

lowerφ_0 values of the set. Fragments **13** and **14** have two HBD displays φ_0 values of 0.523 and 0.547, respectively, that are quite low but similar to the values registered for the molecules having one HBD. Probably, the hydrophobicity index is sensitive to multiple factors and is surely correlated with the descriptors herein found, but these are not enough to fully explain and describe the whole hydrophobic profile of this set of fragments.

4 | CONCLUDING REMARKS

In the context of early-phase drug discovery programs, the rapid knowledge of the physico-chemical properties (including hydrophobicity through the evaluation of the log P/log D value) of new molecular entities is of prior importance. This is also valid for new molecular fragments, dedicatedly synthesized to improve, inter alia, the potency, selectivity, solubility, and permeability of the final molecules. In this framework, the use of a reduced amount of sample is highly desired. Accordingly, the use of derived chromatographic indices plainly satisfies the above requirements. Along this line, in this study, we optimized an HPLC method for the evaluation of log D values of polar molecular fragments investigated in the context of a fragment-based drug discovery program. The new optimized method, despite having been tested on a small array of selected analytes, unlike potentiometric methods is more universal, thus allowing us to determine the log D/log P values for low or non-ionizable substances.

As a result of the study, linear correlations were found between experimental log D and φ_0 values. The best correlation ($r = 0.94$, $p < 0.001$) was obtained with water/MeOH-based mobile phases with a low content of *n*-octanol. The C8-based stationary phase sensitively outperformed the more diffusely employed C18 one. The body of evidence strongly highlighted the importance of introducing H-bond interactions to ameliorate the degree of correlation, implicitly indicating the minor contribution of hydrophobic interactions.

The importance of H-bonds in the quality of log D/φ_0 correlation was furthermore confirmed by *in silico* calculations, indicating that the two molecular descriptors, the number of HBD and the index of cohesive interaction (HBA×HBD$^{1/2}$/SASA) consistently correlated with the experimental data.

Using the φ_0 indexes with the optimized RP-HPLC methods allowed us to estimate the log D of four molecular fragments, which was not allowed by the Sirius T3 apparatus.

Based on these experimental findings, it is our intention to further expand the chemical space of the analytes, with the perspective to use the method for wider applications.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Laura Mercolini https://orcid.org/0000-0002-0644-9461
Roccaldo Sardella https://orcid.org/0000-0001-9856-0230

REFERENCES

1. Landry ML, Crawford JJ. Log D contributions of substituents commonly used in medicinal chemistry. ACS Med Chem Lett. 2019;11:72–6.
2. Hansch C, Maloney PP, Fujita T, Muir RM. Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients. Nature. 1962;4824:178–80.
3. Leo A, Hansch C, Elkins D. Partition coefficients and their uses. Chem Rev. 1971;6:525–616.
4. Fujita T, Iwasa J, Hansch C. A new substituent constant, π, derived from partition coefficients. J Am Chem Soc. 1964;86:5175–80.
5. OECD. Test no. 107: partition coefficient (n-octanol/water): shake flask method, OECD guidelines for the testing of chemicals, section 1. Paris: OECD Publishing; 1995.
6. Lombardo F, Shalaeva MY, Tupper KA, Gao F, Abraham MH. ELogPoct: A tool for lipophilicity determination in drug discovery. J Med Chem. 2000;43:2922–8.
7. Liu X, Tanaka H, Yamauchi A, Testa B, Chuman H. Determination of lipophilicity by reversed-phase high-performance liquid chromatography: Influence of 1-octanol in the mobile phase. J Chromatogr A. 2005;1091:51–9.
8. Tate PA, Dorsey JG. Column selection for liquid chromatographic estimation of the $k_{w'}$ hydrophobicity parameter. J Chromatogr A. 2004;1042:37–48.
9. Horváth C, Melander W, Molnár I. Solvophobic interactions in liquid chromatography with nonpolar stationary phases. J Chromatogr A. 1976;125:129–56.
10. Vailaya A, Horváth C. Solvophobic theory and normalized free energies of nonpolar substances in reversed phase chromatography. J Phys Chem B. 1997;30:5875–88.
11. Valkó K, Bevan C, Reynolds D. Chromatographic hydrophobicity index by fast-gradient RP-HPLC: a high-throughput alternative to log P/Log D. Anal Chem. 1997;69:2022–9.
12. Minick DJ, Frenz JH, Patrick MA, Brent DA. A comprehensive method for determining hydrophobicity constants by reversed-phase high-performance liquid chromatography. J Med Chem. 1988;31:1923–33.
13. Valkó K. General approach for the estimation of octanol/water partition coefficient by reversed-phase high-performance liquid chromatography. J Liq Chromatogr. 1984;7:1405–24.
14. Natalini B, Sardella R, Camaioni E, Macchiarulo A, Gioiello A, Carbone G, et al. Derived chromatographic indices as effective tools to study the self-aggregation process of bile acids. J Pharm Biomed Anal. 2009;50:613–21.

15. Braumann T, Weber G, Grimme LH. Quantitative structure—activity relationships for herbicides: reversed-phase liquid chromatographic retention parameter, Log k_w, versus liquid-liquid partition coefficient as a model of the hydrophobicity of phenylureas, *s*-triazines and phenoxycarbonic acid derivative. J Chromatogr A. 1983;261:329–43.
16. Harnisch M, Möckel H. J, Schulze G. Relationship between Log P_{ow}, shake-flask values and capacity factors derived from reversed-phase high-performance liquid chromatography for *n*-alkylbenzenes and some oecd reference substances. J Chromatogr A. 1983;282:315–32.
17. Valkó K, Slégel P. New chromatographic hydrophobicity index (ϕ_0) based on the slope and the intercept of the log k' versus organic phase concentration plot. J Chromatogr A. 1993;631:49–61.
18. Valkó K. Application of high-performance liquid chromatography based measurements of lipophilicity to model biological distribution. J Chromatogr A. 2004;1037:299–310.
19. Natalini B, Sardella R, Camaioni E, Gioiello A, Pellicciari R. Correlation between CMC and chromatographic index: Simple and effective evaluation of the hydrophobic/hydrophilic balance of bile acids. Anal Bioanal Chem. 2007;388:1681–8.
20. de Esch IJP, Erlanson DA, Jahnke W, Johnson CN, Walsh L. Fragment-to-lead medicinal chemistry publications in 2020. J Med Chem. 2022;65:84–99.
21. Ferri M, Alunno M, Greco FA, Mammoli A, Saluti G, Carotti A, et al. Fragment based drug design and diversity-oriented synthesis of carboxylic acid isosteres. Bioorg Med Chem. 2020;28:115731.
22. Euerby MR, Petersson P. Chromatographic classification and comparison of commercially available reversed-phase liquid chromatographic columns containing polar embedded groups/amino endcappings using principal component analysis. J Chromatogr A. 2005;1088:1–15.
23. Wilson NS, Gilroy J, Dolan JW, Snyder LR. Column selectivity in reversed-phase liquid chromatography: VI. Columns with embedded or end-capping polar groups. J Chromatogr A. 2004;1026:91–100.
24. R Core Team R. A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2002. http://www.R-project.org
25. Kuhn M, Jackson S, Cimentada J. corrr: Correlations in R (version 0.4.4). https://CRAN.R-project.org/package=corrr
26. Wei T, Simko V. R package 'corrplot': Visualization of a Correlation Matrix. (Version 0.92). 2021. https://github.com/taiyun/corrplot
27. Jorgensen WL, Duffy EM. Prediction of drug solubility from monte carlo simulations. Bioorg Med Chem Lett. 2000;11:1155–8.
28. Hauke J, Kossowski T. Comparison of values of pearson's and spearman's correlation coefficients on the same sets of data. Quaestiones Geographicae. 2011;30:87–93.
29. Canals I, Valkó K, Bosch E, Hill AP, Rosés M. Retention of ionizable compounds on HPLC. 8. Influence of mobile-phase pH change on the chromatographic retention of acids and bases during gradient elution. Anal Chem. 2001;73:4937–45.
30. Lewis JA, Dolan JW, Snyder LR, Molnar I. Computer simulation for the prediction of separation as a function of pH for reversed-phase high-performance liquid chromatography: II. Resolution as a function of simultaneous change in pH and solvent strength. J Chromatogr A. 1992;592:197–208.
31. Natalini B, Sardella R, Gioiello A, Rosatelli E, Ianni F, Camaioni E, et al. Fast chromatographic determination of the bile salt critical micellar concentration. Anal Bioanal Chem. 2011;401:267–74.
32. Liu X, Tanaka H, Yamauchi A, Testa B, Chuman H. Lipophilicity measurement by reversed-phase high-performance liquid chromatography (RP-HPLC): A comparison of two stationary phases based on retention mechanisms. Helv Chim Acta. 2004;87:2866–76.
33. Shaw R, Elliott WH, Barisas BG. Estimation of critical micelle concentrations of bile acids by reversed-phase high performance liquid chromatography. Michrochim Acta. 1991;3:137–45.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Carotti A, Varfaj I, Pruscini I, Abualzulof GWA, Mercolini L, Bianconi E, et al. Estimating the hydrophobicity extent of molecular fragments using reversed-phase liquid chromatography. J Sep Sci. 2023;46:2300346. https://doi.org/10.1002/jssc.202300346