High-throughput log $P_{o/w}$ determination from UHPLC measurements: revisiting the Chromatographic Hydrophobicity Index

Xavier Subirats, Martí Rosés, Elisabeth Bosch*

Departament de Química Analítica – Institut de Biomedicina, Universitat de Barcelona, Martí i Franquès 1-11, 08028 Barcelona

*Corresponding Author: E-mail: e.bosch@ub.edu; Tel.: +34 934021284; Fax: +34 934021233

Dedicated to Professor Roman Kaliszan on the occasion of his 70th birthday.

Abstract

A fast and accurate lipophilicity determination is fundamental in the drug discovery process, as long as it is a relevant property in the absorption, distribution, metabolism, excretion and toxicity (ADMET) of a potential drug substance. In the present work, different models based on chromatographic retention values for a large set of compounds and some of their molecular descriptors (calculated by ACD/Labs or CODESSA programs) have been examined in order to establish reliable equations for log $P_{o/w}$ determination from fast chromatographic hydrophobicity index (CHI) measurements. This appears to be a very interesting high-throughput methodology for screening purposes, since CHI values can be measured by UHPLC in very short runs (< 4 min) and molecular descriptors can be easily computed from the structure of any compound. The selected final descriptors were Abraham's hydrogen-bond acidity (*A*) and excess molar refraction (*E*) from ACD/Labs, and hydrogen-bond acidity HDCA-1/TMSA and HOMO-LUMO polarizability descriptors from CODESSA software. The proposed equations allow an accurate determination of log $P_{o/w}$ with standard errors in the range of 0.4 units.

Keywords

Lipophilicity; partition coefficient; partition ratio; log $P_{o/w}$; chromatographic hydrophobicity index; CHI; molecular descriptors.

Highlights

A high-throughput log $P_{o/w}$ determination approach is proposed in the present work.

It is based on fast UHPLC CHI measurements and molecular descriptors.

Molecular descriptors were calculated from ACD/Labs (Abraham) and CODESSA.

Accurate determination of log $P_{o/w}$ with standard errors in the range of 0.4 is achieved.

1. Introduction

It is widely accepted that lipophilicity plays a fundamental role in the estimation of the ADMET properties (absorption, distribution, metabolism, excretion, and toxicity) of a drug candidate. A simple search in SciFinder reveals more than 20 thousand hits for the keyword 'lipophilicity' in the period comprised between 2010 and 2015. Therefore, it seems very convenient to establish high-throughput, reliable and accurate methodologies to determine lipophilicity in the frame of the drug discovery process.

The most widely used lipophilicity index is the logarithm of the 1-octanol/water partition coefficient (log $P_{o/w}$), traditionally measured by equilibrating the compound in this two-phase system in a shake-flask for a relatively long period of time (3 to 24h), followed by the determination of the concentration of the compound in the partition phases by spectrometry or liquid chromatography [1-4]. This method, although being the reference one, is very time consuming and require a high purity of sample. In order to overcome these limitations, a high-throughput Chromatographic Hydrophobicity Index (CHI) was proposed by Valkó [5, 6] based on the retention times measured in a fast gradient reversed-phase HPLC method. The main advantages of CHI in relation to former chromatographic lipophilicity indexes, besides the rapidness of the measurements, were its independency of the column dimensions, the particular fast gradient programmed and the flow-rate. Therefore, when the nature of the stationary phase is maintained (e.g. C18), the CHI value of a compound can be measured in any column, independently of its length, wide or particle size. However, this index depends on the particular organic modifier employed (acetonitrile, methanol), the temperature and, for acidic or basic compounds, the pH of the mobile phase. A few years later and in collaboration with Abraham, the correlations between log $P_{o/w}$ and CHI and lipophilicity indexes were improved using acetonitrile (MeCN) as organic modifier and including experimentally determined hydrogen-bond acidity descriptors in the correlation equations [7, 8].

The method of Abraham relates a solvation property (SP) with the sum of specific interactions terms:

$$\log \mathrm{SP} = c + eE + sS + aA + bB + vV \qquad (1)$$

where E, S, A, B, and V are the solute descriptors, and c, e, s, a, b, and v are the system constants. Briefly, E is the excess molar refraction (i.e. difference between the molar

refraction of a particular solute and that of an alkane of equivalent volume) which models the dispersion force interactions arising from the greater polarizability of π and n electrons, S accounts for the solute dipolarity/polarizability due to interactions between dipoles and induced dipoles, A and B are the solute hydrogen-bond acidity and basicity descriptors, respectively, and V is the McGowan's volume of the molecule. The coefficients c, e, s, a, b, and v reflect the complementary effect of the solute descriptors on the solvent phases, providing chemical information that allows the characterization of the system [9]. Using this approach, a very reliable log $P_{o/w}$ equation was built by Abraham from the experimental data (log $P_{o/w}$, and molecular descriptors) of more than six hundred compounds [10].

In the present study the correlations between log $P_{o/w}$ and the CHI values, measured by Valkó using a C18 HPLC stationary phase and acetonitrile as organic modifier (CHI_{MeCN}) [7], have been revisited using calculated molecular descriptors instead of the experimental ones used in the original work. Computed descriptors, although less accurate than those experimentally obtained, can be calculated for any compound from its structure investing a short time, which seems very convenient for screening purposes in drug discovery processes. Finally, log $P_{o/w}$ values were determined for a set of test compounds by means of the established equations, whose CHI_{MeCN} values were measured by an even faster UHPLC method. Besides a lower consumption of mobile phase solvents (0.5 mL/min) in relation to former HPLC methodologies, UHPLC technology provides high resolutions in short runs (3.1 minutes in the present work) and very short equilibration times within consecutive gradients (0.7 minutes).

2. Material and methods

2.1. Calculation of the descriptors

2.1.1. Abraham two-dimensional descriptors

The values of the solvation parameter model descriptors were calculated using the ACD/Absolv module of the ACD/Percepta platform (Advanced Chemistry Development, Inc. (ACD/Labs), Toronto, ON, Canada). According to this 2D approach, a particular descriptor is obtained from the contributions of molecular fragments to the whole molecule property [11].

2.1.2. CODESSA three-dimensional descriptors

The structures were drawn using the HyperChem Lite software (HyperCube, Gainesville, USA). The geometry optimization was performed with the semiempirical quantum method AM1 using the MOPAC 6.0 and AuxQSPR programs in order to obtain the global energy

minimum. Thus, CODESSA software package (University of Florida, USA) was employed to calculate the numeric values of the structural descriptors. Four different classes of molecular descriptors were computed: hydrogen bonding, both donor and acceptor; geometrical, derived from the 3D coordinates of the atoms; electrostatic, accounting for the charge distribution of the molecule; and quantum-chemical, obtained from MOPAC calculations [12].

2.2 Chromatographic conditions

A Shimadzu (Kyoto, Japan) Nexera UHPLC system was used for CHI measurements. The system was equipped with two LC-30AD high-pressure pumps, a DGU-20A5 online degasser, a CTO-10ASvp oven thermostatized at 25^{o}C, a SIL-30AC autosampler, a SPD-M20A diode array detector and a CBM-20Alite controller. Retention data were obtained from a Waters (Milford, MA, USA) Acquity BEH C18 column, 1.7 μm, 50 mm × 2.1 mm. Samples were prepared in DMSO at a concentration of 0.5 mg/mL, and the injection volume was 0.2 μL. The percentage of MeCN in the fast gradient program was: 0.0-0.4 min, 0%; 0.4-2.5 min, 0-100%; 2.5-2.9 min, 100%; 2.9-3.1 min, 100-0%; 3.1-3.8 0% [13]. Aqueous buffers were, in all cases, 50 mM ammonium acetate at the desired pH value. Acidic buffers were prepared from glacial acetic acid and the pH was adjusted with small volumes of concentrated ammonia (25%), and basic buffers were prepared inversely. Medium acidic and basic buffers were obtained by solving the salt and adjusting the pH with concentrated ammonia or glacial acetic acid. 4-hydroxybenzyl alcohol (pH<8) or caffeine (pH>8), acetanilide, acetophenone, propiophenone, butyrophenone, valerophenone, hexanophenone, and heptanophenone were used as calibration standards for CHI measurements.

2.3 Chemicals

Acetonitrile HPLC gradient grade was purchased from VWR (West Chester, PA, USA). Water was purified by the Milli-Q® plus system from Millipore (Billerica, MA, USA) with a resistivity of 18.2 MΩ·cm. The chemicals used for buffer preparation were anhydrous sodium acetate (>99.6%) and glacial acetic acid from J.T. Baker (Deventer, The Netherlands), and ammonia from Merck (Darmstadt, Germany). Dimethyl sulfoxide (DMSO, >99.9%) was purchased from J.T. Baker. The injected compounds were purchased from Merck, Sigma–Aldrich and Fluka, or provided by Almirall (Barcelona, Spain), all of them in high purity grade.

3. Results and discussion

135 3.1 Training set

136 The training set contained 125 structurally different substances with known log $P_{o/w}$ [14] and
137 CHI_{MeCN} [7] values. It was selected to present a wide range of log $P_{o/w}$, comprised between -
138 2.51 and 5.52, and, consequently, of CHI_{MeCN} values, from -23.0 to 128.4. The list of the
139 substances integrating the training set is available as supplementary material.

140

141 3.2 CHI and Abraham's descriptors

142 Abraham pointed out [15] that for certain solutes in water–solvent systems where the organic
143 layer contains considerable quantities of water, the solute hydrogen bond basicity (*B*) behave
144 in an anomalous way. This is precisely the case of mobile phases commonly used in reversed-
145 phase liquid chromatography. Affected solutes were those containing the S=O and P=O
146 (except sulfones, sulfonamides, sulfonates, and phosphates), anilines, and pyridines. With the
147 aim of solving this issue, Abraham assigned an alternative hydrogen bond basicity descriptor,
148 named B^0. Therefore, in the present work B^0 has been used instead of *B* and, in fact, slightly
149 better correlations have been obtained.

150 According to Eq. (1), independent correlations between both lipophilicity indexes, log
151 $P_{o/w}$ and CHI_{MeCN}, with molecular descriptors of the training set compounds have been
152 established. However, all the coefficients (*e*, *s*, *a*, *b*, and *v*) were normalized dividing each one
153 by the length of the coefficients vector (*l*):

154 $$l = \sqrt{e^2 + s^2 + a^2 + b^2 + v^2} \qquad (2)$$

155 resulting in the following expressions $e_u = e/l$, $s_u = s/l$, $a_u = a/l$, $b_u = b/l$, and $v_u = v/l$,
156 where the subscript *u* denotes normalized coefficients (Table 1). From them, the comparison
157 between lipophilicity systems becomes very easy.

158 In the present work we have considered the measured log $P_{o/w}$ and CHI_{MeCN} of 125
159 substances [7], but in this case the Abraham's molecular descriptors were calculated using the
160 ACD/Labs software. Since computed descriptors are expected to be less accurate than
161 experimental ones, they lead in fact to worse correlation coefficients (*R*) and higher standard
162 errors (*SE*) for lipophilicity equations. However, as presented in Table 1, the normalized
163 coefficients of log $P_{o/w}$ system obtained from a huge set of experimental molecular descriptors
164 [10] are very similar to those calculated in the present work, suggesting that computed
165 descriptors provide a reasonable accuracy. In addition, both lipophilicity systems, log $P_{o/w}$ and
166 CHI_{MeCN}, show similar normalized system coefficients and a very short distance between
167 normalized vectors ($d < 0.2$) [16], confirming the similarity between the studied systems. As

already pointed out by Valkó and co-workers [7], when comparing log $P_{o/w}$ and CHI_{MeCN} coefficients, the major difference lies on the hydrogen-bond basicity of the system (*a*). Notice that the coefficients of Eq. (1) reflect the complementary property of the system related to the solute. Thus, since *A* is the hydrogen-bond acidity descriptor of the solute, *a* is the difference in hydrogen-bond basicity between stationary and mobile phases in the chromatographic system, or between octanolic and aqueous phases in the log $P_{o/w}$ system. Therefore, the inclusion of a hydrogen-bond acidity descriptor of the solute is very convenient in the determination of log $P_{o/w}$ values from CHI_{MeCN} measurements. In addition, a minor difference is found for the *e* coefficient, especially between the log $P_{o/w}$ and CHI_{MeCN} solvation equations built with calculated molecular descriptors.

3.3 CHI and CODESSA descriptors

In order to assay a model similar to that of Abraham, several molecular descriptors were calculated by means of the CODESSA software accounting for (1) hydrogen-bonding donor ability (HDCA, FHDCA, HDCA-1, HDCA-1/TMSA, HDCA-2, HDCA-2/TMSA, HDCA-2/SQRT(TMSA), HDSA-1, HDSA-1/TMSA, HDSA-2, HDSA-2/TMSA, HDSA-2/SQRT(TMSA), FHDSA, and count of H-donors sites); (2) hydrogen-bonding acceptor ability (HACA-1, HACA-1/TMSA, HACA-2, HACA-2/TMSA, HACA-2/SQRT(TMSA), FHACA, HACA, HASA-1, HASA-1/TMSA, HASA-2, HASA-2/TMSA, HASA-2/SQRT(TMSA), HASA, FHASA, and count of H-acceptor sites); (3) polarity and polarizability based intermolecular interactions (DPSA-1, DPSA-2, DPSA-3, PPSA-1, PNSA-1, Polarity parameter, Polarity parameter/square distance (pol/d^2), HOMO - LUMO energy gap, ALFA-, 1X BETA-, (1/2)X BETA-, 1X GAMMA-, and (1/6)X GAMMA polarizability); and (4) molecular volume and surface area (TMSA). Further information about these descriptors is available at the CODESSA PRO project website [17].

To improve the correlations between log $P_{o/w}$ and CHI_{MeCN}, some descriptors accounting for the above mentioned properties were selected because it was proved that they allowed a successful explanation of the relationships between chromatographic retention and octanol/water partition [7, 8, 18-20]. Then, mimicking Eq. (1) and with the aim of selecting the best molecular descriptors, those accounting for hydrogen-bond acidity and basicity, polarity, polarizability and volume, have been tested. Thus, multiple linear correlations including only one descriptor at each time were calculated. The best descriptor of each category has been included in Table 2, which also shows the improvement in the residual sum of squares (SS_{res}) of the multilinear equations in relation to the original linear correlation

between log $P_{o/w}$ and CHI_{MeCN}. Concerning the hydrogen-bond acidity of the solute, two different Zefirov's PC descriptors showed a similar improvement, HDCA-1/TMSA and HDCA-2/TMSA, being the former a 37% and the latter a 35%. In relation to hydrogen-bond basicity, HACA-2/TMSA [Zefirov's PC] reduced the SS_{res} in a 21%. Finally, regarding polarity and polarizability, the descriptors that allowed better correlations were DPSA-3 [Zefirov's PC] and HOMO-LUMO energy gap, with an improvement of SS_{res} of 13% and 10%, respectively. The consideration of the molecular volume or surface area descriptors alone only represented a SS_{res} reduction of the 2% and, consequently, were not further considered.

The above mentioned CODESSA descriptors were individually correlated among them and with CHI_{MeCN}, in order to explore the ortogonality of the information provided by each of the descriptors (Table 3). As expected HDCA-1/TMSA and HDCA-2/TMSA are highly correlated (r^2=0.986), since they account for the same solute property (hydrogen-bond acidity), but in a similar way of Abraham's approach, CHI_{MeCN} is poorly correlated with the rest of descriptors ($r^2 \leq 0.6$). However, acidity and basicity hydrogen-bond descriptors are correlated to some extent ($r^2 \leq 0.8$), suggesting that similar information is provided. It is noteworthy that DPSA-3 and HOMO-LUMO are not correlated at all ($r^2 = 0$), and neither of them are correlated with CHI_{MeCN}.($r^2 < 0.1$).

3.4 Training set correlations: comparison between Abraham's (ACD/Labs) and CODESSA descriptors

Several linear and multilinear correlations between log $P_{o/w}$ vs. CHI_{MeCN} and molecular descriptors performed with the 125 substances of the training set were tested. Note that log $P_{o/w}$ and CHI_{MeCN} correlations were published by Valkó [7] in 2001 using experimental values for a set of 86 substances and a log $P_{o/w}$ range comprised between -0.35 and 5.52, whereas in the present work this range has been extended to -2.51 and molecular descriptors have been calculated by both ACD/Labs and CODESSA programs. The simple linear correlation between log $P_{o/w}$ and CHI_{MeCN} leads to the following expression:

$$\log P_{o/w} = 0.046(\pm 0.002) \cdot CHI_{MeCN} - 0.81(\pm 0.11) \quad (N=125,\ R^2=0.863,\ SE=0.50,\ SS_{res}=30.6) \qquad (3)$$

As expected from the comparison of log $P_{o/w}$ and CHI_{MeCN} solvation equations, the addition of Abraham's hydrogen-bond acidity descriptor (A) significantly enhances the correlation:

$$\log P_{o/w} = 0.055(\pm 0.001)\cdot CHI_{MeCN} + 1.30(\pm 0.10)\cdot A - 1.85(\pm 0.11) \quad (N=125,\ R^2=0.940,\ SE=0.33,\ SS_{res}=13.5) \tag{4}$$

This expression is nearly identical to that proposed by Valkó ($\log P_{o/w} = 0.054\cdot CHI_{MeCN} + 1.319\cdot A - 1.877$), despite Eq. (4) includes about a 50% more of compounds and calculated molecular descriptors instead of the experimental ones, showing in this way the robustness of the correlation. Furthermore, the addition of the excess molar refraction descriptor (E) improves the statistics of the equation, but only slightly:

$$\log P_{o/w} = 0.054(\pm 0.001)\cdot CHI_{MeCN} + 1.24(\pm 0.10)\cdot A + 0.18(\pm 0.05)\cdot E - 2.02(\pm 0.12) \quad (N=125,\ R^2=0.946,\ SE=0.32,\ SS_{res}=12.2) \tag{5}$$

In the case of CODESSA descriptors, the ones proposed in Table 2 were included in multilinear correlations between log $P_{o/w}$ and CHI_{MeCN}, obtaining the following equations:

$$\log P_{o/w} = 0.058(\pm 0.002)\cdot CHI_{MeCN} + 89.7(\pm 18.4)\cdot[HDCA\text{-}1/TMSA] - 11.7(\pm 35.4)\cdot[HACA\text{-}2/TMSA] + 7.5(\pm 5.2)\cdot 10^{-3}\cdot[DPSA\text{-}3] - 0.15(\pm 0.04)\cdot[HOMO\text{-}LUMO] - 0.84(\pm 0.43) \quad (N=125,\ R^2=0.924,\ SE=0.38,\ SS_{res}=17.0) \tag{6}$$

$$\log P_{o/w} = 0.057(\pm 0.002)\cdot CHI_{MeCN} + 427(\pm 99)\cdot[HDCA\text{-}2/TMSA] + 5.5(\pm 38.8)\cdot[HACA\text{-}2/TMSA] + 5.3(\pm 5.4)\cdot 10^{-3}\cdot[DPSA\text{-}3] - 0.15(\pm 0.04)\cdot[HOMO\text{-}LUMO] - 0.69(\pm 0.43) \quad (N=125,\ R^2=0.922,\ SE=0.38,\ SS_{res}=17.6) \tag{7}$$

It should be noticed that the only difference between Eq. (6) and (7) is the hydrogen-bond acidity descriptor. Although hydrogen-bond basicity (HACA-2/TMSA) and polarity (DPSA-3) improved individual correlations between log $P_{o/w}$ and CHI_{MeCN} (Table 2), they become irrelevant in Eqs. (6) and (7) because of the magnitude of the errors associated to the coefficients. Therefore, taking into account that correlations with HDCA-1/TMSA as hydrogen-bond acidity descriptor are slightly better, a new expression was proposed excluding HACA-2/TMSA and DPSA-3:

$$\log P_{o/w} = 0.059(\pm 0.002)\cdot CHI_{MeCN} + 103.0(\pm 12.1)\cdot[HDCA\text{-}1/TMSA] - 0.15(\pm 0.04)\cdot[HOMO\text{-}LUMO] - 0.80(\pm 0.43) \quad (N=125,\ R^2=0.922,\ SE=0.38,\ SS_{res}=17.5) \tag{8}$$

These findings are consistent with the expression reported by Pallicer *et al.* [12] in the determination of log $P_{o/w}$ from chromatographic measurements by means of the polarity model.

In summary, Eq. (5) shows the best correlation between log $P_{o/w}$ and CHI_{MeCN}, with a standard error (*SE*) of only 0.32 log $P_{o/w}$ units. In fact, it is a tiny improvement in relation to Eq. (4), but as long as the time required for the calculation of the *A* descriptor alone or all of them is nearly the same (only seconds of difference), it seems interesting to consider Eq. (4) as well. On the other hand, the correlations using CODESSA descriptors are statistically very similar and, consequently, it seems reasonable to keep the one with a less number of molecular descriptors, that is Eq. (8). Finally, since the suppression of the HOMO-LUMO term provides statistic parameters similar to those of Eq. (4), the simpler expression should be considered too:

$$\log P_{o/w} = 0.060(\pm 0.002)\cdot CHI_{MeCN} + 107.4(\pm 12.7)\cdot\left[HDCA\text{-}1/TMSA\right] - 2.22(\pm 0.19) \quad (9)$$
$$(N=125,\ R^2=0.914,\ SE=0.40,\ SS_{res}=19.3)$$

3.5 Test set correlations ACD vs CODESSA descriptors

A representative set of 43 structurally different compounds of pharmaceutical interest with log $P_{o/w}$ values in the range between -0.07 and 4.45 were selected as test set. In this occasion CHI_{MeCN} were determined by UHPLC, because besides the lower mobile phase consumption it provides better resolutions in shorter analysis times (< 4min). It must be pointed out that CHI_{MeCN} values considered in the training set correlations were measured using a HPLC Luna C18(2) column, whereas CHI_{MeCN} values of the test set in the present work have been determined employing a UHPLC Acquity BEH C18 column. Although both columns have the same octadecilsilane stationary phase, which is the main responsible for the chromatographic behavior of the analytes, the support technology might affect the retention in some extent. The measured CHI_{MeCN} values, together with the Abraham (ACD/Labs) and CODESSA molecular descriptors are shown in Table 4. The log $P_{o/w}$ values were directly calculated from CHI_{MeCN} measurements using multilinear correlations involving solute descriptors according to Eqs. (4), (5), (8), and (9). Figure 1 shows the correlations between literature and obtained log $P_{o/w}$ values, and the built linear regressions present, in all cases, slopes very close to unity and slightly negative intercepts not significantly different to zero at 95% confidence level. Among the solute hydrogen-bond acidity descriptors, calculated *A* (Eq. (4)) and HDCA-1/TMSA (Eq. (9)), conduct to comparable results with similar log $P_{o/w}$ predictive capacity from CHI_{MeCN} measurements, but leading the CODESSA descriptor to a slightly better correlation in terms of intercept and mean of residuals (MR) closer to zero, and lower root mean square error (RMSE). About 75% of the studied compounds are in log $P_{o/w}$ ± 0.5 range according to the

literature values, and major deviations were found for some bases (2,4-lutidine, lidocaine, and papaverine, labeled in Figure 1 as 5, 12, and 14, respectively). In the case of ketoconazole (labeled as 11), the CODESSA descriptor allows a better prediction of log $P_{o/w}$ lipophilicity, probably due to more precisely calculation of the hydrogen-bond acidity of the molecule. Finally, the addition of E (Eq. (5)) or HOMO-LUMO (Eq. (8)) descriptors slightly improve the correlations between reference and estimated log $P_{o/w}$ values. In summary, both ACD/Labs and CODESSA descriptors lead to equations which exhibit comparable predictive capacity, and any of them can be recommended for the high-throughput determination of log $P_{o/w}$ values from CHI_{MeCN} measurements. However, the calculation of Abraham descriptors is simpler and less time consuming, since they do not require the 3D optimization step of the molecular structure before computing the descriptor value and, then, this approach is recommended for everyday work

4. Conclusions

Lipophilicity log $P_{o/w}$ values can be accurately determined from fast gradient chromatographic measurements (CHI_{MeCN}) and the Abraham's hydrogen-bond acidity (A) and excess molar refractivity (E) descriptors. Alternatively, the CODESSA descriptors accounting for hydrogen-bond acidity HDCA-1/TMSA and polarizability HOMO-LUMO also improve the correlations between log $P_{o/w}$ and CHI_{MeCN} in a similar extent. Since CHI_{MeCN} measurements can be performed by UHPLC within 4 minutes and Abraham descriptors can be rapidly computed from molecular structures using the ACD/Labs software, the proposed methodology seems to be very convenient for high-throughput lipophilicity determination in the frame of drug discovery process.

Acknowledgments

This work is supported by the Ministry of Economy and Competitiveness of Spain (project CTQ2014-56253-P). The authors also acknowledge Almirall, S.A. (Barcelona, Spain) for providing the compounds of pharmaceutical interest included in the present work.

Conflict of interest statement

The authors declare no conflict of interest.

Figure caption

Figure 1. Correlations between literature and log $P_{o/w}$ determined in the present work from CHI_{MeCN} measurements and molecular descriptors computed from ACD/Labs (Eqs. (4) and (5)) and CODESSA (Eqs. (8) and (9)) software. Slope, intercept, mean of residuals (MR) and root mean square error (RMSE) of the correlations are also shown, with standard deviations in parentheses. Solid line of unitary slope indicates the total correspondence between pairs of values, dashed and dotted lines represents deviations of ±0.5 and ±1.0 log $P_{o/w}$ units, respectively.

References

[1] OECD, Guideline for the testing of chemicals. Method 107. Partition coefficient (n-octanol/water): shake flask method, 1995.

[2] EPA, Product Properties Test Guidelines. OPPTS 830.7550 Partition coefficient (*n*-octanol/water), shake flask method, United States, 1996.

[3] K. Valkó, Separation methods in drug synthesis and purification, Elsevier, Amsterdam, 2000.

[4] A. Andrés, M. Rosés, C. Ràfols, E. Bosch, S. Espinosa, V. Segarra, J.M. Huerta, Setup and validation of shake-flask procedures for the determination of partition coefficients (log *D*) from low drug amounts, Eur. J. Pharm. Sci., 76 (2015) 181-191.

[5] K. Valkó, C. Bevan, D. Reynolds, Chromatographic Hydrophobicity Index by Fast-Gradient RP-HPLC: A High-Throughput Alternative to log *P*/log *D*, Anal. Chem., 69 (1997) 2022-2029.

[6] M. Rosés, E. Bosch, C. Ràfols, E. Fuguet, in: Advances in Chromatography, CRC Press, Boca Raton, FL, U.S., 2012, pp. 377-414.

[7] K. Valkó, C.M. Du, C. Bevan, D.P. Reynolds, M.H. Abraham, Rapid method for the estimation of octanol/water partition coefficient (log P_{oct}) from gradient RP-HPLC retention and a hydrogen bond acidity term (Σa_2^H), Curr. Med. Chem., 8 (2001) 1137-1146.

[8] R. Kaliszan, P. Haber, T. Baczek, D. Siluk, K. Valko, Lipophilicity and pK_a estimates from gradient high-performance liquid chromatography, J. Chromatogr. A, 965 (2002) 117-127.

[9] M.H. Abraham, H.S. Chadha, Applications of a Solvation Equation to Drug Transport Properties, in: V. Pliska, B. Testa, H. van der Waterbeemd (Eds.) Lipophilicity in Drug Action and Toxicology, Wiley-VCH Verlag GmbH, Weinheim, Germany, 1996, pp. 311-337.

[10] M.H. Abraham, H.S. Chadha, G.S. Whiting, R.C. Mitchell, Hydrogen bonding. 32. An analysis of water-octanol and water-alkane partitioning and the Δlog *P* parameter of Seiler, J. Pharm. Sci., 83 (1994) 1085-1100.

[11] J.A. Platts, D. Butina, M.H. Abraham, A. Hersey, Estimation of Molecular Linear Free Energy Relation Descriptors Using a Group Contribution Approach, J. Chem. Inf. Comput. Sci., 39 (1999) 835-845.

[12] J.M. Pallicer, R. Pascual, A. Port, M. Rosés, C. Ràfols, E. Bosch, The contribution of the hydrogen bond acidity on the lipophilicity of drugs estimated from chromatographic measurements, Eur. J. Pharm. Sci., 48 (2013) 484-493.

[13] J.M. Cabot, X. Subirats, E. Fuguet, M. Rosés, High throughput determination log $P_{o/w}$/pK_a/log $D_{o/w}$ of drugs by combination of UHPLC and CE methods, ADMET & DMPK, 2 (2014) 98-106.

[14] Bio-Loom, BioByte Corp. (Claremont, CA, USA), http://www.biobyte.com, Version 1.5.

[15] M.H. Abraham, A. Ibrahim, A.M. Zissimos, Determination of sets of solute descriptors from chromatographic measurements, J. Chromatogr. A, 1037 (2004) 29-47.

[16] E. Lázaro, C. Ràfols, M.H. Abraham, M. Rosés, Chromatographic Estimation of Drug Disposition Properties by Means of Immobilized Artificial Membranes (IAM) and C18 Columns, J. Med. Chem., 49 (2006) 4861-4870.

[17] The CODESSA PRO project by Alan R. Katritzky, Mati Karelson and Ruslan Petrukhin, University of Florida (USA), http://www.codessa-pro.com/.

[18] R. Kaliszan, Quantitative Structure-Chromatographic Retention Relationships. Chapter 11, New York, 1987.

[19] R. Bosque, J. Sales, E. Bosch, M. Rosés, M.C. García-Álvarez-Coque, J.R. Torres-Lapasió, A QSPR Study of the *p* Solute Polarity Parameter to Estimate Retention in HPLC, J. Chem. Inf. Comput. Sci., 43 (2003) 1240-1247.

[20] J.M. Pallicer, S. Pous-Torres, J. Sales, M. Rosés, C. Ràfols, E. Bosch, Determination of the hydrophobicity of organic compounds measured as log $P_{o/w}$ through a new chromatographic method, J. Chromatogr. A, 1217 (2010) 3026-3037.

TABLES

Table 1. Normalized coefficients for log $P_{o/w}$ and CHI_{MeCN} solvation equation obtained from experimental and calculated molecular descriptors, and *d* distance accounting for system similarity.

	e_u	s_u	a_u	b_u	v_u	*d*	*N*	*R*	*SE*
log $P_{o/w}$ [a]	0.11	-0.20	0.01	-0.65	0.72	0.000	613	0.997	0.12
log $P_{o/w}$ [b]	0.20	-0.24	0.13	-0.69	0.64	0.181	125	0.937	0.50
CHI_{MeCN} [b]	0.12	-0.24	-0.17	-0.67	0.67	0.193	125	0.929	10.3

[a]Calculated from experimental descriptors (ref. [10]).
[b]This work, calculated from computed descriptors.

Table 2. Determination coefficients (R^2), standard errors (*SE*) and residuals sum of squares (SS_{res}) of multiple linear regressions between log $P_{o/w}$, CHI_{MeCN} and the indicated molecular descriptor for the training set (*N*=125). The improvement (%) in relation to simple linear regression between log $P_{o/w}$ and CHI_{MeCN} is also shown.

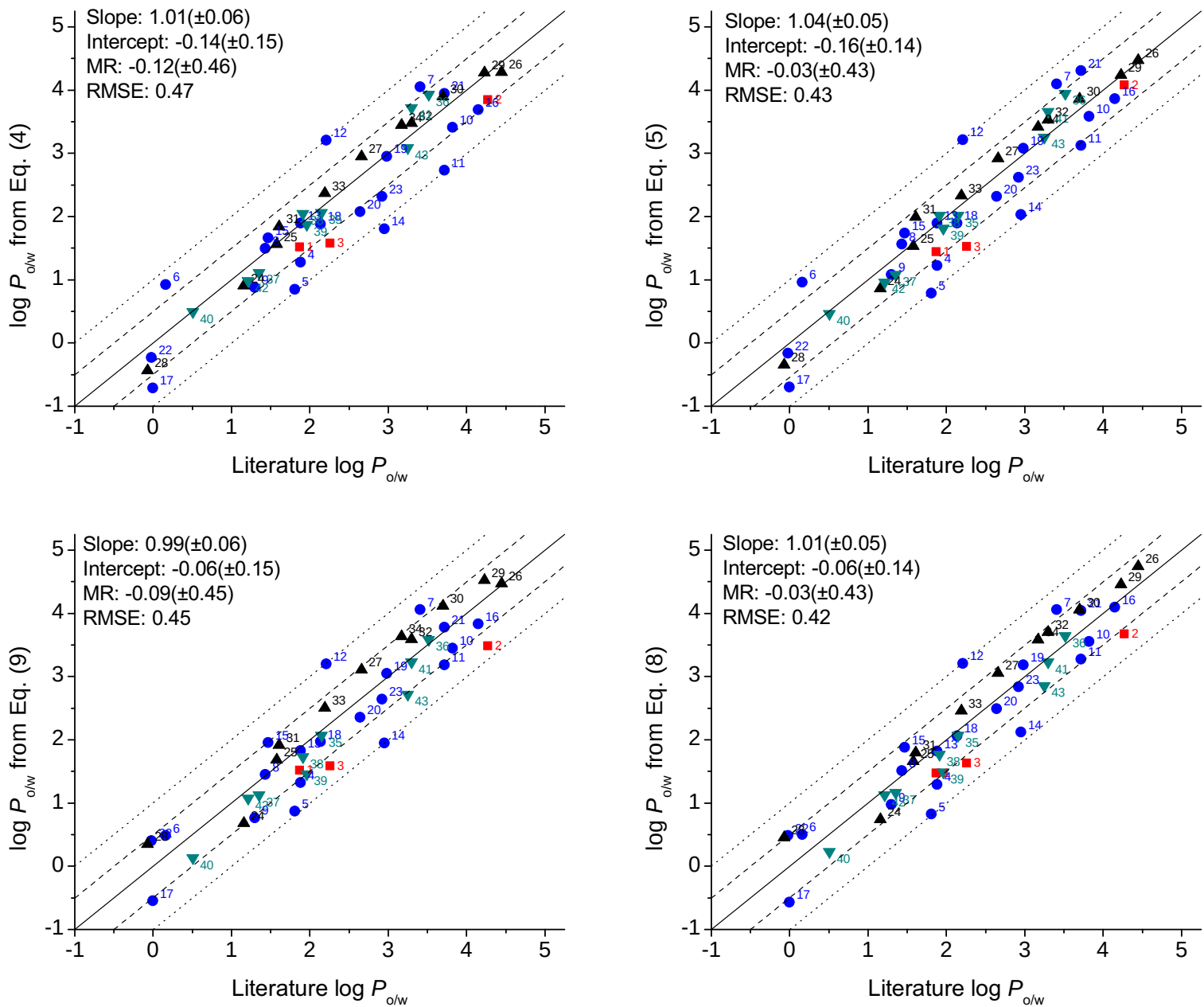
	HDCA-1 /TMSA	HDCA-2 /TMSA	HACA-2 /TMSA	DPSA-3	HOMO-LUMO
R^2	0.913 (5.9%)	0.911 (5.6%)	0.892 (3.4%)	0.881 (2.1%)	0.877 (1.6%)
SE	0.40 (20.2%)	0.40 (19.2%)	0.45 (10.8%)	0.47 (6.3%)	0.48 (4.8%)
SS_{res}	19.6 (36.8%)	20.1 (35.2%)	24.5 (21.1%)	27.0 (12.9%)	27.9 (10.2%)

Table 3. Determination coefficients (r^2) of simple linear regressions between CHI_{MeCN} values and molecular descriptors of the training set (*N*=125) obtained from CODESSA software.

CODESSA	CHI_{MeCN}	HDCA-1 /TMSA	HDCA-2 /TMSA	HACA-2 /TMSA	DPSA-3	HOMO-LUMO
CHI_{MeCN}	1.000	0.593	0.563	0.456	0.083	0.004
HDCA-1/TMSA	0.593	1.000	0.986	0.757	0.240	0.000
HDCA-2/TMSA	0.563	0.986	1.000	0.799	0.289	0.000
HACA-2/TMSA	0.456	0.757	0.799	1.000	0.273	0.002
DPSA-3	0.083	0.240	0.289	0.273	1.000	0.000
HOMO-LUMO	0.004	0.000	0.000	0.002	0.000	1.000

Table 4. Experimental CHI_{MeCN} and calculated molecular descriptors for the test set compounds.

Ref.		Compound	A	E	HDCA-1 ($x10^{-3}$)	HOMO-LUMO	CHI_{MeCN}
1	Acidic	Benzoic acid	0.57	0.75	8.22	9.62	48.0
2		Indomethacin	0.57	2.44	2.85	7.90	90.7
3		Salicylic acid	0.70	0.91	9.92	8.91	46.0
4	Basic	2,4,6-Trimethylpyridine	0.00	0.67	1.25	9.61	57.2
5		2,4-Lutidine	0.00	0.65	1.43	9.79	49.4
6		Atenolol	0.78	1.48	7.38	9.26	32.1
7		Bupivacaine	0.26	1.32	2.01	9.14	101.8
8		Clonidine	0.42	1.48	5.73	8.89	51.3
9		Colchicine	0.26	2.17	3.43	7.99	43.9
10		Haloperidol	0.31	2.00	3.44	8.42	88.9
11		Ketoconazole	0.00	3.14	3.80	8.59	83.9
12		Lidocaine	0.26	1.10	2.52	9.13	86.4
13		Metoprolol	0.29	1.10	3.54	9.35	61.5
14		Papaverine	0.00	2.19	1.70	8.19	66.9
15		Phenobarbital	0.52	1.56	10.13	9.73	51.8
16		Phenothiazine	0.13	1.95	1.77	7.34	98.4
17		Pilocarpine	0.00	1.02	4.04	9.71	20.9
18		Procaine	0.23	1.11	4.10	8.67	62.9
19		Propranolol	0.29	1.76	4.17	8.26	80.9
20		Quinine	0.23	2.41	5.78	8.31	66.4
21		Reserpine	0.31	3.10	1.12	7.38	98.7
22		Theophylline	0.35	1.46	12.53	8.70	21.4
23		Trazodone	0.00	2.64	2.94	7.95	76.3
24	Neutral	Acetanilide	0.41	0.89	4.42	9.01	40.7
25		Acetophenone	0.00	0.79	1.72	9.57	62.5
26		Anthracene	0.00	1.99	0.00	7.28	112.3
27		Butyrophenone	0.00	0.79	0.84	9.58	87.9
28		Caffeine	0.00	1.48	9.55	8.61	25.9
29		Heptanophenone	0.00	0.78	0.61	9.58	112.1
30		Hexanophenone	0.00	0.78	0.66	9.58	105.2
31		Hydrocortisone	0.73	2.04	10.67	10.01	50.3
32		Naphthalene	0.00	1.27	0.00	8.45	97.5
33		Propiophenone	0.00	0.79	1.16	9.58	77.2
34		Valerophenone	0.00	0.79	0.73	9.58	97.0
35	Phenolic	2-Chlorophenol	0.33	0.85	4.66	9.29	63.6
36		3,5-Dinitrophenol	0.83	1.32	6.37	8.68	86.0
37		4-Hydroxybenzaldehyde	0.66	1.04	9.92	9.04	38.4
38		4-Nitrophenol	0.67	1.05	6.07	9.01	55.3
39		Methylparaben	0.66	0.87	5.24	9.14	52.3
40		Paracetamol	0.91	1.12	10.13	8.66	21.2
41		Thymol	0.50	0.84	0.85	9.28	90.0
42		Vanillin	0.44	1.02	7.70	8.92	41.4
43		Warfarin	0.31	1.98	0.01	8.40	82.9


Slope: 1.01(±0.06)
Intercept: -0.14(±0.15)
MR: -0.12(±0.46)
RMSE: 0.47
log $P_{o/w}$ from Eq. (4)
Literature log $P_{o/w}$
Slope: 1.04(±0.05)
Intercept: -0.16(±0.14)
MR: -0.03(±0.43)
RMSE: 0.43
log $P_{o/w}$ from Eq. (5)
Literature log $P_{o/w}$
Slope: 0.99(±0.06)
Intercept: -0.06(±0.15)
MR: -0.09(±0.45)
RMSE: 0.45
log $P_{o/w}$ from Eq. (9)
Literature log $P_{o/w}$
Slope: 1.01(±0.05)
Intercept: -0.06(±0.14)
MR: -0.03(±0.43)
RMSE: 0.42
log $P_{o/w}$ from Eq. (8)
Literature log $P_{o/w}$