Journal of Pharmaceutical and Biomedical Analysis 115 (2015) 183–191

Contents lists available at ScienceDirect

Journal of Pharmaceutical and Biomedical Analysis

journal homepage: www.elsevier.com/locate/jpba

Towards better understanding of lipophilicity: Assessment of *in silico* and chromatographic log*P* measures for pharmaceutically important compounds by nonparametric rankings

Filip Andrić [a], Károly Héberger [b,*]

[a] Faculty of Chemistry, University of Belgrade, Studentski trg 12-16, 11000 Belgrade, Serbia
[b] Research Centre for Natural Sciences, Hungarian Academy of Sciences, H-1117 Budapest XI., Magyar Tudósok krt 2, Hungary

ARTICLE INFO

Article history:
Received 10 April 2015
Received in revised form 6 July 2015
Accepted 7 July 2015
Available online 17 July 2015

Keywords:
Lipophilicity
Natural toxins
Antifungal drugs
Thin-layer chromatography
Sum of ranking differences
Generalized pairwise-correlation method
Multivariate data analysis

ABSTRACT

Lipophilicity is one of the most frequently used physicochemical properties that affects compound solubility, determines its passive transport through biological membranes, influences biodistribution, metabolism and pharmacokinetics. We compared, ranked and grouped chromatographic lipophilicity indices and computationally estimated log*P*–s by sensitive and robust non-parametric approaches: sum of ranking differences (SRD) and generalized pairwise correlation method (GPCM). Chromatographic indices of fourteen neurotoxins and twenty one 1,2,4-triazole compounds have been derived from typical reversed-phase thin-layer chromatography and micellar chromatography. They were compared with *in silico* estimated log*P*–s. Under typical reversed-phase conditions, octadecyl-, octyl-, and cyanopropyl-modified silica have clear advantage over ethyl-, aminopropyl-, and diol-modified beds, *i.e.*, the preferable choice of the stationary phase follows this order: octadecyl > octyl > cyanopropyl > ethyl > octadecyl wettable > aminopropyl > diol. Many of these indices outperform the majority of computationally estimated log*P*–s. Clear distinction can be made based on cross-validation and statistical tests. Oppositely, micellar chromatography may not be successfully used for the lipophilicity assessment, since retention parameters obtained from the typical reversed-phase conditions outperform the parameters obtained by micellar chromatography.

Both ranking approaches, SRD and GPCM, although based on different background, provide highly similar variable ordering and grouping leading to the same, above mentioned conclusions. However, GPCM results in more degeneracy, *i.e.*, in some cases it cannot distinguish the lipophilicity parameters whereas SRD and its cross-validated version can. On the other hand GPCM produces a more characteristic grouping. Both methods can be successfully used for selection of the most and least appropriate lipophilicity measures.

1. Introduction

Lipophilicity is one of the major physical–chemical properties used in pharmaceutical and environmental sciences. Its role is of utmost importance in drug discovery [1] and modeling of the fate of a compound in the environment. It strongly affects compound solubility, and determines passive transport through biological membranes such as gastrointestinal tract or blood to brain barrier [2]. It also influences biodistribution, metabolism and pharmacokinetics [3]. It significantly impacts the protein binding, modeling of drug-receptor interactions, compound-related toxicity or adverse effects [4]. Among other parameters, such as solubility, stability, acid-base character, *etc.*, lipophilicity is determined at the early stages of drug development, and included in identification of starting points, viable chemical leads, and developing candidates [5].

Bioavailability and bioconcentration in the food chain through sorption from water, and soil or sediment, is also affected by lipophilicity [6], which makes it an important factor in risk assessment and management of hazardous materials.

The octanol–water partition coefficient ($\log P_{O/W,}$ or more often written as log*P*) is generally accepted as the golden standard for lipophilicity measurement (assessment) [6]. The experimental

Abbreviations: C18, octadecyl; C18W, octadecyl wettable; C2, ethyl; C8, octyl; CE-PW, conditional exact Fisher test & probability weighted ranking; CMC, critical micellar concentration; CN, cyanopropyl; CRRN, validation of the SRD procedure: comparison of ranks by random numbers; CV, cross-validation; GPCM, generalized pair correlation method; HCA, hierarchical cluster analysis; HPLC, high performance liquid chromatography; IAM, immobilized artificial membrane chromatography; MLC, micellar liquid chromatography; NH_2, aminopropyl; OPLC, overpressured layer chromatography; PC, principal component; PCA, principal component analysis; RP, reversed-phase; RP-TLC, reversed-phase thin-layer chromatography; SRD, sum of ranking (absolute) differences; TLC, thin-layer chromatography.

* Corresponding author.
E-mail address: heberger.karoly@ttk.mta.hu (K. Héberger).

http://dx.doi.org/10.1016/j.jpba.2015.07.006
0731-7085/

measurement of log*P* is described in the guidelines of the Organization for Economic Cooperation and Development (OECD): Test No. 107, Shake flask method [7], and Test No. 123, slow stirring method [8].

However, both methods are time and reagent consuming and they are unsuitable for analysis of impure or degraded compounds or compounds of extremely low or high log*P* values (log*P* < −3 or log*P* > 4). Therefore, in the past decade they have been often replaced by more elegant, far simpler, and more versatile chromatographic methods that give similar and coherent results in the close log*P* range, (the OECD Test No. 117 [9]), which are able to efficiently analyze contaminated compounds and products of their degradation, the latter being of the particular interest in drug analysis.

Among all chromatographic techniques, thin-layer chromatography (TLC) takes specific place because of its simplicity, low costs, and reduced consumptions of solvents and reagents (green analytical technique).

The most frequently employed TLC modality for lipophilicity assessment of drugs and compounds of pharmaceutical importance is a typical reversed-phase one [10–13]. Non-polar stationary phases such as various hydrocarbon modified silica gels (octadecyl, octyl, ethyl, and phenyl, commonly denoted as: C18, C8, C2 and Ph, respectively), [14] or amphiphilic sorbents such as cyanopropyl-, aminopropyl- or diol-modified silica [14,15] in combination with polar mobile phase (binary mixtures of water and miscible organic solvents) are used.

Different stationary phases that imitate biosystems, such as immobilized artificial membranes (IAMs) [16,17], immobilized proteins [18], and cholesterol [19], or other techniques like micellar liquid chromatography (MLC), have also been proposed for studying the lipophilicity of different compounds [20–24].

Micellar chromatography, as a variety of reversed-phase modality, is of particular interest. Addition of surfactants to the mobile phase in the concentration above the critical micelle concentration (CMC) leads to the constellation of specific molecular interactions that form an intricate retention mechanism: solute association with the polar head of the surfactant, solute penetration into the micelle core, adsorption of surfactant monomers on the alkyl-bounded stationary phases, and solute interactions with adsorbed surfactant and alkyl chains [25].

Regardless to the modality used, all lipophilicity indices are derived either directly from retention data or extrapolated from linear or bilinear relationships between retention parameters and mobile phase composition. Most frequently applied lipophilicity indices derived from TLC experiments are: mean R_M, $R_M{}^0$ (R_M values extrapolated to the zero content of organic modifier), b (slope in the linear dependence of R_M against the volume fraction of organic modifier), C_0 (volume fraction of organic modifier in a mobile phase that provides equal distribution of analyte between mobile and stationary phase; $R_M = 0$), and $PC1/R_M$ (score values of the first principal component defined as the linear combination of R_M values) [26]. In the case of micellar chromatography the retention factor extrapolated to the zero micellar concentration of surfactant, *i.e.*, the CMC in the mobile phase, logk_m, is used [20,25]. All chromatographically derived lipophilicity measures used in the scope of this work are summarized in the Table S1 (Supplementary material), accompanied with simple explanations.

Alongside the chromatographic methods, computational approaches for lipophilicity estimation have been extensively utilized. *In silico* predicted log*P*–s, either based on fragmentation approaches, or property-based are often compared with chromatographic lipophilicity indices [27,28]. However many of them exhibit significant differences in predicted log*P* values [29,30], which might question the reliability of these methods.

The computationally estimated log*P* scales used in the scope of the present work are summarized in the Table S2 (Supplementary material), alongside with the short description provided.

Considering the importance of properly selected lipophilicity measure, the primary goal of this study was to compare, rank and group lipophilicity measures using sensitive non-parametric approaches. Therefore, comparison of TLC derived lipophilicity indices were in the main focus of this research. Such indices were obtained under typical reversed-phase conditions and under micellar chromatography, using stationary phases of different polarities. Further comparison of *in silico* estimation approaches, and chromatographic indices based on direct retention measures *vs.* the extrapolated ones was the subject of particular importance, as well. Many lipophilicity measures are derived from chromatographic experiments. For some compounds the chromatographic approaches are the only solutions for experimental lipophilicity determination (*e.g.*, charged, polar molecules resolved under hydrophilic interaction liquid chromatography). To present no systematic, sensitive, and reliable approach for comparison and selection of the most suitable lipophilicity scales is enforced or widely accepted. The lack of systematic solution is especially harmful when it comes to testing of novel or emerging methods for lipophilicity estimation, such as micellar chromatography. Present work is a natural continuation of our previous research regarding the use of the sum of ranking differences (SRD) and generalized pairwise correlation method (GPCM) for selection of the best lipophilicity measures [31].

2. Materials and methods (calculations)

2.1. Lipophilicity data

Lipophilicity data have been taken from the literature [32,33] and are presented in Sections 3.1 and 3.2. We have selected cases providing they cover typical reversed-phase chromatography that combines stationary phases of different polarity [32]. Also, micellar chromatographic indices are studied alongside with typical RP-TLC derived descriptors. [33]. We paid attention to secure significant diversity of studied compounds and target their pharmacological importance (natural toxins and 1,2,4-triazoles as potent fungicides). A special care was taken in order to ensure the wide range of chromatographically derived lipophilicity indices and log*P* computational approaches.

2.2. Data pretreatment and multivariate exploratory statistical analysis

In order to compare different lipophilicity measures all variables were rescaled. Three data transformation approaches have been tested: (a) standardization to unit standard deviation (also called autoscaling), (b) interval scaling between the lowest and the highest computationally estimated log*P* value and (c) rank transformation. Standardized data were further used for the exploratory data analysis employing HCA and PCA, while the sum of ranking differences (SRD) and generalized pairwise correlation method (GPCM) have been performed on all three sets of transformed data separately. In the case of HCA the Euclidian distance was selected as measure of dissimilarity among the variables while the Ward's method was used to define the distances among clusters. PCA has been performed using PCA and multivariate/Batch SPC module as a part of Statistica v.10 (Statsoft Inc., Tulsa, Oklahoma, USA). The number of principal components has been determined using visual evaluation of screen plot.