Consensus-based comparison of chromatographic and computationally estimated lipophilicity of benzothiepino[3,2-*c*]pyridine derivatives as potential antifungal drugs

Sandra Šegan[1], Nina Božinović[2], Igor Opsenica[2], Filip Andrić[2,*]

[1]Institute of Chemistry, Technology and Metallurgy, University of Belgrade, Njegoševa 12, 11000 Belgrade, Serbia

[2]Faculty of Chemistry, University of Belgrade, Studentski trg 12-16, 11000 Belgrade, Serbia

Running head title: *Consensus based comparison of lipophilicity indices*

*Corresponding author: Dr Filip Andrić, assistant professor, Department for Analytical Chemistry, Faculty of Chemistry, University of Belgrade, Studentski trg 12-16, 11000 Belgrade, Serbia,

Phone: +381 11 3336 755

E-mail: andric@chem.bg.ac.rs

Received: 12 25, 2016; Revised: 03 01, 2017; Accepted: 03 11, 2017

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/jssc.201601442.

Abbreviations: **AMA**, Arithmetic mean (average); **CRRN**, Comparison of Ranks by Random Numbers; **CV**, Cross-validation; **HCA**, Hierarchical Cluster Analysis; **MELC**, Microemulsion Liquid Chromatography; **PC**, Principal Component; **PCA**, Principal Component Analysis; **RP18W**, Octadecyl wettable; **Std**, Standardization; **SRD**, Sum of Ranking (absolute) Differences; **Rng**, Range scaling; **Rnk**, Rank transformation

Key words: Benzothiepino[3,2-*c*]pyridines; Lipophilicity; Multicriteria ranking; Multivariate data analysis; Sum of ranking differences

Abstract

Lipophilicity is one of the essential properties influencing drug absorption, excretion and metabolism. It is used for screening of viable drug candidates. Chromatographic behavior of thiepino[3,2-*c*:6,7-*c*']dipyridine and 16 benzothiepino[3,2-*c*]pyridine derivatives as potential antifungal drugs was studied using thin-layer chromatography under typical reversed-phase conditions and two microemulsion chromatographic systems. Seventeen chromatographic and nine in silico lipophilicity measures were estimated. They were compared by classical multivariate approaches: principal component analysis, hierarchical cluster analysis, and ranked and grouped by the non-parametric method—Sum of ranking differences. Two computational and two chromatographic descriptors from the typical reversed-phase conditions using acetone/water mixtures emerged as the best candidates for lipophilicity estimation. The principal component scores related to typical reversed-phase conditions using

dioxane/water were ranked as statistically insignificant (the worst). Microemulsion systems were positioned in between, performing worse than in silico estimates.

Thiepine derivatives were ranked and grouped by sum of ranking differences, fusing multiple lipophilicity measures. In multicriteria maximization ranking, the compound substituted by phenyl group at position 8 was selected as the most lipophilic one. It is also the most active against *Candida albicans*. The ranking confirmed that introduction of phenyl core is essential for increasing the lipophilicity of the studied compounds.

1. Introduction

Lipophilicity is a physicochemical parameter of essential importance for biological activity of compounds. It determines drug absorption, distribution, metabolism, and elimination. As a vital factor it is usually determined in the early stages of the development of viable candidates along with solubility, stability, and acid–base characteristics of target compounds [1,2]. Lipophilicity is commonly expressed as the logarithm of the partition coefficient of a neutral form of a molecule in *n*-octanol/water system (commonly denoted as log*P*, log*P*o/w, or logK_{OW}). For an ionized compound the logarithm of the distribution coefficient at particular pH (logD^{pH}) is usually used.

$$\log P = \log C_{oct}/C_{aq} \quad (1)$$

where C_{oct} and C_{aq} are the concentrations of a solute in organic and aqueous phase, respectively.

Shake-flask method is still considered as the gold standard for experimental determination of lipophilicity, but faces problems like limitation to only pure compounds, formation of stable emulsions, limited measurable range of P (from 10^{-2} to 10^{4}), and significant time and reagent consumption [1].

This is the reason why the RP-HPLC and reversed-phase high-performance thin-layer chromatography (RP-HPTLC), as indirect methods, are largely used for the measurement of log$P$. The chromatographic methods have extended measurable log$P$ range to 10^{6} [1], and are suitable for testing mixtures and impure substances.

Several descriptors resulting from the RP-TLC experiments are well established lipophilicity estimators, such as the intercept (R_M^0) and the slope (b) of the linear dependence of the isocratic retention mobility parameter (R_M) on the volume fraction of the organic component of the mobile phase (φ), Eq. (2) [3]. The R_M is calculated according to the equation (3) [4].

$$R_M = R_M^0 + b\varphi \quad (2)$$

$$R_M = \log (1/R_F - 1) \quad (3)$$

where R_M^0 is the R_M value extrapolated to the pure water. The slope, b, is related to the specific hydrophobic surface area of compounds. However, it is used with less success [5].

The ratio of the intercept R_M^0 and the slope b, denoted as C_0, was introduced by Bieganowska et al. (4) [6].

$$C_0 = - R_M^0/b \quad (4)$$

It is interpreted as the molecular hydrophobicity per unit of the specific hydrophobic surface.

Apart from extrapolated indices (R_M^0, b, and C_0), the ones based on primary retention data are frequently used such as: the first principal component scores (PC1) derived from the principal component analysis (PCA) of multivariate retention data [7–9] and arithmetic means of R_M values, denoted as mR_M [10–12].

In addition to the typical reversed-phase chromatography, methods that tend to mimic biopartitioning such as micellar/microemulsion electrokinetic chromatography (MEKC/MEEKC) and micellar/microemulsion liquid chromatography (MLC/MELC) are becoming popular in estimating pharmacodynamic parameters of pharmaceuticals, including lipophilicity [13–15]. The micelles or micro emulsions are generally comprised of surfactant, water, cosurfactant, and a small amount of organic solvent. The main disadvantage of MELC is long retention times for very lipophilic compounds. Moreover, the preparation of the microemulsion phase generally takes a long time to achieve homogeneity [16,17]. Recently, in a comparative study on several approaches to estimation of lipophilicity Komsta et al. pointed out the importance and universality of the TLC estimations [18].

Alongside with experimental methods, log*P* is estimated by various in silico approaches. Substructure-based methods cut molecular structures into smaller fragments, or to the level of single atoms, that contribute additively to the solute lipophilicity, *e.g*. AlogP, XlogP2, XlogP3, milogP, KOWWIN, ACD/logP, AClogP, and ABLogP. Property-based approaches use descriptors of the molecule as a whole, *e.g.* MLOGP or AlogPs. All of them have significant advantages over experimental techniques simply because they do not require expensive instrumentation, reagents and laborious experimental work. However, difference in log*P* values obtained using different computations can vary 2–3 orders of magnitude for the same molecule, which might question their reliability on a large scale [19–21].

Lipophilicity, as a property determining biological activity of a substance, stands at the basis of quantitative structure-retention/properties/activity relationships (QSRR/QSPR/QSAR) which is valuable assistance in the prognosis of the behavior of new molecules, even before they are actually synthesized.

Thiepines are an important class of biologically active heterocyclic molecules. The tricyclic moiety of dibenzo[*b*,*f*]thiepines is a pharmacophore in a number of drugs, *e.g*., zotepine, an antipsychotic drug that has been used for the treatment of schizophrenia. The recently published results have shown that dibenzo[*b*,*f*]thiepines and benzothiepino[3,2-*c*]pyridines exhibit anticancer [22] and antifungal activity [23,24].

The newly synthesized thiepine derivatives exhibit weak antibacterial activity, but demonstrate pronounced antifungal activity. The most potent derivative against *Candida albicans* possesses phenyl ring in the position 8 of the thiepine core. The position of the phenyl ring was essential for the antifungal activity, while, the electronic effects of the substituents did not have significant influence [25].

The aim of the present work is to evaluate the lipophilicities of a series of thiepine derivatives using typical RP and microemulsion TLC as well as computational methods, and to compare, group and select the best, and the worst approaches.

Several chemometric techniques can extract subsets of similar objects (compounds) or features (lipophilicity measures). PCA, hierarchical cluster analysis (HCA) and simple Pearson's correlation matrix are most frequently applied.

In our previous works we have pointed out that PCA and HCA do not provide any information on statistical significance of such grouping, and the Pearson correlation is applicable if the data are normally distributed, which is frequently not the case [7,20,21]. Also, comparison of Pearson's correlation coefficients tends to be statistically insensitive to small differences among tested methods/objects. Such problems can be overcome by robust non-parametric ranking methods, such as the Sum of ranking differences (SRD) [7,20,21].

Finally, we would like to rank and group thiepine derivatives by fusing information from different chromatographic and in silico lipophilicity estimations. That way information related to pharmacodynamic and pharmacokinetic aspects of their action will be obtained. SRD should provide an insight into small differences in a series of structurally very similar compounds. Therefore, this work is a continuation of our research [7,20,21].

2. Materials and methods

2.1. *Reagents and chemicals*

The synthesis and characterization of the studied compounds was previously reported [23–25]. The structures of the thiepine derivatives are summarized in **Table 1**. Methanol, acetone, dioxane, *n*-heptane, 1-butanol, and SDS were of analytical-grade purity and purchased from Merck (Darmstadt, Germany). Water was purified using Millipore Simplicity 185 S.A., 67120, water purification system (Molshem, France).

2.2. *Chromatographic experiments*

TLC was performed in a horizontal manner using 10 × 10 cm, RP-18W F254s aluminum sheet plates (Art. 5559, Merck, Darmstadt, Germany) and the horizontal HPTLC developing chamber (Camag, Muttenz, Switzerland).

The substances were dissolved in methanol and the plates were spotted with 0.5 mL aliquots of freshly prepared solutions ($C \approx 0.5$ mg/mL). Before isocratic development chromatographic chamber was equilibrated for 15 min with mobile phase vapors. Development distance was 4.5 cm. Individual spots were detected under UV light (254 nm). All experiments were performed at ambient temperature (22 ± 2°C).

In the case of typical RP-TLC the binary mixtures of methanol/water, acetone/water, and dioxane/water were used as mobile phases. The content of organic modifiers was increased in the ranges: methanol, 82–98 Vol% (increment 4%); acetone, 65–90 Vol% (increment 5%); and dioxane, 65–85 Vol% (increment 5%). In the case of MELC, two solvent systems (MELC1 and MELC2) were tested; MELC1 was prepared by mixing 2.8 mL of *n*-heptane, 19.0 mL of 1-butanol, 75 mL water and 7.7 g SDS. MELC2 was prepared in the same way as MELC1 with the difference of using 24.0 mL of 1-butanol and 70 mL of water. After mixing, the solutions were sonicated for 30 min to produce transparent microemulsions.

2.3. *In silico lipophilicity estimation*

The computational log*P* values (AlogPs, AClogP, AlogP, MLOGP, XlogP2, XlogP3), and water solubility (AlogpS) were calculated from the molecular structures by the Virtual Computational Chemistry Laboratory (VCCLAB, http://www.vcclab.org/) [26,27]. The miLogP was calculated using Molinspiration property engine v2014.11 (http://www.molinspiration.com/); KOWWIN log*P* values were estimated using KOWWIN v. 1.68 (EPI Suite package v.4.1, U.S. EPA).

2.4. *Data pretreatment and statistical analysis*

Before comparison chromatographic data (17 indices) and computational log*P* values (8 indices) were put on the same scale using three data pretreatment methods: (*i*) standardization (Std) which assumes mean centering and scaling to the unit SD, (*ii*) range scaling (Rng) between the highest and the lowest in silico log*P* value, and (*iii*) rank transformation (Rnk) using arithmetic mean for ties.

All statistical computations, data pretreatments, PCA and HCA, were performed using Statistica v. 10 (Statsoft, Tulsa, Oklahoma, USA), and Microsoft Excel, Microsoft office 2010.

PCA is used to reduce dimensionality and to explore a multivariate space. It calculates new latent variables (principal components) which are linear combinations of the original variables. The first principal component represents a direction in a multivariate space that accounts for the maximum variability of the projection values of objects. The remaining principal components are selected following conditions of mutual orthogonality and maximum variance of object projections. Object projections are called scores and linear combination coefficients are called loadings. The results of PCA are usually depicted using score and loading plots, from which similarities among the objects and variables are observed.

In the case of agglomerative HCA, distances among objects are calculated first. Then, the pair of objects with the shortest distance forms a cluster. Distances among objects and the new cluster (or other clusters) are recalculated using amalgamation rules such as: complete linkage, single linkage, cluster centroids, or Ward's method *etc*. Objects (clusters) that are closest to each other, form a new cluster. Procedure is repeated until all objects fall under one cluster. Result is usually presented in a form of a dendrogram. In this particular case Euclidian metric was used as a distance measure and Ward's amalgamation rule for building dendrograms. Both, PCA and HCA have been carried out on standardized data.

2.4.1. *Non-parametric ranking*

Comparison, grouping and ranking of chromatographic and *in silico* lipophilicity scales were done by the Sum of Ranking Differences (SRD) [28,29]. For this purpose freely available Microsoft Excel visual basic macros were used (http://aki.ttk.mta.hu/srd/). The SRD method requires an input data matrix composed of variables (lipophilicity indices) that should be ranked, arranged in columns, and objects (compounds) arranged in rows. Then a reference is added which can be golden standard, row maximum, minimum or arithmetic mean average (AMA). For comparing lipophilicity scales row-wise arithmetic mean was used, instead of any preferred method. Such consensus based approach is justified from at least two points: *i*) according to the maximum likelihood principle the arithmetic mean is the estimator for which the observation is the most probable and *ii*) the average has ability to cancel out systematic as

well as random errors, at least partially. After selecting the appropriate reference, objects are ranked for each variable separately including the reference. Then the ranks are subtracted from the reference (object pairwise) and the sum of absolute differences is assigned to each variable as the SRD score. SRD score is usually scaled between 0 and 100 according to the equation (4)

$$SRD\ [\%] = 100 \times SRD/SRD_{max} \quad (4)$$

where SRD_{max} is the maximum possible SRD value.

Variables are arranged in ascending order of SRD-s. The lower the SRD, the closer is the variable to the reference. Such ranking can be validated in two ways. One way uses randomization test which builds a random distribution of SRD values. If the variable falls under the 95% confidence area of the random distribution, then ranking of that particular variable can be considered statistically insignificant. The other way assumes introduction of a certain portion of variability into data. Approximately 1/7 of objects are omitted and ranking is performed on the truncated data set. Procedure is repeated seven times, on different subsets giving seven slightly different ranking results. In such a way variability is assigned to the SRD value for each variable. Then, it is possible to test whether the SRD values associated with different variables differ statistically significantly. If so, similar variables can be grouped together and separated from the rest.

3. Results and discussion

Chromatographic retention parameters of sixteen benzothiepino[3,2-*c*]pyridines and one thiepino[3,2-*c*:6,7-*c*']dipyridine are given as the R_F values (**Table S1**, Supporting information). Chromatographic lipophilicity indices and computational log*P* estimates are summarized in the **Tables S2a** and **S2b** (Supporting information).

3.1. *Exploratory data analysis*

To explore the data structure, reveal possible similarities among lipophilicity indices or thiepine derivatives and detect the presence of outliers, PCA and HCA have been carried out on the standardized data from the **Tables S2a and S2b**. Arithmetic mean average (AMA), was included as a reference point PCA resulted in two principal components describing 81.16 and 7.38% of the data variability respectively (in total 88.54%, see eigenvalues scree plot, **Figure S1**, Supporting information). Score plot (**Figure 1a**) reveals no outliers among solutes, tough compound **1** drifts away from the rest. Evidently, PC1 mostly encodes the hydrophobic (lipophilic) character of compounds. Therefore, the higher the PC1 score of a particular solute is, the more lipophilic compound should be. In that sense three groups of thiepines could be observed: I – high lipophilicity solutes (**10**, **4**–**6**, **12**–**15**), II – intermediate lipophilicity (**3**, **9**, **8**, **16**, **17**, **2**) and III – low lipophilicity compounds (**1**).

The loading plot (**Figure 1b**) reveals that R_M^0 and *b* obtained using MeOH as organic modifier, and AlogPs are the closest to the AMA. The rest of the in silico and the majority of

chromatographic measures dissipate moderately from the consensus point, following circular line. Extrapolated chromatographic properties obtained mostly from the TLC experiments involving acetone and dioxane deviate from the AMA more. Among them, C_0 calculated from the methanol/water mixtures is the most distant one, and together with PC1 component calculated from the dioxane based experiments makes the farthest pair of points. Such deviations of extrapolated measures from consensus are expected, because of uncertainties associated with the extrapolation itself. R_M values obtained from microemulsion chromatographic experiments are located together with computationally estimated log*P* values and should be considered fairly close to the AMA value.

HCA reveals complementary relationships among clustering of lipophilicity indices and grouping of compounds, as in the case of PCA. Detailed explanation is provided in the Supporting information (pages 6 and 7, **Figures S2 and S3**).

3.2. *Comparison of chromatographic and in silico lipophilicity indices by sum of ranking differences*

Although PCA and HCA provide information regarding similarities among lipophilicity descriptors or the thiepine derivatives, they do not test whether such similarity is statistically significant or not, making selection of the most suitable lipophilicity estimates difficult. This is easily overcome in the consensus based non-parametric comparison. According to the

SRD-CRRN of the standardized lipophilicity estimates (**Figure 2a**) the best lipophilicity measures, *i.e.*, the closest to the consensus point are: XlogP2 and PC1 derived from the TLC experiments using acetone as a mobile phase modifier. They are closely followed by miLogP, AlogP and acetone/water mR_M values. The PC1 component obtained from the TLC measurements using dioxane could be regarded as the worst lipophilicity measure. It falls under the Gaussian distribution curve of random SRD-s, which makes it incapable to rank thiepine derivatives better than a chance ($p = 0.05$).

MELC retention parameters are located in a distant part of the SRD graph, and perform equally well as the extrapolated descriptors from the typical RP conditions (dioxane and acetone used as organic modifiers). Although being ranked worse than any of in silico estimators, MELC descriptors are statistically significantly different from the random ranking.

The ranking could be altered by different preprocessing methods. In that case, each method should be assessed and the best one should be applied. However, range scaling (Rng) and rank transformation (Rnk) preserve the main pattern, *i.e.* the position of the few best indices, the place of MELC estimators, and selection of the worst log*P* measure remain the same (**Table S3**, Supporting information).

To test whether some of the lipophilicity indices differ among themselves, the sevenfold cross-validation was performed as described in Section 2.4.1. Indices are then arranged in ascending order of the median values of SRD scores. The results are depicted in a form of a box and whisker plot indicating the maximum, minimum, median and interquartile ranges of the SRD scores (**Figure 2b**). Significance of a difference among the pairs of variables is tested by the Wilcoxon matched pair test and the sign test, leading to the four distinct sections marked with vertical dashed lines. Length of the lines reflects the magnitude of the difference. XlogP2, AlogP, miLogP, and acetone/water based mR_M and *PC*1 are in the first section, ranked as the best estimates with the lowest, yet statistically indistinguishable median values of the SRD scores. The rest of in silico estimates are located in the second group with some of the extrapolated RP chromatographic measures such as C_0. The third group gathers purely chromatographic measures, among them MLC retention parameters. The last section separates *PC*1_Diox, as the worst descriptor. This statistically justified solution is impossible to achieve by using PCA, HCA and correlation matrices, which are the most frequently applied methods to this kind of problems [2,8–13].

3.3. *Non-parametric ranking and grouping of thiepine derivatives by fusion of multiple lipophilicity measures*

Selection of the most and the least lipophilic compounds based on the lipophilicity indices recommended by consensus driven SRD should be straightforward. In that way miLogP, AlogP, XlogP2 as well as mR_M_Acet, and *PC*1_Acet unanimously identified compound **1** as the least lipophilic one. On the other hand in silico log*P* values suggest **10** for the most lipophilic derivative, while chromatographic measures select **4**. Clearly, different lipophilicity indices rank the same compound differently. It is, therefore, important to group and rank thiepines using a fusion of all statistically significant lipophilicity measures. This could be considered as a problem of multicriteria optimization with maximization as a goal. To find a solution, all statistically significant lipophilicity indices were used as an input for the SRD ranking. The data matrix was arranged with compounds in columns and lipophilicity measures in rows. Then the reference vector was calculated as a row-wise maximum. This vector should correspond to a virtual compound with maximum values of all lipophilicity parameters. Compound with the lowest SRD score (the closest to the virtual one) should be considered most lipophilic.

According to the SRD-CRRN (**Figure 3a**) compound **4** was selected as the most lipophilic, while **2**, **8**, **1** and **17** ranked the last. Compounds **3**, **5**, **9–12**, **14**, are in the middle. However, all of them are located in a very narrow range of SRD scores (4–17 units). Therefore it is important to test whether such differences are statistically significant. After the sevenfold cross-validation procedure, compounds are arranged in increasing order of the SRD

medians. Pairwise comparison (performed by Wilcoxon matched pair test and sign test) resulted in four distinct sections of similar compounds (**Figure 3b**). Compound **4** is again selected as the most lipophilic (phenyl substituted thiepine derivative in a position 8). This is also the most active compound against *C. albicans* strain [25]. The rest of 4-methoxy substituted phenyl compounds as well as phenyl substituted derivatives in positions 10 and 11 (**14**, **5**, **6**, **11**, and **12**) are located in the second group with slightly decreased lipophilicity. These compounds also have lower activity towards *C. albicans*. Introduction of cyano group to the phenyl core (**13**, **15** and **8**) significantly decreases lipophilicity, placing these compounds in the third and the fourth group along with methoxy and fluoro thiepine derivatives substituted in the position 8 (**16** and **17**), as well as unsubstituted compounds (**1** and **2**). Obviously, that introduction of the phenyl ring lead to increase of lipophilicity depending on the substitution position on the thiepine core (position 8 > 10 > 11). It also has a significant impact on the antifungal activity of the studied compounds.

4. Concluding remarks

The SRD ranking and grouping of chromatographic and in silico lipophilicity measures demonstrated advantages over the classical multivariate approach The best lipophilicity estimates, *i.e.*, the closest to the consensus point, were selected among computational (XlogP2, miLogP, AlogP) and chromatographic experiments carried out under typical RP conditions using acetone/water mixtures. Descriptors obtained from microemulsion

chromatography were statistically significantly ranked; however, they perform worse than all *in silico* measures. The worst lipophilicity index was PC1 calculated from TLC parameters based on dioxane/water mixtures.

The SRD ranking, based on multicriteria maximization, selected the thiepine derivative with phenyl ring in the position 8 as the most lipophilic compound. This is also a compound with the highest antifungal activity against *C. albicans*. Grouping of compounds into four distinct sections was observed in decreasing order of their lipophilicity (SRD-s). Introduction of a phenyl substituent increased lipophilicity significantly. However, cyano groups attached to the phenyl core alleviate such increase.

jssc201601442-sup-0001-SuppMat.pdf

Supplementary material

Acknowledgments

This study has been supported by the Ministry of Education, Science and Technological Development, Republic of Serbia (Grant nos. 172008 and 172017).

References

[1] Poole S., Poole C., Separation methods for estimating octanol–water partition coefficients. *J. Chromatogr. B* 2003,*797*, 3–19.

[2] Janicka M., Pachuta-Stec A., Retention–property relationships of 1,2,4-triazoles by micellar and reversed-phase liquid chromatography. *J. Sep. Sci.* 2014, *37*, 1419–1428.

[3] Soczewinski E., Matysik G., Two types of RM-composition relations in liquid partition chromatography. *J. Chromatogr.* 1968, *32*, 458-471.

[4] Bate-Smith E.C., Westall R.G., Chromatographic Behavior and Chemical Structure. Some Naturally Occurring Phenolic Substances. *Biochim. Biophys. Acta* 1950, *4*, 427–440.

[5] Andrić F., Šegan S., Dramićanin A., Majstorović H., Milojković-Opsenica D., Linear modeling of the soil-water partition coefficient normalized to organic carbon content by reversed-phase thin-layer chromatography. *J. Chromatogr. A* 2016, *1458*, 136–144.

[6] Bieganowska M.L., Doraczynska A., Petruczynska A., The retention behavior of some sulfonamides on different TLC plates. 2. Comparison of the selectivity of the systems and quantitative determination of hydrophobicity parameters. *J. Planar Chromatogr.—Mod. TLC* 1995, *8(2)*, 122–128.

[7] Andrić F., Bajusz D., Ráczc A., Šegan S., Héberger K., Multivariate assessment of lipophilicity scales—computational and reversed phase thin-layer chromatographic indices. *J. Pharm. Biomed. Anal.* 2016, *127*, 81–93.

[8] Sarbu C., Kuhajda K., Kevresan, S., Evaluation of the lipophilicity of bile acids and their derivatives by thin-layer chromatography and principal component analysis, *J. Chromatogr. A* 2001, *917*, 361–366.

[9] Onisor C., Posa M., Kevresan S., Kuhajda, K., Sarbu, C., Estimation of chromatographic lipophilicity of bile acids and their derivatives by reversed-phase thin layer chromatography, *J. Sep. Sci.* 2010, *33*, 3110–3118.

[10] Casoni D., Sarbu C., Comprehensive evaluation of lipophilicity of biogenic amines and related compounds using different chemically bonded phases and various descriptors, *J. Sep. Sci.* 2012, *35*, 915–921.

[11] Briciu R.D., Kot-Wasik A., Wasik A., Namiesnik J., Sarbu C., The lipophilicity of artificial and natural sweeteners estimated by reversed-phase thin-layer chromatography and computed by various methods, *J. Chromatogr. A* 2010, *1217*, 3702–3706.

[12] Nascu-Briciu R.D., Sarbu C., A comparative study concerning the chromatographic behaviour and lipophilicity of certain natural toxins, *J. Sep. Sci.* 2012, *35*, 1059–1067.

[13] Janicka M., Pietras-Ozga D., Chromatographic evaluation of the lipophilicity of N-phenyltrichloroacetamide derivatives using micellar TLC and OPLC. *J. Planar Chromatogr.- Mod. TLC* 2010, *23*, 396–399.

[14] Medina-Hernandez M.J., Sagrado S., Chromatographic evaluation of hydrophobicity using micellar mobile phases. *J. Chromatogr. A* 1995, *718*, 273–282.

[15] Nishi H., Pharmaceutical application of micelles in chromatography and electrophoresis. *J. Chromatogr. A* 1997, *780*, 243–264.

[16] Xu L., Li L., Huang J., Yu S., Wang J., Determination of the lipophilicity (logPo/w) of organic compounds by microemulsion liquid chromatography. *J. Pharm. Biomed. Anal.* 2015, *102*, 409–416.

[17] Xuan X., Xu L., Li L., Gao G., Li N., Determination of drug lipophilicity by phosphatidylcholine-modified microemulsion high-performance liquid chromatography. *Int. J. Pharm.* 2015, *490*, 258–264.

[18] Komsta L., Skibinski R., Berecka A., Gumieniczek A., Radkiewicz B., Radon M., Revisiting thin-layer chromatography as a lipophilicity determination tool—A comparative study on several techniques with a model solute set. *J. Pharm. Biomed. Anal.* 2010, *53*, 911–918.

[19] Sobanska A.W., Wojcicka K., Brzezinska E., Evaluation of the lipophilicity of selected sunscreens—A chemometric analysis of thin-layer chromatographic retention data. *J. Sep. Sci.* 2014, *37*, 3074–3081.

[20] Andrić F., Héberger K., Chromatographic and computational assessment of lipophilicity using sum of ranking differences and generalized pair-correlation, *J. Chromatogr. A* 2015, *1380*, 130–138.

[21] Andrić F., Héberger K., Towards better understanding of lipophilicity: assessment of in silico and chromatographic logP measures for pharmaceutically important compounds by nonparametric rankings. *J Pharm. Biomed. Anal*. 2015, *115*, 183–191.

[22] Ansari M.I., Hussain M.K., Arun A., Chakravarti, B., Konwar R., Hajela K., Synthesis of targeted dibenzo[b,f]thiepines and dibenzo[b,f]oxepines as potential lead molecules with promising anti-breast cancer activity. *Eur. J. Med. Chem.* 2015, *99*, 113–124.

[23] Božinović N., Novaković I., Kostić Rajačić S., Opsenica I., Šolaja B., Synthesis and antimicrobial activity of azepine and thiepine derivatives. *J. Serb. Chem. Soc.* 2015, *80(7)*, 839–852.

[24] Božinović N., Opsenica I., Šolaja B.A., Double Palladium-Catalyzed Synthesis of Azepines. *Synlett* 2013, *24*, 49–52.

[25] Božinović N., Šegan S., Vojnović S., Pavić A., Šolaja B., Nikodinović-Runić J., Opsenica I., Synthesis and anti-Candida activity of novel benzothiepino[3,2-c]pyridine derivatives. *Chem. Biol. Drug Des.* 2016, *88*, 795–806.

[26] Tetko I.V., Gasteiger J., Todeschini J.R., Mauri A., Livingstone D., Ertl P., Palyulin V.A., Radchenko E.V., Zefirov N.S., Makarenko A.S., Tanchuk V.Y., Prokopenko V.V., Virtual computational chemistry laboratory - design and description. *J. Comput. Aid. Mol. Des.* 2005, *19*, 453–463.

[27] Tetko I.V., Computing chemistry on the web. *Drug Discov. Today* 2005, *10*, 1497–1500.

[28] Héberger K., Sum of ranking differences compares methods or models fairly, *TrAC - Trends Anal. Chem.* 2010, *29*, 101–109.

[29] Héberger K., Kollár-Hunek K., Sum of ranking differences for method discrimination and its validation: comparison of ranks with random numbers, *J. Chemometr.* 2010, *25*, 151–158.

Figure captions

Figure 1 Principal component analysis (PCA) of lipophilicity indices; (**a**) Score plot with characteristic disposition of thiepine derivatives in the PC1/PC2 space; (**b**) Loading diagram of the lipophilicity indices including the consensus point (AMA).

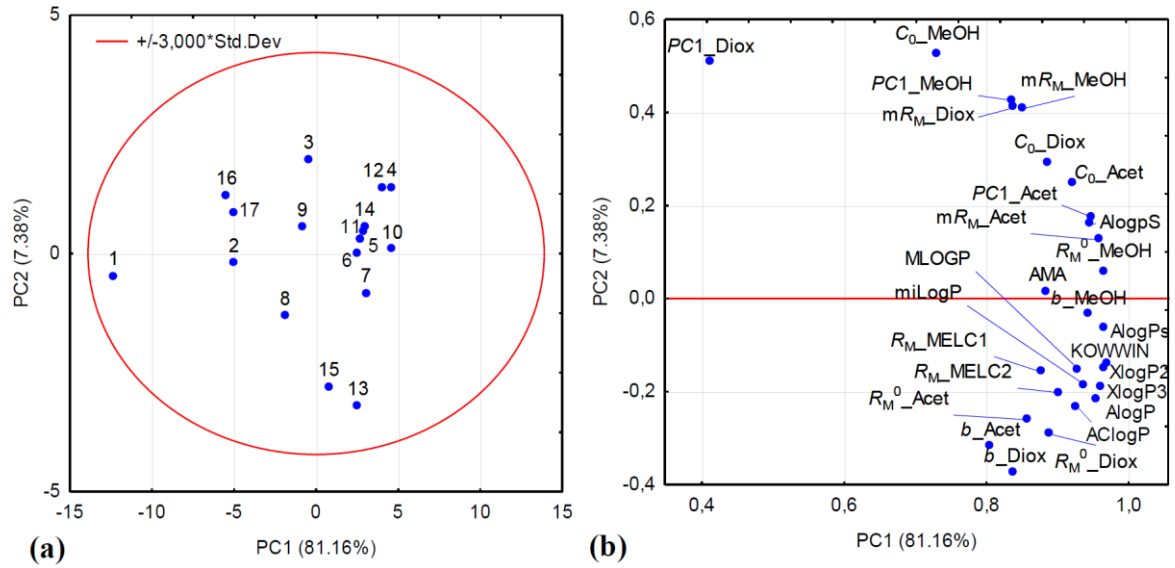


Figure 2 SRD of the standardized chromatographic and in silico lipophilicity measures; (**a**) Comparison with random numbers: *x* axis and left side *y* axis depicts the interval scaled SRD values (%), right side *y* axis represent relative frequencies of the fitted Gaussian distribution, XX1 denotes the lower 95% confidence limit; (**b**) Box and whisker plot of the SRD-s obtained by the sevenfold cross-validation. Dashed lines separate statistically significantly different variables following the Wilcoxon matched pair test and the sign test ($p = 0.05$).

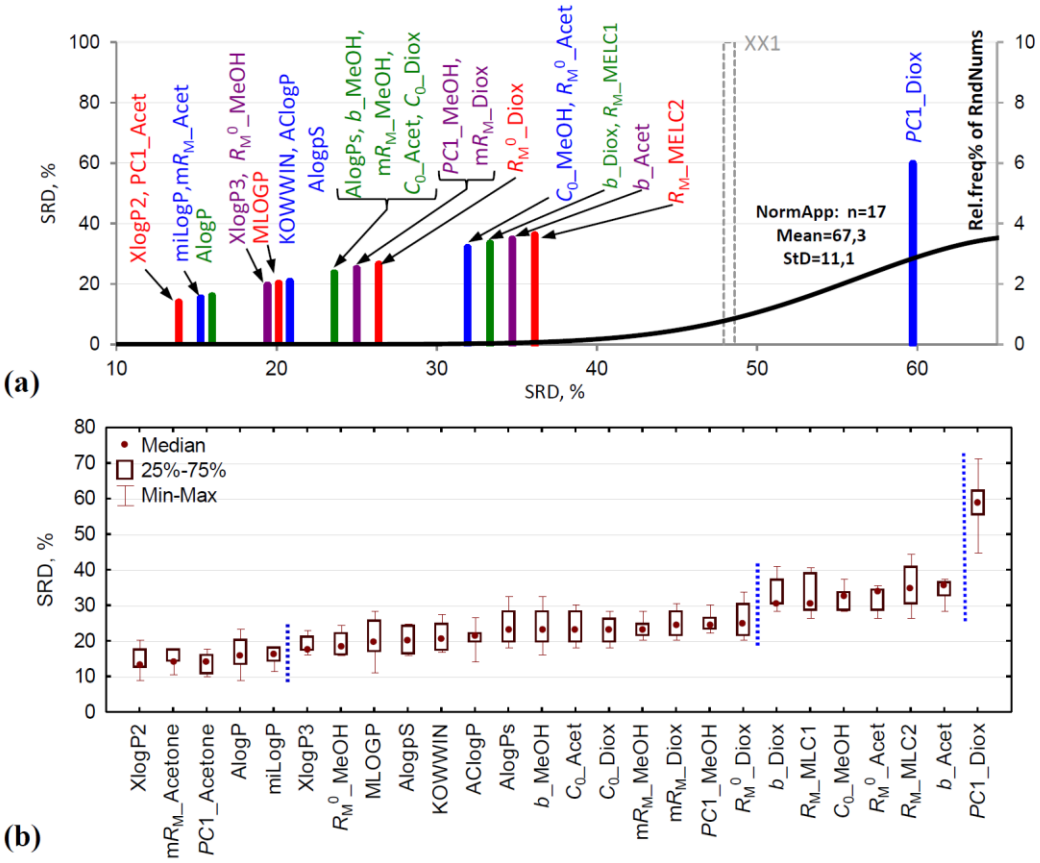


Figure 3 SRD of thiepine derivatives; (**a**) Comparison with random numbers: *x* axis and left side *y* axis stands for the interval scaled SRD values (%), right side *y* axis represent relative frequencies of the fitted Gaussian distribution of SRD-s, XX1 denotes the lower 95% confidence limit; (**b**) Box and whisker plot of the SRD-s obtained by the sevenfold cross-validation. Dashed lines separate statistically significantly different compounds following the Wilcoxon matched pair test and the sign test ($p = 0.05$).

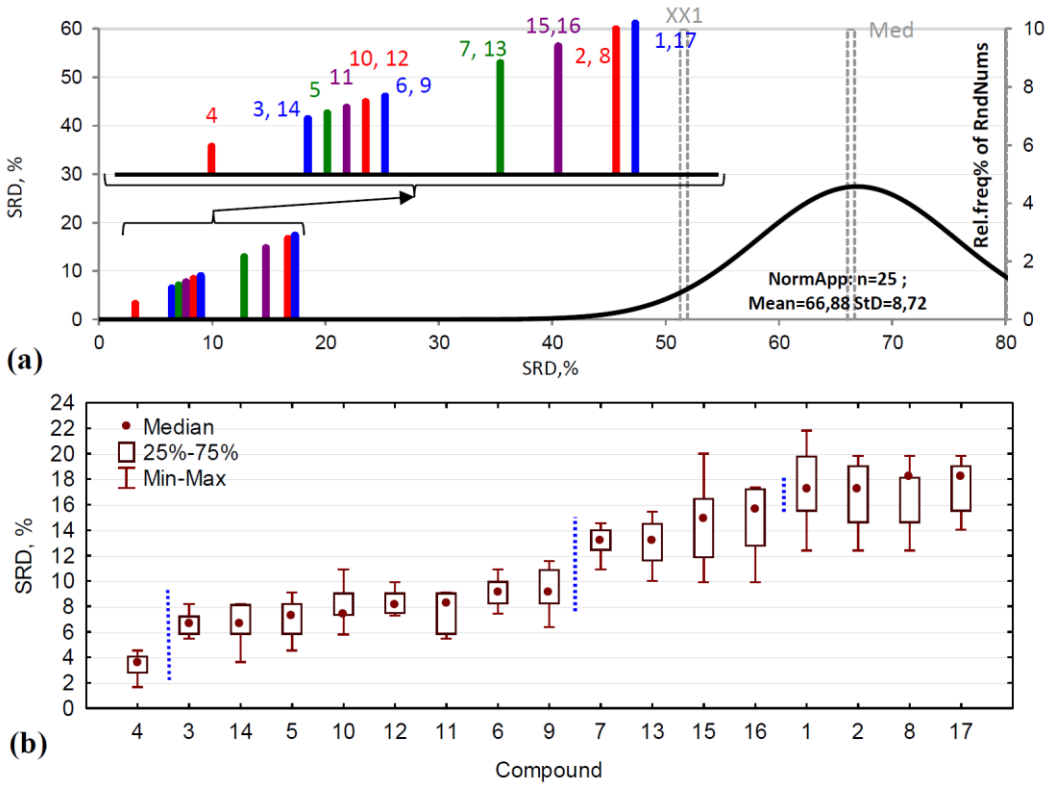
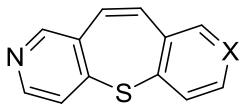
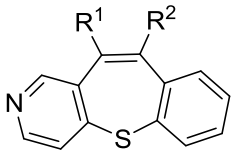


Table 1. Structural formulas of the studied benzothiepino[3,2-c]pyridine derivatives





Comp. no.	X	Comp. no.	R^1	R^2
1	N	**5**	Ph	H
2	CH	**6**	4-MeO-Ph	H
3	CCl	**7**	4-F-Ph	H
4	CPh	**8**	4-CN-Ph	H
12	C-(4-MeO-Ph)	**9**	Br	H
15	C-(4-CN-Ph)	**10**	SPh	H
16	$COCH_3$	**11**	H	Ph
17	CF	**13**	S-(4-CN-Ph)	H
		14	H	4-MeO-Ph