Revisiting Thin-Layer Chromatography as a Lipophilicity Determination Tool.
Part IV — A Closer Look on Extrapolation and Averaging

Ł. KOMSTA* AND E. GOWIN

Chair and Department of Medicinal Chemistry, Faculty of Pharmacy, Medical University of Lublin
Jaczewskiego 4, 20-090 Lublin, Poland
E-mail: lukasz.komsta@am.lublin.pl

Summary. The paper compares linear, quadratic, and cubical regression together with several weighted and robust approaches in the context of lipophilicity determination. The comparison is done on 35 model compounds on data from different modifiers used on RP18, CN, and silica plates. It can be concluded that the use of weighted and moderately robust regression technique increases correlation between extrapolated retention and real lipophilicity, whereas polynomial and very robust techniques give visibly worse results due to their excessive flexibility and higher extrapolation uncertainty. Additionally, we have compared averaging retention from different modifiers by $R_F$, $k$, and $R_M$ values. The results are similar; however, surprisingly, $R_F$ averaging performs slightly better to the other approaches.

Key Words: lipophilicity, extrapolation, robust regression, polynomial regression

Introduction

Thin-layer chromatography is very often used as a powerful tool for lipophilicity estimation. As there are no standardized conditions for such analysis, various approaches are used in literature. There are almost all water-miscible solvents used as modifiers and many computational approaches — from simple linear extrapolation to advanced chemometric methods. Newer ideas are also proposed — for example, use of ionic liquids [1] or micellar mobile phases [2] with high-performance liquid chromatography (HPLC), which was not done in thin-layer chromatography (TLC). There have been dozens of papers published since the first part of our study. To mention only several examples, in reversed phase, methanol is used mainly as a modifier [3–7]; however, there are also studies done with acetone [8–10], acetonitrile and tetrahydrofuran [11–13] or dioxane [14]. Several new papers report use of micellar TLC [15–16] or even reversed phase thin-layer chromatography (RP-TLC) done on silica plates impregnated with oils or...
fats [17–20]. So TLC is still seen as an interesting alternative, having its advantages and disadvantages. The main advantage is low cost, use of low solvent volumes, and time saving. The disadvantage is a difficulty to use gradient development, which is often used in HPLC to enhance range of lipophilicity determined in single run [21–22].

Table I. Comparison of the regression methods used in the investigation

<table>
<thead>
<tr>
<th>Name</th>
<th>Linear</th>
<th>Weighted</th>
<th>Robust</th>
<th>Minimized objective function</th>
</tr>
</thead>
<tbody>
<tr>
<td>LM</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Sum of squared residuals</td>
</tr>
<tr>
<td>WLM</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Sum of squared residuals</td>
</tr>
<tr>
<td>RLM</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Robust measure (M-estimator) done on residuals</td>
</tr>
<tr>
<td>WRLM</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Robust measure (M-estimator) done on residuals</td>
</tr>
<tr>
<td>LQS</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Quantile of squared residuals</td>
</tr>
<tr>
<td>LMS</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Quantile of squared residuals (quantile value is different than in LQS)</td>
</tr>
<tr>
<td>LTS</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Sum of quantile smallest squared residuals</td>
</tr>
<tr>
<td>C</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Sum of squared residuals</td>
</tr>
<tr>
<td>SQ</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Sum of squared residuals</td>
</tr>
</tbody>
</table>

This fact inspired us to previous parts of the study, where we decided to compare different modifiers, adsorbents and computational techniques on 35 model compounds with simple molecules. We have performed such study on RP18 (Part I, [23]), silica gel (Part II, [24]), and cyanopropyl adsorbent (Part III, [25]). The reader is referred to these papers for detailed literature background regarding older papers. From the previous parts of the study, it can be generally concluded that, in reversed phase TLC, acetoneitrile should be avoided as modifier because it gives visibly worse results. From well-behaving modifiers, methanol should be recommended not only due to our results, but also in theoretical context as it is most water-like in its molecular structure. RP18 is the best stationary phase for lipophilicity determination among the three investigated previously (Part I).
The previous parts of the study did not show any visible advantage of advanced chemometric methods used for lipophilicity computation from retention matrix, such as principal component analysis or PARAFAC (parallel factor analysis). The classical extrapolation of the retention seems to be still the preferred method; however, our findings recommended also computation of lipophilicity from single chromatographic plate using reference compounds. This method had similar or even better performance than linear extrapolation on our model compounds data set.
The main problem with retention extrapolation is a slight nonlinearity of the retention dependence on modifier concentration. This dependence is almost always slightly concave (curvilinear or piecewise linear). Therefore, linear extrapolation from full data set results in visibly lower extrapolated value than in the case of lower part of modifier concentrations. This concavity often encourages chromatographers to fit the data to other nonlinear equations proposed by many authors [26].

However, nonlinear extrapolation of retention can be rarely treated as a good lipophilicity estimate due to high uncertainty of predicted value and sensitivity to huge changes as a response to small changes to retention. On the contrary, classical least squares linear method is sensitive to outliers and seems to underestimate extrapolated value due to concavity. Therefore, we have decided to discuss popular robust and weighted regression methods in lipophilicity estimation, comparing them with classical least squares method and also to quadratic and cubical equation.

**Theory**

The classical least squares regression method (LM) minimizes the sum of squared residuals. Each object is treated in the same way (the method is unweighted), and therefore, this method does not “prefer” any part of the data and it is extremely sensitive to outliers. In general, we can divide the modifications of linear regression to three groups [27]:

1. Robust methods, which fit the line to “majority” of the data points, neglecting all data not conforming to the major trend. These methods automatically detect the major trend, i.e. which observations should be taken into account and which should be neglected. The main aim to use these methods is to neglect the outliers which can affect the fitted dependence significantly.
2. Weighted methods, which allow the user to define weights. The points with higher weights have higher impact onto the fitted line and user decides which points are more important. In the current study, we use $1/x$ weighting, which gives higher impact to points with smaller modifier concentration.
3. Combined (weighted and robust) methods.

The open source GNU R computational environment (used in the study and available for free from www.r-project.org) has all often used robust regression methods implemented in base system and MASS package. We have used the following modified regression methods: WLM (linear with weights set to $1/x$), RLM (robust linear model based on M-estimator, iteratively re-
weighting the fit to robust one), WRLM (combined approach — RLM with $1/x$ weights), LTS (least trimmed squares, minimizing sum of smaller half of residuals), LMS (least median of squares, minimizing not the sum but the median itself), LQS (analogous method but minimizing higher quantile estimated automatically), SQ (classical quadratic equation fitted by ordinary least squares), and C (cubic equation also classically fitted). The differences between methods are summarized in Table I.

Experimental

The chromatographic data set used in this paper is collected from the results of Parts I, II and III of our study [23–25] and the interested reader is referred to these papers for experimental conditions. The data set consisted of $R_M$ values of 35 compounds (Table II) for 54 TLC systems on RP18, 54 TLC systems on silica and 30 systems on CN plates. The used modifiers were: acetonitrile (ACN), acetone (ACT), dioxane (DIO), isopropanol (IPA), methylethylketone (MEK), methanol (MET), ethyl acetate (ETA), and tetrahydrofuran (THF). Six of them were studied in RP18 and Silica and five on CN plates. Thus, the regression values were arranged into the matrix containing 9 rows (techniques) and 5–6 columns (modifiers).

All computations were done within GNU R 2.14.0 under Windows (open source software available freely at www.r-project.org).

Results and Discussion

Fig. 1 presents typical plot of retention with three annotated lines. The classical least squares regression (A) treats equally all data points. It results with an extrapolated retention, which seems to be underestimated (too low) in the case of concavity. If one prefers to fit the line to lower modifier concentrations (increase the extrapolated value), a weighted regression is needed (B). Setting weights to $1/x$ makes this method sensitive to lower modifier concentrations. An opposite behavior is observed in the case of robust regression methods (C). These methods fit line to majority of data points, which in typical case fits a line to higher modifier concentrations, lowering the extrapolated value below that of classical least squares retention. This figure does not contain polynomial curves, as they are easy to imagine and would obscure the three lines. In the case of quadratic or cubic equation, the extrapolated value would be even higher, but some disturbances in data or outlying points make the obtained value almost unpredictable.
We have used in our comparison four robust methods: RLM, LQS, LTS, and LMS; one weighted method: WLM, one combined method: RWLM, and two polynomial methods: quadratic SQ and cubical C. In the first group, RLM method has a different behavior than the others. It is based on robust M-estimator, not minimizing quantile of residuals. Therefore, it is less robust and neglects only obvious outliers.

For each regression method, six correlation values for each used modifier were obtained. This formed matrix of dimension $9 \times 6$ (9 regression techniques and 6 modifiers) for each adsorbent. Therefore, we ended with three such matrices, for RP18, silica and CN, respectively.

The easiest way to explore the similarities between correlation is scaled principal component analysis on such matrix. The results for three adsorbents are presented in Fig. 2. The loading vectors (arrows) indicate directions responsible for increasing the correlation values (from the analyzed matrices) with particular modifier (annotated along the arrows), and one can see...
that there is no strict correspondence between technique efficiency and used modifier.

Fig. 2. Scaled PCA comparison of resulted $R_{M0}$ correlations with lipophilicity for (A) RP18, (B) CN, and (C) silica as adsorbent. See text for abbreviations.
In the case of RP18 (Fig. 2A), the PCA plot explains 77.5% of total variance (50.2% in first PC, 27.3% in second PC). One can conclude that RLM method (regardless of weighting) is the best technique. It outperforms classical LM probably due to neglecting single individual outliers. WLM is slightly worse than classical LM, whereas other techniques give worse results. The other robust methods perform especially poorly in the case of acetonitrile and isopropanol, whereas cubical equation behaves poorly with acetone, dioxane, and methanol. The similar situation can be concluded with CN plates (Fig. 2B), but the loading vectors group themselves differently. Acetone, acetonitrile, and isopropanol gave worst results for cubical extrapolation than dioxane and methanol for quadratic one. It should be noted that quadratic extrapolation performs visibly worse on CN plates than on RP18 adsorbent. On CN plates, weighting results in slightly better correlation.

On silica gel (Fig. 2B), first two components model 81.9% of whole variance (55.3% in PC1, 27.4% in PC2). One can note that quadratic extrapolation performs similar to linear weighted and unweighted method, both classical and RLM. Other robust techniques and cubical extrapolation perform visibly worse. The interesting difference can be seen in the case of robust techniques, LQS behaves similarly to cubical extrapolation but in different manner than LMS and LTS. The first two are worst choice in the case of acetone and ethyl acetate, whereas they perform better with dioxane, isopropanol, and methylethylketone. LMS and LTS perform in reverse manner.
In the case of CN adsorbent (Fig. 2C), the modelled variance ratio is 94.3% (83.9% in PC1, 10.4% in PC2). The conclusions are slightly similar, but the main difference is good performance of quadratic extrapolation, opposite to two previous adsorbents.

The weighted methods allow user to choose weights according to his/her wish. This raises the question if the efficiency can be improved by increasing disproportion between modifier concentrations, using for example $1/x^2$ or $1/x^3$ weights instead of $1/x$. We have checked this idea; however, this did not improve the results. Raising denominator to the power 2 has similar impact onto resulting correlation, whereas further increasing worsens the efficiency. Therefore, weighting on $1/x$ or $1/x^2$ can be treated as optimal.

These results conform to uncertainty theory, as RLM method has the smallest uncertainties at $x = 0$ among all investigated methods (except of LMS, LQS, and LTS, where the uncertainty cannot be computed). Therefore, the results are in good agreement to theoretical background of regression methods.

Another main point, not addressed in the literature, is averaging the extrapolated retention obtained with several modifiers to compute better lipophilicity estimator. This can be done in three ways, giving different results, summarized in Table III with final evaluated formulas:

1. Averaging $R_F$ values then converting them to $R_M$ and extrapolation;
2. Computing $k$ values, averaging them, then converting to $R_M$ and extrapolation;
3. Averaging $R_M$ values, which is the most intuitive approach.

### Table III. Comparison of averaging methods used in this paper

<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
<th>Final average $R_M$ formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_F$</td>
<td>Averaging $R_F$, computing $k$ from average value and converting it to $R_M$</td>
<td>$R_M = \log \left[ \frac{n}{\sum R_F} - 1 \right]$</td>
</tr>
<tr>
<td>$k$</td>
<td>Computing $k$ for each $R_F$, averaging $k$ values and converting average value to $R_M$</td>
<td>$R_M = \log \left[ -\frac{n}{nR_F} + 1 \right]$</td>
</tr>
<tr>
<td>$R_M$</td>
<td>Computing $R_M$ for each $R_F$ and averaging the results</td>
<td>$R_M = \log \left[ \frac{1 - R_F}{\Pi R_F} \right]$</td>
</tr>
</tbody>
</table>

Although averaging $R_M$ values (doing this before and after extrapolation is mathematically equivalent) is the most natural approach here, there is no theoretical evidence what method should be preferred. This assumption results from initial assumptions of arithmetic mean, that it is a good
location measure in the case of normal distribution. One cannot prove which variable ($R_F$, $k$, $R_M$) has a distribution closest to normal. Therefore, we have decided to perform additional comparison of averaging between all modifiers on RP18 plates on $R_F$, $k$, and $R_M$ values.

Fig. 3 shows the correlation between averaged extrapolated retention and real known lipophilicity (determined by shake-flask method and taken from www.vcclab.org database). One can surprisingly conclude that,

![Fig. 3. Correlation with $R_{M0}$ values averaged from six modifiers on RP18 adsorbent obtained by different averaging techniques, i.e., by their (A) $R_F$ value, (B) $k$ value, and (C) $R_M$ value. See Table II for outlying compound annotations and Table III for three methods of averaging.](image-url)
although all three methods result in similar correlation, $R_F$ values averaging performs slightly better (correlation is 0.8 in this case, comparing to 0.7 in the other two approaches). In this case, further studies should be performed to explain this behavior.

**Conclusions**

The use of weighted regression based on M-estimator (RLM with weights equal to $1/x$) or RLM without weighting is a good compromise between classical regression and other more flexible techniques in lipophilicity estimation. When averaging several modifiers between extrapolation, one can consider averaging $R_F$ values and compare to averaging by $R_M$ or at least compare the proposed three methods of averaging.

**References**


Accepted by TK