CROATICA CHEMICA ACTA
CCACAA, ISSN 0011-1643, e-ISSN 1334-417X
*Croat. Chem. Acta* **87** (3) (2014) 261–269.
http://dx.doi.org/10.5562/cca2382

*Original Scientific Article*

# HPLC Retention Behavior of Triacylglycerols Extracted from Soybean Oil by Supercritical $CO_2$

**Stela Jokić,[a] Sanja Podunavac Kuzmanović,[b] Lidija Jevrić,[b] Rezica Sudar,[c] Senka Vidović,[d] Krunoslav Aladić,[e] and Strahinja Kovačević[b,*]**

[a]*University of Josip Juraj Strossmayer in Osijek, Faculty of Food Technology Osijek, Franje Kuhaca 20, 31000 Osijek, Croatia*
[b]*University of Novi Sad, Faculty of Technology, Department of Applied and Engineering Chemistry, Bulevar cara Lazara 1, 21000 Novi Sad, Serbia*
[c]*Agricultural Institute Osijek, Juzno predgrade 17, 31000 Osijek, Croatia*
[d]*University of Novi Sad, Faculty of Technology, Department of Biotechnology and Pharmaceutical Engineering, Bulevar cara Lazara 1, 21000 Novi Sad, Serbia*
[e]*Croatian Veterinary Institute, Branch-Veterinary Institute Vinkovci, Josipa Kozarca 24, 32100 Vinkovci, Croatia*

RECEIVED NOVEMBER 8, 2013; REVISED JULY 9, 2014; ACCEPTED JULY 24, 2014

*Abstract.* Soybean oil fractions were obtained by collecting extract at different time intervals during supercritical $CO_2$ extraction. Relationships between retention behavior of seventeen TAGs and their molecular characteristics were studied using chemometric approach. Quantitative structure-retention relationship (QSRR) analysis was carried out on retention time values ($t_r$) obtained by high pressure liquid chromatography to identify structural requirements of different TAGs for their retention. Principal component analysis (PCA) was performed in order to select molecular descriptors that best describe retention behavior of the compounds investigated, and to determine the similarities among molecules. The accurate mathematical models were developed for predicting the retention behaviour of some TAGs. The validity of the models was evaluated by suitable statistical and cross-validation parameters.

*Keywords:* soybean oil, HPLC, triacylglycerols, supercritical $CO_2$, extraction

*Abbreviations:* LnLnLn - trilinolenoin, LnLnL - dilinolenolinolein, LnLnO - dilinolenoolein, LnLL - linolenodilinolein, LnLnP - dilinolenopalmitin, LLL - trilinolenoin, LnLO - linolenolinoleoolein, LLO - dilinoleoolein, LLP - dilinoleopalmitin, LLO - linoleodiolein, LOP - linoleooleopalmitin, PLnP - linolenodipalmitin, OOO - triolein, LOS - linoleooleostearin, OOP - dioleopalmitin, OOS - dioleostearin, SOP - stearinoleopalmitin

## INTRODUCTION

Pressing and extraction with organic solvents is widely used in the production of vegetable oils. For soybean oil, hexane has been the preferred extraction solvents for a long time but in recent year's supercritical $CO_2$ extraction, as environmentally friendly technique, appeared to be alternative to current extraction methods. $CO_2$ is non-toxic, non-explosive, inflammable, cheap, readily available solvent with recoverable characteristics.[1–3] Supercritical $CO_2$ extraction as a replacement of organic solvents in soybean oil extraction was considered in the last years by few researchers.[3–13] It has been proven that the oil extracted from soybeans with supercritical $CO_2$ is much higher quality than the same oil extracted by hexane. Furthermore, the refinement stages are simplified significantly and the solvent distillation stage is completely removed.

Edible oils are composed of mainly triacylglycerols (TAGs) and analysis of TAGs is a critical step to understand physicochemical properties of vegetable oil. TAGs are esters of fatty acids and glycerol. Triacylglycerol chain of soybean oil contains five different fatty acids: palmitic, (16 : 0); stearic, (18 : 0); oleic, (18 : 1); linoleic, (18:2) and linolenic acid, (18:3).[14]

The mechanisms of chromatographic separation are very complex and depend on many factors such as experimental conditions, type of chromatographic system, physicochemical characteristics of analytes, *etc.* In order to understand chromatographic processes, it is

* Author to whom correspondence should be addressed. (E-mail: strahko@uns.ac.rs)

very useful to establish mathematical models which can predict the retention behavior of analytes on the basis of their structural characteristics in applied chromatographic system. Determination of the correlations between molecular structure and retention behavior of molecules in different chromatographic systems is the main task of quantitative structure–retention relationship (QSRR) chemometric method. Chemometric analysis is undoubtedly of great importance in modern science. It means performing calculations on measurements of chemical data.

QSRR analysis is very often applicable for prediction of the retention behavior of newly synthesized molecules and quantitative comparison of separation properties of individual types of chromatographic layers. QSRR studies are widely applied in high-pressure liquid chromatography (HPLC), gas chromatography (GC) and thin-layer chromatography (TLC).[15–19]

In this context, the goal of the present study was to evaluate the HPLC retention data by QSRR analysis. A central object of this study was to establish the possible relationships between retention characteristics and the structural descriptors of the investigated triacylglycerols in order to predict the retention behavior of this class of molecules.

## EXPERIMENTAL

### Materials

The supercritical $CO_2$ extraction was performed on the soybean cultivar "Ika" created at the Agricultural Institute Osijek in Croatia in 2009. The samples were cleaned from impurities. The material was grounded and sieved using sieve sets (Erweka, Germany) and the average particle size was determined. The prepared samples were then stored at +4 °C prior to extraction. Moisture was determined by oven drying to the constant weight at 105 °C.[20]

Commercial $CO_2$ (M/s, Novi Sad, Serbia) was used. HPLC grade acetone and acetonitrile were purchased from Baker J. T., Milan (Italy). Lipid standards (LnLnLn, LLL, OOO) were obtained from Sigma chemical, St. Louis (USA). Other used chemicals were of analytical reagent grade.

### Supercritical $CO_2$ Extraction of Soybean Oil

The experiments were performed on the laboratory-scale high pressure extraction plant (HPEP, NOVA-Swiss, Effertikon, Switzerland) given in detail elsewhere.[8,9,21] The main plant parts and properties, according to manufacturer specifications, were: the diaphragm type compressor (with pressure range up to 1000 bar), extractor with internal volume 200 mL ($p_{max}$ = 700 bar), separator (with internal volume 200 mL, $p_{max}$ = 250 bar), and maximum $CO_2$ mass flow rate of 5.7 kg $h^{-1}$.

The ground soybean sample of 120 g was placed into an extractor vessel. The extracts were collected in previously weighed glass tubes. The amount of extract obtained at regular intervals of time was established by weight using a balance with a precision of ±0.00001 g. Separator conditions were 15 bar and 298 K.

At the different extraction conditions of pressure (300, 400, 500 bar), temperature (40, 50, 60 °C), $CO_2$ flow rate (0.194, 0.436 kg $h^{-1}$) and characteristic particle size (0.238, 0.383, 1.059 mm), extraction process was carried out until extraction yield become constant. Different fractions, depending on extraction conditions, were obtained by collecting extract every two hours during the extraction process. After each extraction, the obtained extract was placed into glass vials (25 mL), sealed and stored at +4 °C to prevent any possible degradation.

### HPLC Analysis of Studied Compounds

TAGs were analysed by the IUPAC method[22] using a Perkin-Elmer High Performance Liquid Chromatography system series 200 equipped with isocratic pump, refractive index detector and TotalChrom Navigator (HPLC software). The separation was performed on two serial connected PE Pecosphere C18 columns (83×4.6). The analysis was carried out with acetone / acetonitrile (70 : 30) as a mobile phase. Standard and oil samples (5 %) were dissolved in HPLC-grade acetone and 20 μL aliquots were injected into the column and eluted at a flow rate of 2.5 mL $min^{-1}$. Furthermore, TAGs were identified by comparing their retention time to standards. Experiments were conducted in triplicate.

### Molecular Modeling and *in silico* Molecular Descriptors

The derivation of *in silico* molecular descriptors proceeds from the chemical structure of the compounds. In order to calculate the molecular descriptors, all molecules were drawn into ChemBioDraw Ultra version 12.0 program. The 3D modeling of examined molecules was carried out using ChemBio3D Ultra version 12.0 software running on AMD Sempron Processor 3000+. The obtained 3D models were subjected to energy minimization using molecular mechanics force field method (MM2). The cutoff for structure optimization was set at a gradient of 0.1 kcal $Å^{-1}$ $mol^{-1}$. The Austin Model 1 (AM-1) was used for full geometry optimization of all structures until the root mean square (RMS) gradient reached a value smaller than 0.0001 kcal $Å^{-1}$ $mol^{-1}$ using MOPAC.

**Table 1.** Retention times and molecular parameters of investigated triacylglycerols

| Compounds | $t_r$ | *mi*log*P* | Clog*P* | Alog*Ps* | AClog*P* | log$P_{Kow}$ | Xlog*P*2 | Xlog*P*3 | Alogp*S* | AClog*S* | MV |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Series I | | | | | | | | | | | |
| LnLnP | 3.800 | 10.487 | 21.876 | 10.590 | 19.580 | 21.660 | 18.580 | 19.500 | −8.050 | −12.120 | 932.420 |
| LLP | 5.200 | 10.643 | 22.844 | 10.750 | 20.300 | 22.090 | 19.240 | 20.870 | −8.100 | −12.580 | 944.790 |
| LOP | 6.460 | 10.694 | 23.328 | 10.770 | 20.660 | 22.310 | 19.560 | 21.560 | −8.110 | −12.810 | 950.80 |
| OOP | 8.570 | 10.742 | 23.812 | 10.740 | 21.020 | 22.520 | 19.890 | 22.240 | −8.050 | −13.030 | 957.160 |
| SOP | 10.750 | 10.772 | 24.296 | 10.710 | 21.370 | 22.740 | 20,410 | 23.180 | −7.890 | −13.260 | 963.350 |
| Series II | | | | | | | | | | | |
| LnLnLn | 2.520 | 10.380 | 21.482 | 9.980 | 19.440 | 22.000 | 18.340 | 18.280 | −8.200 | −11.980 | 947.460 |
| LnLnL | 2.910 | 10.470 | 21.966 | 10.220 | 19.800 | 22.220 | 18.670 | 18.960 | −8.260 | −12.210 | 953.650 |
| LnLnO | 3.070 | 10.530 | 22.450 | 10.510 | 20.160 | 22.430 | 18.990 | 19.650 | −8.230 | −12.430 | 959.830 |
| LnLL | 3.430 | 10.553 | 22.450 | 10.500 | 20.160 | 22.430 | 18.990 | 19.650 | −8.210 | −12.430 | 959.830 |
| LLL | 4.030 | 10.629 | 22.934 | 10.650 | 20.510 | 22.650 | 19.320 | 20.340 | −8.050 | −12.660 | 966.020 |
| LnLO | 4.360 | 10.608 | 22.934 | 10.670 | 20.510 | 22.650 | 19.320 | 20.340 | −8.050 | −12.660 | 966.020 |
| LLO | 4.950 | 10.680 | 23.418 | 10.740 | 20.870 | 22.860 | 19.650 | 21.020 | −8.150 | −12.890 | 972.210 |
| LOO | 6.220 | 10.729 | 23.902 | 10.780 | 21.230 | 23.080 | 19.980 | 21.710 | −8.140 | −13.120 | 978.340 |
| OOO | 7.940 | 10.775 | 24.386 | 10.820 | 21.590 | 23.290 | 20.300 | 22.390 | −8.160 | −13.350 | 984.580 |
| LOS | 8.120 | 10.760 | 24.386 | 10.820 | 21.590 | 23.290 | 20.300 | 22.390 | −8.160 | −13.350 | 984.580 |
| OOS | 10.360 | 10.805 | 24.870 | 10.750 | 21.940 | 23.510 | 20.820 | 23.330 | −8.070 | −13.570 | 990.770 |
| PLnP[(a)] | 6.880 | 10.584 | 22.270 | 10.760 | 19.730 | 21.330 | 18.830 | 20.720 | −8.090 | −12.270 | 917.340 |

[(a)] Does not belong to any series.

The values of molecular descriptors (Table 1) for each molecule in the data set were calculated using the software ChemBio3D Ultra version 12.0 and ALOGPS 2.1. Determined descriptors of examined compounds were solubility descriptors (Alogp*S* and AClog*S*), molecular volume (MV) and the lipophilicity parameters, logP values, calculated by use of different theoretical procedures from the internet data (*mi*log*P*, Clog*P*, Alog*Ps*, AClog*P*, log$P_{Kow}$, Xlog*P*2, Xlog*P*3) (Table 1).

## Chemometric Analysis and Model Validation

In chemometric analysis the main problem is how to reduce the number of variables. This can be done by various statistical methods of explorative analysis, classification methods and regression methods. Principal component analysis (PCA) is the most often used explorative statistical methods.[23–26]

PCA is a technique for reducing the amount of data when there is correlation present. It is worth stressing that it is not a useful technique if the variables are uncorrelated. PCA calculates latent, new variables by a combination of the original variables, representing the multidimensional data structure in an optimal way. In a multidimensional space, where the variables define the axes, the data are projected into a few principal components (PCs) that are linear combinations of the original variables and describe the maximum variation within the data. Each PC is characterized by scores and loadings. Scores are the new coordinates of the projected objects, and loadings reflect the direction with respect to the original variables. The loadings plot displays relationships between variables and can be used to identify variables (molecular descriptors in this study) which contribute to the positioning of the objects on the scores plot. The scores plot provides a data overview displaying patterns or groupings within the data.

Model validation is a very important aspect of any QSRR analysis. The statistical quality of the generated models was measured by using the standard statistical parameters (Pearson's correlation coefficient ($r$), $F$-test (Fisher's value), and the standard error of estimation ($s$)), and cross-validation parameters (cross-validated coefficient of determination ($r^2_{CV}$), adjusted determination coefficient ($r^2_{adj}$), predicted residual sum of squares (PRESS), total sum of squares (TSS), and standard deviation based on predicted residual sum of squares ($S_{PRESS}$)).[27–30] The correlation coefficient values closer to 1.0 represent the better fit of the regression, and high values of the $F$-test indicate that the model is statistically significant. Standard deviation expresses the variation of the residuals or the variation about the regression line, and should have a low value for the regression to be significant. The lower PRESS value is, the better the predictability of the model.[31] If PRESS value is less than TSS value, the model predicts better and can be considered statistically significant. TSS

values are in terms of the dependent variable $y$. In many cases, $r^2_{CV}$ and $r^2_{adj}$ are taken as a proof of the high predictive ability of estimated mathematical models in QSRR. High values of these statistical characteristics ($r^2_{CV}$, $r^2_{adj} > 0.5$) indicate high predictivity of the equations. Unlike $r^2$, $r^2_{CV}$ may be negative, indicative of a very poor mathematical model, also unlike $r^2$, which tends to increase upon the addition of any descriptor, $r^2$CV will decrease upon the addition of irrelevant descriptors.

### Statistical Methods

The complete regression analysis was carried out by PASS 2005, GESS 2006, NCSS Statistical Software, as well as Statistica v. 8 software.

## RESULTS AND DISCUSSION

In our previous paper[10] we explain in detail how different extraction parameters (pressure, temperature, $CO_2$ mass flow rate and characteristic particle size) influenced on the extraction yield of soybean oil. The increase in pressure, temperature and $CO_2$ flow rate improved the extraction yield while decrease in particle size show higher extraction yield because of the increase in oil amount outside the particles, due to the enhancement of surface area with particle size reduction. The maximum obtained yield at different supercritical extraction conditions was 19.33 % which is very close to oil yield obtained by *n*-hexane (20.19 %). Furthermore, we investigated also the tocopherols content in soybean oil obtained by supercritical $CO_2$ at different extraction process conditions.[11] Chemometric analysis was successfully applied on different tocopherols isomers to model the relationships between the contents of different tocopherols isomers in soybean oil. Accurate mathematical models were developed for predicting the total tocopherols contents, as well as the contents of $\delta$-tocopherole isomer.

In this study, during the extraction of soybean oil by supercritical $CO_2$, different number of fractions, depending on extraction conditions, was collected every two hours. At the pressure of 300 bar and temperature of 40 °C extraction process was carried out for 12 hours, so six different fractions were collected. At the pressure of 400 bar extraction process was carried out for 8 hours for every set of temperature (40, 50 or 60 °C) and four different fractions has been collected. At the pressure of 500 bar and temperature of 40 °C extraction process was the shortest, 6 hours, so three different fractions were collected. In all collected fractions the concentration of TAG was determined using reversed phase high performance liquid chromatography. The application of this method resulted in successful separation of the

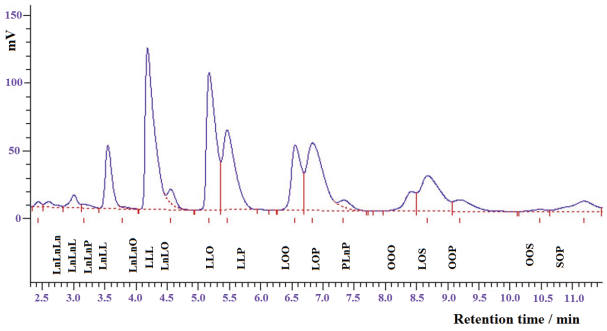


**Figure 1.** The representative chromatogram of the HPLC analysis of the analysed extract.

triacylglycerols in 15 min, with very simple sample preparation.[32] Chromatogram of every injected sample showed 17 individual triacylglycerol peaks and their concentrations were calculated from peak area. Retention times and molecular parameters of investigated TAGs are given in Table 1. The representative chromatogram of the analysed extract is presented in Figure 1. The major TAG was LLL (trilinolein), followed by LLO (dilinoleoolein), LLP (dilinoleopalmitin), and LOP (linoleooleopalmitin). The concentration of each TAG, depending on the different investigated extraction conditions was as follows: LLL (16.34–23.62 %), LLO (14.61–17.07 %), LLP (10.86–16.82 %) and LOP (11.82–15.44 %). Furthermore, the levels of LnLnLn (trilinolenoin), LnLnL (dilinolenolinolein), LnLO (linolenolinoleoolein), PLP (linoleodipalmitin), OOP (dioleopalmitin), OOO (triolein) and SOP (stearinoleopalmitin) were relatively low (less than 4 %). Similar data for soybean oil TAGs composition was reported previously[32–34] with specific differences due to use of different soybean cultivars.

## PCA

In order to overview the data for similarities and dissimilarities, PCA has been applied on calculated descriptors of studied compounds and resulted in a two-component model that explains 93.61 % of total variance. The first PC explains 78.61 % of the variability, and the second accounts for up to 15.00 %. Score values and the mutual projections of the loading vectors for the first two PCs are presented in Figure 2.

The obtained results show that PC2 separate examined compounds in two big groups, already presented in Table 1. Scores plot revealed that the classification of the studied triacylglycerols was achieved according to presence of the palmitic acid in their structure. Unlike molecules in second series, molecules in first series contain one molecule of palmitic acid in their structure. Compound 17 contains two palmitic acid molecules. The loading plot highlights the most influential descriptors responsible for such compounds order.

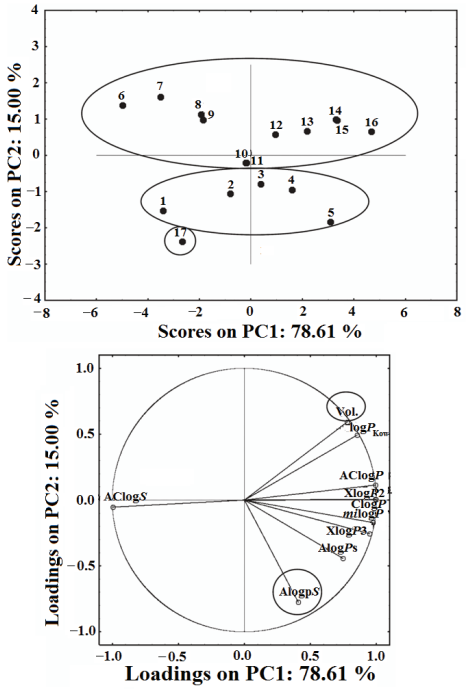


**Figure 2.** Score values and factor loadings of the calculated descriptors for the first two PCs.

Alogp*S* has the highest negative impact on the PC2, while the molar volume expresses the highest positive impact on the mentioned PC. Therefore, PC2 could be considered as a discriminating factor between compounds according to their solubility and molecular size.

## QSRR

In the second step, we focused our efforts on developing the chemometric models which relate the retention characteristics and the structural descriptors of the investigated triacylglycerols. To obtain the quantitative effects of the triacylglycerols molecular structure on their retention behavior QSRR analysis was operated. The regression analyses including non-linear regression for two series of triacylglycerols were carried out. The specifications for the derived mathematical models are shown in Table 2.

The statistical quality of the resulting models, as depicted in Table 2, were determined by squared correlation coefficient ($r^2$), standard error of estimation ($s$) and sequential Fischer test ($F$).[34–37] $F$-value was specified to evaluate the significance of a variable. The higher $F$-value, the more stringent was the significance level. It is noteworthy that all these equations were derived using the entire data set of compounds and no outliers were identified. The $F$-value presented in Table 2 is found statistically significant at 99 % level since all the calculated $F$-values are higher as compared to tabulated values.

Also, all the models show high squared correlation coefficient greater than 0.9600. But, only high correlation coefficient is not enough to select the equation as a model and hence various statistical approaches were used to confirm the robustness and practical applicability of the equations. There are three important

**Table 2.** Statistical parameters of the relationships between the retention time and lipophilicity of the investigated compounds

| Dependent variable | Independent varialble | $y = A + B_1x + B_2x^2$ | | | $r^2$ | $s$ | $F$ | Model |
|---|---|---|---|---|---|---|---|---|
| | | $A$ | $B_1$ | $B_2$ | | | | |
| Series I | | | | | | | | |
| $t_r$ | *mi*log*P* | 14019.00 | −2660.30 | 126.24 | 0.9924 | 0.3393 | 130.60 | 1 |
| | Clog*P* | 479.08 | −43.89 | 1.01 | 0.9990 | 0.1243 | 979.22 | 2 |
| | AClog*P* | 716.12 | −73.28 | 1.89 | 0.9992 | 0.1127 | 1191.30 | 3 |
| | log$P_{Kow}$ | 2356.35 | −218.25 | 5.06 | 0.9985 | 0.1527 | 648.42 | 4 |
| | Xlog*P*2 | 357.68 | −39.98 | 1.13 | 0.9887 | 0.4137 | 87.49 | 5 |
| | Xlog*P*3 | 117.93 | −12.42 | 0.37 | 0.9932 | 0.3213 | 145.70 | 6 |
| | AClog*S* | 675.86 | 111.75 | 4.64 | 0.9985 | 0.1505 | 667.60 | 7 |
| | MV | 5364.23 | −11.53 | 0.01 | 0.9990 | 0.1248 | 971.60 | 8 |
| Series II | | | | | | | | |
| $t_r$ | *mi*log*P* | 6278.50 | −1200.40 | 57.40 | 0.9672 | 0.5194 | 118.10 | 9 |
| | Clog*P* | 307.66 | −28.36 | 0.66 | 0.9942 | 0.2183 | 686.68 | 10 |
| | AClog*P* | 460.51 | −47.05 | 1.21 | 0.9938 | 0.2263 | 639.27 | 11 |
| | log$P_{Kow}$ | 1618.18 | −146.78 | 3.33 | 0.9946 | 0.2115 | 732.28 | 12 |
| | Xlog*P*2 | 343.18 | −37.80 | 1.05 | 0.9942 | 0.2193 | 681.00 | 13 |
| | Xlog*P*3 | 90.49 | −9.78 | 0.28 | 0.9960 | 0.1818 | 993.30 | 14 |
| | AClog*S* | 430.90 | 71.41 | 2.98 | 0.9935 | 0.2307 | 615.24 | 15 |
| | MV | 3622.97 | −7.64 | 0.01 | 0.9943 | 0.2165 | 698.80 | 16 |

**Table 3.** *Cross*-validation parameters of the relationships between the retention time and lipophilicity of the investigated compounds

| PRESS | TSS | PRESS/ TSS | $S_{PRESS}$ | $r^2_{cv}$ | $r^2_{adj}$ | Model |
|---|---|---|---|---|---|---|
| Series I | | | | | | |
| 37.9600 | 30.29 | 1.2500 | 2.7554 | - 0.2532 | 0.9848 | 1 |
| 0.1555 | 30.29 | 0.0051 | 0.1764 | 0.9949 | 0.9980 | 2 |
| 0.1060 | 30.29 | 0.0035 | 0.1456 | 0.9965 | 0.9983 | 3 |
| 0.2436 | 30.29 | 0.0080 | 0.2207 | 0.9920 | 0.9969 | 4 |
| 9.7568 | 30.29 | 0.3221 | 1.3969 | 0.6779 | 0.9774 | 5 |
| 4.6287 | 30.29 | 0.1528 | 0.9622 | 0.8472 | 0.9864 | 6 |
| 0.2340 | 30.29 | 0.0077 | 0.2163 | 0.9923 | 0.9970 | 7 |
| 0.1569 | 30.29 | 0.0052 | 0.1771 | 0.9948 | 0.9979 | 8 |
| Series II | | | | | | |
| 7.5800 | 65.89 | 0.1150 | 0.8300 | 0.8850 | 0.9590 | 9 |
| 0.9941 | 65.89 | 0.0151 | 0.3006 | 0.9849 | 0.9928 | 10 |
| 1.0833 | 65.89 | 0.0164 | 0.3138 | 0.9836 | 0.9922 | 11 |
| 0.9181 | 65.89 | 0.0139 | 0.2889 | 0.9861 | 0.9932 | 12 |
| 0.9471 | 65.89 | 0.0144 | 0.2934 | 0.9856 | 0.9927 | 13 |
| 0.4371 | 65.89 | 0.0066 | 0.1993 | 0.9934 | 0.9950 | 14 |
| 1.1554 | 65.89 | 0.0175 | 0.3241 | 0.9825 | 0.9919 | 15 |
| 0.9774 | 65.89 | 0.0148 | 0.2981 | 0.9852 | 0.9929 | 16 |

components in any chemometric analysis: development of models, validation of models and utility of developed models. Validation is a crucial aspect of any chemometric analysis.[38] For the testing the quality of the predictive power of selected models leave-one-out (LOO) procedure was used (Table 3).

The PRESS value above can be used to compute an $r^2_{CV}$ statistic, called $r^2$ *cross*-validated, which reflects the prediction ability of the model. This is a good way to validate the prediction of a regression model without selecting another sample or splitting data. It is very possible to have a high $r^2$ and a very low $r^2_{CV}$. When this occurs, it implies that the fitted model is data dependent. This $r^2_{CV}$ ranges from below zero to above one. When outside the range of zero to one, it is truncated to stay within this range.

Adjusted *r*-squared ($r^2_{adj}$) is an adjusted version of $r^2$. This parameter shows the statistical significance of incorporated variable in model. Adjustable $r^2$ takes into account the adjustment of conventional correlation coefficient ($r$). Therefore, if an independent variable is added that does not contribute its fair share, the $r^2_{adj}$ will actually decline. Adjustable correlation coefficient is a measure of the percentage explained variation in the dependent variable that takes into account the relationship between the number of cases and the number of independent variable in the regression model. Whereas $r^2$ will always increase when an independent variable is added, adjustable correlation coefficient will decrease if the added variable does reduce the unexplained variation enough the loss of degrees of freedom.

In many cases $r^2_{CV}$ and $r^2_{adj}$ are taken as a proof of the high predictive ability of chemometric models. A high value of these statistical characteristic (> 0.5) is considered as a proof of the high predictive ability of the model. But, recent reports have proved the opposite.[39] Although, the low value of $r^2_{CV}$ for the training set can indeed serve as an indicator of a low predictive ability of a model, the opposite is not necessarily true. Indeed, the high $r^2_{CV}$ does not imply automatically a high predictive ability of the model. Thus, the high value of LOO $r^2_{CV}$ is the necessary condition for a model to have a high predictive power, but it is not a sufficient condition.

The only way to estimate the true predictive power of the models is to test their ability to predict accurately the retention times of the triacylglycerols investigated. To confirm our finding, $t_r$ values were calculated from the selected models 2−4, 7, 8, 10−16, and graphically compared with experimental data (Figure 3). Low scattering of points around the linear relationship, significant slope (> 0.99), and intercept close to zero (< 0.005), indicate very good concurrence between experimental values of retention parameters and values obtained by defined mathematical models. It proves the usefulness of the derived models.

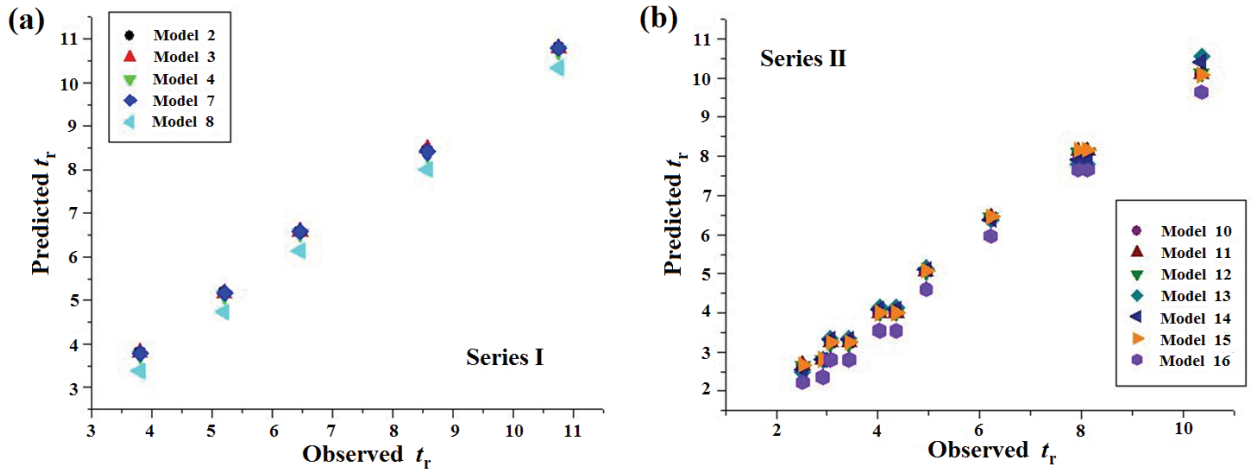


**Figure 3.** Graphs of observed *vs.* predicted $t_r$ values according to the equations: (a) 2−4, 7,8; and (b) 10−16.

Also, on the basis of the magnitude of the individual percentage deviation (IPD %) there is close agreement between observed and calculated retention constants (Table 4, Figure 4). The results of this investigation indicate that these models can be successfully applied in prediction of the retention times of analysed triacylglycerols. The use of chemometric models for prediction of retention behavior of these triacylglycerols reduces cost and time of determination.

As a result of the detailed statistical validation, it can be concluded that model 3 and model 14 have the best statistical performance and should preferably be used in prediction of retention behavior of studied compounds in the applied chromatographic system.

## CONCLUSION

QSRR study has been carried out for training set of 17 TAGs from soybean oil to correlate and predict the HPLC retention time of studied compounds. Soybean oil fractions were obtained by collecting extract at different time intervals during supercritical $CO_2$ extraction at different process parameters. Molecular modeling and QSRR analysis were performed to find the quantitative effects of the lipophilicity of the compounds on their retention behavior. Accurate mathematical models were developed for predicting the HPLC retention time of some TAGs. The validity of the models has been established using LOO *cross*-validation. The established models were used to predict the retention time of the investigated compounds and close agreement between experimental and predicted values was obtained. It indicates the retention time of series of TAGs can be successfully modeled using different lipophilicity descriptors, log*Ps*.

*Acknowledgements.* This work was financially supported by the Ministry of Science, Education and Sport of the Republic of Croatia, project: 073-0730489-0344. The authors are also grateful to the Josip Juraj Strossmayer University of Osijek, Republic of Croatia for financial support. Furthermore, these results are part of the projects No. 172012 and No. 172014 financially supported by the Ministry of Science and Technological Development of the Republic of Serbia, 2011−2014.

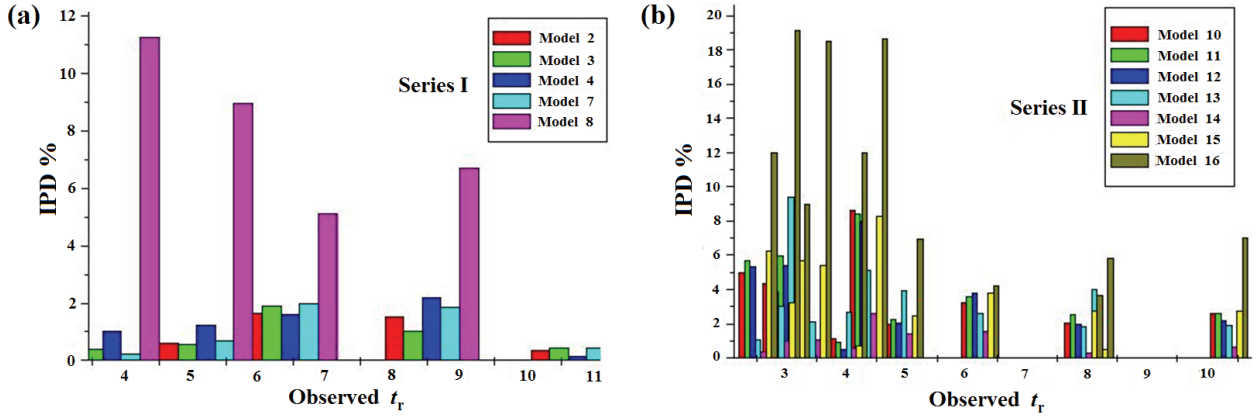


**Figure 4.** Plot of the IPD % values against the experimentally observed $t_r$ values according to equations: (a) 2−4, 7,8; and (b) 10−16.

**Table 4.** Predicted values of the retention time of investigated triacylglicerols

| Series I | Model 2 | | Model 3 | | Model 4 | | Model 7 | | Model 8 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $t_{r\ pred.}$ | IPD % | $t_{r\ pred.}$ | IPD % | $t_{r\ pred.}$ | IPD % | $t_{r\ pred.}$ | IPD % | $t_{r\ pred.}$ | IPD % |
| LnLnP | 3.7946 | 0.14 | 3.8152 | 0.40 | 3.7380 | 1.01 | 3.7920 | 0.21 | 3.3717 | 11.27 |
| LLP | 5.1684 | 0.61 | 5.1719 | 0.54 | 5.1118 | 1.23 | 5.1648 | 0.68 | 4.7348 | 8.95 |
| LOP | 6.5674 | 1.66 | 6.5831 | 1.91 | 6.5385 | 1.60 | 6.5883 | 1.99 | 6.1294 | 5.12 |
| OOP | 8.4410 | 1.51 | 8.4829 | 1.02 | 8.3575 | 2.18 | 8.4098 | 1.87 | 7.9959 | 6.70 |
| SOP | 10.7893 | 0.37 | 10.7982 | 0.45 | 10.7418 | 0.16 | 10.7949 | 0.42 | 10.3403 | 3.81 |

| Series II | Model 10 | | Model 11 | | Model 12 | | Model 13 | |
|---|---|---|---|---|---|---|---|---|
| | $t_{r\ pred.}$ | IPD % | $t_{r\ pred.}$ | IPD % | $t_{r\ pred.}$ | IPD % | $t_{r\ pred.}$ | IPD % |
| LnLnLn | 2.6456 | 4.98 | 2.6623 | 5.65 | 2.6540 | 5.32 | 2.4928 | 1.08 |
| LnLnL | 2.7828 | 4.37 | 2.8006 | 3.76 | 2.7976 | 3.86 | 2.8219 | 3.03 |
| LnLnO | 3.2290 | 5.18 | 3.2523 | 5.94 | 3.2358 | 5.40 | 3.3591 | 9.42 |
| LnLL | 3.2290 | 5.86 | 3.2523 | 5.18 | 3.2358 | 5.66 | 3.3591 | 2.07 |
| LLL | 3.9910 | 1.14 | 3.9919 | 0.95 | 4.0103 | 0.49 | 4.1380 | 2.68 |
| LnLO | 3.9910 | 8.62 | 3.9919 | 8.44 | 4.0103 | 8.02 | 4.1380 | 5.09 |
| LLO | 5.0479 | 1.98 | 5.0616 | 2.25 | 5.0507 | 2.03 | 5.1452 | 3.94 |
| LOO | 6.4207 | 3.23 | 6.4446 | 3.61 | 6.4560 | 3.79 | 6.3808 | 2.58 |
| OOO | 8.1024 | 2.05 | 8.1410 | 2.53 | 8.0985 | 2.00 | 7.7969 | 1.80 |
| LOS | 8.1024 | 0.22 | 8.1410 | 0.26 | 8.0985 | 0.27 | 7.7969 | 3.98 |
| OOS | 10.0930 | 2.58 | 10.0907 | 2.60 | 10.1346 | 2.18 | 10.5562 | 1.89 |

| Series II | Model 14 | | Model 15 | | Model 16 | |
|---|---|---|---|---|---|---|
| | $t_{r\ pred.}$ | IPD % | $t_{r\ pred.}$ | IPD % | $t_{r\ pred.}$ | IPD % |
| LnLnLn | 2.5593 | 0.38 | 2.6764 | 6.21 | 2.2183 | 11.97 |
| LnLnL | 2.8076 | 0.98 | 2.8160 | 3.23 | 2.3524 | 19.16 |
| LnLnO | 3.3176 | 2.35 | 3.2444 | 5.68 | 2.7945 | 8.97 |
| LnLL | 3.3176 | 1.07 | 3.2444 | 5.41 | 2.7945 | 18.53 |
| LLL | 4.0875 | 0.54 | 4.0003 | 0.74 | 3.5458 | 12.01 |
| LnLO | 4.0875 | 2.57 | 4.0003 | 8.25 | 3.5458 | 18.67 |
| LLO | 5.1005 | 1.41 | 5.0713 | 2.45 | 4.6059 | 6.95 |
| LOO | 6.3865 | 1.55 | 6.4572 | 3.81 | 5.9602 | 4.18 |
| OOO | 7.9081 | 0.30 | 8.1581 | 2.75 | 7.6497 | 3.66 |
| LOS | 7.9081 | 1.97 | 8.1581 | 0.47 | 7.6497 | 5.79 |
| OOS | 10.4272 | 0.62 | 10.0798 | 2.70 | 9.6359 | 6.99 |

## REFERENCES


1. G. Brunner, *J. Food Eng*. **67** (2005) 21−33.
2. F. Sahena, I. S. M. Zaidul, S. Jinap, A. A. Karim, K. A. Abbas, N. A. N. Norulaini, and A. K. M. Omar, *J. Food Eng*. **95** (2009) 240−253.
3. J. P. Friedrich and E. H. Pryde, *J. Am. Oil Chem. Soc*. **61** (1984) 223−228.
4. W. E. Artz, T. Kinyanjui, and M. Cheryan, *J. Food Lipids* **12** (2005) 91−102.
5. L. Brühl and B. Matthäus, *Fresenius' J. Anal. Chem*. **364** (1999) 631−634.
6. J. P. Friedrich and G. R. List, *J. Agric. Food Chem*. **30** (1982) 192−193.
7. I. K. Hong, S. W. Rho, K. S. Lee, W. H. Lee, and K. P. Yoo, *Korean J. Chem. Eng*. **7** (1990) 40−46.
8. S. Jokić, S. Svilović, Z. Zeković, S. Vidović, and D. Velić, *Eur. J. Lipid Sci. Technol*. **113** (2011) 644−651.
9. S. Jokić, S. Svilović, Z. Zeković, and S. Vidović, *Int. J. Food Sci. Technol*. **46** (2011) 1031−1037.
10. S. Jokić, B. Nagy, Z. Zeković, S.Vidović, M. Bilić, D. Velić, and B. Simándi, *Food Bioprod. Process*. **90** (2012) 693−699.
11. S. Jokić, S. Vidović, Z. Zeković, S. Podunavac Kuzmanović, L.

Jevrić, and B. Marić, *J. Supercrit. Fluids* **72** (2012) 305−311.
12. M. D. Nodar, A. M. Gómez, and E. Martínez de la Ossa, *Food Sci. Technol. Int.* **8** (2002) 337−341.
13. S. L. Taylor, F. J. Eller, and J. W. King, *Food Research International* **30** (1997) 365−370.
14. A. Jakab, K. Heberger, and E. Forgacs, *J. Chromatogr. A* **976** (2002) 255−263.
15. K. Héberger, *J. Chromatogr. A* **1158** (2007) 273−305.
16. N. U. Perišić Janjić, S. O. Podunavac Kuzmanović, J. S. Balaž, and Đ. Vlaović, *J. Planar Chromatogr*. **13** (2000) 123−129.
17. N. U. Perišić Janjić and S. O. Podunavac Kuzmanović, *J. Planar Chromatogr*. **21** (2008) 135−141.
18. S. Šegan, F. Andrić, A. Radoičić, D. Opsenica, B. Šolaja, M. Zlatović, and D. Milojković Opsenica, *J. Sep. Sci*. **34** (2011) 2659−2667.
19. J. Trifković, F. Andrić, P. Ristivojević, D. Andrić, Ž. Lj. Tešić, and D. M. Milojković Opsenica, *J. Sep. Sci.* **33** (2010) 2619−2628.
20. AOAC, *Official Methods of Analysis,* 17th edition, No 925.40 (2000), Association of Official Analytical Chemists: Washington.
21. S.Vidović, I. Mujić, Z. Zeković, Ž. Lepojević, S. Milošević, and S. Jokić, *J. Am. Oil Chem. Soc*. **88** (2011) 1189−1196.
22. IUPAC, *Standard Methods for the Analysis of Oils, Fats and Derivatives*, 7th edition, No. 2.325 (1986), Blackwell Scientific Publications, Oxford, UK.
23. A. Z. Dudek, T. Arodz, and J. Gálvez, *Comb. Chem. High Throughput Screening* **9** (2006) 213−228.
24. Z. Királyová, K. Šnuderl, F. Kraic, and J. Mocak, *Acta Chimica Slovaca* **1** (2008) 143−152.
25. J. N. Miller and J. C. Miller, *Statistics and Chemometrics for Analytical Chemistry*, 6th edition, Pearson Education Limited, Harlow, UK, 2010, p. 224.
26. K. Šnuderl, M.Simonič, J. Mocak, and D. Brodnjak Vončina, *Acta Chim. Slov*. **54** (2007) 33−39.
27. S. O. Podunavac-Kuzmanović, S. L. Markov, and D. J. Barna, *J. Theor. Comput. Chem.* **6** (2007) 687−698.
28. S. O. Podunavac-Kuzmanović, D. D. Cvetković, and D. J. Barna, *Int. J. Mol. Sci.* **10** (2009) 1670−1682.
29. S. O. Podunavac-Kuzmanović and D. D. Cvetković, *Chem. Ind. Chem. Eng. Q.* **17** (2011) 9−15.
30. S. O. Podunavac-Kuzmanović and D. D. Cvetković, *Chem. Ind. Chem. Eng. Q.* **17** (2011) 33−38.
31. O. Farkas, J. Jakus, and K. Heberger, *Molecules* **9** (2004) 1079−1088.
32. R. Sudar, Z. Jurković, M. Vratarić, A. Sudarić, and T. Duvnjak, *Eur. Food Res. Technol*. **217** (2003) 115−119.
33. W. E. Neff, E. Selke, T. L. Mounts, W. Rinsch, E. N. Frankel, and M. A. M. Zeitoun, *J. Am. Oil Chem. Soc*. **69** (1992) 111−118.
34. M. Holčapek, P. Jandera, P. Zderadička, and L. Hruba, *J. Chromatogr. A* **1010** (2003) 195−215.
35. S. Chaltterjee, A. S. Hadi, and B. Price, *Regression Analysis by Examples*, Wiley VCH, New York, 2000.
36. M. V. Diudea, *QSPR/QSAR Studies for Molecular Descriptors*, Nova Science, Huntingdon, New York, 2000.
37. G. W. Snedecor and W.G. Cochran, *Statistical Methods*, Oxford and IBH, New Delhi, India, 1967.
38. J. G. Topliss and R. P. Edwards, *J. Med. Chem*. **22** (1979) 1238−1244.
39. A. Golbraikh and J. Tropsha, *J. Mol. Graphics Modell*. **20** (2002) 269−276.