# We are IntechOpen,
# the world's leading publisher of
# Open Access books
# Built by scientists, for scientists

## 6,900
Open access books available

## 186,000
International authors and editors

## 200M
Downloads

## 154
Countries delivered to

Our authors are among the

## TOP 1%
most cited scientists

## 12.2%
Contributors from top 500 universities

**CLARIVATE ANALYTICS**
**BOOK CITATION INDEX**
**INDEXED**

**WEB OF SCIENCE™**

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

## Interested in publishing with us?
## Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

# Chemometrics-Based TLC and GC-MS for Small Molecule Analysis: A Practical Guide

Juan Vázquez-Martínez and Mercedes G. López

Additional information is available at the end of the chapter

**Abstract**

Nowadays, thin-layer chromatography (TLC) and gas chromatography/mass spectrometry (GC-MS) instruments can produce more data than even before. At this point, the use of mathematical and statistical tools has provided the key to resolve the information overload. In this chapter, a practical guide is provided for the TLC and GC-MS analysis of short-chain fatty acids (SCFAs), amino acids, and monosaccharides. A methodology for extracting and transforming the chromatographic data to a suitable format for chemometrics is described. Furthermore, a procedure for chemometric analysis based on principal components analysis and clustering analysis is suggested.

**Keywords:** thin layer chromatography, gas chromatography coupled to mass spectrometry, amino acids, short-chain fatty acids, monosaccharides, principal components analysis, clustering analysis

## 1. Introduction

Chemometrics can be defined as the application of mathematical and/or statistical methods to chemical analysis [1]. Like any aspect of mathematics, it requires the use of numbers obtained from measured values of chemical variables [1, 2]. There exist a large number of instrumental techniques for analytical chemistry, which have certain advantages for specific metabolites or matrices. *Grosso modo*, we can mention spectroscopic (infrared, ultraviolet-visible, X-ray, etc.), mass spectrometric, chromatographic, electrochemical and thermal methods, and hybrid techniques [3, 4]. The advances in electronics and computer-assisted data processing have provided powerful instruments that obtain more information than can be analyzed by using traditional data analysis methods. Thus, at this point, chemometrics appears as the preferred choice for the analysis of these complex data [1, 5].

Chromatography comprises the separation techniques based on the partition of the analytes between a mobile phase and a stationary phase. It is extensively used in food industry, pharmaceutical sciences, and natural products science and technology [3, 4] due to its high sensitivity, selectivity, and reproducibility. The versatility of chromatographic techniques allows analyzing a wide range of metabolites, from low-molecular-weight aliphatic gases to complex high-molecular-weight polymeric substances.

In this chapter, we will describe the state-of-the-art in the use of chemometrics-based thin-layer chromatography (TLC) and gas chromatography coupled to mass spectrometry (GC-MS), paying special attention to the analysis of low-molecular-weight metabolites such as amino acids, short-chain fatty acids (SCFAs), and monosaccharides (referred herein as small molecules). In brief, schematic representations of the advantages and some characteristics of TLC and GC-MS for small molecule analysis are shown in **Figure 1**.
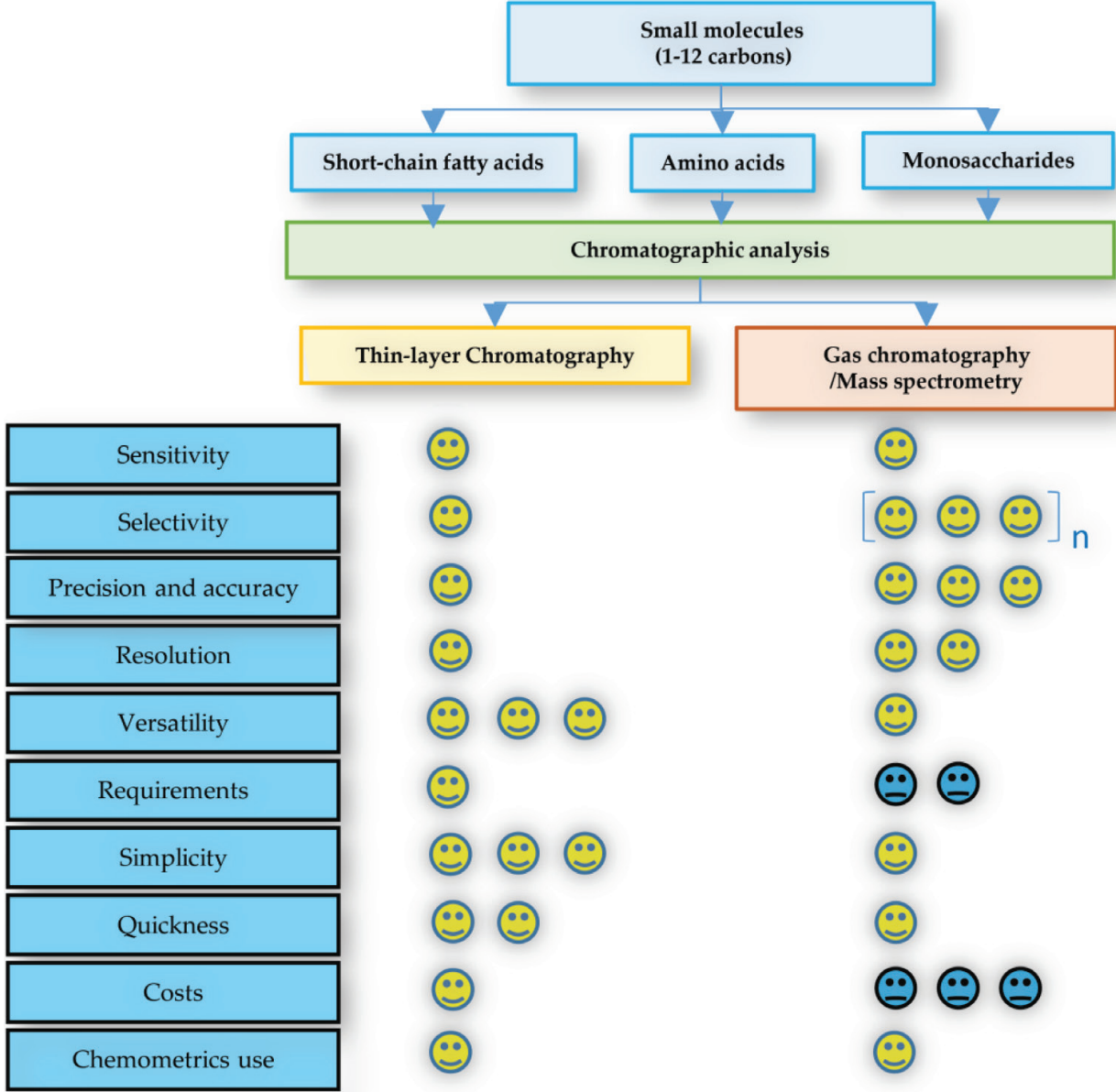


**Figure 1.** Comparison of some characteristics between TLC and GC-MS for small molecule analysis.

## 2. TLC analysis of small molecules

### 2.1. Generalities

TLC is a planar chromatographic technique extensively used due to its rapidity, versatility, and affordable equipment [6]. Bit by bit, it has been relegated as a chromatographic technique for screening tests; however, TLC analysis of some metabolites may represent the best option. The versatility of TLC results from a wide variety of stationary phases used for separation. The common stationary phases include silica with different physical (e.g., pore diameter) and chemical traits (e.g., silanized silica, $C_n$ alkyl-bonded silica) as well as other sorbents such as celluloses, aluminas, and polyamides [6–9]. On the other hand, the improvement of separation of specific metabolites could be largely done by varying the mobile phases from pure nonpolar solvents to complex mixtures of solvents with different polarities [7–9].

The principal measure that can be obtained from a TLC analysis is the retention factor (*Rf*). This parameter is used to describe the migration of components over a TLC plate and is defined as the ratio of the distance traveled by the center of a spot to the distance traveled by the solvent front, both distances are measured from the starting point [7]. By definition,

$$Rf = \frac{\text{component traveled distance}}{\text{mobile phase front traveled distance}} \qquad (1)$$

*Rf* values are always in the range from 0 to 1, or from 0 to 100 if multiplied by 100 (*hRf*) to avoid the decimal point. This parameter is used for component identification, since every compound has a specific *Rf* value for every specific mobile phase. If other compounds comigrate and appear with the same or similar *Rf* value, it is preferable to improve the separation by changing the mobile phase or using a specific visualization reagent [6–8]. Although *Rf* value is very useful for identifying components in TLC and in some cases gathers enough information to qualify and semiquantify a sample [7, 10], it is inadequate for chemometrics-based analysis. For a more precise TLC data handling, it is necessary to complement *Rf* data with numerical values that correspond to the abundance of retained components [10].

An elementary and simple approach for quantitative purpose is the visual comparison of the spot/band intensity of a known sample aliquot with the intensities of a concentration series of known standards, all developed in the same TLC plate. This approach offers semiquantitative results, with precision (expressed as absolute deviation) and accuracy (expressed as experimental error) ranged from 10 to 30% [6–9]. Concentration estimation via visual comparison depends on the interpretation of the analyst. To reduce the interanalyst interpretation, the standard concentration series should be very close to the sample aliquot intensity. Other simple approaches for quantitative TLC determination include measuring the area of the spot or band directly by approximating the spot/band to a regular geometric figure. However, there is still an error range dependent on the precision of the approximation with these approaches.

### 2.2. TLC method development

SCFAs are fatty acids with 2–12 carbon atoms, although other authors differentiate SCFAs with 2–6 carbon atoms [11] from medium-chain fatty acids with 7–12 carbon atoms [11, 12].

These metabolites are produced by dietary fiber fermentation in the colon [13]. Some SCFAs, like butyric acid, have medical relevance since they contribute to colon health by working as anti-inflammatory and energy-source compounds [11–13]. For TLC analysis of SCFAs, it is necessary to consider the autoesterification and autopolymerization reactions of some organic acids because of their dehydration from a dilute aqueous solution, particularly for lactic acid [14]. These reactions are evidenced by the presence of two or more spots in a chromatogram of the respective acid.

The term "amino acids" is used to group the organic compounds that contain amine and carboxyl functional groups, along with side chain group. Some of these compounds are the building units of proteins. Commonly, this term refers only to the 20 amino acids of the genetic code, but there exist about 500 naturally occurring compounds that are chemically amino acids [15, 16]. The TLC techniques to analyze this group of compounds exploit the physicochemical characteristics of the amino and carboxylic moieties, although the separation on a TLC plate is due to the characteristics of the side chain [17–19]. It is noteworthy that two-dimensional TLC development should be required to better separate a complex mixture of more than 10 amino acids [19–21].

Monosaccharides belong to a large family of natural products (i.e., carbohydrates) with the general formula $C_n(H_2O)_{n'}$ the basic structures consisting of five carbons (pentoses) or six carbons (hexoses). They are either polyhydroxyaldehydes or polyhydroxyketones. The *alpha* hydroxyl group of some monosaccharides can be replaced by another substituent such as hydrogen in deoxy sugars and amino group in amino sugars. Furthermore, they can be oxidized to acidic sugars or reduced to polyols [8, 22]. Although they mainly exist as their cyclic hemiacetals or hemiketals, it is necessary to consider the equilibrium of both cyclic and acyclic forms for appropriate chromatographic analysis [8]. Monosaccharides are the most polar compounds of the small molecules here mentioned, and the TLC techniques to analyze this group exploit this feature.

**Table 1** summarizes the condition and procedure for TLC analysis of abovementioned small molecules. Because most of these molecules are colorless and nonfluorescent under ultraviolet and visible light, the use of a derivatization reagent is a must for their visualization.

### 2.3. TLC data treatment

Direct optical quantification in TLC can be realized by using a slit-scanning densitometer. In this technique, the absorbance or emitted fluorescence of the components separated in a TLC plate is measured. According to the compound nature or their derivatives' spectral characteristics, halogen or tungsten (for the visible range) and deuterium lamps (for the UV region) can be used; nevertheless, better results are usually obtained with absorption of UV light on regular layers. On the other hand, on layers with incorporated phosphor, the compound abundance can be quantified upon UV absorption by dark zones on a fluorescent background (fluorescence quenching) [7]. This type of equipment can be computer-controlled, performing automated and accurate data acquisition and processing. At present, commercially available scanning TLC densitometers have common technical characteristics such as spectral range

| Technical parameters | SCFAs | Amino acids | Monosaccharides |
|---|---|---|---|
| Stationary phase | 20 × 20 cm aluminum-backed plates precoated with silica gel 60 (0.2 mm layer) [14] | 20 × 20 cm silica gel 60 plates [7, 18–21] | (1) 10 × 20 cm high-performance thin-layer cellulose plates [8, 25]<br><br>(2) 10 × 20 cm aluminum-backed plates precoated with silica gel 60 [26, 27] |
| Mobile phase | (1) *n*-Hexane-acetone (4:1, *v/v*) for C6 to C8 SCFAs [14]<br><br>(2) Acetone-water-chloroform-ethanol-aqueous ammonia (30:1:3:5: 11, *v/v*) for C2 to C5 SCFAs and lactic acid [14, 23] | Two-dimensional development: (i) *n*-butanol-acetic acid-water (4:1:1, *v/v*), and (ii) phenol-water (75:25, *w/w*) [7, 18–21] | For stationary phase 1: ethyl acetate-pyridine-acetic acid-water (A, composed of 35:15:1:7, *v/v* and B, composed of 35:15:1:9, *v/v*) [25]<br><br>For stationary phase 2: C, composed of *n*-propanol-*n*-butanol-water (12:3:4, *v/v*) and D, composed of ethyl acetate-*n*-propanol-acetic acid-water (4:2:2:1, *v/v*) [25] |
| Chamber and TLC plate's postseparation treatment | Saturate chamber 30 min [14]<br><br>After development, dry TLC plate at room temperature for 24 h or until a complete mobile phase evaporation | Saturate chamber 20 min [18–21]<br><br>After each development, dry TLC plate at 100°C | For stationary phase 1: Saturate 20 min before each respective development<br><br>After development, dry TLC plate at 105°C for 5 min [8, 25, 27]<br><br>For stationary phase 2: Saturate chamber 10 min before each respective development. A two- or three-time development recommended at a proportional distance in TLC plate [26, 28]<br><br>Dry TLC plate at 105°C for 5 min [28] |
| Derivatization agent and visualization | 0.25 g of methyl red and 0.25 g of bromophenol blue in 100 mL of 70% methanol<br><br>Derivatized SCFAs (from two to eight carbons) appeared as blue spots, lactic acid as pink-red spot [14, 23] | 0.25% ninhydrin in acetone [7, 18]<br><br>Optional double derivatization: 0.20% picric acid acetone and 0.25% ninhydrin acetone [24]<br><br>Heat TLC plate at 110°C for 10 min | For stationary phase 1: 1.6 g of *O*-phthalic acid in 100 mL of water-saturated *n*-butanol, containing 0.9 mL of aniline [25]<br><br>For stationary phase 2: (i) 250 mg of orcinol dissolved in 95 mL of ethanol and 5 mL of sulfuric acid; or (ii) 4 g of diphenylamine, 4 mL of aniline, and 20 mL of 85% phosphoric acid in 180 mL of acetone<br><br>Afterward, dry TLC plate at room temperature and heat it at 85–100°C for 5–15 min [26, 28] |
| Limit of detection (LOD)* | 8 µg [14, 23] | 0.05–0.1 µg [18]<br><br>0.04–0.08 µg [24] | For aniline-*O*-phthalic acid reagent: 10 µg [8]<br><br>For aniline-diphenylamine-phosphoric acid reagent: 10 ng (fluorescence) [8] |

*LOQ is calculated by multiplying LOD by 3.333 [29].

**Table 1.** Technical parameters for TLC analysis of small molecules.

(190–900 nm), data step resolution (25–200 µm), and scanning speed (1–100 mm/s) that translate in a more representative, accurate, and precise data matrix. The principal disadvantage of these densitometers could be the cost of the instrument itself (30,000–60,000 USD) and of its spare parts [30].

A TLC densitometer chromatogram consists of one axis representing the *Rf* values and the other representing the measured absorbance or fluorescence, which can be extracted as a two-column matrix for posterior chemometric analysis. A well-resolved component is characterized by a well-shaped (taller-than-wide) and normally distributed peak in the chromatogram, typically leading to RSD values in the range of 0.5–3% in quantitative high-performance TLC, using the peak area corresponding directly to the compound concentration [7, 10].

Apart from slit-scanning densitometry, a compound can be quantified based on the analysis of its spot/band image obtained from a TLC plate. The instrument used for this purpose is known as video scanner densitometer, and it is certainly coupled to a computer system. For chromatogram data acquisition, image densitometers obtain a picture of the TLC and subsequently measure the color brightness of visible spots. Commonly, the image is obtained under white light and/or UV light (short-wave and/or long-wave radiation) [30]. Since the first step for this technique is the obtaining of a good-resolution colorful image, homemade instruments equipped with a high-quality digital camera can be adapted for image acquisition.

The disadvantages of image densitometers are related to their lower sensitivity and chromatographic resolution than slit-scanner densitometers. As a result, the data matrix loses a lot of information of the components that cannot be detected under a few wavelengths of light applied and may contain more background interference. However, due to their simplicity and lower cost, these instruments are the most popular for densitometric evaluation nowadays [7, 30].

## 3. GC-MS analysis of small molecules

### 3.1. Generalities

Basically, GC is a more sophisticated technique than TLC; for a better understanding, some references are suggested [3, 4, 31]. This is the method of choice for separation and detection of permanent inorganic gases and volatile organic compounds in a mixture. It is based on the partitioning of vaporized or gaseous compounds between an inert gas mobile phase and a stationary liquid or solid phase. Helium is the most commonly used carrier gas, but others such as nitrogen or hydrogen can be used. The separation column is packed with a finely divided solid or coated with a thin film of liquid (typically <1 µm). In the market, there is a wide variety of capillary columns, which can be grouped by the polarity of their stationary phases, for achieving better separation and resolution [31].

The affinity of the components of a mixture for the stationary phase depends on their physicochemical characteristics and it impacts directly on their separation and resolution.

Components' separation is based on the "like-dissolves-like" rule that explains the different interaction strengths between the compounds and stationary phase. A stronger compound-stationary phase interaction provokes a longer compound-stationary phase contact, and more time is needed for compounds' migration through the column. This migration time is known as retention time $Rt$ (commonly expressed in minutes) and represents one of the principal numeric values that can be obtained by GC for compound characterization. The other value also obtained by GC is the relative abundance of the components that corresponds with the height or area of the respective chromatogram peak. The unit of the relative abundance depends on the type of detector coupled to GC [4, 31].

GC-MS is a hybrid technique, in which a gas chromatograph is coupled to a mass spectrometer via an interface (i.e., a heated metal tube equipped with a temperature controller, connecting the column exit in the gas chromatograph and the entrance to the ion source of the mass spectrometer). GC alone can separate volatile compounds with great resolution, but it cannot identify them properly [4]. MS uses the difference in mass-to-charge ratio ($m/z$) of ionized atoms or molecules, providing structural information by identification of distinctive fragmentation patters. Thus, after separation in the GC column, analytical species are transported to the mass spectrometer to be ionized for subsequent mass filtration and detection. GC-MS is a potent tool for modern analytical chemistry that allows separating the compounds in complex mixtures and identifying them effectively with some considerations [3, 31].

Theoretically, MS is based on the analysis of ions moving through a vacuum, and a mass spectrometer must include the ion source (electron or chemical ionization), ion analyzer (quadrupole, ion trap, or time-of-flight), and ion detector [3, 31]. For GC-electron impact-quadrupole-mass spectrometry (GC-EIMS), immediately after the compounds leave the capillary column, they are bombed by an electron beam and fragmented in ions that correspond to molecule fractions. Then, these ions are separated in the quadrupole according to their mass-to-charge ratio ($m/z$) before being further sensed and quantified by the detector.

A GC-EIMS chromatogram (named total ion chromatogram or TIC) is a graph showing the relationship between the retention time ($x$-axis) and ion abundance or total ion current data ($y$-axis). Since the TIC comes from the convolution of the individual abundances of all the monitored ions, it can be deconvoluted to obtain each ion distribution. It can be also displayed in three dimensions simultaneously (i.e., a 3D chromatogram recording the number of ions created along with their masses over time). By examining TIC and "slicing" along the third dimension ($m/z$) of a chosen peak, the mass spectrum can be evaluated at a given time. When all of these ions come from the same compound, the ion distribution can be plotted in an abundance *versus* $m/z$ graph (aka the compound mass spectrum or fragmentation pattern). In the case that mass spectra are obtained under standard conditions, the distribution of ions is always the same independently of GC step and/or MS instrument. This important feature has enabled constructing big mass spectra libraries that serve for precise compound identification [31].

The first step to analyze small molecules or any metabolite by GC-EIMS is to make sure that all analytes can be volatilized under the injection port and oven temperatures. It is known that the vaporization of a specific compound occurs at a given temperature and pressure that, at the time, depends on the number of carbons and the polarity of the compounds (among other

properties). Amino acids, SCFAs, and monosaccharides are polar compounds due to their amino, hydroxyl, carbonyl, and carboxylic groups. These functional groups allow intramolecular interactions such as hydrogen and van der Waals bonds, increasing the compounds' boiling point and making their analysis by GC-EIMS difficult. Thus, only some of these compounds can be GC-EIMS analyzed directly, and for the remaining ones, a derivatization step must be required for better analysis.

## 3.2. GC-MS method development

### 3.2.1. Direct GC-MS analysis of small molecules

Among small molecules (i.e., amino acids, monosaccharides, and SCFAs), SCFAs are suitable for direct GC-EIMS analysis due to their small aliphatic acyl chain and relatively low vapor pressure [32]. The critical issue in analyzing SCFAs directly is the correct stationary phase selection. Excellent results could be obtained by using a polar-phase capillary column such as Nukol (acid-modified poly-ethylene-glycol phase, Supelco, Sigma-Aldrich). Since these compounds do not have more than 12 carbons, neither a large column nor a long chromatographic method is required. For direct GC-EIMS analysis of SCFA, the chromatographic conditions are as follows: injection port temperature set at 250°C; GC oven temperature initially set at 90°C for 3 min, then subjected to a three-step program [(i) increased to 150°C with 15°C/min ramp rate, (ii) increased to 170°C with 5°C/min ramp rate, and (iii) increased to 200°C with 20°C/min ramp rate and hold for 10 min]; transfer line temperature set at 250°C; stationary phase: a 30 m × 320 μm × 0.25 μm Nukol capillary column (Supelco, Sigma-Aldrich); carrier gas: a constant helium flow of 1 mL/min; standard MS parameters applied (electron energy of 70 eV, ion source temperature set at 230°C, quadrupole analyzer temperature set at 150°C). This method takes around 16 min for analysis of C2 to C12 SCFAs, with measurements obtained at a 25–300 $m/z$ range and approximately 3 scans per second.

### 3.2.2. Derivatization of small molecules for GC-EIMS analysis

The goal of derivatization before GC-MS analysis is to obtain chemical derivatives being more volatile and less reactive than the interest compounds, thus presenting improved chromatographic characteristics [29, 33].

The labile hydrogens of amino acids, SCFAs, and monosaccharides are commonly the target of derivatization procedures. In practice, these hydrogens of the carboxyl, amino, and hydroxyl groups of the abovementioned small molecules can be substituted by trimethylsilyl groups [33] according to the following general derivatization.

First, it is needed to prepare a completely solvent-free sample since water and any solvent (as trace impurities) with labile hydrogens can interfere with the derivatization. To each 5–50 mg of sample, 20 μL of pyridine and 80 μL of $N,O$-bis(trimethylsilyl)trifluoroacetamide supplied with 1% of trimethylchlorosilane (BSTFA + TMCS) are added. For pure standards, a final concentration of 0.5 μg/μL is prepared. Then mixed and heated at 85°C for 15–30 min. After that, pyridine and BSTFA are evaporated under nitrogen flux of the sample; a known volume of

isooctane (ca. 200 µL) is subsequently added and 1 µL of this resultant solution can be injected into the GC-MS system in split or splitless mode.

The following GC-EIMS conditions could enable the individual and simultaneous analysis of amino acids, SCFAs, and monosaccharides. The injector temperature is set at 260°C. An HP-5-MS capillary column (30 m × 25 µm × 0.25 µm) is used with helium as carrier gas at a constant flow rate of 1 mL/min. The oven program begins at 45°C (hold for 5 min), then increases at a rate of 10°C/min until 300°C (hold for 25 min). The transfer line temperature is set at 280°C. The mass spectrometer operates at 70 eV of electron energy; the quadrupole and ion-source temperatures are set at 150 and 230°C, respectively. The scan mode is used in the range 40–550 $m/z$. Using this method, SCFAs, amino acids, and monosaccharides can be chromatographically separated if their mixture contains appropriate amounts. SCFAs elute from the GC column during the first 10–15 min of analysis, followed by amino acids and monosaccharides later. Due to the interference of solvent and derivatizing reagent traces, this method is not suitable for the analysis of compounds containing less than three to four carbons, for example, for C2-C3 SCFAs, a direct analysis is preferable.

For amino acids, it is important to consider that in the equilibrium, different forms of derivatization products can be found, i.e., the totally derivatized compound (i.e., all the labile hydrogens are derivatized) and partially derivatized compound (i.e., conserving one or more labile hydrogens). If the same conditions are applied and the equilibrium is reached in all samples, the proportion of totally and partially derivatized amino acids remains constant and they are both commonly included in the mass spectra libraries.

A critical point for the GC-EIMS analysis of monosaccharide silyl derivatives is the presence of natural isomers. D-Hexoses naturally have five isomers: two furanoses, two pyranoses, and one linear structure, and their silyl derivatives possess the same isomers too. These derivatives can be separated by GC and the mass spectra of pyranoses and furanoses can be differentiated [34]. In a comparable way to amino acids, in the equilibrium, the proportion of each isomer remains constant. It is important to note that in TLC analysis, the problem of isomers generation after derivatization process is avoided.

For better GC-EIMS results, the use of internal standards is advisable, for example, synthetic methylated SCFAs, synthetic or nonprotein amino acids, and nonbiological polyols or glycosides, as well as carbon- and hydrogen-labeled compounds [32, 35, 36].

### 3.3. GC-MS data treatment

Nowadays, the extraction of GC-MS data is a simple and fast task, thanks to convenience and availability of the commercial and free software products. For GC-MS data analysis of small molecules, ChemStation Data Analysis and MassHunter software products (Agilent Technologies, Inc.) are commonly used [29].

Using these software packages, TIC data could be extracted to CSV format, that is, a two-column data matrix is obtained, in which are stored the retention time and ion relative abundance values.

Since TIC data contain the distribution of all measured ions, the data of each individual ion can be extracted and merged to obtain a consensus chromatogram that includes the peaks, in which are present all the selected ions. This feature is very useful when characteristic ions have been detected for the compound(s) of interest. For example, for silylated monosaccharides, the 204 and 217 $m/z$ ions can be used as a diagnostic tool and by monitoring both ions a complex chromatogram can be simplified to analyze only monosaccharides.

## 4. Chemometrics-based chromatographic data analysis

At this point, it can be recognized that the data obtained by TLC and GC-EIMS are in the same format, that is, a numeric matrix for subsequent chemometric analysis.

The following paragraphs are considered a practical guide as more than a theoretical description of the chemometric concepts that can be widely reviewed in Refs. [1, 2, 5, 37]. It describes a workflow for obtaining a chemometric analysis of the TLC and GC-MS data (**Figure 2**), which can be divided into three distinct stages: data preprocessing, data processing, and model validation.

### 4.1. Data preprocessing

This stage includes the baseline correction, retention time or retention factor correction, and noise reduction [38]. It can be applied before or after numeric data extraction. For a prior data extraction, the software products ImageJ, winCATC, or VisionCATS can be used for TLC analysis, whereas OpenChrom, ChemStation Data Analysis, or MassHunter Workstation can be used for GC-MS analysis. For postdata extraction, the data matrices can be edited with statistical software products such as R or MATLAB. Normalization of data is recommended to minimize systemic variation in the data due to changes in instrumental response; this can be done by using internal standards and a subsequent data correction. A data cleanup step can be added to remove artifact peaks or peaks with low repeatability; this can be done by deleting suspected peaks. After doing this, the data must be presented in a single table all under the same conditions and sifted through the same filters.

### 4.2. Data processing and model validation

Data can be processed by unsupervised [principal component analysis (PCA) and clustering analysis] and supervised multivariate statistical methods [partial least square discriminant analysis (PLS-DA) and between-group analysis (BGA)] [39].

Regarding unsupervised methods, chromatographic data can be analyzed by the visualization of grouping trends and inspection of atypical values using PCA [38, 40]. A way to do a good PCA analysis is to utilize the FactoMineR package [41] in the R software, using R studio as an interface [42, 43] and considering eigenvalues to obtain suitable plots.

Clustering analysis uses resemblance or dissemblance measures between the samples to be analyzed. The goal of this analysis is to obtain a symbolic description of the data and an identification pattern [44]. The most commonly used clustering algorithm is the hierarchical
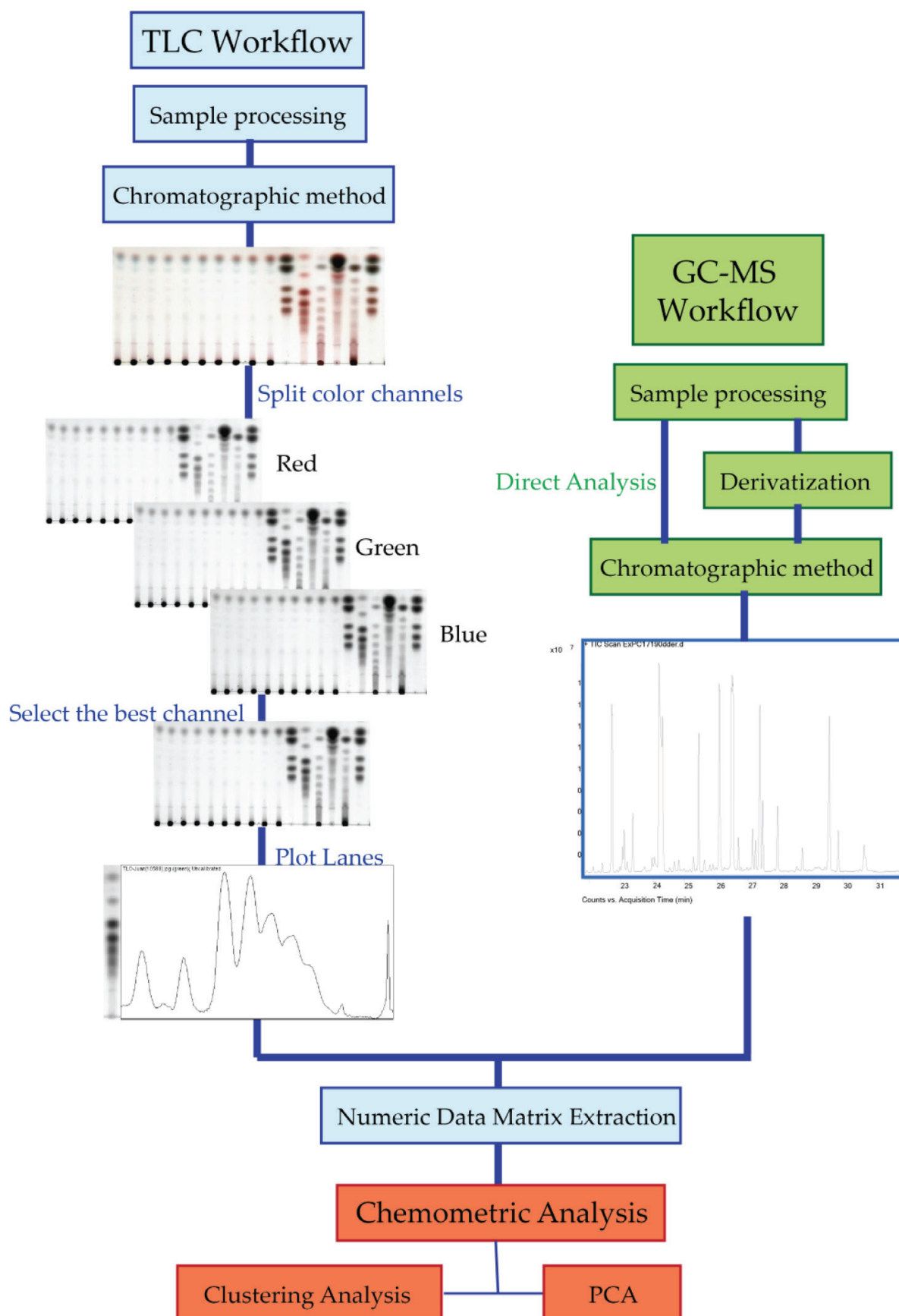
**Figure 2.** Workflow for chemometric analysis of TLC and GC-MS data.

method [44, 45]. For a hierarchical clustering procedure, the pvclust package in R could be used to provide approximately unbiased and bootstrap probability $p$ values [46]. Other R packages could be also used to improve the appearance and analysis of the dendrograms [47]. In addition, model validation can be performed by an internal or external method [48].

### 4.3. Chemometrics-based TLC and GC-MS analysis in practice

#### 4.3.1. Example of chemometrics-based TLC analysis

This example illustrates a TLC method using image analysis densitometry but with some consideration is also applicable to slit-scanning densitometry too. Four extracts of *Agave* (S2, S3, and S4) and *Cichorium* (S1) species containing considerable amounts of carbohydrates (mainly fructans) were chromatographed and derivatized by using diphenylamine-aniline-phosphoric acid reagent. Using this derivatization step, fructose and fructans appear as red-dish spots, whereas glucose and maltooligosaccharides as bluish ones. After chromatographic development and visualization, the acquisition of an image of the TLC plate was done by using a commercial image densitometer TLC-Visualizer 2 (CAMAG), as shown in **Figure 3**. Once acquired, the TLC image was processed in the free software ImageJ (National Institutes of Health, NIH), and it was split into red, green, and blue channels. For this example, the red and blue channels were selected because the spots of interest were enriched and appear more defined in these channels. For noise removal, a median filter with an appropriate resolution
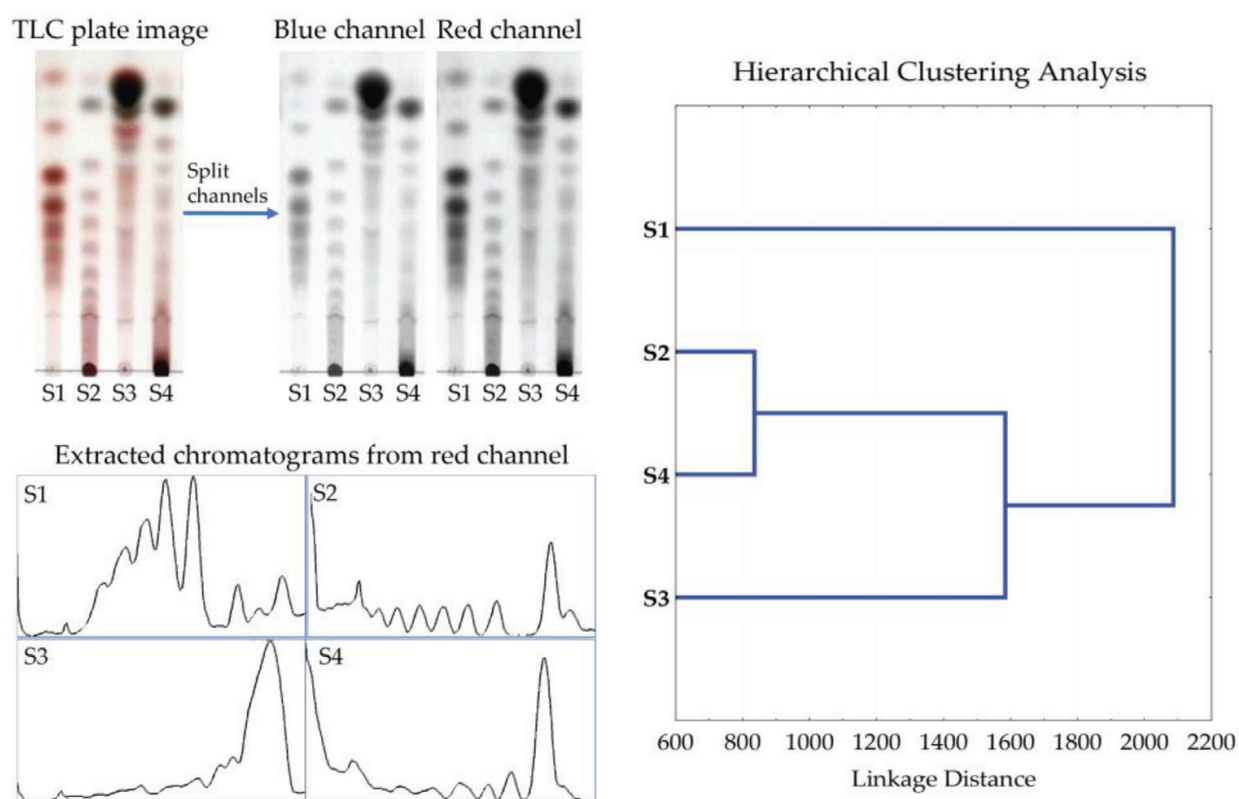


**Figure 3.** Hierarchical clustering analysis of oligosaccharide composition of four plant extracts: a TLC plate image with four samples (S1–S4) derivatized with diphenylamine-aniline-phosphoric acid reagent, the split color channel into blue and red, and the extracted chromatograms obtained for the red channel lanes.

was applied. A track containing the component(s) of interest was selected, designated as lane, and then plotted. The numeric data from each plot could be extracted to select exclusively the area of the plot line (avoid selecting any text or another data). In this study, a Cartesian graph was also obtained to extract single numeric data matrix for further chemometric analysis [49–51]. Besides the use of retention factor/relative abundance matrix, the relative peak area of all or part of the mixture components could also be used to construct data matrices.

The data matrix was used to construct a hierarchical clustering analysis with the software STATISTICA (StatSoft, Inc.), as displayed in **Figure 3**. This analysis allows to group the samples according to their monosaccharide and oligosaccharide composition without identification requirement. It is observed that the most closely related samples are S2 and S4, which form a clade and have the smallest distance among all the samples. S3 appears as sister of the S2–S4 clade, but with a distance equivalent to four times the distance between S2 and S4. S1 appears as an external group of the clade formed by S2–S4–S3, having a greater distance as compared to the other samples. According to this dendrogram, it is possible to conclude that based on the carbohydrate composition, S2 and S4 have roughly similar compositions among all the samples, S1 has the most different composition, and S3 has a composition most similar to S2–S4 than S1; this corresponds with the vegetal origin of the samples. It is observable that different polymerization degrees of the carbohydrates correspond with *Rf* values, that is, the higher the *Rf* value, the lower the polymerization degree. Thus, besides the grouping according to the type of carbohydrate (spots' color), the classification is also influenced by the size of carbohydrate molecules. The advantage of using chemometrics to analyze these TLC data consists of avoiding wasting time in compound identification and standard-based quantification, making the grouping and sample classification a fast but robust process.

### 4.3.2. Example of chemometrics-based GC-MS analysis

To demonstrate the application of the GC-MS methods herein described, a dataset obtained from the monosaccharide analysis of plant tissue extracts by GC-EIMS was studied. A total of nine samples obtained from three tissues by triplicate measurement (A = tissue 1, B = tissue 2, and C = tissue 3) were analyzed, and the total ion chromatograms were processed to extract the 204 and 217 *m/z* ion chromatograms with the aim of filtering data for silylated monosaccharides. After that, the chromatographic data were extracted and the numeric data matrix was subjected to principal component analysis. The resulting PCA data are shown in **Figure 4**; principal component 1 (PC1) and principal component 2 (PC2) explain 61.6 and 16.3% of the total variance, respectively. It can be observed that all samples are grouped according to the tissue from which they come. Samples of tissue B are closer, and samples of tissue C are more scattered. For B and C, the PC1 has a major effect on the data dispersion, whereas for A, the PC2 does the same thing. This nondirected analysis, in which no compounds have been identified, allowed grouping the data according to their origin, demonstrating the strength of the chemometric analysis. Although, in this example, only nine samples are used and their origin is known, this methodology has sufficient sturdiness to be applied to datasets of tens or even hundreds of samples with diverse sources as long as the same methodology is applied to all samples. In this chemometric analysis, our samples were grouped according to their monosaccharide content in a simple way, as well as for the above TLC chemometric analysis, avoiding wasting time in compound identification and standard-based quantification.
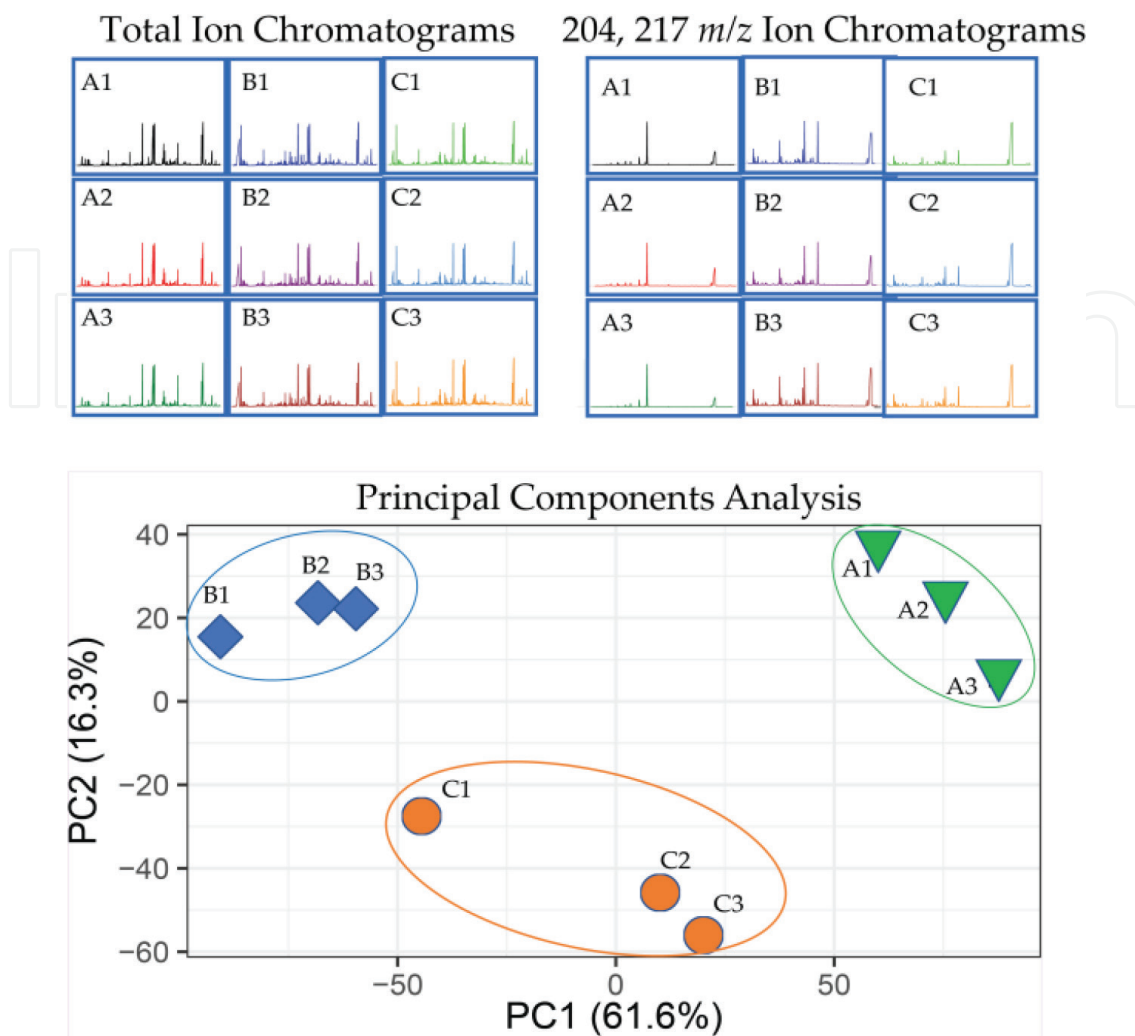
**Figure 4.** Chemometric analysis of GC-EIMS data of monosaccharides in plant tissues: the GC-EIMS total ion chromatograms (TICs), chromatograms obtained from the TIC extraction of the 204 and 217 *m/z* ions, and PCA constructed with the numeric data obtained from extracted chromatograms. Letter labels correspond to the tissue type and a label numeric fraction represents the sample replicate.

## 5. Conclusions

This chapter represents a practical guide for chemometric analysis of amino acids, short-chain fatty acids, and monosaccharides by thin-layer chromatography and gas chromatography coupled to mass spectrometry in complex mixtures. Furthermore, it provides a workflow to convert chromatographic data to numerical values, for which are described some chemometric analysis methods such as principal component analysis and clustering analysis.

## Conflict of interest

Authors declare no conflict of interest.

## Author details

Juan Vázquez-Martínez and Mercedes G. López*

*Address all correspondence to: mercedes.lopez@cinvestav.mx

Center for Research and Advanced Studies of IPN Campus Irapuato (Centro de
Investigación y de Estudios Avanzados del IPN Unidad Irapuato), Guanajuato, Mexico

## References

[1]  Deming SN. Chemometrics: An overview. Clinical Chemistry. 1986;**32**:1702-1706

[2]  Roussel S, Preys S, Chauchard F, Lallemand J. Multivariate data analysis (chemometrics). In: O'Donnell CP, Fagan C, Cullen PJ, editors. Process Analytical Technology for the Food Industry. 1st ed. New York, USA: Springer-Verlag; 2014. pp. 7-17

[3]  Gallo M, Ferranti P. The evolution of analytical chemistry methods in foodomics. Journal of Chromatography A. 2016;**1428**:3-15

[4]  Siddiqui MR, AlOthman ZA, Rahman N. Analytical techniques in pharmaceutical analysis: A review. Arabian Journal of Chemistry. 2017;**10**:S1409-S1421

[5]  Guidetti R, Beghi R, Giovenzana V. Chemometrics in food technology. In: Varmuza K, editor. Chemometrics in Practical Applications. 1st ed. Rijeka, Croatia: InTech; 2012. pp. 2017-2052

[6]  Ciura K, Dziomba S, Nowakowska J, Markuszewski MJ. Thin layer chromatography in drug discovery process. Journal of Chromatography A. 2017;**1520**:9-22

[7]  Sherma J, Fried B, editors. Handbook of Thin Layer Chromatography. 1st ed. New York, USA: Marcel Dekker, Inc.; 2003. 997 p

[8]  Maloney MD. Carbohydrates. In: Sherma J, Fried B, editors. Handbook of Thin Layer Chromatography. 1st ed. New York, USA: Marcel Dekker, Inc.; 2003. pp. 579-610

[9]  Sherma J. Biennial review of planar chromatography: 2013-2015. Journal of AOAC International. 2016;**99**:323-331

[10] Stroka J, Spangenberg B, Anklam E. New approaches in TLC-densitometry. Journal of Liquid Chromatography and Related Technologies. 2002;**25**:1497-1513

[11] Schönfled P, Wojtczak L. Short- and medium-chain fatty acids in energy metabolism: The cellular perspective. Journal of Lipid Research. 2016;**57**:943-954

[12] Nishi Y, Hiejima H, Hosoda H, et al. Ingested medium-chain fatty acids are directly utilized for the acyl modification of ghrelin. Endocrinology. 2005;**146**:2255-2264

[13] Koh A, De Vadder F, Kovatcheva-Datchary P, Bäckhed F. From dietary fiber to host physiology: Short-chain fatty acids as key bacterial metabolites. Cell. 2016;**165**:1332-1345

[14] Pyka A, Bober K. Visualizing agents for short-chain fatty acids in TLC. Journal of Planar Chromatography—Modern TLC. 2005;**18**:145-150

[15] Xie J, Schultz PG. Adding amino acids to the genetic repertoire. Current Opinion in Chemical Biology. 2005;**9**:548-554

[16] Wagner I, Musso H. New naturally occurring amino acids. Angewandte Chemie International Edition. 1983;**22**:816-828

[17] Mohammad A, Moheman A, El-Desoky GE. Amino acid and vitamin determinations by TLC/HPTLC: Review of the current state. Central European Journal of Chemistry. 2012;**10**:731-750

[18] Bhawani SA, Mohamad-Ibrahim MN, Salaiman O, et al. Thin-layer chromatography of amino acids: A review. Journal of Liquid Chromatography and Related Technologies. 2012;**35**:1497-1516

[19] Sleckman BP, Sherma J. A comparison of amino acid separations on silica gel, cellulose, and ion exchange thin layers. Journal of Liquid Chromatography. 1982;**5**:1051-1068

[20] Dolowy M, Pyka A. Application of TLC, HPLC and GC methods to the study of amino acid and peptide enantiomers: A review. Biomedical Chromatography: BMC. 2014;**28**:84-101

[21] Dale T, Court W. Improved separation of amino acids by thin-layer chromatography. Chromatographia. 1981;**14**:617-620

[22] Danishefsky SJ, DeNinno MP. Totally synthetic routes to the higher monosaccharides. Angewandte Chemie International Edition English. 1987;**26**:15-23

[23] Lee KY, So JS, Heo TR. Thin layer chromatographic determination of organic acids for rapid identification of bifidobacteria at genus level. Journal of Microbiological Methods. 2001;**45**:1-6

[24] Sinhababu A, Basu S, Dey H. Modified ninhydrin reagents to detect amino acids on TLC plates. Research on Chemical Intermediates. 2015;**41**:2785-2792

[25] Cheong K, Wu D, Hu D, et al. Comparison and characterization of the glycome of *Panax* species by high-performance thin-layer chromatography. Journal of Planar Chromatography—Modern TLC. 2014;**27**:1-5

[26] Mancilla-Margalli NA, López MG. Water-soluble carbohydrates and fructan structure patterns from *Agave* and *Dasylirion* species. Journal of Agricultural and Food Chemistry. 2006;**54**:7832-7839

[27] Ferey J, Da Silva D, Bravo-Veyrat S, et al. Validation of a thin-layer chromatography/densitometry method for the characterization of invertase activity. Journal of Chromatography A. 2016;**1477**:108-113

[28] Waksmundzka-Hajnos M, Sherma J, Kowalska T, editors. Thin Layer Chromatography in Phytochemistry. 1st ed. Boca Raton, USA: CRC Press; 2008. pp. 255-276

[29] Gutierrez-Villagomez JM, Vázquez-Martínez J, Ramírez-Chávez E, et al. Analysis of naphthenic acid mixtures as pentafluorobenzyl derivatives by gas chromatography-electron impact mass spectrometry. Talanta. 2017;**162**:440-452

[30] Hess AVI. Digitally enhanced thin-layer chromatography: An inexpensive, new technique for qualitative and quantitative analysis. Journal of Chemical Education. 2007;**84**:842-847

[31] Settle F, Pleva M, editors. Gas Chromatography-Mass Spectrometry, a Knowledge Base. 1st ed. Vol. 12. Amsterdam: Elsevier Science Publishers B.V.; 1989. 287 p

[32] García-Villalba R, Giménez-Bastida JA, García-Conesa MT, Tomás-Barberán FA, Espín JC, Larrosa M. Alternative method for gas chromatography-mass spectrometry analysis of short-chain fatty acids in faecal samples. Journal of Separation Science. 2012;**35**:1906-1913

[33] Orata F. Derivatization reactions and reagents for gas chromatography analysis. In: Ali Mohd M, editor. Advanced Gas Chromatography—Progress in Agricultural, Biomedical and Industrial Applications. 1st ed. Rijeka, Croatia: InTech; 2012. pp. 83-108

[34] Ruiz-Matute AI, Hernández-Hernández O, Rodríguez-Sánchez S, Sanz ML, Martínez-Castro I. Derivatization of carbohydrates for GC and GC-MS analyses. Journal of Chromatography B. 2011;**879**:1226-1240

[35] Grimm F, Fets L, Anastasiou D. Gas chromatography coupled to mass spectrometry (GC–MS) to study metabolism in cultured cells. In: Koumenis C, Coussens L, Giaccia A, Hammond E, editors. Tumor Microenvironment. Advances in Experimental Medicine and Biology. Vol. 899. Cham: Springer; 2016. pp. 59-88

[36] Zweckmair T et al. Improved quantification of monosaccharides in complex lignocellulosic biomass matrices: A gas chromatography-mass spectrometry based approach. Carbohydrate Research. 2017;**446**:7-12

[37] Brereton RG. Applied Chemometrics for Scientists. Chichester, UK: John Wiley & Sons, Ltd; 2007. 379 p

[38] Yip LY, Yong Chan EC. Gas chromatography/mass spectrometry-based metabonomics. In: Issaq HJ, Veenstra TD, editors. Proteomic and Metabolomic Approaches to Biomarker Discovery. 1st ed. Waltham, USA: Academic Press; 2013. pp. 131-144

[39] Wallace IM, Higgins DG. Supervised multivariate analysis of sequence groups to identify specificity determining residues. BMC Bioinformatics. 2007;**8**:1-12

[40] Pereira AC, Reis MS, Saraiva PM, Marques JC. Analysis and assessment of Madeira wine ageing over an extended time period through GC-MS and chemometric analysis. Analytica Chimica Acta. 2010;**660**:8-21

[41] Lê S, Josse J, Husson F. FactoMineR: An R package for multivariate analysis. Journal of Statistical Software. 2008;**25**:1-18

[42] R-Core-Team. R: A Language and Environment for Statistical Computing [Internet]. 2015. Available from: https://www.r-project.org/ [Accessed: March 15, 2018]

[43] RStudio-Team. RStudio: Integrated Developmet for R [Internet]. 2016. Available from: https://www.rstudio.com/ [Accessed: March 15, 2018]

[44] Diday E, Simon C. Clustering analysis. In: Fu KS, editor. Digital Pattern Recognition. 2nd ed. Berlin, Germany: Springer-Verlag; 1980. pp. 47-92

[45] Messai H, Farman M, Sarraj-Laabidi A, Hammami-Semmar A, Semmar N. Chemometrics methods for specificity, authenticity and traceability analysis of olive oils: Principles, classifications and applications. Food. 2016;**5**:1-35

[46] Suzuki R, Shimodaira H. Pvclust: An R package for assessing the uncertainty in hierarchical clustering. Bioinformatics. 2006;**22**:1540-1542

[47] Galili T. Dendextend: An R package for visualizing, adjusting and comparing trees of hierarchical clustering. Bioinformatics. 2015;**31**:3718-3720

[48] Enot DP, Hass B, Weinberger KM. Bioinformatics for mass spectrometry-based metabolomics. In: Mayer B, editor. Bioinformatics for Omics Data—Methods and Protocols. 1st ed. New York, USA: Humana Press; 2011. pp. 351-375

[49] Milojković Opsenica D, Ristivojević P, Trifković J, Vovk I, Lušić D, Tešić Ž. TLC fingerprinting and pattern recognition methods in the assessment of authenticity of poplar-type propolis. Journal of Chromatographic Science. 2016;**54**:1-7

[50] Jain R, Singh R, Sudhaker S, Kumar Barik A, Kumar S. Coupling microextraction with thin layer chromatography-image processing analysis: A new analytical platform for drug analysis. Toxicology and Forensic Medicine. 2017;**2**:17-25

[51] Olech M, Komsta Ł, Nowak R, Cieśla Ł, Waksmundzka-Hajnos M. Investigation of antiradical activity of plant material by thin-layer chromatography with image processing. Food Chemistry. 2012;**132**:549-553