

This is a post-peer-review, pre-copyedit version of an article published in Pérez-Gálvez A., Viera I., Roca M. (2020) Acquisition of Mass Spectrometry Data of Carotenoids: A Focus on Big Data Management. In: Rodríguez-Concepción M., Welsch R. (eds) Plant and Food Carotenoids. Methods in Molecular Biology, vol 2083. Humana, New York, NY. The final authenticated version is available online at: https://doi.org/10.1007/978-1-4939-9952-1_10

Acquisition of Mass Spectrometry Data of Carotenoids: a focus on big data management

Antonio Pérez-Gálvez, Isabel Viera and María Roca*

Food Phytochemistry Department, Instituto de la Grasa (CSIC), Building 46, 41013, Sevilla, Spain.

*Corresponding author. E-mail: María Roca, mroca@ig.csic.es

i. Running title: Mass Spectrometry of Carotenoids

ii. Abstract

Accurate determination of carotenoid profile in plant tissues and food samples requires the application of hyphenated analytical resources including HPLC with high resolution hybrid mass spectrometers. The high analytical power of modern MS equipment means the generation of *big-data* resulting from the independent and complementary physicochemical properties of the target compounds that requires a complex processing to unravel the results. The present protocol describes a complete pipeline methodology for high-throughput analysis of carotenoids based on mass spectrometry (MS). After an exhaustive extraction, the workflow includes the acquisition of HPLC-hr-MS and MS² spectra assisted step by step by specific post-processing software.

iii. Key words: carotenoids, mass spectrometry, product ions, exact mass, isotopic pattern, HPLC-MS, qQ-TOF, APCI, tandem mass spectra, software.

1. Introduction

The carotenoids are a family of natural pigments with a steadily increasing number of compounds and expanding research interest. Probably due to the even improved analytical techniques, new carotenoid structures are continuously characterized in addition to the exciting identification of derivative products resulting from their metabolism. Several public databases have compiled part of the chemical knowledge of carotenoids, relieving the work devoted to the survey of certain structures. For example, the Carotenoids Database (<http://carotenoiddb.jp/>), includes information of almost 1200 different carotenoids in 700 organisms [1], while the Prokaryotic Carotenoid Database (<http://bioinfo.imtech.res.in/servers/procardb/>) is specific for prokaryote organisms, and contains more than 1800 entries [2]. To face this high diversity, it is necessary to apply current accurate metabolomics tools to unravel the carotenoid profile in a specific tissue. Existing methodologies are based on pioneer studies on mass spectrometry of carotenoids [3-8], which have been recently reviewed [9]. Hence, fast atom bombardment (FAB) ionization has been replaced by atmospheric pressure ionization (API) methods for the ionization of most natural products, and such improvement reached the carotenoids. Although electrospray ionization (ESI) and atmospheric pressure chemical ionization (APCI) show both advantages and drawbacks for ionization of carotenoids, the use of APCI source has probably become the standard choice. Similarly, ion trap (IT) and time of flight (TOF) are the conventional hardware designs for mass analysis at the ionization downstream, IT for structural elucidation through ion accumulation of selected protonated molecular ions which are subjected to MSⁿ analysis (tandem-in-time), and TOF for acquisition of accurate mass and isotopic pattern values by filtering protonated molecular ions (tandem-in-space). Anyhow, with the modern soft ionization techniques, tandem mass spectrometry has emerged as an essential strategy, as in addition to increase the accuracy of mass measurements, it allows to figure out certain reaction patterns during the fragmentation process of the protonated molecular ions. Accordingly, the final aim is to allocate qualifier ions that simplify the identification process. Furthermore, it is possible to reach high-resolution mass spectrometry (*hr*-MS) with the modern equipment, so that the possibility of

inaccurate identifications is considerably reduced [10]. These analytical advancements empower the performance of metabolomics analysis, and two approaches can be pursued: targeted or untargeted. The former applies when the study is focused in *known* carotenoids, while the latter is performed if the analysis comprised the study of all carotenoids present, either known or unraveled structures. In any case, the complexity of metabolomics studies demands the existence of specific workflows to consistently analyze carotenoids from different sources. The present protocol describes a specific workflow for mass spectrometry analysis of carotenoids, with special emphasis in the assistance of post-processing software. Although the protocol has been focused on licensed software, free applications with comparable capabilities can be downloaded on the internet.

2. Materials

2.1 Equipment

1. Homogenizer, filter paper, separating funnel
2. Rotary evaporator.
3. LC-MS system and software for instrument control.
4. Analytical column.
5. Computer and software for MS data analysis.

2.2 Reagents

1. For carotenoid extractions, the quality of the solvents should reach HPLC-grade, while for MS analysis the solvents must be LC-MS-grade (*see Note 1*).
2. Diethyl ether, sodium chloride 10% (w/v), anhydrous Na₂SO₄, acetone
3. 0.22 µm PTFE syringe filter
4. Calibration mix of synthetic polymer poly(ethylene glycol) (PEG 400, 10 µg/mL).

3. Methods

3.1 Carotenoid extraction

The user should note that the extraction is the sum of simple steps, but slight modifications can be introduced at any of them that result in an improved success of the entire extraction procedure. Hence, each sample material requires almost a tailored protocol, so that the present description should be considered as a sequence of criteria and description of critical points to diminish common drawbacks affecting the subsequent HPLC-MS analysis. Carotenoids are usually stable in diminished light and avoiding excessive working temperature and continuous contact with air (*see Note 2*). Extracts are stored at $-20\text{ }^{\circ}\text{C}$ for 1 month.

1. Mix 10 grams of fresh tissue with 30 mL of extracting solvent and break up the tissue with a homogenizer, filter the solution and pour the solvent in a separating funnel. Repeat the operation until no more color is extracted from the remaining tissue (*see Note 3*).
2. Mix all the combined extracts with 100 mL of diethyl ether and 100 mL of 10% (w/v) sodium chloride. After gentle shaking let the phases to separate (*see Note 4*).
3. Discard the water layer and filter the remaining organic phase through a solid bed of anhydrous Na_2SO_4 (25 g).
4. Evaporate the diethyl ether in a rotary evaporator and dissolve the residue in 100 μL acetone.
5. Filter the extract through a 0.22 μm PTFE syringe filter before chromatographic analysis.

3.2 LC-MS Data Acquisition

1. Set up the LC-MS system with the isocratic or gradient conditions of the selected HPLC method. Once the system is equilibrated, analyze the sequence of samples in the autosample tray. The following description is the practice routinely used in our laboratory [11]. A binary solvent system consisting of methyl-*tert*-butyl ether:methanol:water 85:15:4 (solvent A) and methyl-*tert*-butyl methyl:methanol:water 7:90:3 (solvent B) is used at a flow rate of 1 mL min^{-1} .

Carotenoid separation is performed on a C30 analytical column (3 μm particle size, 25 \times 0.46 cm i.d.). Injection volume is 20 μL per sample. LC separation is performed using the following gradient setting: hold at 100% A for 10 min, then gradient from 100 to 50% A in 30 min, 50 to 100% B in 10 min, 100% A in 5 min, and finally isocratic 100% A for 5 min for re-equilibrating column.

2. Set the photodiode array detector to acquire the common spectrum range for carotenoids (300-700 nm), although monitoring at 400-480 nm is commonly made during the run.

3. A split post-column of 0.4 mL min^{-1} is directly introduced on the APCI source, which is operated in positive mode.

4. A calibration routine of the MS instrument is made on each HPLC run (start-and-end mode) with the calibration mix that yields product ions in the entire mass range of the instrument. Hence, it is not necessary to create different calibration files, and any modification of both the ion source and the optics is individually collected for each sample.

5. The MS scan modes are tailored depending on the mass analyzer available at the lab of the user (Q, IT, Q-IT, TOF, Q-TOF). The strategy to follow at this point is whether the acquisition of 'only' protonated molecular ions is required (full-scan mode, FS) or the choice of the user is to acquire both protonated molecular ions and product ions (broad band collision-induced dissociation mode, bbCID). Additionally, a specific scan mode could be selected, the precursor ion scan mode (PI) where some protonated molecular ions are designated for transmitting their corresponding product ions. Depending on the MS scan mode, additional screening tasks are subsequently required (*see Note 5*). The m/z scan in the UHR-TOF is in the 50-1200 Da range. In our laboratory, MS data are acquired in bbCID mode to obtain MS and tandem MS spectra simultaneously. The conditions of the APCI source are optimized for each LC-

MS system. However, the user should consider that the subsequent settings of the mass analyzer affect the precondition of the APCI source (*see Note 6*).

3.3 Peak prediction and annotation

To get the effective information that finally drives the user to build a scenario for the extract within a biological context, it is necessary to apply a systematic workflow for peak annotation and identification. Fig. 1 depicts the MS analysis of an extract from human milk with the automatically annotated protonated molecular ions to indicate the challenge of the study of MS data means. The pipeline methodology is represented in Fig. 2, allowing the user to take advantage of the complementary action of the inclusive analytical approach, and the application of bioinformatics tools for a comprehensive study of data [10].

1. The user should collect in an *in-house* database the calculated m/z values and the elemental composition of the predicted carotenoids in the analyzed extract (targeted strategy) or to compile the comprehensive information regarding the current knowledge of carotenoids and the anticipated structures the user proposes (untargeted strategy). Additionally, the calculated m/z values and the elemental composition of the product ions should be included (*see Fig. 3 and Note 7*).

2. Perform the mass screening with the software to all the raw data (FS) in a batch mode. To increase the accuracy of the screening, the software allows the user to include filtering rules (*see Note 8*). Commonly, the mass filtering is the first rule applied for the protonated molecular ions, setting a mass error value for MS peak match between the experimental and the theoretical values. The lower the mass error value (5 ppm, 3 ppm, 1 ppm...) the higher the accuracy of the predicted match. Additionally, modern MS software includes a filtering rule for the isotopic pattern value of the calculated and experimental MS data. This filtering rule is managed by the software with the application of an algorithm that yields a numerical value. The

lower the deviation between the calculated and experimental isotopic patterns, the higher the accuracy of the predicted match. (Fig. 4)

3. A first list is obtained including the tentatively identified carotenoids in the extract.

The report is commonly exported to a Microsoft Excel based document where the user can obtain additional information regarding the peak area, retention time, UV-visible maxima as columns.

4. A second screening is performed with the data obtained with the bbCID or PI mode.

This screening is made with specific applications built in the software for data managing (*see Note 8*).

3.4 Peak identification

Peak identification is subjected to the best practices and expertise of the user, including the tentative lists generated in section 3.3 and the final compilation of chromatographic data, UV-visible spectra and MS information. The aim at this point of the data processing is to confirm that specific product ions arise from the protonated molecular ions, so that the experimental fingerprint is correlated with the common fragmentation reactions observed for carotenoids [12]. This step is critical for the identification of unanticipated carotenoid structures.

1. Draw the structure of the carotenoid (*see Note 9*).
2. Apply the software for the *in-silico* generation of product ions (*see Note 10*).
3. Compare the list of the *in-silico* product ions with the list of experimental product ions for the selected carotenoid structure.
4. Review the structures proposed for each product ion and select the correct one. The common fragmentation rules of carotenoids should be followed.
5. Review the proposed mechanisms for each product ion. The software also suggests different mechanisms to generate the same fragment ion structure. Revise and select the most appropriate mechanism proposed for each fragment ion. The mechanism

should be consistent with the chemical behavior of carotenoids. In principle, the simplest should be the most probable [10].

6. Examine the consistency of the isotopic pattern between product ions and molecular ion, taking advantage of different software tools. For example, starting from a tandem MS experiment and selecting a precursor ion, the SmartFormula3D™ yields a list of product ions with their elemental composition and exact mass, the fragment loss, mass error and isotopic pattern matching. Only those fragment ions with the same isotopic pattern as the precursor ion should be selected as true fragment ions derived from the molecular ion.

7. Compare the *in-silico* and experimental data of the tentatively identified carotenoids with experimental data of authentic standards (if available) to increase the accuracy of the identification.

4. Notes

1. Purity of the solvent is a critical issue to get the appropriate signal-to-noise ratio and avoid the presence of interferences in the MS analysis. Our experience has shown that even those LC-MS-grade solvents need to be checked in the applied experimental conditions. If none of them yields the required standard, then the use of pesticide residue analysis (PRA)-grade solvents is advised.

2. Avoid the extensive use of plastic lab supplies as contaminants dissolve with the extract and appear as artifacts in the MS analysis.

3. The extracting solvent and the ratio tissue/solvent depends on the pigment polarity range and concentration in the tissue, as well as on the kind of matrix for extraction. It is convenient to perform some preliminary tests to establish amount of solvent and number of extractions. Methanol and acetone are frequently applied in this step, although solvent mixtures with water are also used. In any case, the user should consider safety and ease of use in the solvent selection.

4. If the sample lacks esterified carotenoids, chemical hydrolysis step could be applied at this point to remove interference with chlorophylls and triacyl glycerides [5].
5. The filtering rules are independent and complementary physicochemical properties available at the experimental and theoretical level, including chromatographic behavior, UV-visible spectrum, m/z values (for both $[M+H]^+$ and the product ions), isotopic pattern, (for both $[M+H]^+$ and the product ions) that once organized and compared (experimental vs. theoretical values) allow the characterization within a database of targeted compounds.
6. In our laboratory, the APCI conditions commonly applied are: ion spray voltage set -4,500 eV and the turbo gas temperature at 700 °C. The nebulizing gas is set at 70 psi, the heating gas at 30 psi, the curtain gas at 30 psi and the collision gas at 10 psi.
7. The library is a Microsoft Excel document that contains information from literature and data generated in the lab of the user. This library should be continuously updated and enriched. Hence, for a specific carotenoid the user can include new proposed fragmentations in addition to the already known product ions.
8. Producers of MS equipment offer up-to-date software for instrument control and tools for data analysis. These tools may be embedded as a module or option in the main software distributed with the MS instrument. In our laboratory we apply the DataAnalysis™ software with the TargetAnalysis™ and SmartFormula3D™ utilities from Bruker, but other alternatives exist (Xcalibur, Analyst, MassHunter, MSFact, XCMS2).
9. Several chemical drawing software tools are free available on the internet (ChemDraw Prime, ChemSketch, BKChem, Avogadro).
10. Predictive software tools commonly applied in this step are MassFrontier™, HighChem, Mass Fragmenter, Advanced Chemistry Development, and EPIC. Alternatively, open versions can be found in internet, as OpenMS-Simulator, MetFragCL, CFM-ID, MAGMa+ or MS-FINDER.

11. Particularly useful could be to visit the webpage ms-utils.org. It is a repository and updated list of free software for analysis of mass spectrometry data.

Acknowledgements

This work was supported by the Ministerio de Ciencia, Investigación y Universidades, Agencia Estatal de Investigación y Fondo Europeo de Desarrollo Regional (FEDER), project numbers RTI2018-095415-B-I00 and AGL2017-87884-R. Antonio Pérez-Gálvez is a tenured scientist at the Spanish National Research Council (CSIC).

Figure Legends

Fig. 1 Full scan MS chromatogram of an extract from human milk with the automatically annotated protonated molecular ions.

Fig. 2 Complete pipeline methodology for *hr*-MS analysis of carotenoids.

Fig. 3 Example of an *in-house* database including m/z and elemental composition of parent compound and product ions, retention times and UV-Vis maxima.

Fig. 4 Carotenoid MS screening according to UV-Vis spectrum, mass accuracy and isotopic pattern for $[M+H]^+$ at 537.4465 Da.

5. References

1. Yabuzaki, J (2017) Carotenoids database: structures, chemical fingerprints and distribution among organisms. Database 2017:2017-bax004.
2. Nupur LNU, Vats A, Dhanda SK, Raghava, GPS, Pinnaka AK, Kumar A (2016) ProCarDB: a database of bacterial carotenoids. BMC Microbiol 16:96.
3. van Breemen RB (1997) Liquid chromatography/mass spectrometry of carotenoids. Pure & Appl Chem 69:2061-2066.
4. Schweiggert U, Kammerer DR, Carle R, Schieber A (2005) Characterization of carotenoids and carotenoid esters in red pepper pods (*Capsicum annuum* L.) by high-performance liquid chromatography/atmospheric pressure chemical ionization mass spectrometry. Rapid Commun Mass Spectrom 19:2617-2628.

5. de Rosso VV, Mercadante AZ (2007) Identification and quantification of carotenoids, by HPLC-PDA-MS/MS, from Amazonian fruits. *J Agric Food Chem* 55:5062-5072.
6. Dugo P, Herrero M, Kumm T, Giuffrida D, Dugo G, Mondello L (2008) Comprehensive normal-phase x reversed-phase liquid chromatography coupled to photodiode array and mass spectrometry detection for the analysis of free carotenoids and carotenoid esters from mandarin. *J Chromatogr A* 1189:196-206.
7. Ornelas-Paz JJ, Yahia, EM, Gardea AA (2008) Changes in external and internal color during postharvest ripening of 'Manila' and 'Ataulfo' mango fruit and relationship with carotenoid content determined by liquid chromatography-APCI+-time-of-flight mass spectrometry. *Postharvest Biol Technol* 50:145-152.
8. van Breemen RB, Dong L, Pajkovic, ND (2012) Atmospheric pressure chemical ionization tandem mass spectrometry of carotenoids. *Int J Mass Spectrom* 312:163-172.
9. Rivera SM, Christou P, Canela-Garayoa R (2014) Identification of carotenoids using mass spectrometry. *Mass Spectrom Rev* 33:353-372.
10. Pérez-Gálvez A, Roca M (2019). Recent developments in the analysis of carotenoids by mass spectrometry. In: Zepka L, Jacob-Lopes J, De Rosso VV (eds) *Progress in carotenoid research*. IntechOpen, London, pp. 17-44.
11. Ríos JJ, Xavier AAO, Díaz-Salido E, Arenilla-Vélez I, Jarén-Galán M, Garrido-Fernández J, Aguayo-Maldonado J, Pérez-Gálvez A (2017) Xanthophyll esters are found in human colostrum. *Mol Nutr Food Res* 61:1700296.
12. Pérez-Gálvez A, Sánchez-García A, Garrido-Fernández J, Ríos JJ (2018) MS tools for a systematic approach in survey for carotenoids and their common metabolites. *Arch Biochem Biophys* 650:85-92.

Figure 1

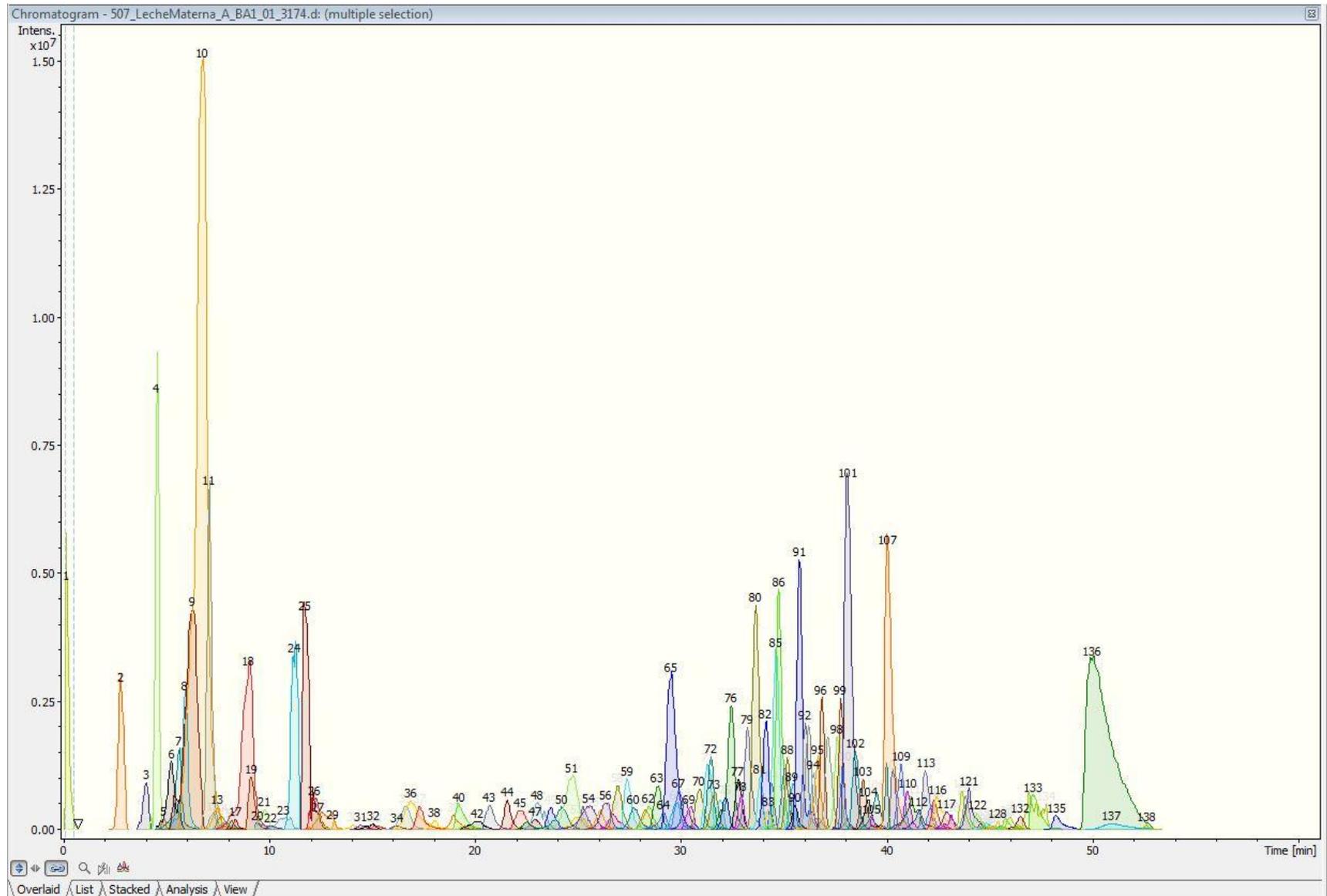


Figure 2

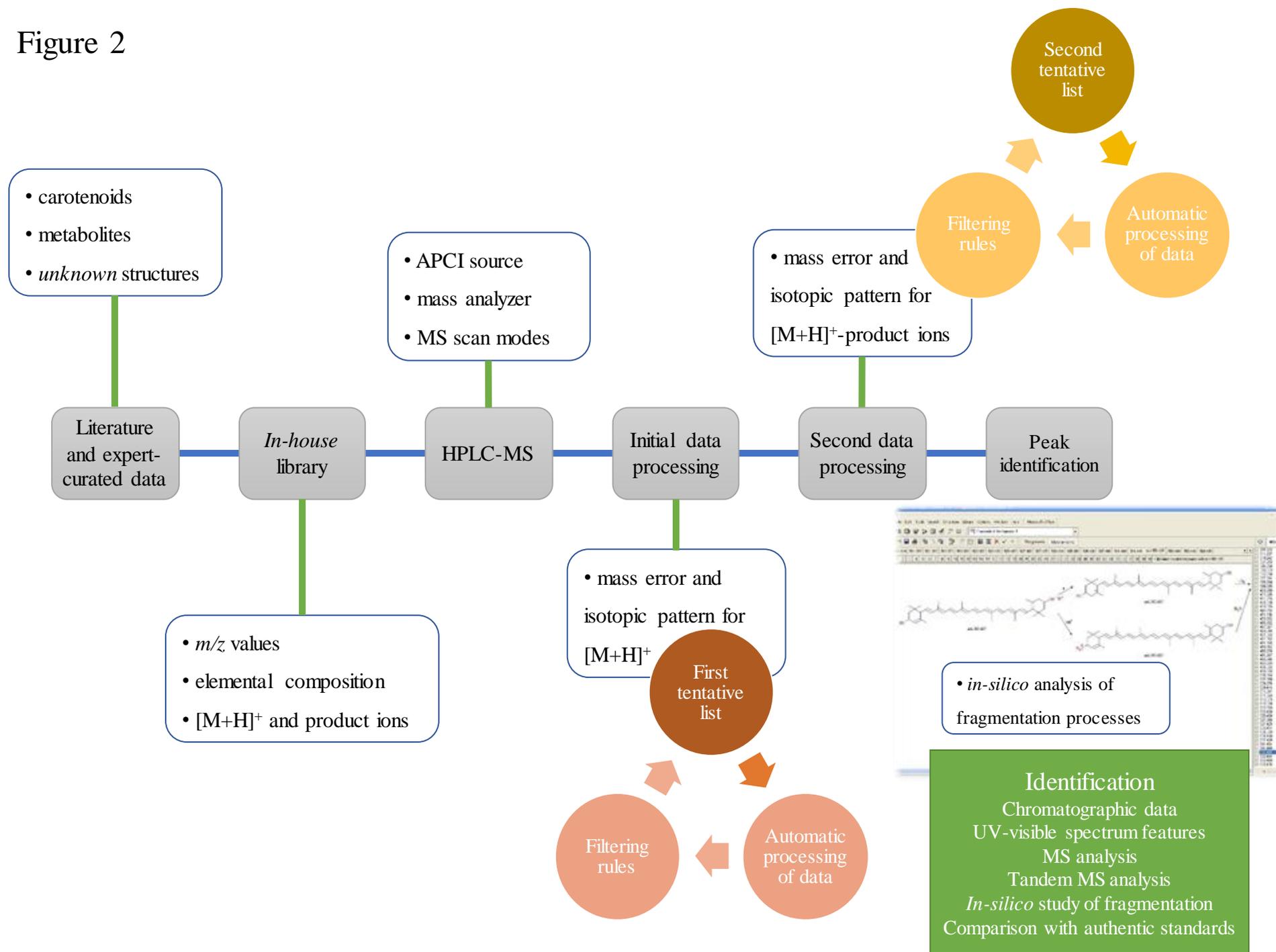


Figure 3

Carotenoid	Elemental composition	[M+H] ⁺	RT (min)	UV-Vis max	[M+H-18] ⁺	[M+H-H ₂ O] ⁺	[M+H-80] ⁺	[M+H-C ₆ H ₈] ⁺	[M+H-92] ⁺	[M+H-C ₇ H ₈] ⁺	[M+H-106] ⁺	[M+H-C ₈ H ₁₀] ⁺	[M+H-124] ⁺	[M+H-C ₉ H ₁₆] ⁺
Lutein	C ₄₀ H ₅₆ O ₂	569,4370	14,50	420, 444, 472	551,4247	C ₄₀ H ₅₄ O			477,3727	C ₃₃ H ₄₈ O ₂	463,357	C ₃₂ H ₄₆ O ₂		
Zeaxanthin	C ₄₀ H ₅₆ O ₂	569,4370	17,20	425, 450, 476	551,4247	C ₄₀ H ₅₄ O	489,3727	C ₃₄ H ₄₈ O ₂	477,3727	C ₃₃ H ₄₈ O ₂	463,357	C ₃₂ H ₄₆ O ₂		
β- carotene	C ₄₀ H ₅₆	537,4434	36,60	424, 451, 477			457,3829	C ₃₄ H ₄₈	445,3829	C ₃₃ H ₄₈			413,3203	C ₃₁ H ₄₀

Figure 4

