BUILDING AN EMPIRICAL MASS SPECTRA LIBRAY FOR NON-TARGETED METABOLOMICS ANALYSIS IN PLANTS BY UHPLC-Q TOF MS

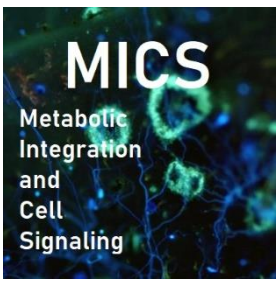
Escola Superior de Tecnologia i Ciències Experimentals

FINAL DEGREE PROJECT, JULY 2018

DEGREE IN CHEMISTRY

DAVID IZQUIERDO SANDOVAL

MENTOR: MARÍA IBAÑEZ MARTÍNEZ



Index:

Index

1. Introduction:

1.1. Metabolomics analysis. Targeted and non-targeted analysis

Metabolomics refers to research that aims at the characterization of all metabolites, which are small size molecules (<1500 Da), intermediates or end-products of metabolic processes presents in biological systems [1]. The central dogma of molecular biology is based in the explanation of the flow of genetic information within a biological system and it is summarized as "DNA makes RNA and RNA makes protein" [2]. However, this definition is outdated and incomplete, since it should be included that proteins, acting as enzymes, are responsible for the synthesis of metabolites (Figure 1). Therefore, since metabolites represent the end point of the gene expression and cell activity, metabolomics offers a holistic approach for understanding the phenotype of an organism [3] and is a valuable tool for understands plant system [4].

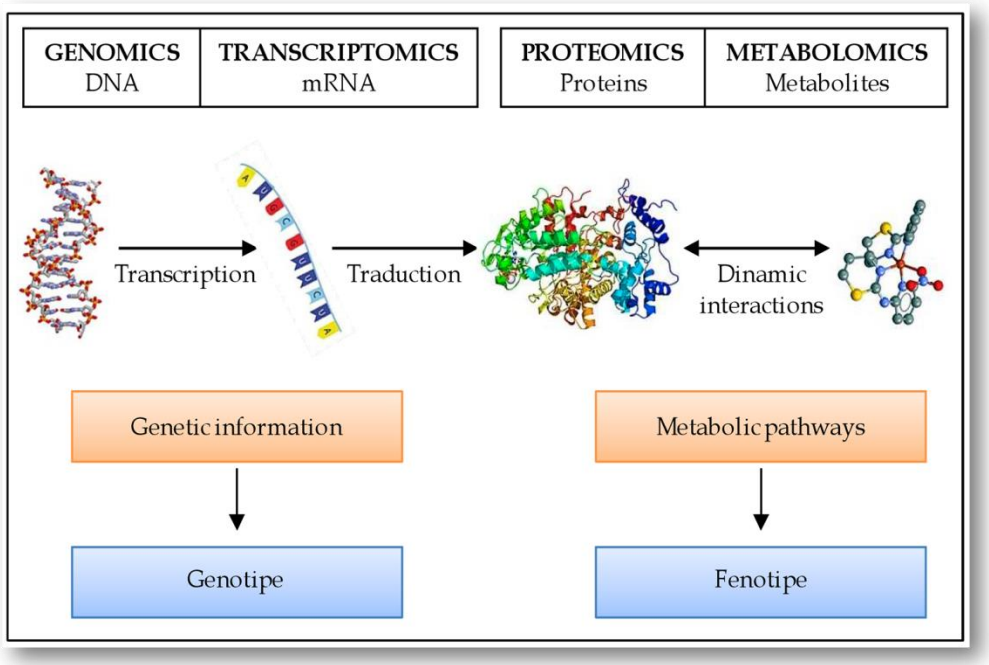


Figure 1. Central dogma of molecular biology. The genes are transcribed to mRNA, which translates to porteins responsible for catalyzing metabolic pathways. (Source: https://www.intechopen.com/)

There are two main ways to work on metabolites: targeted and non-targeted analysis. When the goal of an analysis is to obtain quantitative data of known metabolites, whose chemical structure is also known, is called targeted [5]. On the other hand, the goal of non-targeted analysis is perform a semi-quantitative profiling of plant metabolome[1]. Plant untargeted metabolomics is a great challenge because of the rich chemical diversity of metabolites which are presented in a huge range of concentrations; the amount of metabolites in a plant matrix round between 100000 and 200000 [6]. Therefore, high resolution separation techniques, mainly gas or liquid chromatography (GC or LC), are coupled to other analytical methods as nuclear magnetic resonance (NMR) or mass spectrometry (MS) in order to add qualitative variable.

1.2. Ultra high performance liquid chromatography: UHPLC

Chromatography comprises an important set of different methods. The goal of chromatography is the separation of the components of a complex mixture. In all chromatographic separations sample is transported by a *mobile phase*, which could be a gas, liquid or supercritical fluid. Mobile phase passes through *stationary phase*, the chemical composition of both phases is selected in order to get a different distribution of the mixture components. The components which are more retained by stationary phase move slowly in mobile phase flow; however, those which retention with stationary phase is weaker move faster. As a consequence of different mobility, components are separated in bands which can be analysed quantitatively or qualitatively. [7].

In liquid chromatography (LC), stationary phase is located inside of a column, sample is injected by automatic injector and mobile phase consist in liquid or a mix of different solvents. After the process, a detector measures a property of analytes, for example mass spectrometry measures *m/z* ratio. The introduction of high pressure systems (HPLC) has been an enormous advance in LC due to the possibility of carrying out mobile phase and sample through a well-packaged column whose particle size is minuscule, ultra-high performance liquid chromatography (UHPLC) is referred to HPLC with particle size <2μm.

[1] The complete set of metabolites.

Several types of analytical columns have been developed depending on the physical properties of studied analytes: reversed phase (RP) with non-polar liquid, octadeyl chains (C18), supported on silica particle is the most suitable technique for polar metabolites [7].

Figure 1 is shows an example of a UHPLC system scheme. The mobile phase can be formed by one or more components which are stored in the solvent reservoirs. Normally, two miscible components are used: the component A is an aqueous solution and the component B an organic solution (typically methanol or acetonitrile). The objective of using two different eluents is the creation of a gradient capable of eluting all the components. Reversed-phase columns employ non-polar stationary phase, therefore a high proportion of A is used to elucidate the polar compounds retained in the column. Later, the proportion of component B increases with respect to that of A to elucidate the more non-polar. As it can be seen at Figure 2, elution gradient is achieved regulating the proportion of solvents A and B using valves. The next step is a mixing vessel where the two solvents are mixed. Finally the mix is pumped with a high-pressure pump. This is the typical scheme which low pressure gradient operates. If a high pressure gradient is desired, the eluents pumping must be performed before the mixing vessel [7].

Sample is injected within the chromatographic system through an automatic injector. In UHPLC there are three requirements demanded for sample injection. Sample injected must have little and reproducible volumes, physic dispersion must be minimal and, hydrodynamic features of chromatographic system must not be changed. The transition between introducing mobile phase and sample without violate any of this requirements is achieved due to switching valve. [7]

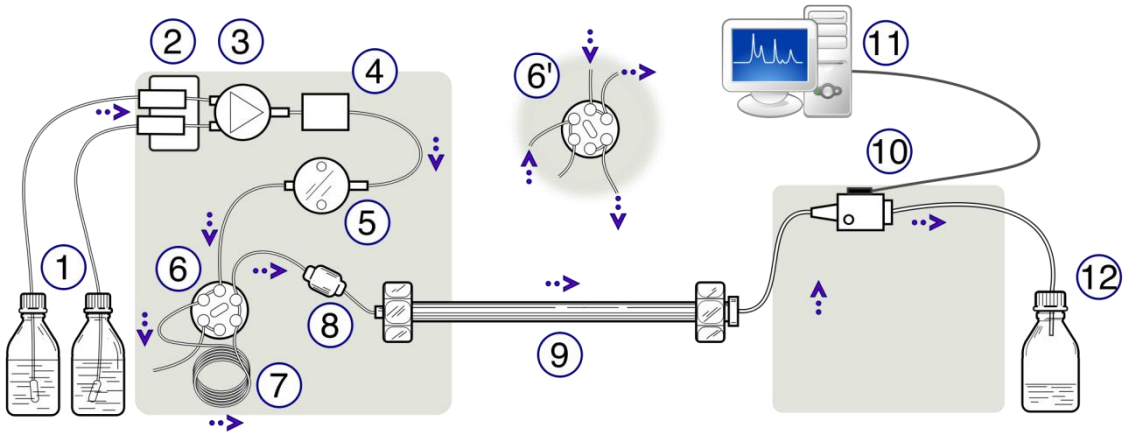


Figure 2. Basic components and flow paths in a conventional HPLC/UHPLC system. (1) Solvent reservoirs, (2) Solvent degasser, (3) Gradient valve, (4) Mixing vessel for delivery of the mobile phase, (5) High-pressure pump, (6) Switching valve in "sample inject position", (6′) Switching valve in "sample load position", (7) Sample injection loop (10–100 μL), (8) Pre-column (guard column), (9) Analytical column, (10) Detector (i.e., IR, UV, MS), (11) Data acquisition, (12) Waste or fraction collector. (Source: http://upload.wikimedia.org/wikipedia/commons/thumb/a/a0/HPLC_apparatus.svg/2000px-HPLC_apparatus.svg.png.)

1.3. Mass spectrometry: QTOF

UHPLC-RP coupled to MS has become the most powerful analytical method for plant metabolomics because of its great sensitivity, selectivity and robustness, adding its capacity for analysing polar and thermally unstable metabolites without complex sample pre-treatment [9-12].

MS consist in the detection of a determinate mass-to-charge ratio (*m/z*) and their intensity in an ionized sample which in this case is a chromatography fraction. Mass spectrometer has three main parts: an ionization source, a mass analyser and a receptor; all of them are kept under vacuum in order to allow the transmission of ions [13,14]. The main problem of LC-MS is that mass spectrometry works in gas phase, so, it is necessary the use of an interface between the two techniques.

There are several types of ionisation sources; the most used is electrospray (ESI). The success of ESI resides in its facility to couple with HPLC and its high sensibility. An electrospray is produced by applying a strong electric field, under

atmospheric pressure, to a liquid passing through a capillarity tube with a weak flux(Fig 3). The electric field is obtained by applying a potential difference from 3 to 6 kV between this capillary and the counter-electrode, separated by 0.3-2 cm, producing electric fields of the order of 10^6 V/m. This field induces a charge accumulation at the liquid surface located at the end of the capillary, which will break to form charged droplets. A gas injected coaxially at low flow rate allows dispersion of the spray to be limited in space. These droplets then pass through a curtain of heated inert gas (usually nitrogen) to remove the last solvent molecules [15]. While solvent molecules are being evaporated, the repelling coulomb forces in the droplets increase until match cohesion forces (Rayleigh limit). As a consequence, droplets are divided in the ions, isolated and charged. Analytes could be charged on positive or negative in accordance with its nature [16]. The potential difference carries the charged analytes into the mass analyser.

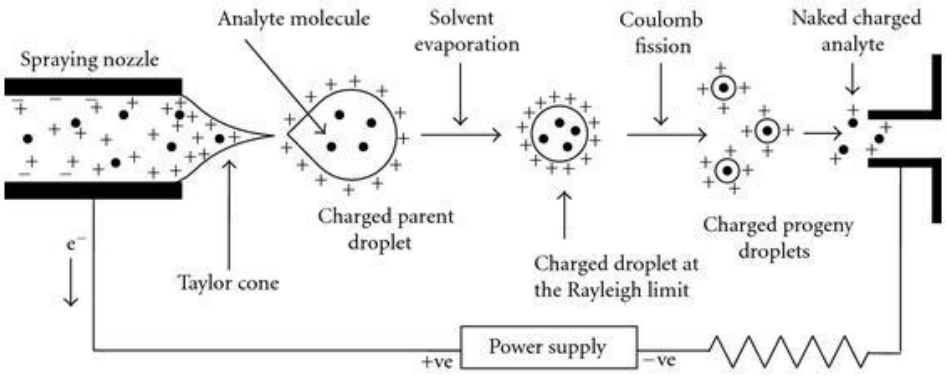


Figure 3. Schematic representation of the electrospray ionization process.

ESI has several key advantages, including ionisation across a large mass range appropriate for metabolome analysis, good sensitivity, soft ionisation and high adaptability [17]. However, ESI also has some disadvantages, such as adduct formation, quenching, low tolerance towards salts and suppression of the ionisation of metabolite species due to the presence of high concentration of a different species (matrix effect), both leading to ionisation suppression or enhancement and failure to detect certain metabolite classes/species [18].

Once the gas-phase ions have been produced, it is necessary to separate them according to their *m/z*; there are several types of mass analyser. All mass analysers use static or dynamic electric and magnetic fields which can be single or combined; the

difference between the different types of mass analysers lies in the way in which these fields are pledged to achieve separation [15]. The two main mass analysers used in plant metabolomics are TQD and TOF.

Triple quadrupole mass detector (TQD) has been arisen as the best option for quantification of known metabolites (targeted analysis). A quadrupole analyser is a device which used the stability of the trajectories in oscillating electric fields in order to separate ions according to their *m/z* ratios. TQD consists in a linar combination of a quadrupole, hexapole and quadrupole. The first quadrupole isolates the selected ion; hexapole or collision cell hits the ion selected with an inert gas at high energy in order to break it down into different fragments (daughter ions). Finally, the second quadrupole isolates the daughter ions providing specific transitions for each targeted plant metabolite. To perform this analysis, the compounds of interest must be infused into the system in order to know the most intense daughter ions [5].

However, to perform a non-targeted full scan of a plant metabolome the best option is a time-of-flight (TOF) analyser. The principle of TOF is quite simple: ions of different *m/z* are dispersed in time during their flight along a field-free drift path of known length. Provided all the ions start journey at the same time or at least within a sufficient short time interval, the lighter ones will arrive earlier at the detector than the heavier ones [14].

The physic principle of TOF instruments relies in the conversion of potential energy of a charged particle in an electric field into kinetically energy.

$$E_{el} = qU = ezU = \frac{1}{2} m_i v = E_{kin}$$

q: electric charge of an ion equal to z electrons of charge e.

U: voltage

m_i: electron mass

v: electron velocity

As time is equal to s/v (where s is the distance travelled by electron):

$$t = \frac{s}{\sqrt{2eU}}\sqrt{\frac{m}{z}}$$

Hybrid quadrupole time-of-flight mass spectrometry (Q TOF MS) provides good sensitivity, high resolution, mass accuracy and scan rate, enabling MS profiling and MS/MS analysis within a single experiment [19]. This technique allows the profiling of intact precursor ions generated from metabolites by ESI and perhaps represents the most appropriate MS instruments to apply LC separations for this objective [4]. The first quadrupole isolates ions of interest, a collision cell T-wave is used for collision induced dissociation (CID) where the filtered precursor ion is fragmented via a stream of collision gas, finally fragment ion mas is obtained from TOF mass analyser (Figure 4) [15, 19]. The analyser allows the sequential acquisition of low and high collision energy spectra thus, (de)protonated molecule and fragment ions accurate mass are obtained (MS^E).

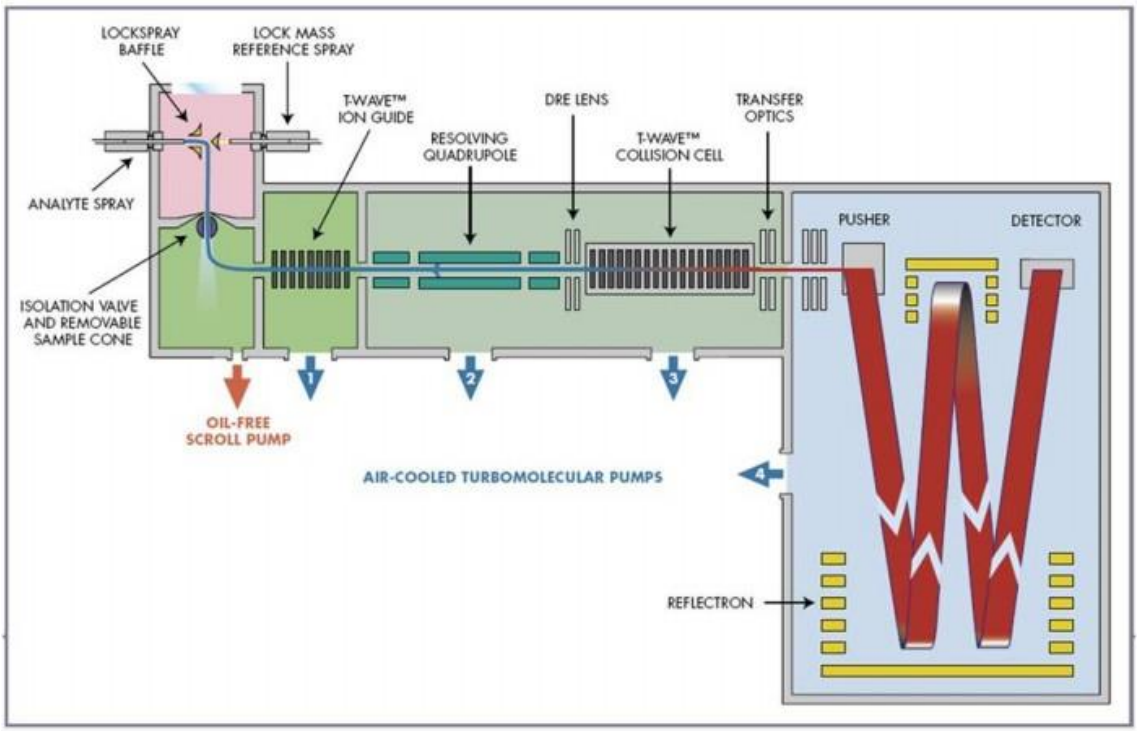


Figure 4. Experimental setup of Q TOF MS. (Source: http://penyfan.ugent.be/labo/joelv/Qtof.html)

1.4. Post-acquisition data processing and metabolite identity assignment

Interpretation of spectral data is a really challenging task due to the large amount of metabolites, the complexity of the mixture and the existence of isomers, adducts and in-source fragment ions [1]. The goal of non-targeted metabolomics in plants is to obtain semi-quantitative information of the changes produced in the metabolome of a set of samples. There are different types of algorithms for identification and comparison of features in metabolome. XCMS was developed as a metabolomics data processing algorithm to extract metabolic features from raw MS data (retention time, parental *m/z* and peak area) and perform semi-quantitative statistical analysis [19]. In addition, there are some biological softwares capable of working with the large volume of data obtained by the XCMS algorithm and reach satisfactory biological interpretation [5]. One example is Mar-Vis Suite which was introduced for the extraction, clustering and visualization of metabolic markers form data originating from non-targeted experiments [20].

Compound identification is the most critical step in high resolution mass spectrometric non-targeted analysis. Different levels of identification are distinguished in the literature (Figure 5).

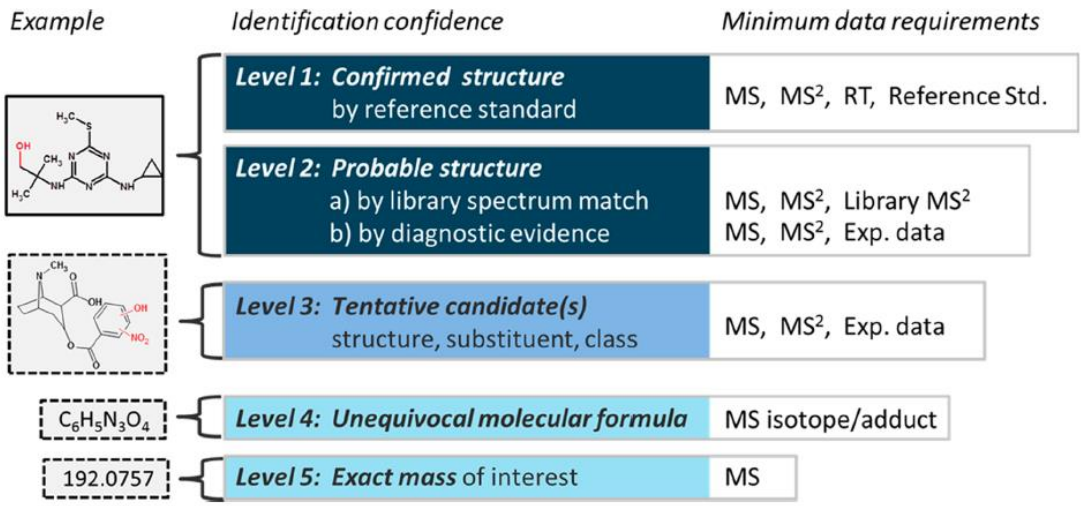


Figure 5. Proposed identification confidence levels in high resolution mass spectrometric non-targeted analysis [21].

MarVis-Suite considers isotope abundance and adducts formation in order to identify metabolites in databases as KEGG[2] or home-made libraries [20]; thus, it will be at a level 4. As shown in Figure 5, the highest confirmation levels in non-targeted analyses required mass spectral databases with MS^E information. There are several freely available databases that provide MS^E information for a large number of compounds, such as Metlin or MassBank. However, these do not incorporate information on the retention time since the chromatographic conditions, such as mobile phase and stationary phase; they change depending on the type of metabolome that is being made. In addition, there may be changes in the conditions and type of ionization source which would affect to fragmentation patterns. Therefore, an ideal characterization of the compounds requires the building of home-made mass spectral libraries using standard reference and the same experimental and instrumental conditions.

[2]The Kyoto Encyclopaedia of Genes and Genomes (KEGG) is a pathway map from public database which contains information about reactions as well as the associated enzymes, genes and metabolites.

1.5. Phenolic compounds

Phenols are one of the most important families of secondary metabolites found in plants. They are generally characterized as aromatic metabolites which possess one or more acidic hydroxylic groups attached to the aromatic arene ring. [22].

There are three main groups of phenolic compounds according to the biosynthetic pathway [23]:

1. The shikimate/chorizmate or succinylbenzoate pathway wich produces phenyl propanoid derivates. The compounds selected in this group are: ferulic acid, caffeic acid, chlorogenic acid, 3-dehidroshikimate, 4-coumaryl alcohol, gentisic acid, coumarine and sinapic acid.
2. The acetate/malonate or polyketide pathway, which produces the side-chain-elongated phenyl propanoids: flavonoids and some quinones. Flavonoids are really interesting for plant *defense* studies because is widely known its benefit in plant resistance against pathogen in spite of the lack of knowledge about the mechanism which perform this action [24]. Flavonoids selected to this project were: kaempferol, apigenin, neohesperedin, (-) epicatechin, ampelopsin, taxifolin, quercitrin, epigallocatechin, catechin and naringin. Capsaicin and skimmianine form part of this metabolic pathway but they are not flavonoids.
3. The acetate/mevalonate pathway, which produces terpenoids. No compound was selected from this group.

The importance of phenols in plant defense is evidenced in a broad range of publications, for example, there are studios dealing with their accumulations in epidermal layers of plant tissues exposed to a pathogen attack [25]. Moreover, it was proved that phenolic levels vary along time in front of external factors [26]. Therefore, for those cases in which a phenotypic plant response is observed but the metabolites involved are unknown, incorporating phenolic mass spectrum database in full profiling metabolomics could be of great interest.

1.6. Plant *defense,* induced resistance and hormonal metabolomics analysis

Plants must continuously defend against attacks from bacteria, viruses, fungi, invertebrates, and even other plants; thus, plants have evolved to survive to different situations and unfavourable environments. Mechanism of resistance in plants has little in common with the vertebrate's one, and its knowledge is rather new and continues in development. For two decades, the number of published research related to plant defense has increased markedly year by year. The importance of this endeavour relies in two main points. First of all, the study of interactions between plant-pathogen should help to elucidate signalling mechanisms by which plant cells cope with a stress situation. Secondly, such studies should provide sustainable practical solutions for the control of plant disease in agricultural crops. Modern agriculture consist in great extensions of monoculture of genetically uniform crops which are really sensitive to epidemics, such situation decrease crop yield and quality and safety of the end product. Nowadays agrochemicals are used in order to avoid that; however they cause serious problems of pollution and increase the cost of production. Gaining knowledge about how metabolites are involved in plant resistance would help to reduce the use of agrochemicals [22].

To provide a satisfactory understanding of plant resistance this phenomenon must be studied as the interaction between plant and pathogen. Recent advances have showed that several pytohormones control the main plant responses against them [5]. Plant hormones or phytohormones are signal molecules, present in trace quantities. Changes in hormone concentration mediate a whole range of developmental processes in plants, many of which involve interactions with environmental factors [22]. Plants have developed adaptive and defensive mechanisms against biotic and abiotic stress. They are provided of a basal immunity which contributes to slowing down the colonization process of a pathogen; however, it is too weak to prevent the disease in some cases. The level of basal immunity of a plant can be enhanced through application of appropriate stimuli, this phenomena is known as induced resistance (IR)[27]. In the literature there are various examples of treatments which have been shown to successfully induce resistance [28,29].

1.7. Case study: Indole-3-carboxylic (I3CA) acid as mediator of induced resistance against *Plectosphaerella cucumerina*

Plectosphararella cucumerina is a necrotrophic fungus which is widespread in tropical and subtropical regions and affects to cucurbits and other crops. Its mechanism of infection is to kill the plant cell where the infection occurs.

Triptophan derivates are an important family of secondary metabolites, a specific branch of the Triptophan pathway generates auxines which are hormones that regulate many developmental process of the plant. Tryphtophan-derived secondary metabolites are important for the plant *defense* responses to necrotrophic fungus as *Plectosphaerella cucumerina* and *Botrytis cinerea* [30]. Auxin precursor I3CA was elucidated as a mediator of induced resistance against *Plecosphaerella cucumerina* in BABA-treated plants. B-aminobutyric acid (BABA) is able to induce effective resistance against *Plecosphaerella cucumerina* by stimulating callose-rich cell wall [30]. I3CA-treated adults and seedling plants presented greater resistance to *P. cucumerina* than untreated plant [32], it was discovered that I3CA is related to greater callose deposition in adult plants [33]. Callose deposition acts as a constitutive barrier against the action of the fungus [22]. However, the resistance mechanism employed by I3CA-treated seedling treated plants is different from that of adults. From this, a new line of research has been opened that compares the differences in both the metabolome and the phenotype between I3CA adult and seedling treated plants.

2. Objectives

This project is based on the building of an empirical mass spectra library with the aim of confirming hormone structures in semi-quantitative non-targeted metabolomics analysis in plants. The library was composed by twenty phenolic compounds: kaempferol, apigenin, neohesperedin, ferulic acid, caffeic acid, chlorogenic acid, 3-dehidroshikimate, 4-coumaryl alcohol, capsaicin, (-) epicatechin, ampelopsin, taxifolin, gentisic acid, coumarine, quercitrin, epigallocatechin, skimmianine, catechin, sinapic acid and naringin. Ultra-high performance-liquid chromatography coupled to hybrid quadrupole time-of-flight mass spectrum (UHPLC-QTOF MS) has been used in order to acquire retention time at the same time that low and high collision energy spectra.

On the other hand, the effectiveness of the extraction method used for this type of analysis in the analytes of the library is checked. For this purpose, samples of *Arabidopsis thaliana* and *Solanum lycopersicum* (tomato) spiked at several levels were analysed and their spectra compared with those of pure standards at the same levels.

Finally, the library has been applied to an on-going research consisting in the metabolomic comparison between adult and seedlings plants of *Arabidopsis thaliana* treated with I3CA against the infection of *P. cucumerina*.

3. Material and methods

3.1 Reagent and chemicals

Reference stardards of kaempferol, apigenin, neohesperedin, ferulic acid, caffeic acid, chlorogenic acid, 3-dehidroshikimate, 4-coumaryl alcohol, capsaicin, (-) epicatechin, ampelopsin, taxifolin, gentisic acid, coumarine, quercitrin, epigallocatechin, skimmianine, catechin, sinapic acid and naringin were purchased from Sigma-Aldrich (St Louis, MO, USA). Structure of phenolics compounds can be found attached to spectrums in *Supplementary information 1*.

Standard stock solutions of 500 ppm of each compound were prepared; a mixture of 90:10 water:methanol was used in most cases as a solvent, adding more percentage of methanol in more hydrophobic compounds such as apigenin and skimmianine. Intermediate solution of 7.5 ppm of each compound were prepared in water:methanol (90:10); finally, 300 ppb solution of each compound were injected in the UHPLC-QTOF MS system

Indole-3-carboxylic acid was obtained from VWR (Barcelona, Spain). HPLC-grade methanol (MeOH) and formic acid eluent additive for LC-MS (HCOOH, content>98%) were purchased from Scharlab (Barcelona, Spain). HPLC-grade water was obtained from deionized water passed through Mili-Q Gradient A10 (18.2 MΩ cm) water purification system (Milipore, Bedford, MA, USA). Growth and culture medium were purchased from (Duchefa).

3.2 Biological material

Adults and seedlings *Arabidopsis* plants were the subject of study. A treatment with indole-3-carboxilic acid (I3CA) was made in order to find a different metabolic behaviour between them against the fungus *Plectospaerella cucumerina.* The experiment consisted in four treatments: control plants, plants infected without

treatment, plants treated without infection and plants treated with infection. The method of treatment and infection has been described previously in the literature [31,32].

Arabidopsis thaliana accession Col-0 seeds were germinated in 50 mL Jiffy pellets, maintained at 21 ºC day/18ºC night with 9 h of light (125 μE m-2 s-1) and 60 % of relative humidity. After 4-5 weeks in adults and 2-3 weeks in seedlings, 5 mL of I3CA 1500 μM is injected to each Jiffy (soil treatment) in order to get a final concentration of 150μM; relative humidity was maintained at 100%. Control plants were treated with 5 mL of water.

Some spores of *Plectosphaerella cucumerina* were put in a petri dish with a potato dextrose agar (PDA) culture medium; when the fungus had growth enough, spores were recollected scrapping the dish with a spatula. Then, spores were deposited in a potato dextrose broth (PDB) ½ solution. Spores amount was estimated counting some of them in a determinate area using a microscope. Solution was diluted in order to get 10^5 spores/mL. The plants which did not need to be infected were treated with a PDB spray ½ since PDB may have an effect on plant metabolism. Infection is carried out two days post-treatment and solution is applied to the leaves via spray in order to get homogenous humification.

3.3 Sampling and metabolite extraction

Arabidopsis leaves were sampled 48 hours post-infection by freezing in liquid nitrogen. The experiment was performed twice in different growing rooms and with three biological replicas per treatment in each room, so that, there are a whole of 6 replica per treatment. Leaves collected are stored at -80 ºC.

Before the extraction plants must be lyophilised. The extraction process follows the next scheme as it was described in [31,32]:

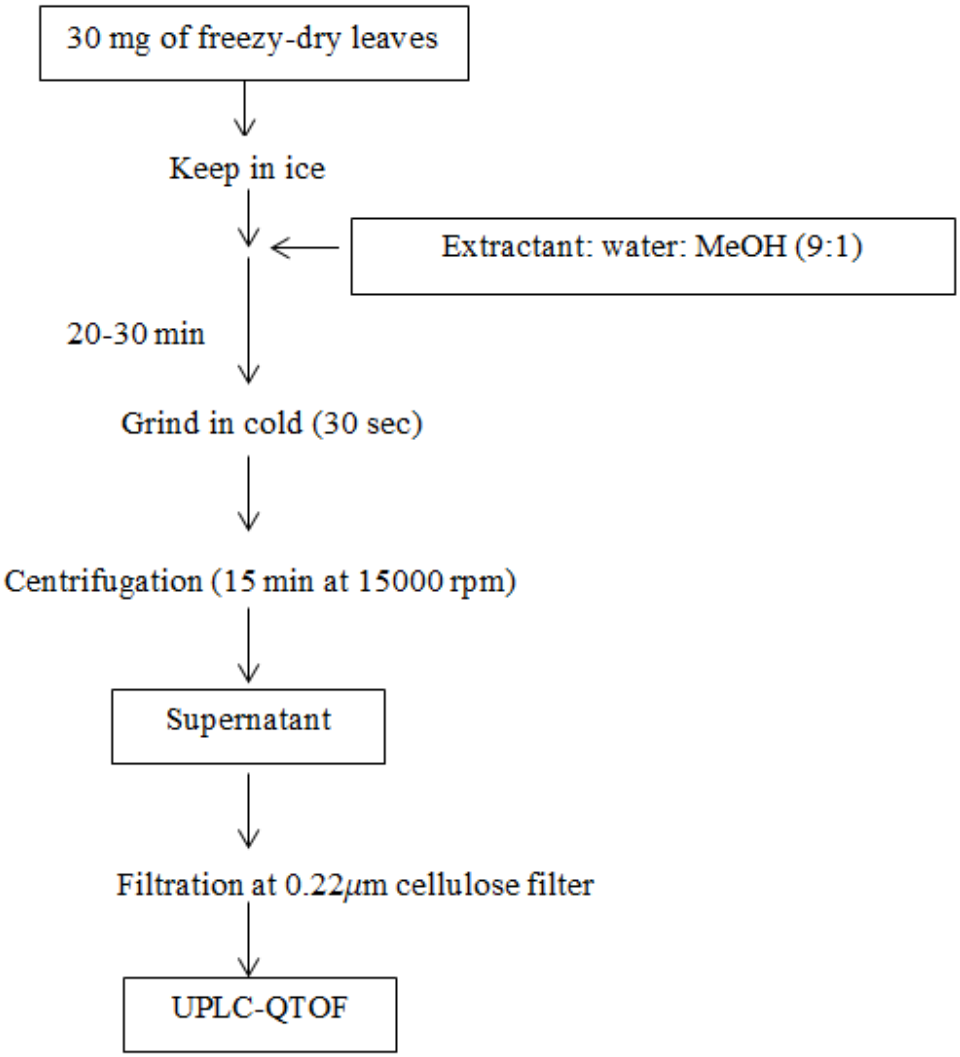


Figure 6. Extraction procedure scheme.

3.4 Instrumentation: UHPLC-QTOF MS

Chromatographic conditions and QTOF MS parameters were optimized for hormone plant metabolomics [30,31]. Metabolome Analysis was performed using an Acquity ultra-performance liquid chromatography (UHPLC) system (Waters, Mildford, MA, USA) hyphenated to hybrid quadrupole time-of flight mass spectrometer (QTOF Premier, Waters Micromass, Manchester, UK) using an orthogonal Z-spray-electrospray interface. The LC separation was performed using a UHPLC SunFire C18 analytical column, 5μm particle size, 2.1 x 100 mm (Waters). Analytes were eluted with a gradient of methanol and water containing 0.01 % of formic acid (HCOOH), at a flow rate of 0.3 μL min^{-1}. The eluent methanol component was maintained constant the first minute at 5% (v/v) and then was increased linearly between intervals: 1-5 min, from 5 to 30 %; 5-7 min, from 30 to 50%; 7-8 min, from 50 to 95%; finally is maintained at 95 % during 2 minutes. Total run time was 12 minutes. The injection volume was 20 μL and the column was set at 40 ºC.

ESI capillary voltages of 3.3 and -3.3 kV were used in positive and negative ionization modes, respectively. Nitrogen (Praxair, Valencia, Spain) was used as nebulizing and drying gas, and desolvation gas flow was set at 1000 L h^{-1}. Cone voltage was set up at 25 V in both functions, and cone gas flow was set at 60 L h^{-1}. Quadrupole was set off in order to do a full scan. The collision gas was argon (99.995 % Praxair). In MS^E there are two functions with different collision energies: the first function works at low energy (LE) with collision energy of 4 eV; in the second function, high energy (HE), a collision energy ramp is promoted from 5 to 40 eV.

The resolution of the TOF mass spectrometer was about 10000 at full width half maximum (FWHM) at *m/z* 554.2615. MS data were acquired over an *m/z* range of 40-1000 in scan time of 0.2 s.

A lock-spray probe was employed in order to assure an automated accurate mass measurement. Leucine enkephalin (10 ppm) in ACN:water (50:50) with 0.1% of HCOOH pumped at 20 μL min^{-1} was used as lock mass. The leucine enkephalin $[M+H]^+$ ion (*m/z* 556.2771) and its fragment ion (*m/z* 278.1141) in positive ionization mode; and $[M-H]^-$ ion (*m/z* 554.2625) and its fragment ion (*m/z* 236.1035) in negative ionization,

were used with the goal of recalibration the mass axis and obtain a robust accurate mass measurements over the time.

Masslynx version 4.1 (Waters) was the data station operation software.

3.5. Data processing

Raw data obtained from MassLynx software was transformed to .CDF using Databridge, which is provided by MassLynx package. CFD data are processed with R using XCMS package for peak integration and relative quantification.

3.6. Statistical analysis

For statistical purposes data were analysed with the STATGRAPH 4.0 software by a one way ANOVA and an LSD test $p < 0.05$. Six replicas were used per treatment. Grubbs test $p<0.05$ was performed in order to detect outliers.

4. Results and discussion

4.1 In-house empirical mass spectrum library

Reference standard of phenolic compounds were injected in UHPLC-Q TOF MS system in optimized conditions for hormone plant non-targeted metabolomics analysis, using the optimized conditions was described in section 3.

Figure 7 shows an example of all the information obtained for each analyte in the library: the spectra at both low energy (LE) and high energy (HE) at the retention time of the compound. There is a set of 20mDa narrow-windows extracted ion chromatograms (nwXICs) which have been generated for the analyte (de)protonated molecule at LE and for fragments at HE in order to check that all *m/z* appear at the same retention time and with same chromatographic peak shape.

However, this information is not enough; it is necessary to show that these *m/z* signals correspond to real chemical structures. MassLynx V 4.1 allows to calculate the elemental composition of a determinate *m/z* signal, thus, chemical formula of fragments is obtained. The problem is that in most cases several elemental composition options are possible and only one is suitable to propose a fragmentation chemical structure. Structure elucidation is not an easy task overall in flavonoid compounds; fortunately, fragmentation patterns of some compounds studied in this work were found in literature: apigenin, ampelopsin, taxifolin and quercitrin [34] ; naringin and neohesperidin [35]; epicatechin, epigallocatechin and (+) catechin [36]. Fragmentation patterns of the remaining compounds have been proposed.

The chromatograms, spectra and structures of the proposed fragments of the other compounds are found in *Supplementary information 1*.

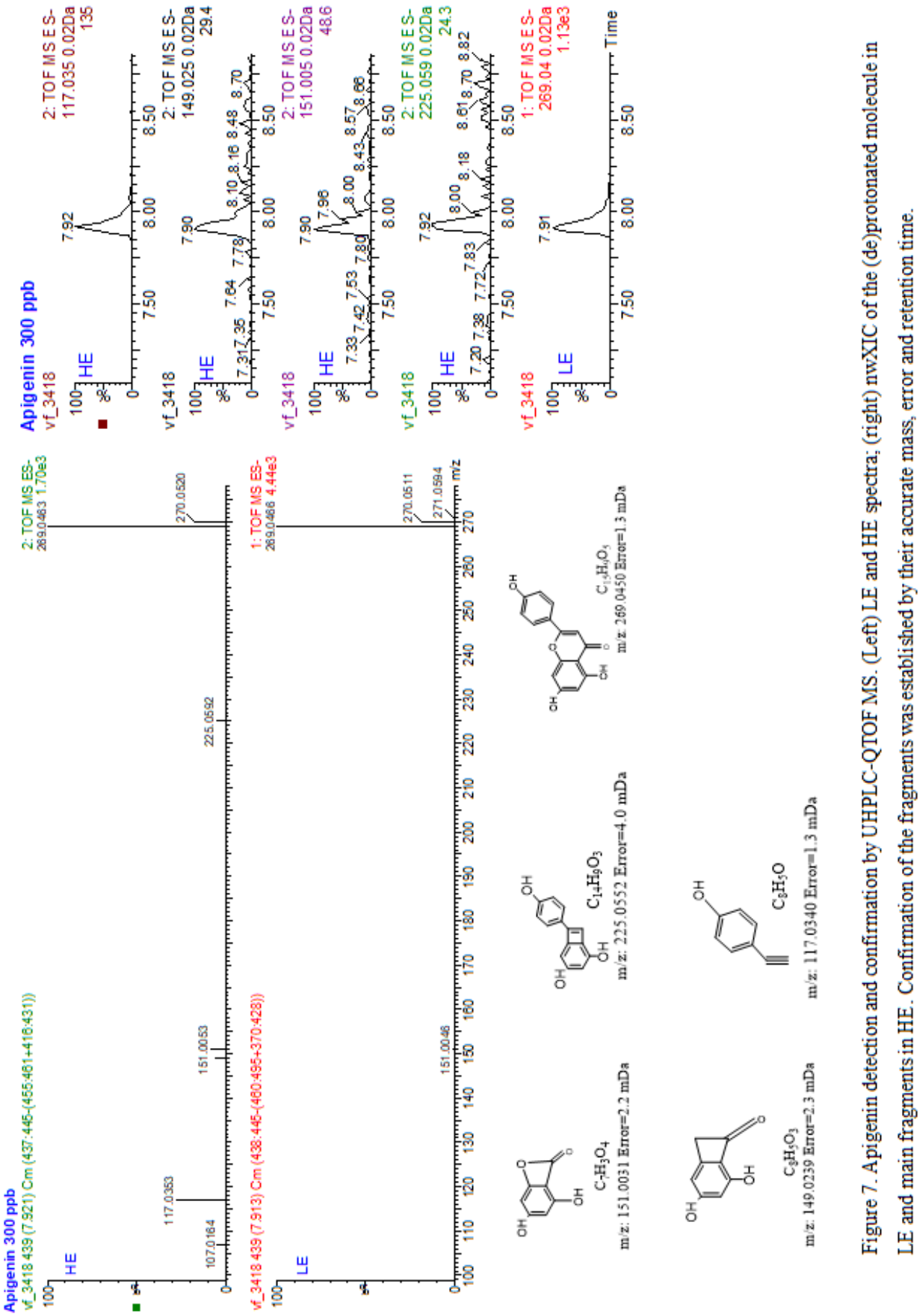


Figure 7. Apigenin detection and confirmation by UHPLC-QTOF MS. (Left) LE and HE spectra; (right) nwXIC of the (de)protonated molecule in LE and main fragments in HE. Confirmation of the fragments was established by their accurate mass, error and retention time.

Tipically, maximum error allowed between theoretical mass fragments and empirical *mass* is 5 ppm. In *Supplementary information 1* this error is shown in labels of the compounds structures pictures, expressed in mDa.

$$Error\ (ppm) = \frac{\Delta m\ (mDa)}{m} \cdot 10^6$$

This limit has not been into account since it was too much restrictive for results obtained at these MS conditions. In some cases, the sensitivity was very low, which markedly increased the error. It has been accepted in negative mode until 5.0 mDa and in positive until 15 mDa.

All the information presented in the *Supplementary information 1*, is summarized in Table 1 to facilitate a rapid identification of these compounds in a metabolome.

Table 1. List of compounds included in the empirical spectral library and accurate mass measurements for up to the five most abundant ions (higher than 5%) in the HE spectra (mass measurements are ordered according to their normalised ion abundance). Positive or negative ion mode (ESI), accurate mass measurement for (de)protonated molecule, retention time and molecular formula are also shown. The main fragment ions in the LE function are marked with *

Noun	Molecular formula	ESI	(de)protonated *m/z* ion	rt (min)	Fragment *m/z* ion
Kaempferol	$C_{15}H_{10}O_6$	-	285.0399	7.88	-
Apigenin	$C_{15}H_{10}O_5$	-	269.0450	7.91	117.0340>151.0031>149.0239>225.0552
Neohesperedin	$C_{28}H_{34}O_{15}$	-	609.1819	6.4	301.0712>323.0556>286.0477
Ferulic Acid	$C_{10}H_{10}O_4$	-	193.0501	4.92	134.0368*>133.0290>149.0603>89.0391
Caffeic Acid	$C_9H_8O_4$	-	179.0344	3.40	135.0446>134.0368
Chlorogenic Acid	$C_{16}H_{18}O_9$	-	353.0873	3.75	191.0556>>>85.0290>133.0290>179.0344
3-Dehidroshikimate	$C_7H_8O_5$	-	171.0293	0.74	109.0290>>108.0211>127.0395>111.0082>153.0188
4-Coumaryl alcohol	$C_9H_{10}O_2$	-	149.0603	3.57	130.0419>131.0497
Capsaicin	$C_{18}H_{27}NO_3$	+	306.2069	8.24	328.1898*>182.1565>137.0624>122.0399
(-) Epicatechin	$C_{15}H_{14}O_6$	-	289.0712	4.00	245.0814>137.0239>123.0446>109.0290>205.0501
Ampelopsin	$C_{15}H_{12}O_8$	-	319.0454	3.73	193.0137>151.0031>178.9980>125.0239>137.0239
Taxifolin	$C_{15}H_{12}O_7$	-	303.0505	5.01	125.0239>285.0399>175.0395>199.0395>301.0348
Gentisic Acid	$C_7H_6O_4$	-	153.0188	3.46	108.0239>>152.0110
Coumarine	$C_9H_6O_2$	-	147.0446	5.1	103.0548>91.0548
Quercitrin	$C_{21}H_{20}O_{11}$	-	447.0927	6.69	300.0270>301.0348>271.0243>255.0293>151.0031
Epigallocatechin	$C_{15}H_{14}O_7$	-	305.0661	2,98	125.0239>137.0239>165.0188>173.0344
Skimmianine	$C_{14}H_{13}NO_4$	+	260.0923	7.68	227.0582>199.0633>184.0399>245.0688>216.0661
Catechin	$C_{15}H_{14}O_6$	-	289.0712	2.97	245.0814>137.0239>203.0708>123.0446>109.0290
Sinapic Acid	$C_{11}H_{12}O_5$	-	223.0606	5.17	121.0290>149.0239>164.0473>193.0137>208.0372
Naringin	$C_{27}H_{32}O_{14}$	-	579.1704	6.13	271.0606>151.0031>459.1139>293.0450>119.0497

4.2. Validation. Extraction efficiency for the compounds under study

The selection of an extraction method for a full profiling metabolomics is not an easy decision due to the complexity of plant matrix, the great variety of compounds present, and their different concentration levels. The more appropriate choice is the selection of a method which would be used for a greatest variety of hormones [4,5]. The extraction method used in this project is the same used in the literature for similar experiments [31, 32]. However, it is unknown if this extraction method is valid for the compounds studied in this work.

In this section we will study two issues: on the one hand, we will study if the matrix of the plant affects the identification and detection of the compounds of the library; on the other hand we will study the process efficiency.

The process efficiency was determined by comparing plant material samples fortified with a stock solution of a mix of the standards before extraction at 150 and 500 μg/kg (n= 3 for each standard level), with solutions containing a mix of pure standards at the same levels [37]. Zero point extraction is made in order to substrate compound intensity to fortified sample in case of having basal level of this hormone. The extraction was made as explained in Section 3.3 and experiment was performed with two kinds of plant matrix, *Arabidopsis thaliana* and *Solanum lycopersicum* (tomato).

Once the extraction has been carried out, it must be verified that the compounds are present in the fortified samples (*Supplementary information 2).* On the one hand, the chromatograms of a 500 ppb standard and the fortified sample at the corresponding concentration level were compared; an example is shown in Figure 8. If both chromate present a peak at the same retention time, the identification and determination of the compound is done as shown in section 4.1. Information obtained in *Supplementary information 2* is collected in Table 2

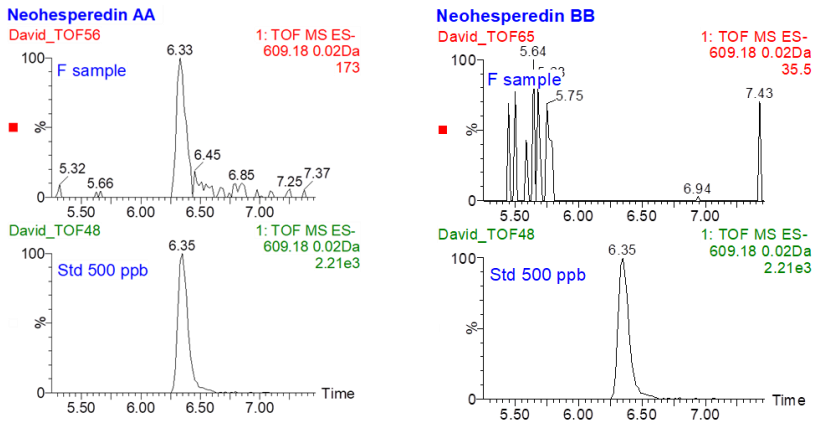


Figure 8. Comparison between nwXICs for neohesperidin in the fortified plant and the standard at the same level. AA is referred to *Arabidopsis thaliana* and BB to *Solanum lycopersicum*. As it can be seen, the peak only appears in the fortified sample of *Arabidopsis thaliana.*

Table 2. List of compounds included in the empirical library. Accurate mass measurements for confirmed fragment ions with its mass error (in ppm). Hyphens represent no identified compounds and asterisk indicates that only (de)protonated molecule *m/z* was shown (no fragmentation)

Compound	**Fragment *m/z* ion**	
	Arabidopsis thaliana	*Solanum lycopersicum*
Kaempferol	-	-
Apigenin	-	-
Neohesperedin	301.0712 (4 ppm)	-
Ferulic Acid	134.0368 (15 ppm) >178.0266 (14 ppm)>149.0603 (8 ppm)	-
Caffeic Acid	135.0446 (9 ppm)	-
Chlorogenic Acid	191.0556 (5 ppm)	191.0556 (0 ppm)
3-Dehidroshikimate	-	-
4-Coumaryl alcohol	131.0497 (0 ppm)	131.0497 (11 ppm)
Capsaicin	137.0603 (15 ppm)	137.0603 (9 ppm)
(-) Epicatechin	205.0501 (7 ppm)>245.0812 (8 ppm)	-
Ampelopsin	-	-
Taxifolin	285.0399 (-19 ppm)	285.0399 (13 ppm)
Gentisic Acid	-	-
Coumarine	103.0548(13 ppm)>91.0548(13 ppm)	103.0548 (11 ppm)
Quercitrin	301.0348 (2 ppm)	*
Epigallocatechin	125.0239 (-12 ppm)	
Skimmianine	227.0582(-4 ppm) >199.0633(2 ppm) >184.0399(5 ppm)>245.0688 (1 ppm)>216.0661(-2 ppm)	227.0582(7 ppm)>199.0633(4 ppm)>184.0399 (9 ppm)>245.0688(-21 ppm)>216.0661(5 ppm)
Catechin	245.0812 (4 ppm)	-
Sinapic Acid	193.0137 (4 ppm) >164.0473 (7 ppm)>149.0250 (7 ppm)> 208.0372(3 ppm)> 121.0300 (8 ppm)	-
Naringin	271.0606 (5 ppm) >459.1139 (4 ppm)>151.0031(10ppm)	271.0606(10 ppm) >459.1139 (9 ppm)

In Table 3 it is shown what compounds have been extracted after the fortification. Extraction procedure works fairly well in *Arabidopsis* compared to *Solanum;* however, kaempferol, apigenin, 3-dehydroshikimate and gentisic acid could not been identified in any of the both matrix. On the other hand, the identification criterion that is usually used consists of the presence of two masses (de)protonated and fragmentss) in either the LE or HE function measured with a mass error lower than 5 ppm). In most of the detected compounds the two masses are found, however, the mass errors are greater than those established by the criterion. High errors are usually linked to low peak intensity.

XCMS package of R was used for peak area quantification. Results are collected at *Supplementary information 3*. Efficiency extraction results are collected in Table 3.

Table 4. Extraction efficiency at the two different concentration levels assayed.

Compound	Arabidopsis				Solanum			
	150 ppb		500 ppb		150 ppb		500 ppb	
	Ext eff (%)	RSD (%)	Ext eff (%)	RSD (%)	Ext eff (%)	RSD (%)	Ext eff (%)	RSD (%)
Kaempferol	-	-	-	-	-	-	-	-
Apigenin	-	-	-	-	-	-	-	-
Neohesperedin	8.23	11.76	7.54	1.95	-	-	-	-
Ferulic Acid	23.61	43.24	37.42	1.51	-	-	-	-
Caffeic Acid	36.56	48.25	49.63	11.14	-	-	-	-
Chlorogenic Acid	27.26	34.25	29.27	12.18	*	*	*	*
3-Dehidroshikimate	-	-	-	-	-	-	-	-
4-Coumaryl alcohol	54.95	18.26	51.15	13.15	33.33	15.64	21.24	4.35
Capsaicin	37.23	38.31	29.33	38.50	23.47	38.86	24.75	4.38
(-) Epicatechin	14.72	52.38	16.41	21.77	-	-	-	-
Ampelopsin	-	-	-	-	-	-	-	-
Taxifolin	23.22	27.00	23.35	7.96	8.19	38.65	6.44	7.73
Gentisic Acid	-	-	-	-	-	-	-	-
Coumarin	95.36	17.19	78.18	14.07	94.68	19.73	96.51	7.52
Quercitrin	17.65	7.63	16.95	14.32	14.40	19.54	13.48	18.21
Epigallocatechin	10.85		15.14		-	-	-	-
Skimmianine	45.62	35.99	39.38	39.02	35.63	36.74	34.85	25.53
Catechin	10.85	66.34	15.14	20.76	-	-	-	-
Sinapic Acid	*	*	*	*	-	-	-	-
Naringin	35.32	7.42	36.53	4.66	37.23	15.53	27.35	1.31

- Unidentified compounds
- The "blank" sample has more analyte concentration than the spiked, so that these compounds could not be validated at this level

In view of the results obtained it is necessary to stress some points. Extraction efficiency results were really variable between compounds; however, different levels of the same compound present the same efficiency. Therefore, it could be claimed that the concentration level seems not affect to the process efficiency results but nature of compound it does. However, RSD (%) values are really high; thus, it could be interesting repeat the experiment with a greater number of samples.

In section 2 it was explained that phenolic compounds usually form adducts with proteins which difficult the analysis, so that, compounds with less extraction efficiency could be more affected by matrix. This leads to another point, flavonoinds (kaempferol, apigenin, neohesperedin, (-) epicatechin, ampelopsin, taxifolin, quercitrin, epigallocatechin, catechin and naringin) were prone to have very low extraction efficiency and even, in some cases, were not detected. In *Solanum* only 8 compounds have been semi-quantified, there were not a correlation between extraction efficiency values in *Arabidopsis* and *Solanum*, so that, compound-matrix interaction seems different in both plant.

The results of this work proved that flavonoids presented extraction efficiencies lower than 35%. Therefore, for a flavonoid targeted analysis, other extraction methods should be used. For example, an extraction method using fractional methanol extractions with ultra sounds and solid extraction (SE) with LC-18 cartridge works very well for hesperidin, hesperetin, quercetin, kaempferol and caffeine [38].

4.3. Case study. The research of phenolic compounds in I3CA treated adults and seedling metabolome

Planting of *Arabidopsis* plants, treatment and infection is explained at Section 3.2. and metabolome extraction at Section 3.3.

Peaks intensities were quantified and separated according to treatment by XCMS package in R. Mar-Vis Suite was used deploying phenolic library in order to find which compounds present more variability between treatments. Nevertheless, any compound of the list was matched by Mar-Vis Suite, so that, a manual research was carried out with data extracted of XCMS package.

Sinapic acid (Figure 9) and caffeic acid (Figure 10) were detected and confirmed. In sinapic acid three fragments were detected together with the (de)protoned molecular with low mass error (>2 mDa). There are two fragments that were in doubt because their chromatogram present different shape compared to the (de)protoned molecular peak. In caffeic acid, only one fragment was detected and confirmed and its mass error is acceptable.

Figures 11 and 12 show a bar graph, where the bar height represent the average intensity of each treatment. Uncertainty is also represented on the graph as error bar. There are two graphic bars per compound one for adult plants and one for seedings since the aim of the work is to see different metabolite behaviour between both.

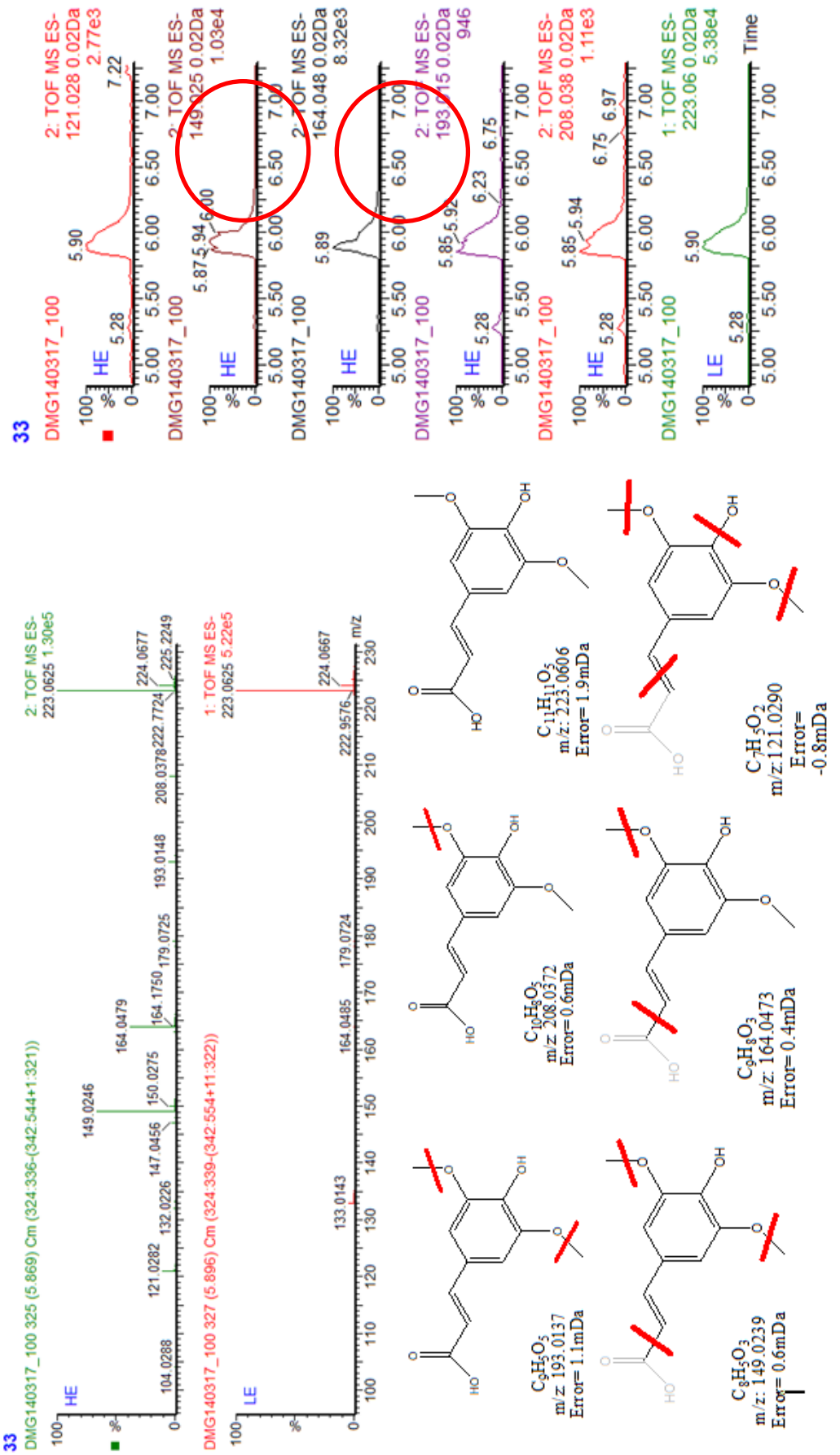


Figure 9. Sinapic acid detection and confirmation by UHPLC-QTOF MS. (Left) LE and HE spectra; (right) nw-XIC of the (de)protonated molecular in LE and main fragments in HE. Confirmation of the fragments was established by their exact mass, mass error and retention time.

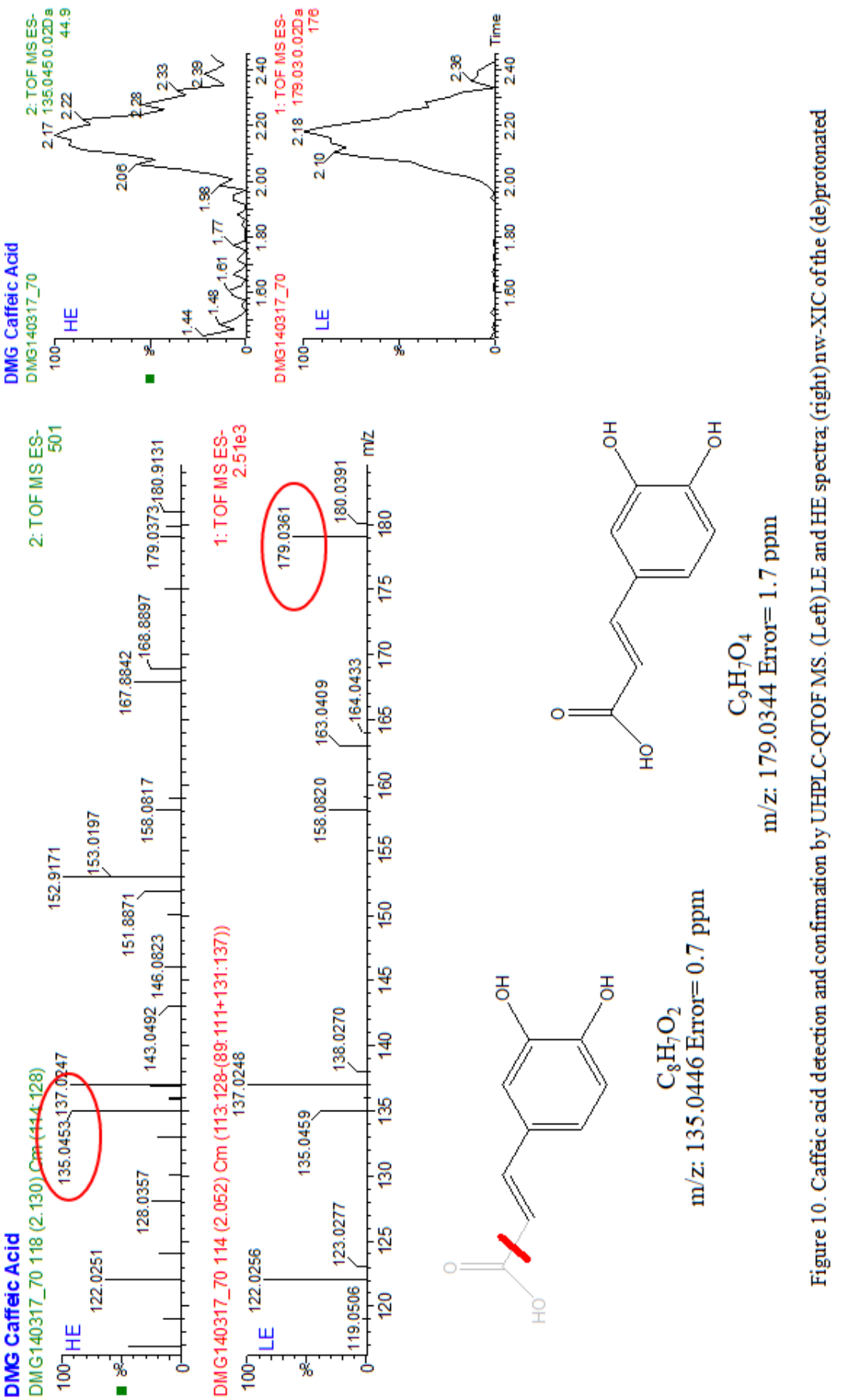


Figure 10. Caffeic acid detection and confirmation by UHPLC-QTOF MS. (Left) LE and HE spectra; (right) nw-XIC of the (de)protonated molecular in LE and fragment in HE. Confirmation of the fragment was established by their exact mass, mass error and retention time.

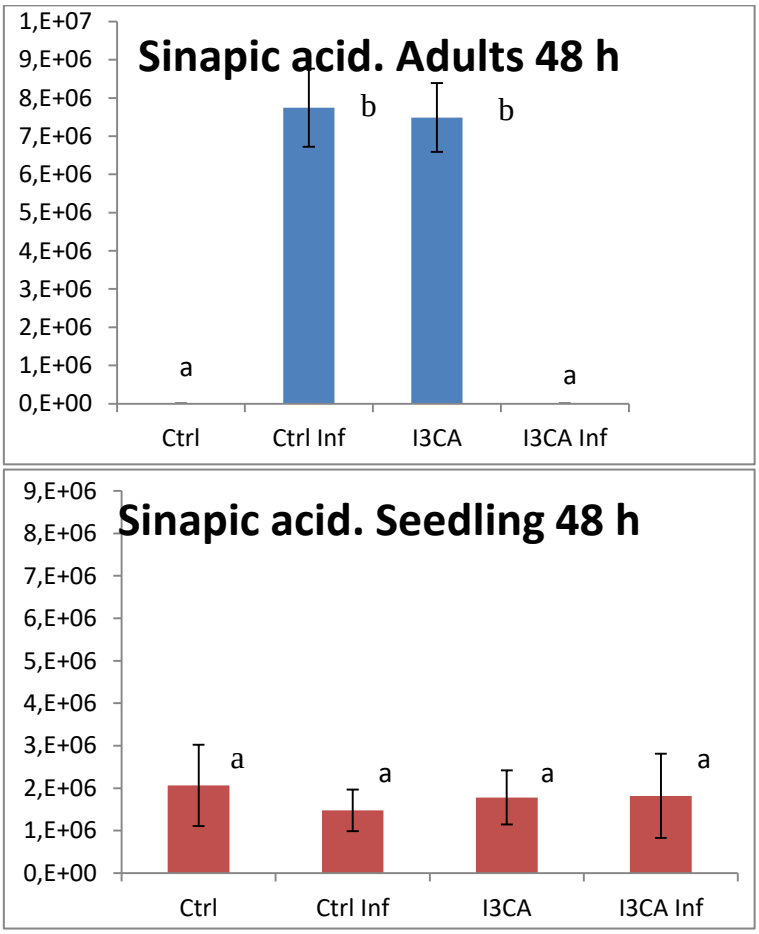


Figure 11. Semi-quantification of sinapic acid against *P.cucumerina* in water and I3CA treated plants. Figure shows the average of six chromatographics measuraments (three biological and three technical) for Col-0 water treated mock plants (Ctrl), Col-0 water treated *P.cucumerina* infected plants (Ctrl Inf), I3CA treated mock plants (I3CA) and Col-0 I3CA treated *P.cucumerina* infected plants (I3CA Inf). The lowercase letters represent the classes obtained by the ANOVA analysis.

Semi-quantitative data of sinapic acid were statistically analysed by ANOVA. In adults, mock plants had low concentration levels of sinapic acid in comparision with the infected plant. However, when plant is treated with I3CA, control plant have high levels in comparision with infected plants. In seeding, in contrast, there were not differences between treatments. Sinapic acid could play an important role in plant

mechanism defense aganist *P. cucumerina* in adults. I3CA could increase sinapic acid levels in uninfected plant as a reservoir for rapid response when infection occurs. It is known that I3CA is involved in a metabolic pathway which induce callose deposition in adults after infection; however in seedling this callose levels do not change in comparison with no treated plants. It could be interesting to check if sinapic acid have any link with metabolic pathway related with the building of this constitutive barrier.

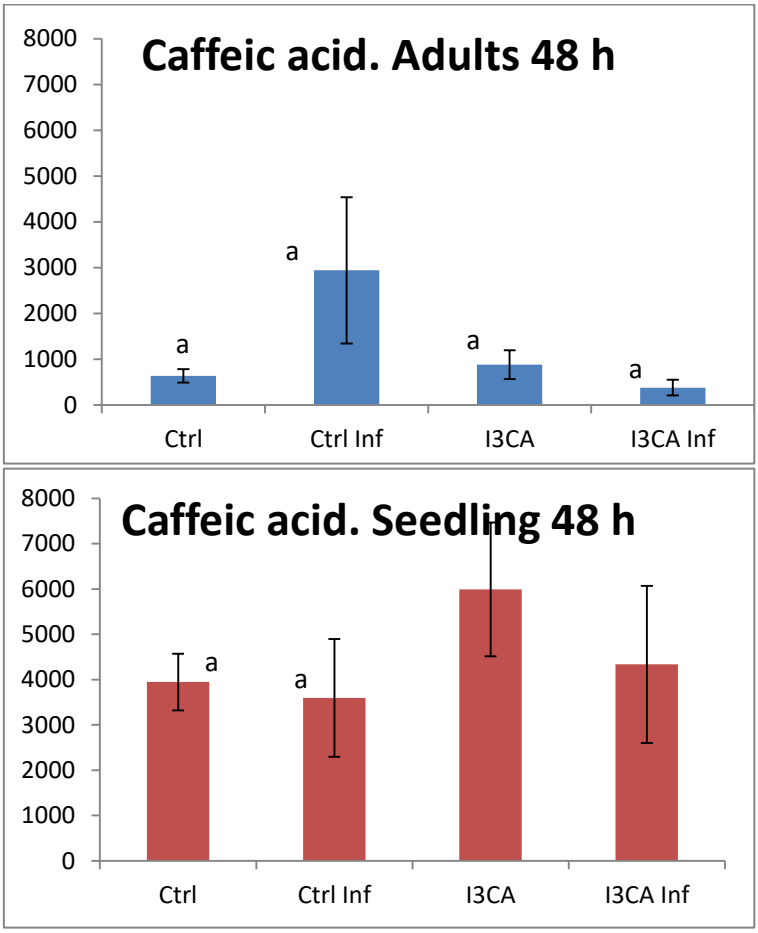


Figure 12. Semi-quantification of caffeic acid against *P.cucumerina* in water and I3CA treated plants. Figure shows the average of six chromatographics measuraments (three biological and three technical) for Col-0 water treated mock plants (Ctrl), Col-0 water treated *P.cucumerina* infected plants (Ctrl Inf), I3CA treated mock plants (I3CA) and Col-0 I3CA treated *P.cucumerina* infected plants (I3CA Inf). The lowercase letters represent the classes obtained by the ANOVA analysis.

The ANOVA analysis shows no difference in the levels of caffeic acid in the treatments in either case.

5. Conclusions

An home-made compound spectral library has been built, with the main fragment ions reported, for 20 phenolic compounds. Compounds of this list could be rapidly detected and confirmed since this library is included MS^E which understand mass spectrums at low and high collision energy. Incorporating the information gotten to a more extensive hormone spectral library and using biological software processing data, a rapid screening could be made in non-targeted metabolomics analysis; without the need of analyse reference standard in each analysis. However, the library should be repeated at a higher concentration and the well-calibrated instrument to avoid large mass errors.

A validation was performed in order to check if extraction procedure has any effect on the detection and confirmation of compounds, the study of extraction efficiency was made in *Arabidopsis thaliana* and *Solanum* (object of analysis of this kind of metabolome). In *Arabidopsis* fortified plants, 17 out of 20 compounds were detected and confirmed; however, in *Solanum* plants only 8. The results of extraction efficiency presented a wide range of values which depended on the interaction compound-matrix, generally, really high for flavonoid compounds. The limitation of a full profiling non-targeted metabolome is that the extraction method selected have to be useful for a huge range of compounds of different chemical nature. Therefore, if targeted flavonoids analysis method is desired, another extraction methods should be used as solid extraction (SE). On the other hand, detection and confirmation of compunds in plant matrix was complicated due to the high mass error which do not fulfil with identification criteria; and repetiblity was really low (high RSD(%)). Validation should be repeated with greater number of replica.

Finally, the in-house spectral library was used in a full profiling non-targeted metabolome analysis on a real experiment which consisted in the different behaviour between adults and seeding I3CA treated plants infected with *Plectosphorela cucumerina*. Sinapic acid was detected and confirmed and semi-quantified in the different treatments in both cases. While in seeding plants there are no differences

between treatments, sinapic acid seems to play a role in *defense* mechanism on I3CA treated adult plants.

6. Acknowledgements

I am very grateful for the help received from Dr. María Victoria Pastor Fuentes when structuring the project. I would also like to thank PhD student Diego Mateu García for helping me with the biological part and Dr. Victor Flors Herrero for supervising the work.

Thank also to Dr. Maria Ibañez Martinez for its dedication and support during the development of this work and for advising me in the whole analytical part.

7. References

[1] Rathahao-Paris, E., Alves, S., Junot, C., & Tabet, J. C. (2016). High resolution mass spectrometry for structural identification of metabolites in metabolomics. Metabolomics, 12(1), 10.

[2] Leavitt, Sarah A. (June 2010). "Deciphering the Genetic Code: Marshall Nirenberg". Office of NIH History.

[3] Kuehnbaum, N. L., & Britz-McKibbin, P. (2013). New advances in separation science for metabolomics: resolving chemical diversity in a post-genomic era. *Chemical Reviews*, *113*(4), 2437-2468.

[4] Allwood, J. W., & Goodacre, R. (2010). An introduction to liquid chromatography–mass spectrometry instrumentation applied in plant metabolomic analyses. *Phytochemical Analysis*, *21*(1), 33-47.

[5] Gamir, J., Agut B., Pastor, V., & Flors, V. (2013). Targeted and non-targeted chromathographic analysis applied to plant resistance research. *Recent Res. Plant Physol., 5*

[6] Oksman-Caldentey K-M, Inzé D. 2004. Plant cell factories in the post-genomics era: new ways to produce designer secondary metabolites. *Trends Plant Sci 9*: 433–440.

[7] Skoog, Holler & Nieman. *Principios de Análisis Instrumental.* Mc Graw Hill 2003

[8] Shushan, B. (2010). A review of clinical diagnostic applications of liquid chromatography–tandem mass spectrometry. *Mass spectrometry reviews*, *29*(6), 930-944.

[9] Müller, M., & Munné-Bosch, S. (2011). Rapid and sensitive hormonal profiling of complex plant samples by liquid chromatography coupled to electrospray ionization tandem mass spectrometry. Plant methods, 7(1), 37.

[10] Tolstikov VV, Lommen A, Nakanishi K, Tanaka N, Fiehn O. 2003. Monolithic silica-based capillary reversed-phase liquid chromatography/electro-spray mass spectrometry for plant metabolomics. Anal Chem 75:

[11] Rischer H, Orešič M, Seppänen-Laakso T, Katajamaa M, Lammertyn F,Ardiles-Diaz W, Van Montagu MCE, Inzé D, Oksman-Caldentey KM,Goossens A. 2006. Gene-to-metabolite networks for terpenoid indolealkaloid biosynthesis in Catharanthus roseus cells. Proc Natl Acad SciUSA 103: 5614–5619.

[12] De Vos CHR, Moco S, Lommen A, Keurentjes JJB, Bino RJ, Hall RD. 2007. Untargeted large-scale plant metabolomics using liquid chromatog- raphy coupled to mass spectrometry. Nat Protoc 2: 778–791.

[13] Niessen WM. 1998. Advances in instrumentation in liquid chromatogra-phy–mass spectrometry and related liquid-introduction techniques.J Chromatogr A 794: 407–435

[14]. J.H. Gross, Mass Spectrometrhy: a textbook. Springer 2004

[15] E.Hoffmann and V. Stroobant, Mass Spectrometry. Viley, 2002.

[16] Banerjee, S., & Mazumdar, S. (2012). Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte. *International journal of analytical chemistry*, *2012*.

[17] Gaskell SJ. 1997. Electrospray: principles and practice. J Mass Spectrom 32:677–688.

[18] King R, Bonfi glio R, Fernandez-Metzler, Miller-Stein C, Olah T. 2000.Mechanistic investigation of ionization suppression in electrosprayionization. J Am Soc Mass Spectrom 11: 942–950.

[19] Lee, D. Y., Bowen, B. P., & Northen, T. R. (2010). Mass spectrometry—based metabolomics, analysis of metabolite-protein interactions, and imaging. *Biotechniques*, *49*(2), 557-565.

[20] Kaever, A., Landesfeind, M., Feussner, K., Mosblech, A., Heilmann, I., Morgenstern, B., ... & Meinicke, P. (2015). MarVis-Pathway: integrative and exploratory pathway analysis of non-targeted metabolomics data. Metabolomics, 11(3), 764-777.

[21] Schymanski, E. L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H. P., & Hollender, J. (2014). Identifying small molecules via high resolution mass spectrometry: communicating confidence.

[22] Buchanan, Gruissen & Jones. *Biochemistry & Molecular Biology of Plants*. American Society of Plants Physiology 2000.

[23] Bhattacharya, A., Sood, P., & Citovsky, V. (2010). The roles of plant phenolics in *defense* and communication during Agrobacterium and Rhizobium infection. *Molecular plant pathology*, *11*(5), 705-719.

[24] Treutter, D. (2005). Significance of flavonoids in plant resistance and enhancement of their biosynthesis. *Plant biology*, *7*(6), 581-591.

[25] Schmitz-Hoerner, R., & Weissenböck, G. (2003). Contribution of phenolic compounds to the UV-B screening capacity of developing barley primary leaves in relation to DNA damage and repair under elevated UV-B levels. *Phytochemistry*, *64*(1), 243-255.

[26] Kefeli, V. I., Kalevitch, M. V., & Borsari, B. (2003). Phenolic cycle in plants and environment. *J. Cell Mol. Biol*, *2*(1), 13-18.

[27] Pastor, V., Balmer, A., Gamir, J., Flors, V., & Mauch-Mani, B. (2014). Preparing to fight back: generation and storage of priming compounds. *Frontiers in plant science*, *5*, 295.

[28] Conrath, U., Beckers, G. J., Flors, V., García-Agustín, P., Jakab, G., Mauch, F., ... & Pugin, A. (2006). Priming: getting ready for battle. *Molecular Plant-Microbe Interactions*, *19*(10), 1062-1071.

[29] Balmer, A., Pastor, V., Gamir, J., Flors, V., & Mauch-Mani, B. (2015). The 'prime-ome': towards a holistic approach to priming. Trends in plant science, 20(7), 443-452.

[30] Llorente, Francisco, et al. "Repression of the auxin response pathway increases Arabidopsis susceptibility to necrotrophic fungi." Molecular plant 1.3 (2008): 496-509.

[31] Gamir, J., Pastor, V., Cerezo, M., & Flors, V. (2012). Identification of indole-3-carboxylic acid as mediator of priming against Plectosphaerella cucumerina. *Plant physiology and biochemistry*, *61*, 169-179.

[32] Gamir, J., Pastor, V., Kaever, A., Cerezo, M., & Flors, V. (2014). Targeting novel chemical and constitutive primed metabolites against P lectosphaerella cucumerina. *The Plant Journal*, *78*(2), 227-240.

[33] Gamir, J., Sánchez-Bel, P., & Flors, V. (2014). Molecular and physiological stages of priming: how plants prepare for environmental challenges. *Plant cell reports*, *33*(12), 1935-1949.

[34] Fabre, N., Rustan, I., de Hoffmann, E., & Quetin-Leclercq, J. (2001). Determination of flavone, flavonol, and flavanone aglycones by negative ion liquid chromatography electrospray ion trap mass spectrometry. *Journal of the American Society for Mass Spectrometry*, *12*(6), 707-715.

[35] Guo, X., Yue, Y., Tang, F., Sun, J., Wang, J., Yao, X., & Xun, H. (2013). Fragmentation Characterization and Differentiation of Isomeric Diglycosyl Flavonoids Using Ultra-High-Performance Liquid Chromatography Electrospray Ionization Quadrupole Time-of-Flight Tandem Mass Spectrometry in Negative Ion Mode. *European Journal of Mass Spectrometry*, *19*(1), 49-55.

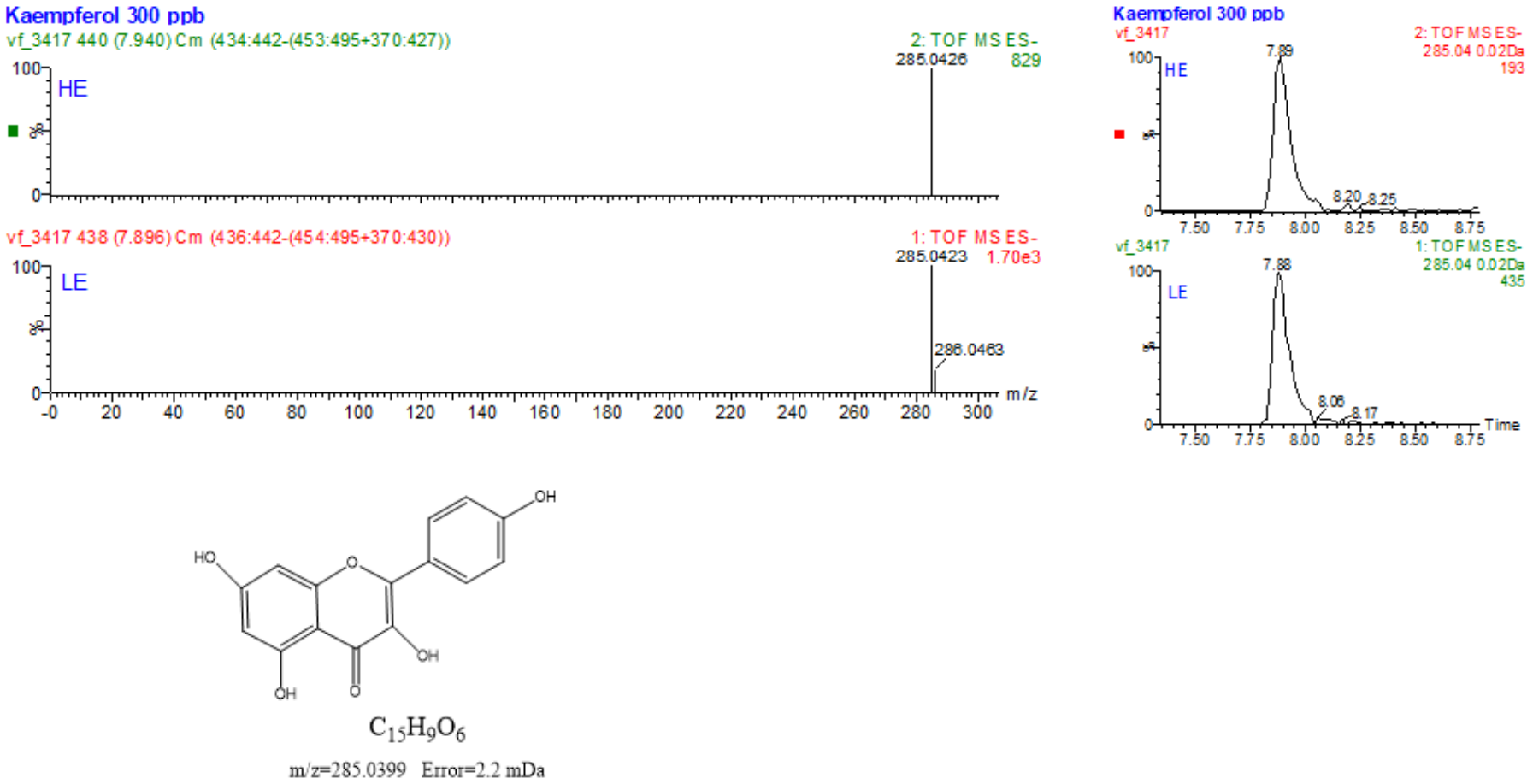
[36] Fecka, I., Kucharska, A. Z., & Kowalczyk, A. (2015). Quantification of tannins and related polyphenols in commercial products of tormentil (Potentilla tormentilla). *Phytochemical analysis*, *26*(5), 353-366.

[37] Thevenet, D., Pastor, V., Baccelli, I., Balmer, A., Vallat, A., Neier, R., ... & Mauch-Mani, B. (2017). The priming molecule β-aminobutyric acid is naturally present in plants and is induced by stress. *New Phytologist*, *213*(2), 552-559.

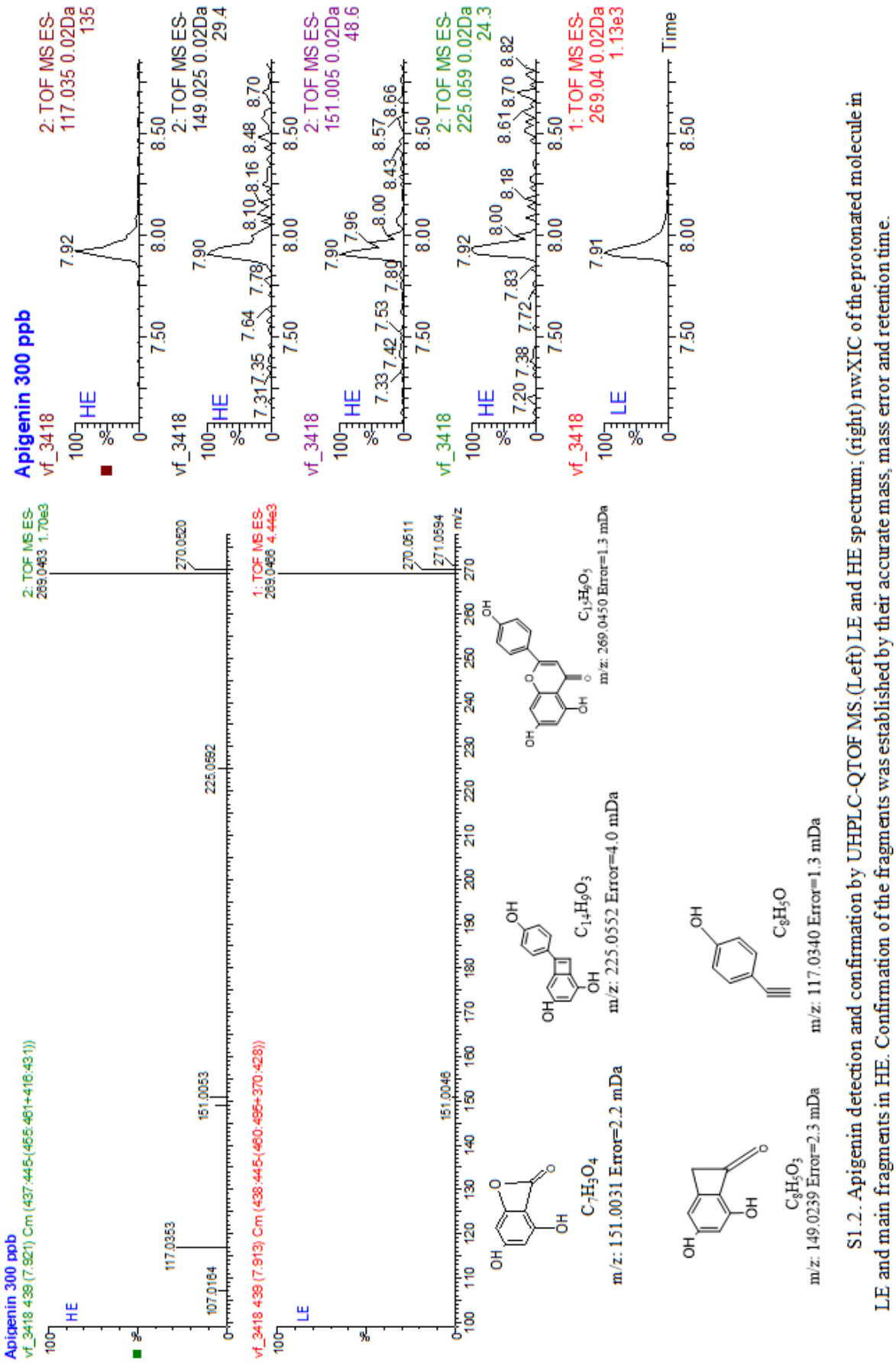
[38] Ćirić, A., Prosen, H., Jelikić-Stankov, M., & Đurđević, P. (2012). Evaluation of matrix effect in determination of some bioflavonoids in food samples by LC–MS/MS method. *Talanta*, *99*, 780-790.

SUPPLEMENTARY INFORMATION

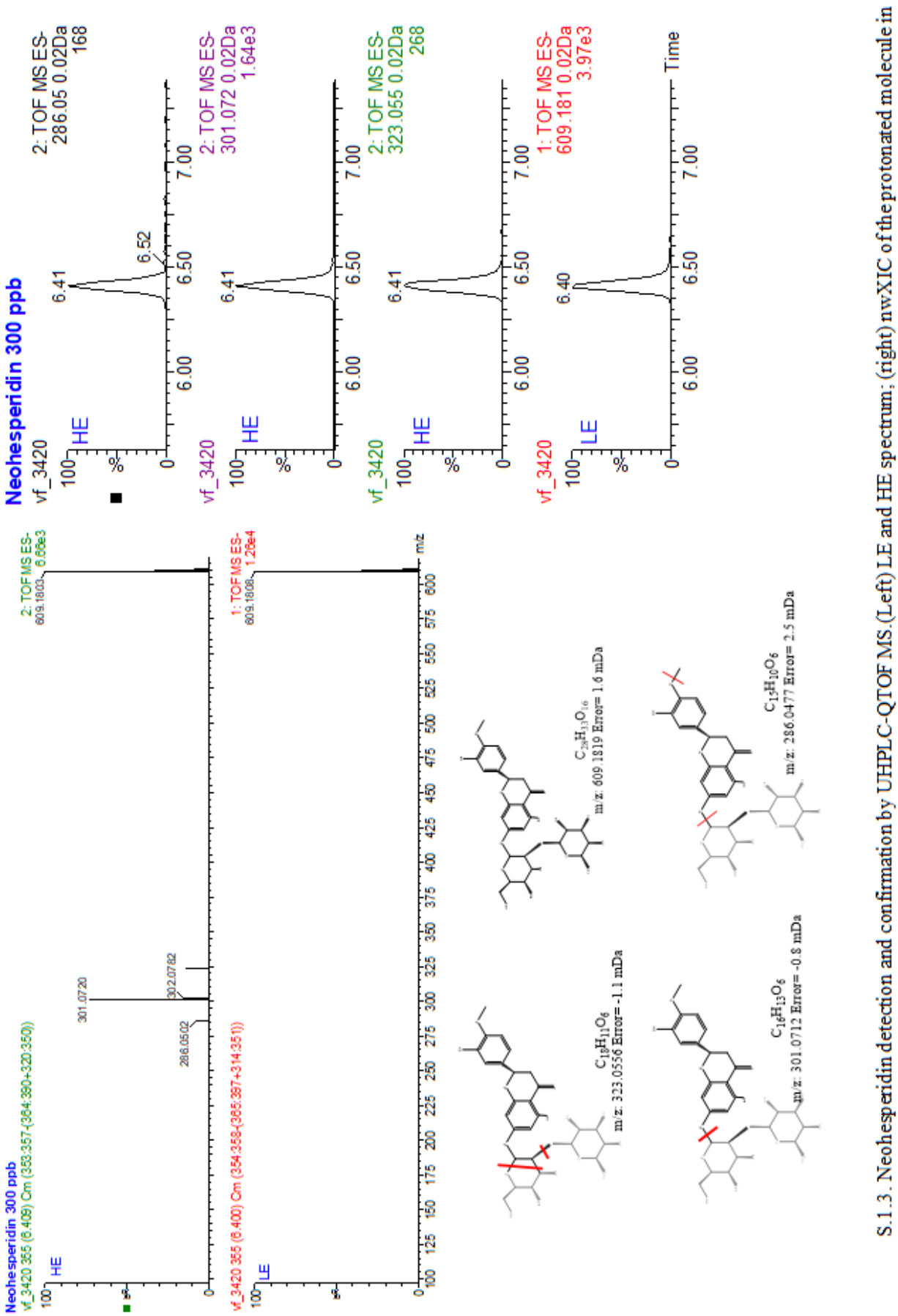
S1. Phenolic in-house database mass spectra, chromatograms and fragments structures



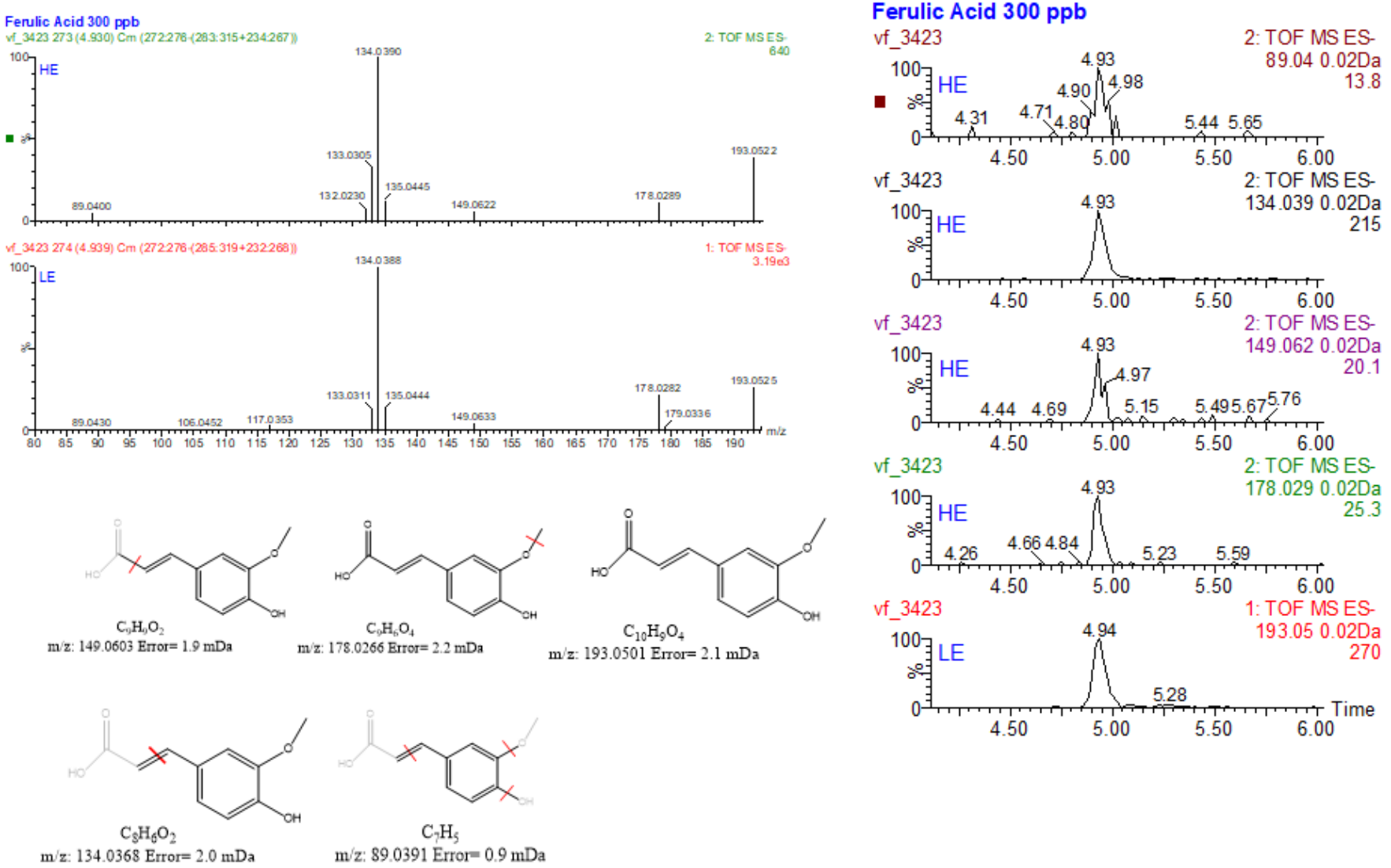
S1.1. Kaempferol detection and confirmation by UHPLC-QTOF MS.(Left) LE and HE spectra; (right) nwXIC of the protonated molecule in LE and HE. No fragmentation was observed.



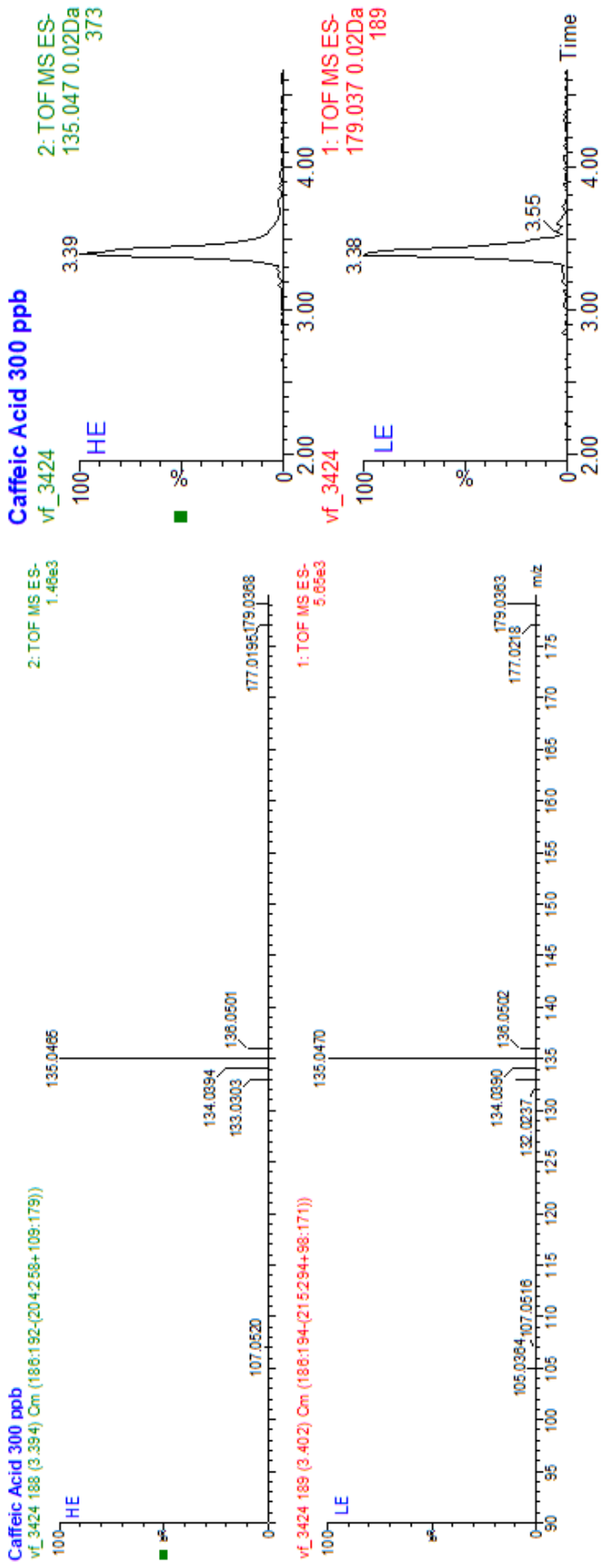
S1.2. Apigenin detection and confirmation by UHPLC-QTOF MS. (Left) LE and HE spectrum; (right) nwXIC of the protonated molecule in LE and main fragments in HE. Confirmation of the fragments was established by their accurate mass, mass error and retention time.



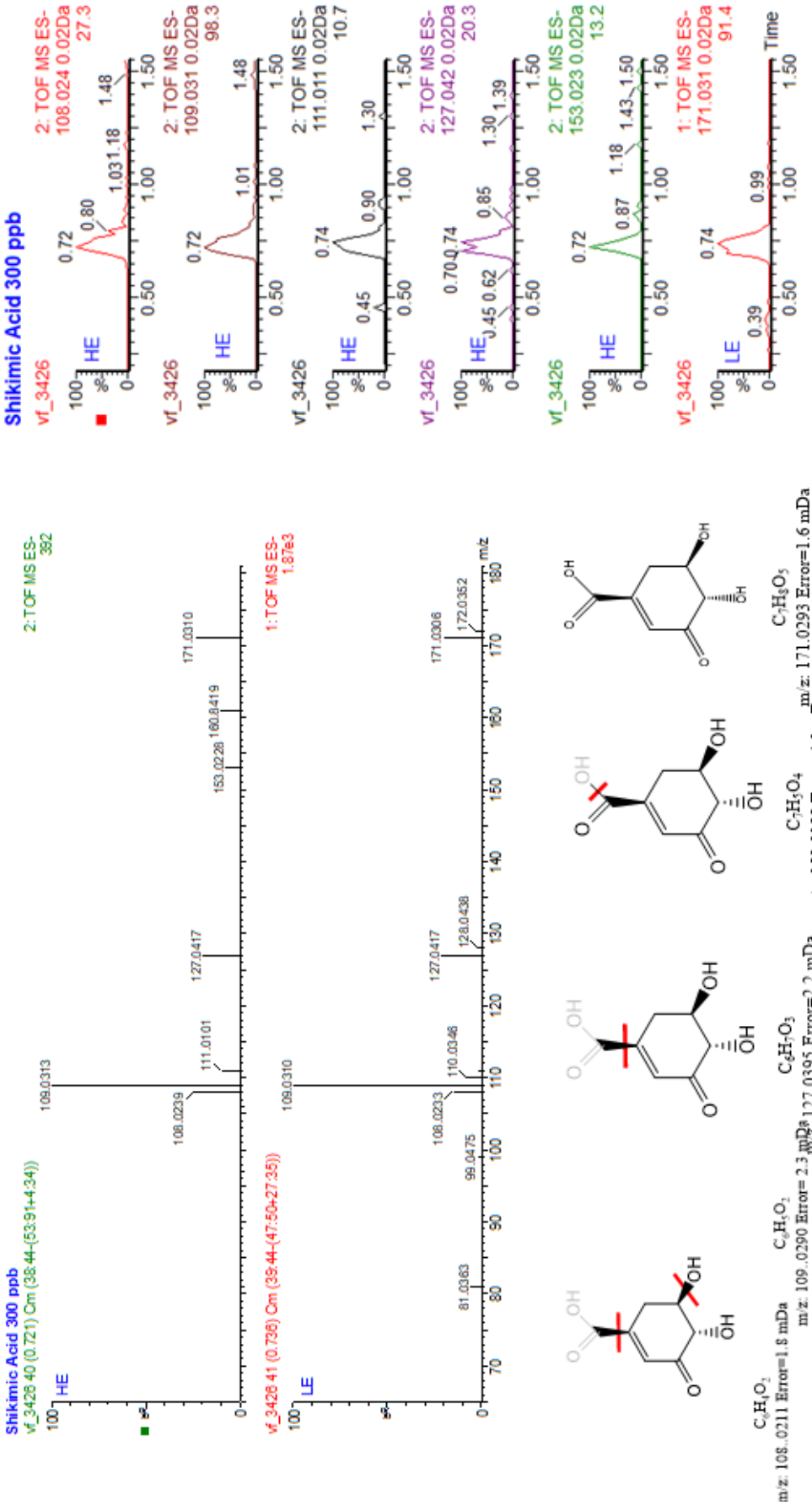
S.1.3. Neohesperidin detection and confirmation by UHPLC-QTOF MS. (Left) LE and HE spectrum; (right) nwXIC of the protonated molecule in LE and main fragments in HE. Confirmation of the fragments was established by their accurate mass, mass error and retention time.



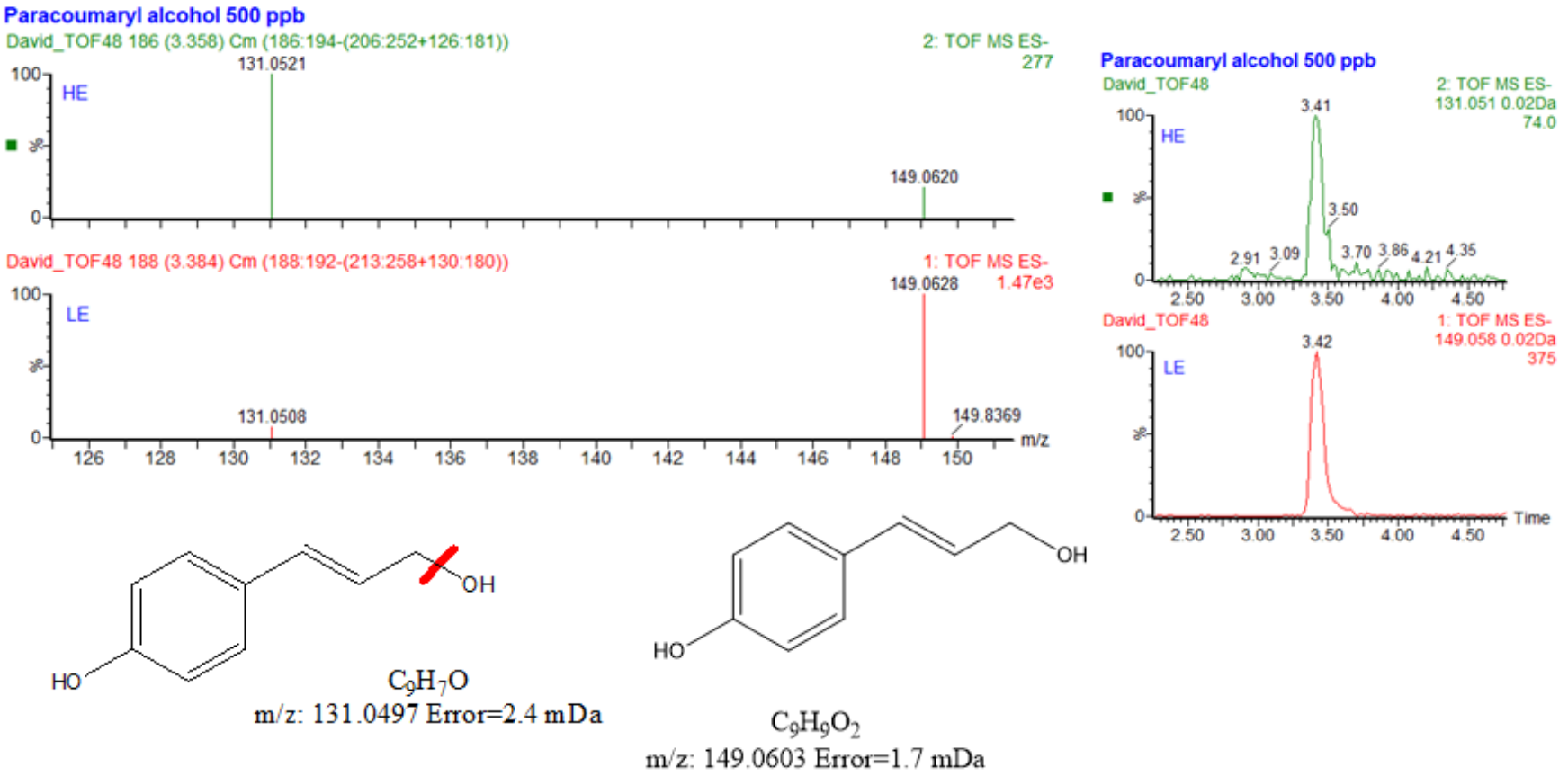
S.1.4. Ferulic acid detection and confirmation by UHPLC-QTOF MS.(Left) LE and HE spectrum; (right) nwXIC of the protonated molecule in LE and main fragments in HE. Confirmation of the fragments was established by their accurate mass, mass error and retention time.



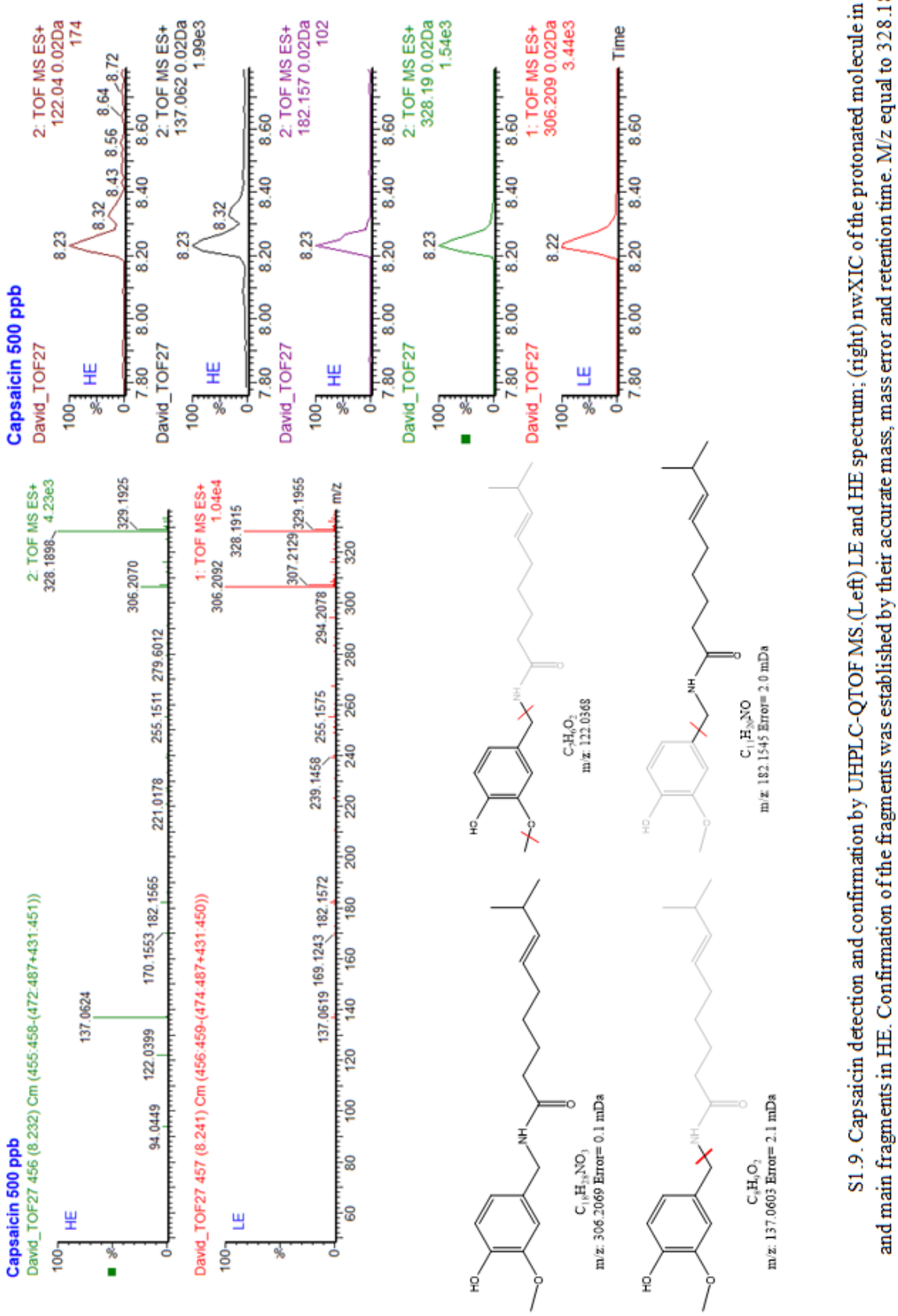
S1.5. Caffeic acid detection and confirmation by UHPLC-QTOF MS.(Left) LE and HE spectrum; (right) nwXIC of the protonated molecule in LE and fragment in HE. Confirmation of the fragment was established by their accurate mass, mass error and retention time.



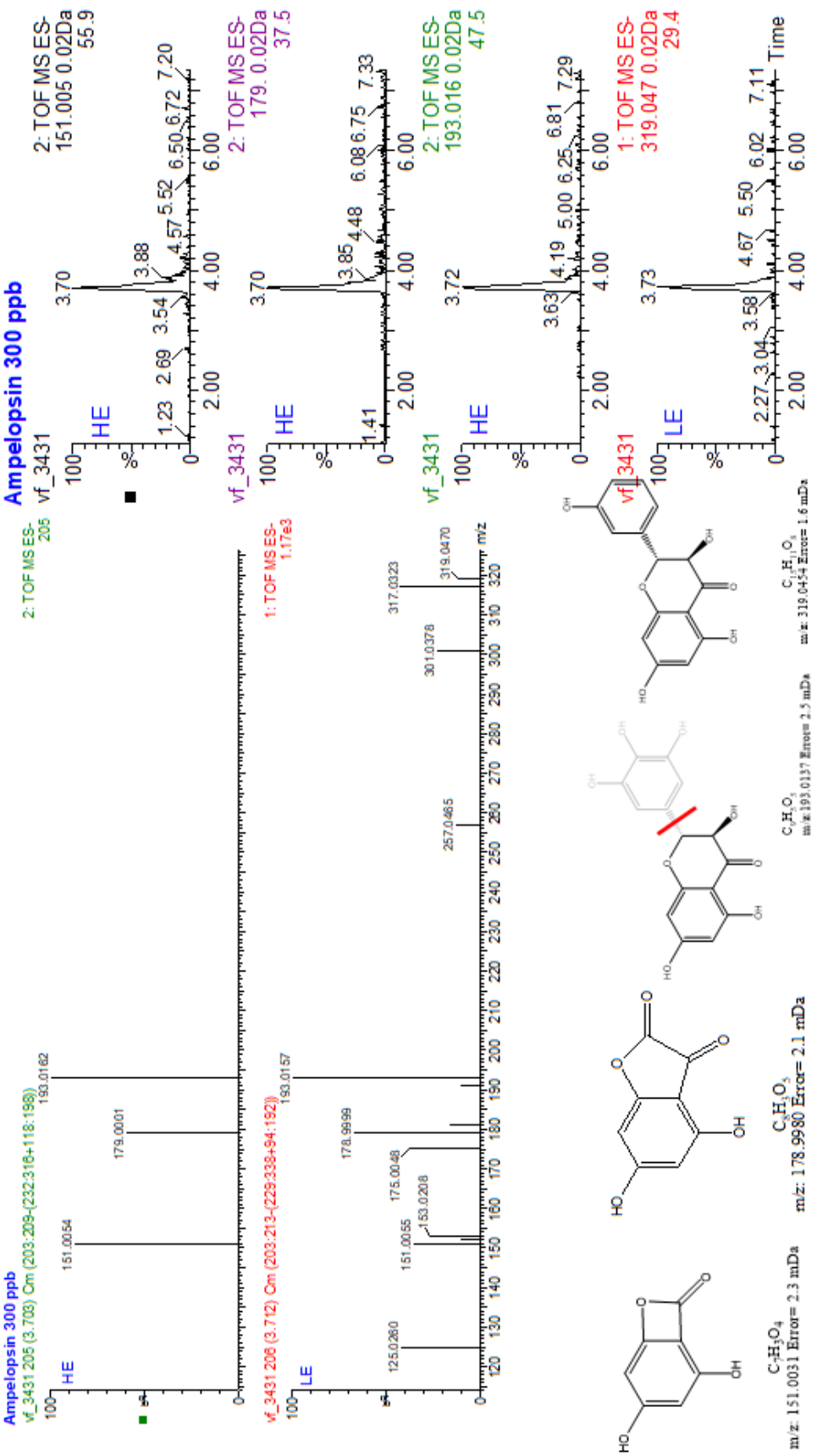
S1.7. Shikimic Acid detection and confirmation by UHPLC-QTOF MS.(Left) LE and HE spectrum; (right) nwXIC of the protonated molecule in LE and main fragments in HE. Confirmation of the fragments was established by their accurate mass, mass error and retention time.



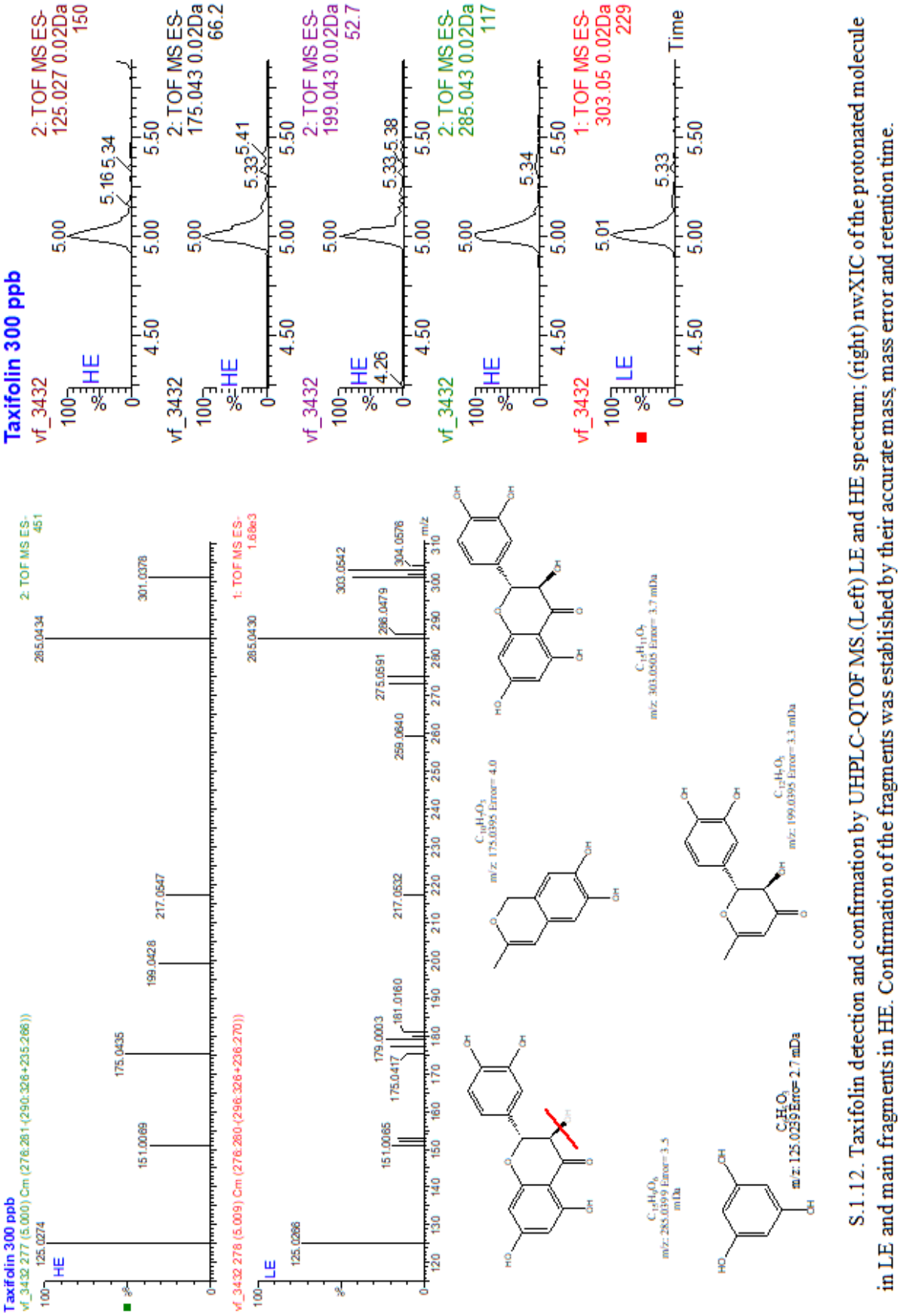
S1.8. 4-Coumaryl alcohol detection and confirmation by UHPLC-QTOF MS.(Left) LE and HE spectrum; (right) nwXIC of the protonated molecule in LE and fragment in HE. Confirmation of the fragment was established by their accurate mass, mass error and retention time.



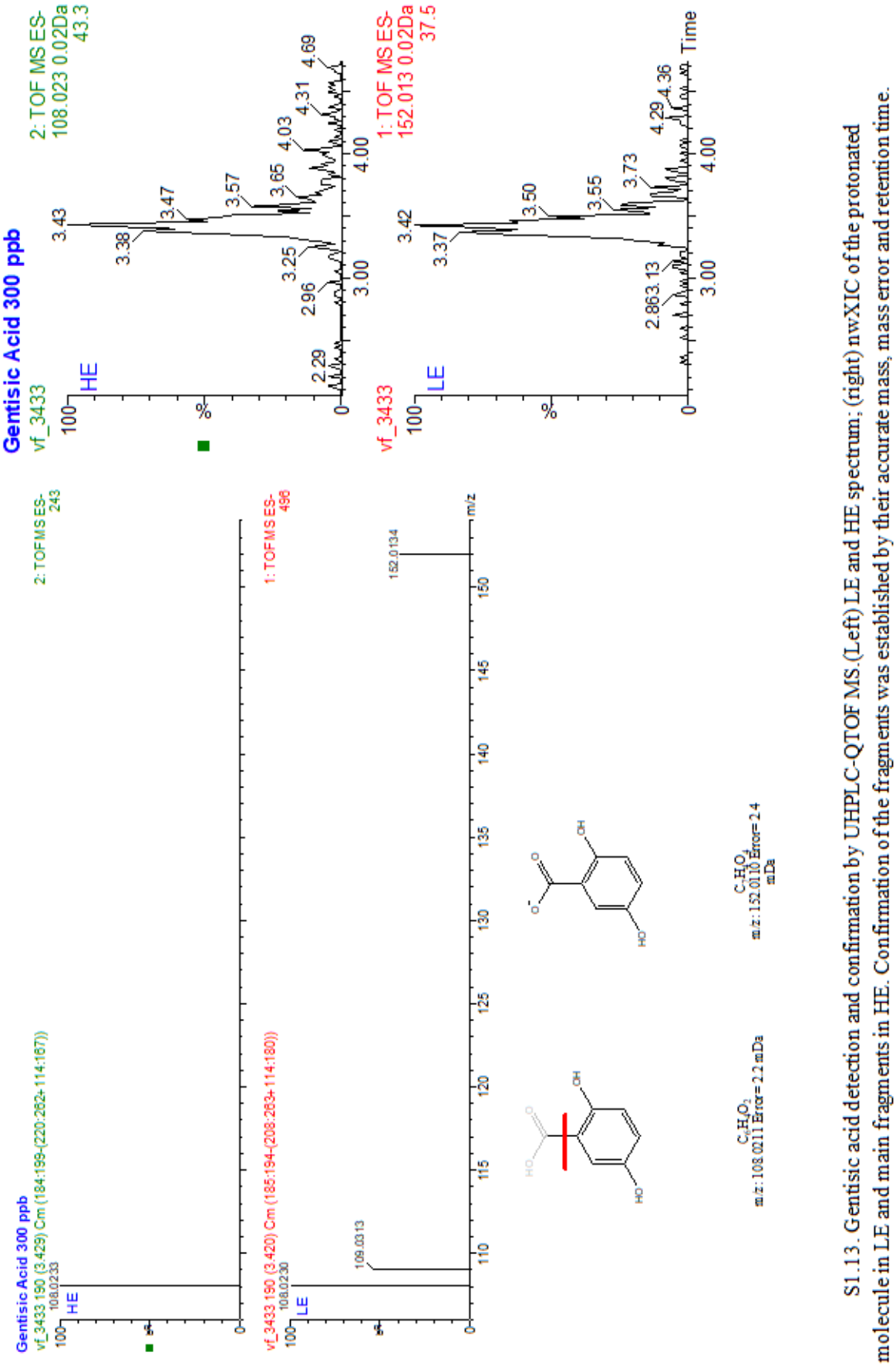
S1.9. Capsaicin detection and confirmation by UHPLC-QTOF MS. (Left) LE and HE spectrum; (right) nwXIC of the protonated molecule in LE and main fragments in HE. Confirmation of the fragments was established by their accurate mass, mass error and retention time. M/z equal to 328.1898 is referred to $[M+Na]^+$ adduct.



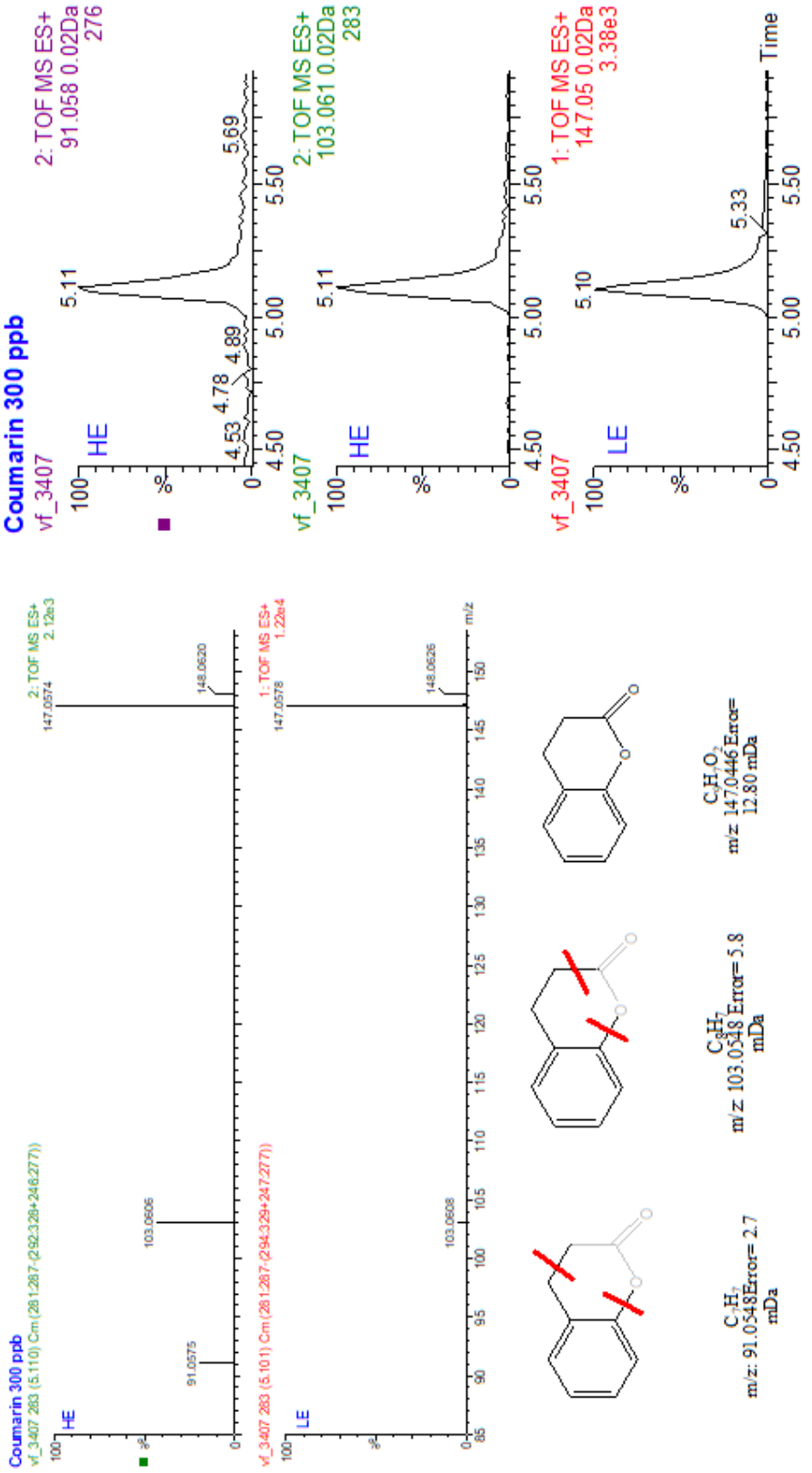
S.1.11. Ampelopsin detection and confirmation by UHPLC-QTOF MS. (Left) LE and HE spectrum; (right) nwXIC of the protonated molecule in LE and main fragments in HE. Confirmation of the fragments was established by their accurate mass, mass error and retention time.



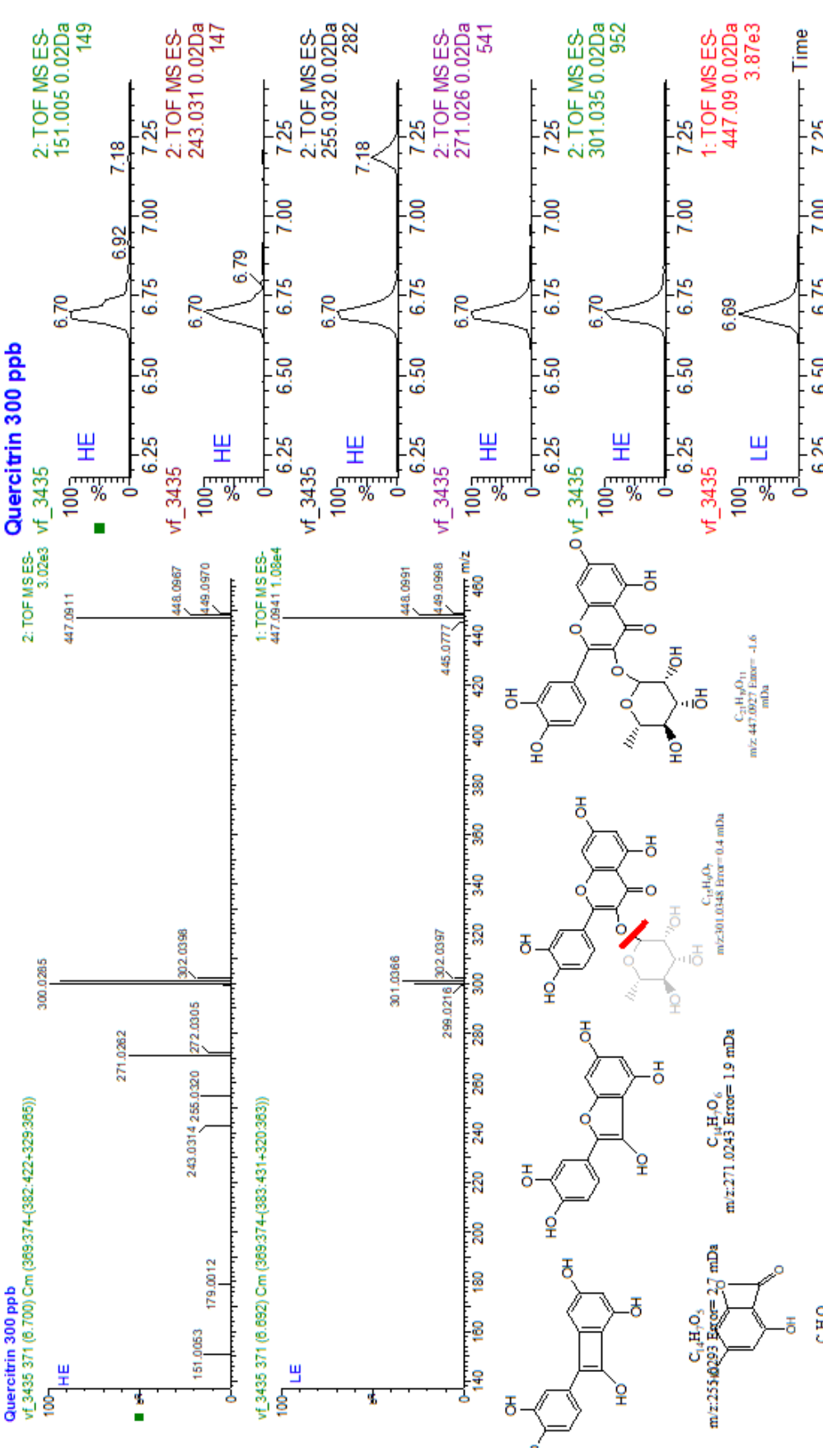
S.1.12. Taxifolin detection and confirmation by UHPLC-QTOF MS. (Left) LE and HE spectrum; (right) nwXIC of the protonated molecule in LE and main fragments in HE. Confirmation of the fragments was established by their accurate mass, mass error and retention time.



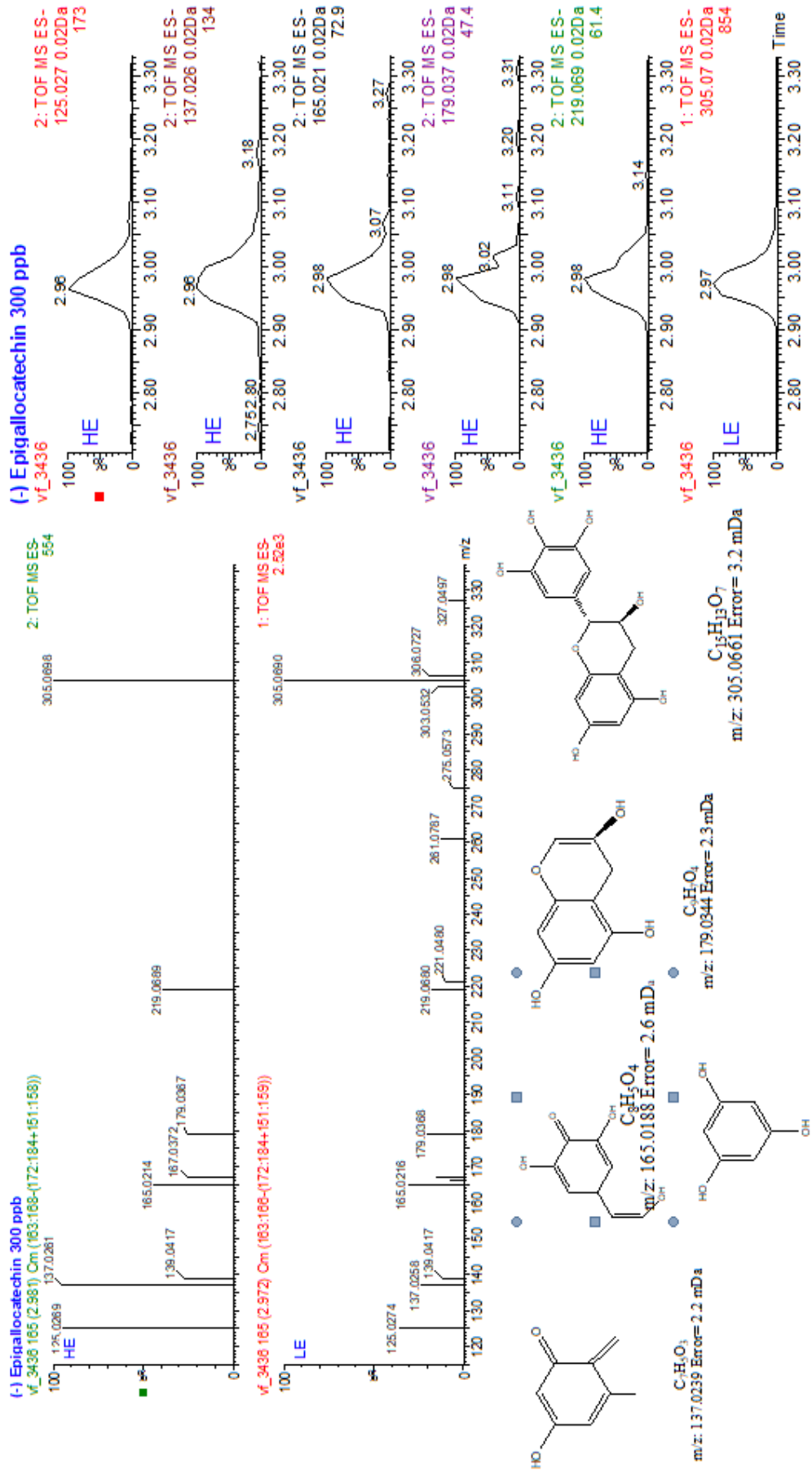
S1.13. Gentisic acid detection and confirmation by UHPLC-QTOF MS.(Left) LE and HE spectrum; (right) nwXIC of the protonated molecule in LE and main fragments in HE. Confirmation of the fragments was established by their accurate mass, mass error and retention time.



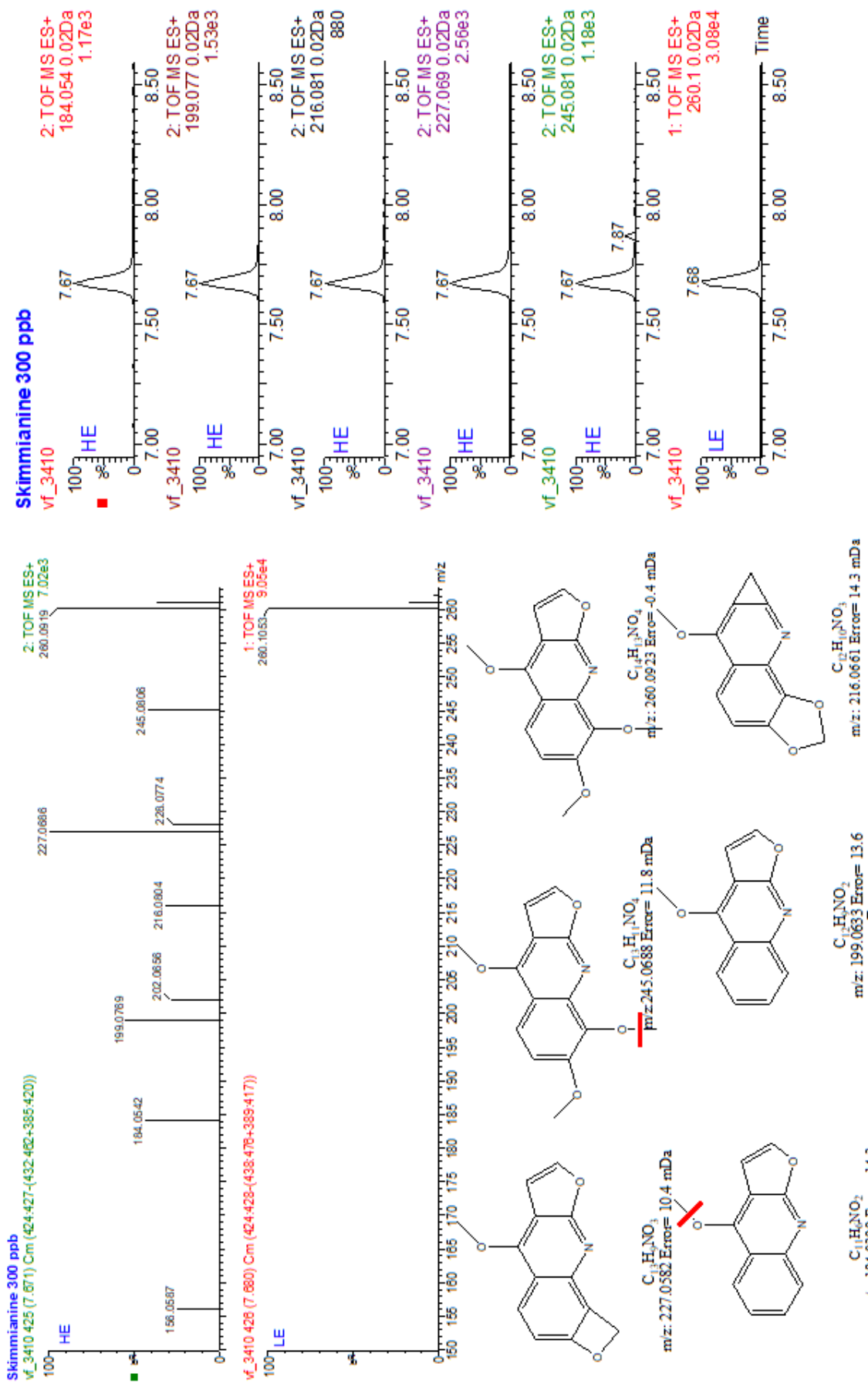
S.1.14. Coumarin detection and confirmation by UHPLC-QTOF MS.(Left) LE and HE spectrum; (right) nwXIC of the protonated molecule in LE and main fragments in HE. Confirmation of the fragments was established by their accurate mass, mass error and retention time.



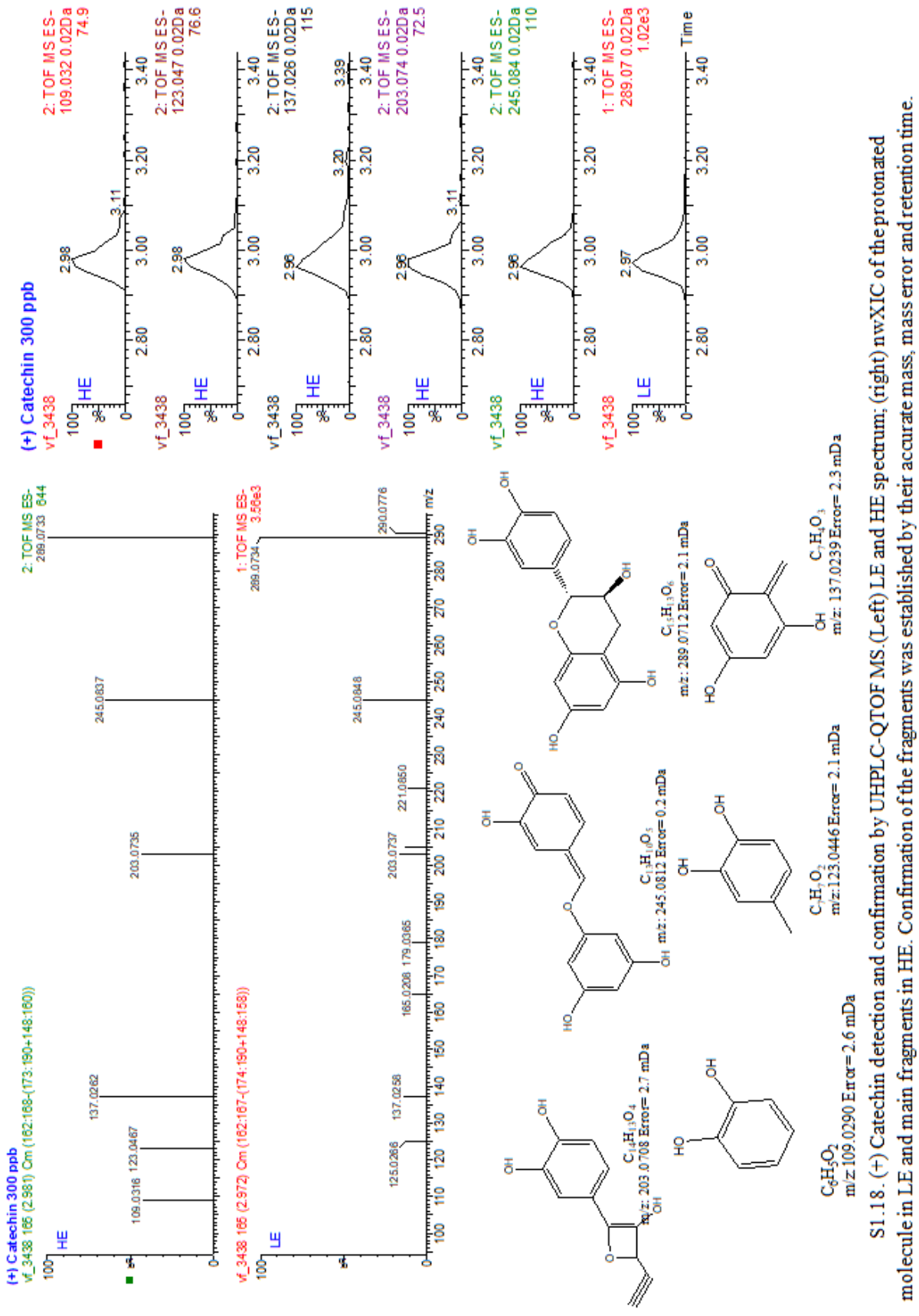
S1.15. Quercitrin detection and confirmation by UHPLC-QTOF MS.(Left) LE and HE spectrum; (right) nwXIC of the protonated molecule in LE and main fragments in HE. Confirmation of the fragments was established by their accurate mass, mass error and retention time.



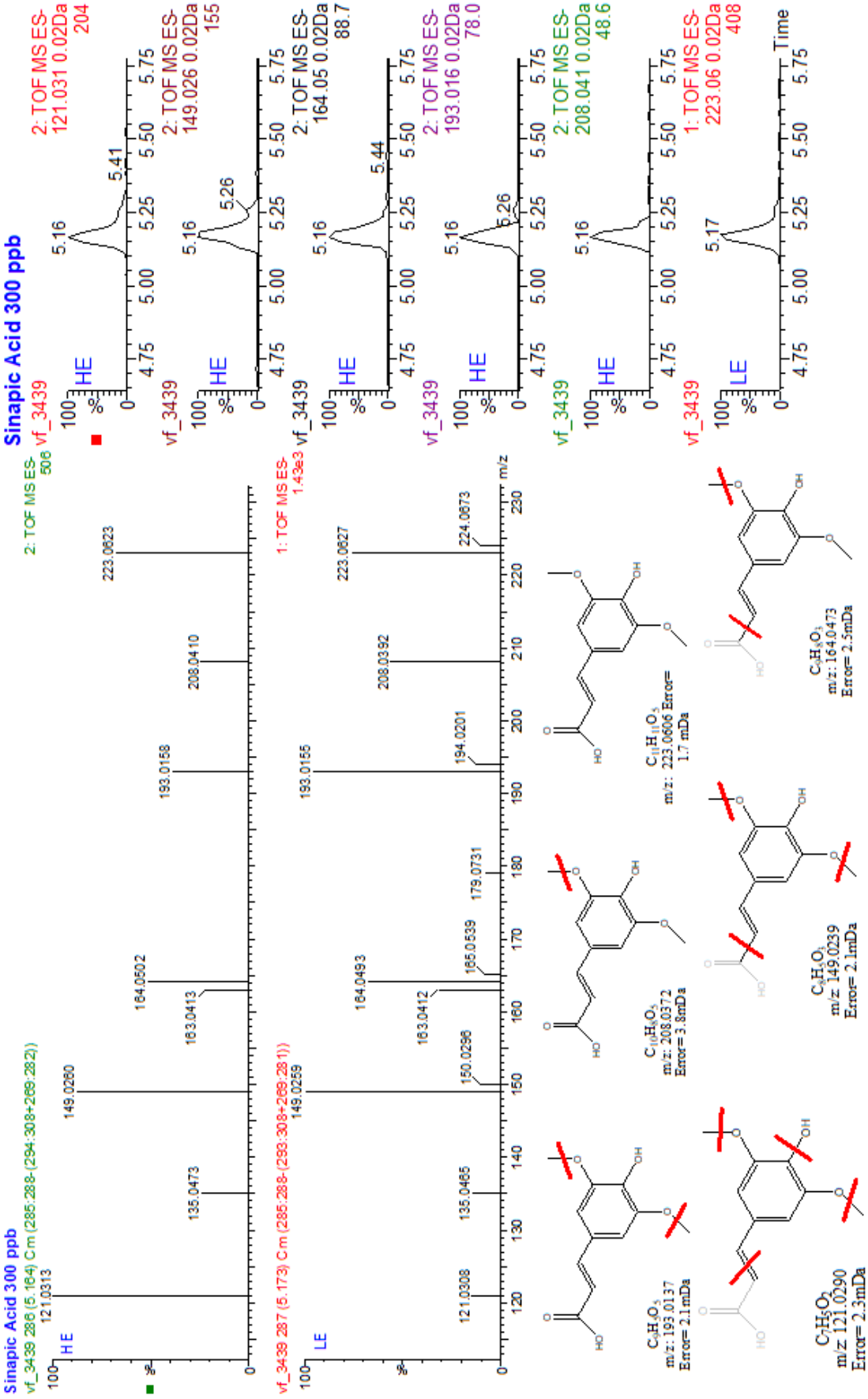
S1.16. (-) Epigallocatechin detection and confirmation by UHPLC-QTOF MS. (Left) LE and HE spectrum; (right) nwXIC of the protonated molecule in LE and main fragments in HE. Confirmation of the fragments was established by their accurate mass, mass error and retention time.



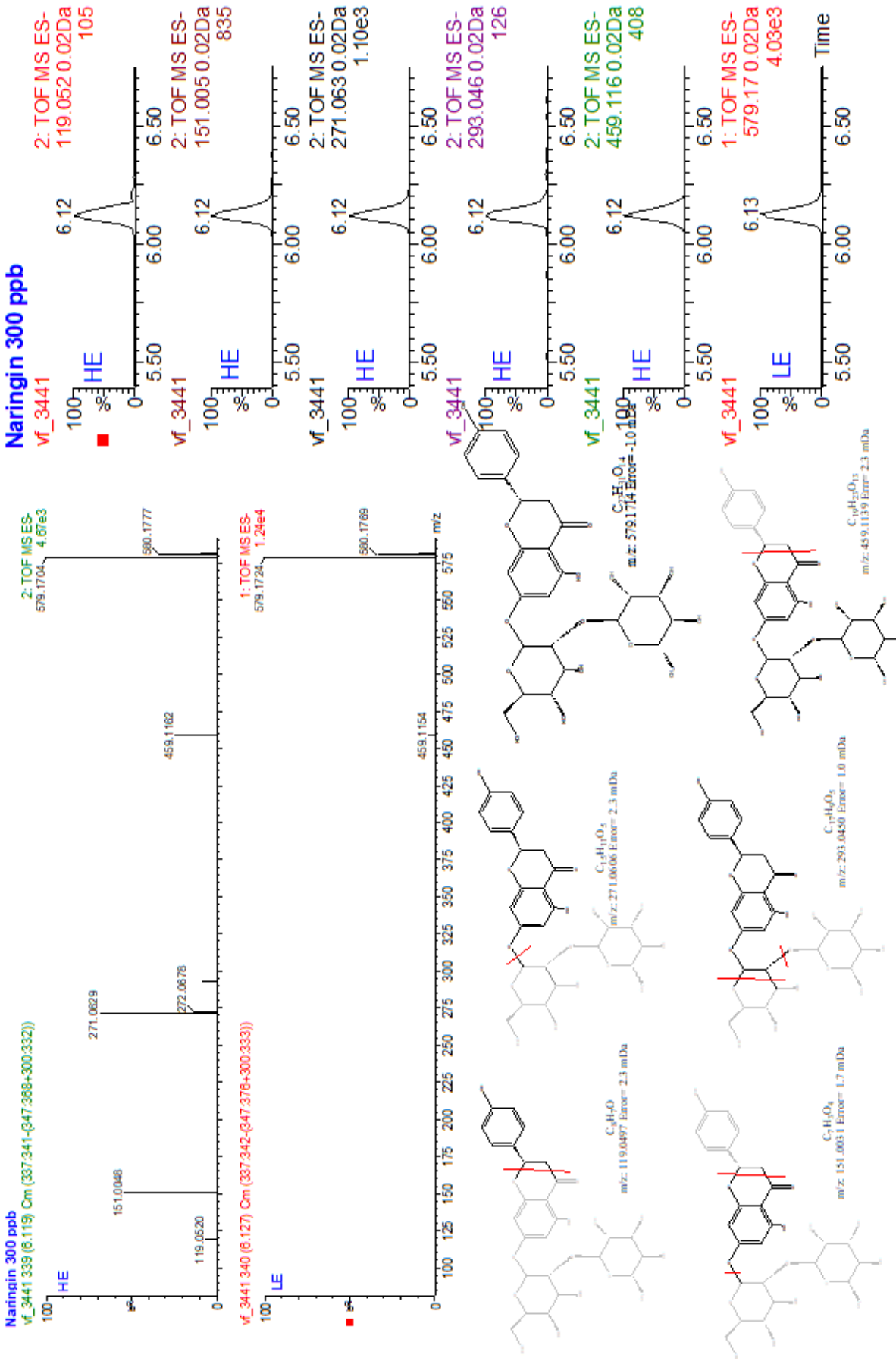
S1.17. Skimmianine detection and confirmation by UHPLC-QTOF MS.(Left) LE and HE spectrum; (right) nwXIC of the protonated molecule in LE and main fragments in HE. Confirmation of the fragments was established by their accurate mass, mass error and retention time.



S1.18. (+) Catechin detection and confirmation by UHPLC-QTOF MS.(Left) LE and HE spectrum; (right) nwXIC of the protonated molecule in LE and main fragments in HE. Confirmation of the fragments was established by their accurate mass, mass error and retention time.

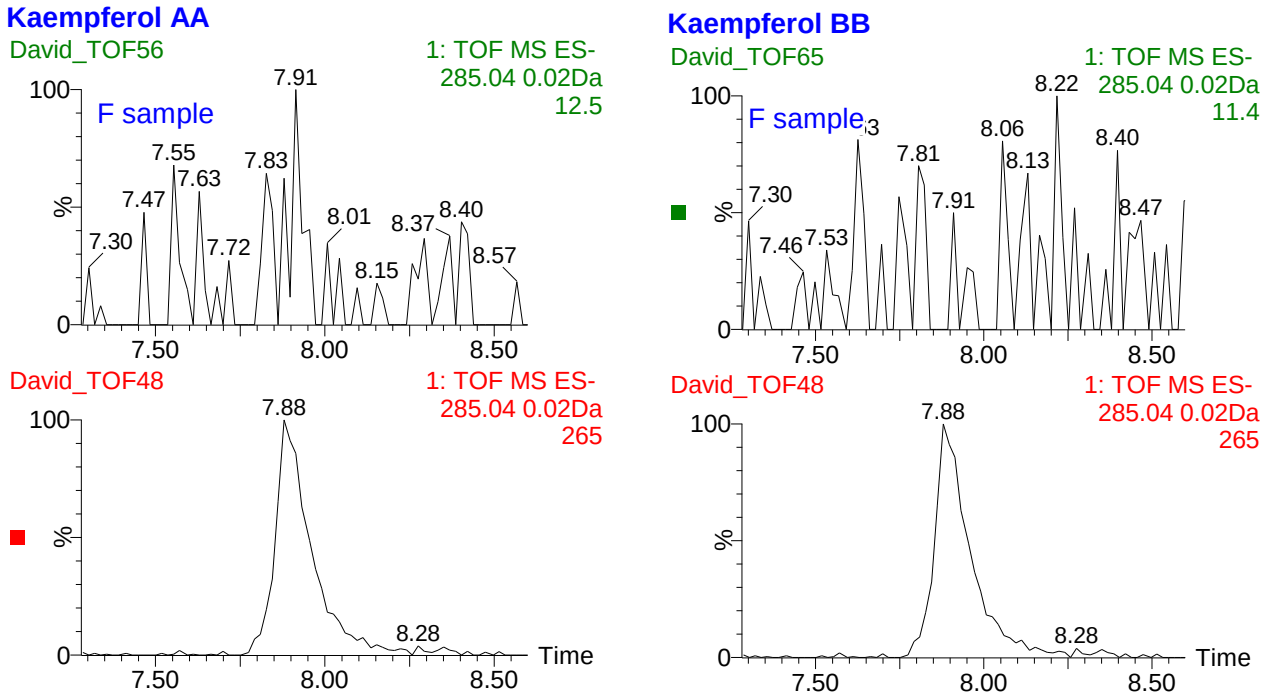


S1.19. Sinapic acid detection and confirmation by UHPLC-QTOF MS.(Left) LE and HE spectrum; (right) nwXIC of the protonated molecule in LE and main fragments in HE. Confirmation of the fragments was established by their accurate mass, mass error and retention time.

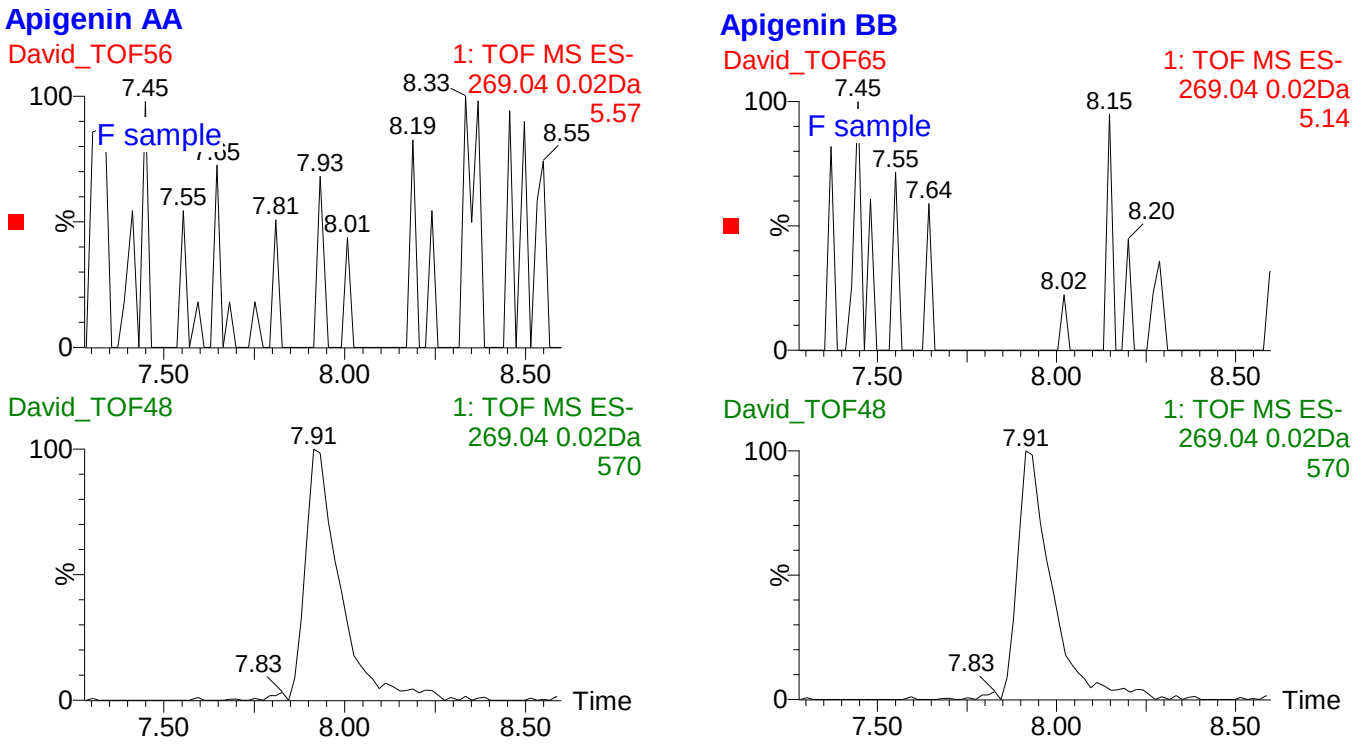


S1.20. Naringin detection and confirmation by UHPLC-QTOF MS. (Left) LE and HE spectrum; (right) nwXIC of the protonated molecule in LE and main fragments in HE. Confirmation of the fragments was established by their accurate mass, mass error and retention time.

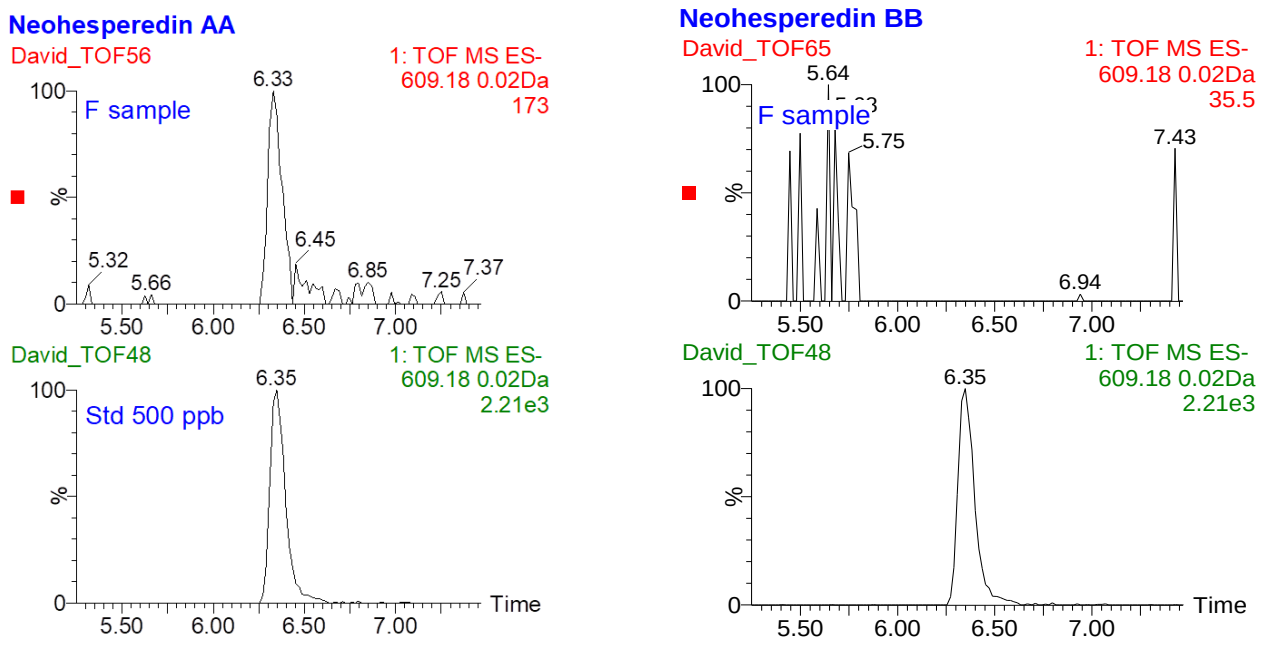
S2. UHPLC-QTOF MS chromatograms, spectra and tentative fragmentation pathway of studied compounds in fortified sample



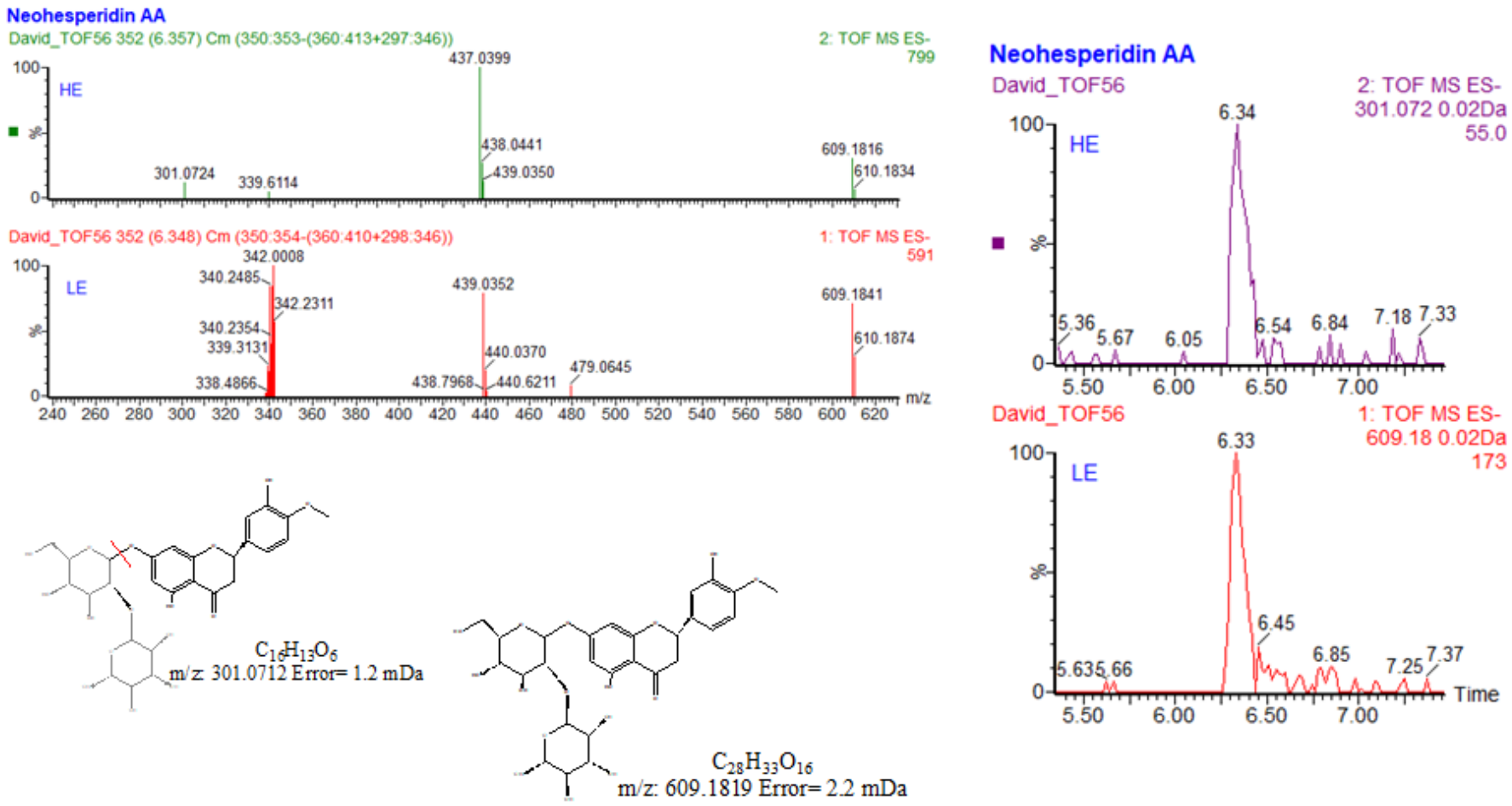
S2.1. Comparison between nwXICs for kaempferol in the fortified plant and in a reference standard at the same level. AA is referred to *Arabidopsis thaliana* (left) and BB to *Solanum lycopersicum* (right). As it can be seen, the peak does not appear in the fortified sample in either case.



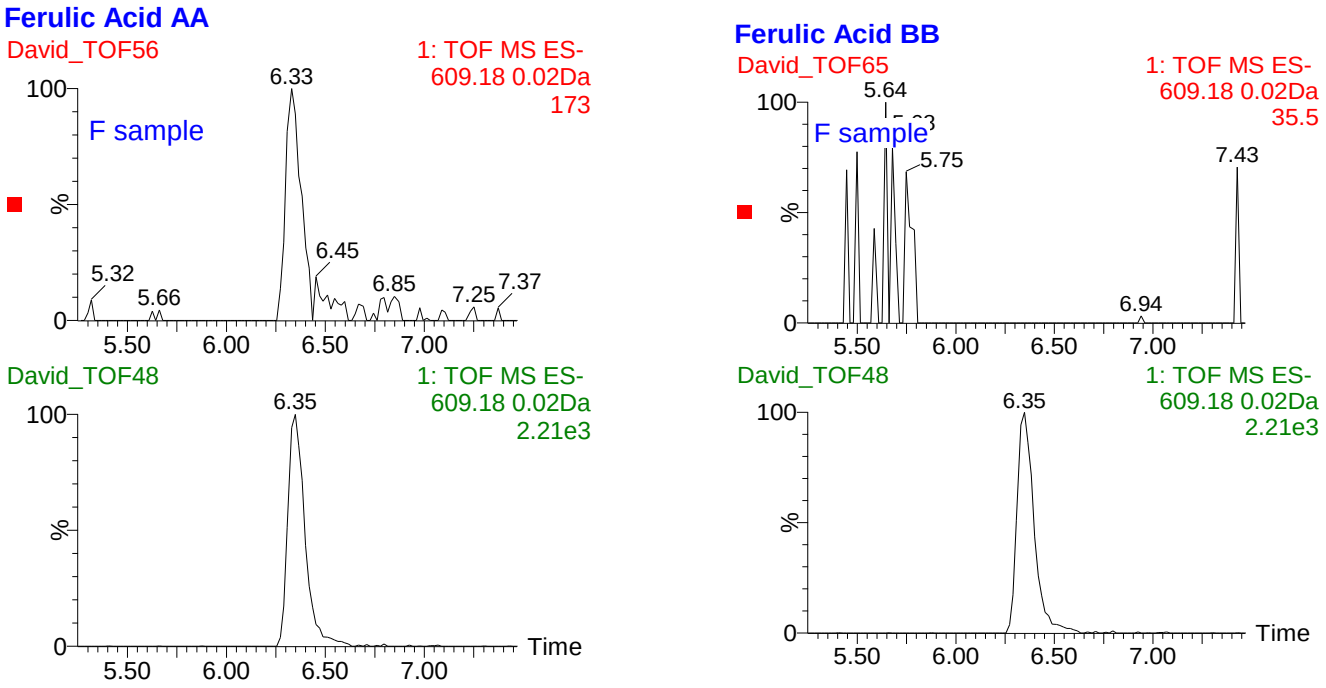
S2.2. Comparison between nwXICs for apigenin in the fortified plant and in a reference standard at the same level. AA is referred to *Arabidopsis thaliana* (left) and BB to *Solanum lycopersicum* (right). As it can be seen, the peak does not appear in the fortified sample in either case.



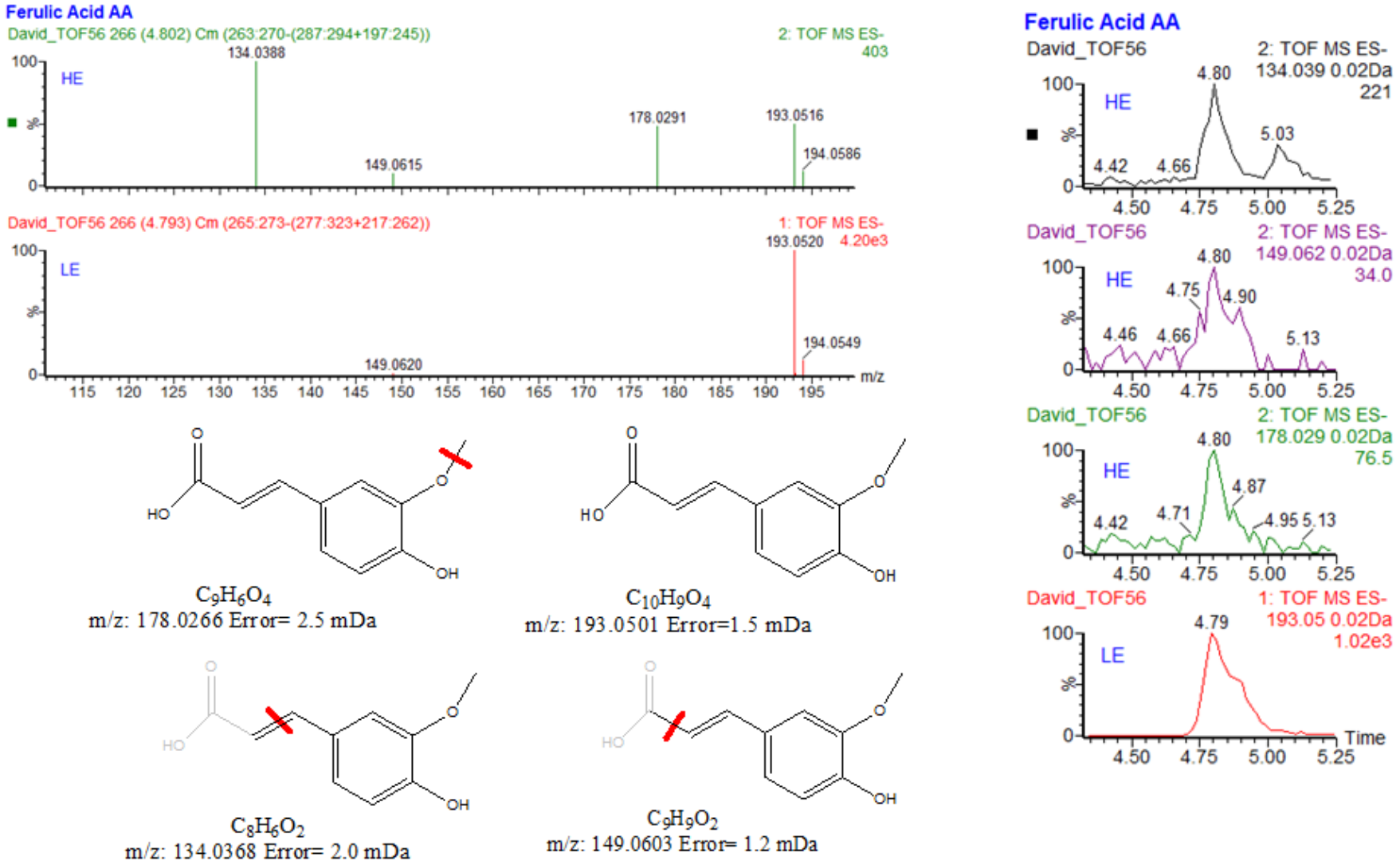
S2.3. Comparison between nwXICs for neohesperidin in the fortified plant and in a reference standard at the same level. AA is referred to *Arabidopsis thaliana* (left) and BB to *Solanum lycopersicum* (right). As it can be seen, the peak only appears in the fortified sample of Arabidopsis thaliana.



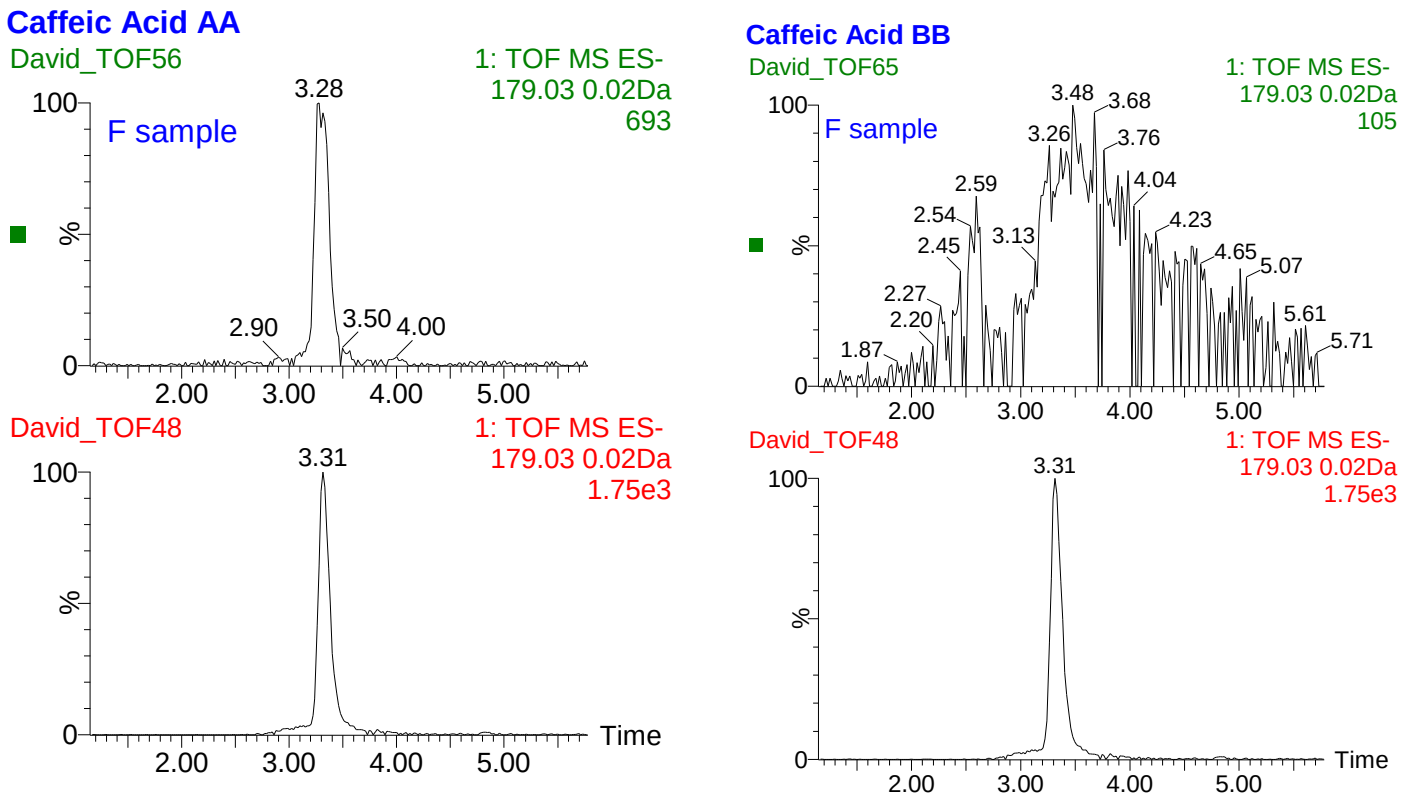
S2.4. Neohesperidin detection and confirmation confirmation by UHPLC-QTOF MS in an example of fortified *Arabidopsis thaliana* at 500ppb. (Left) LE and HE spectrum; (right) nwXIC of the protonated molecule in LE and the fragments in HE. Confirmation of the fragment was established by their accurate mass, mass error and retention time.



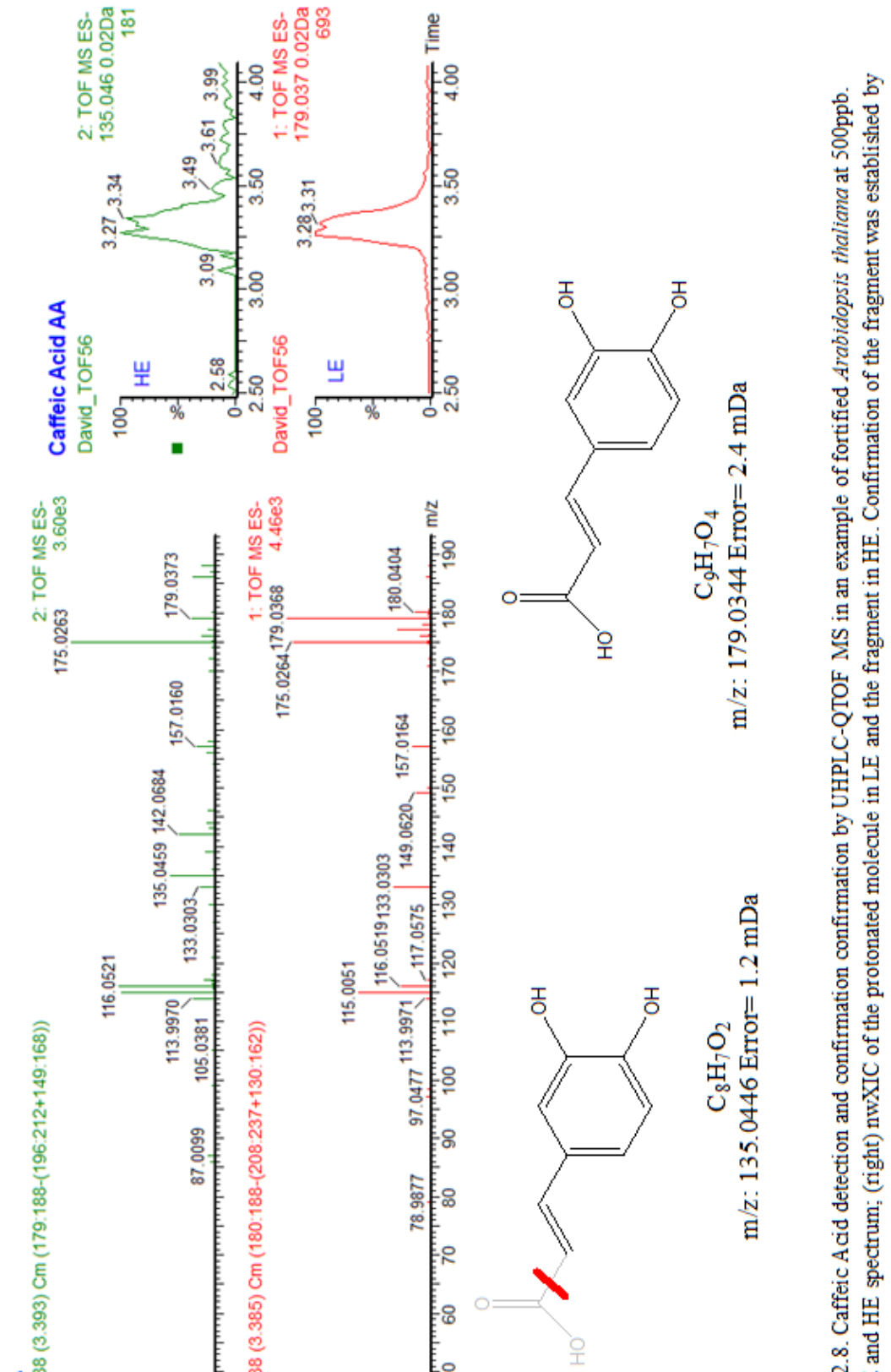
S2.5. Comparison between nwXICs for ferulic acid in the fortified plant and in a reference standard at the same level. AA is referred to *Arabidopsis thaliana* (left) and BB to *Solanum lycopersicum* (right). As it can be seen, the peak only appears in the fortified sample of Arabidopsis thaliana.



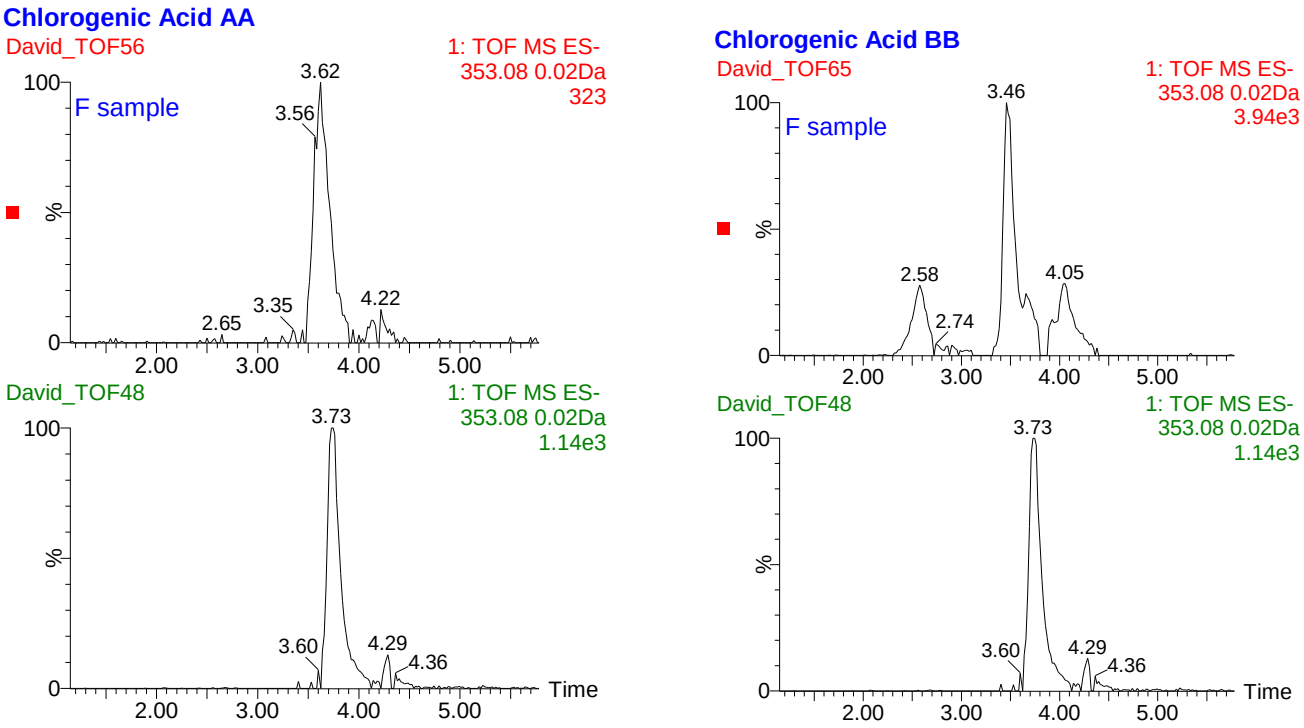
S2.6. Ferulic Acid detection and confirmation confirmation by UHPLC-QTOF MS in an example of fortified *Arabidopsis thaliana* at 500ppb. (Left) LE and HE spectrum; (right) nwXIC of the protonated molecule in LE and main fragments in HE. Confirmation of the fragments was established by their accurate mass, mass error and retention time.



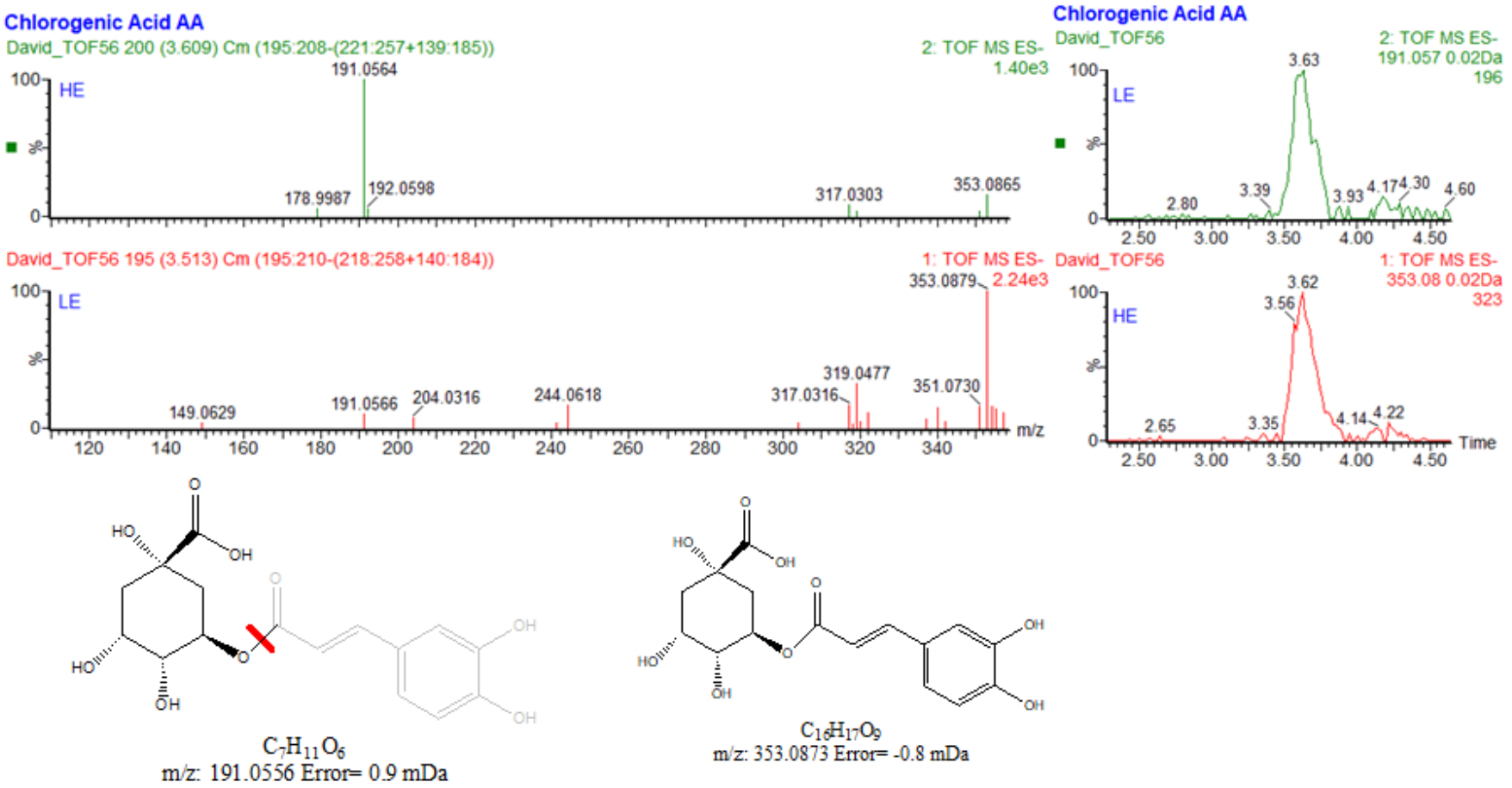
S2.7. Comparison between nwXICs for caffeic acid in the fortified plant and in a reference standard at the same level. AA is referred to *Arabidopsis thaliana* (left) and BB to *Solanum lycopersicum* (right). As it can be seen, the peak only appears in the fortified sample of Arabidopsis thaliana.



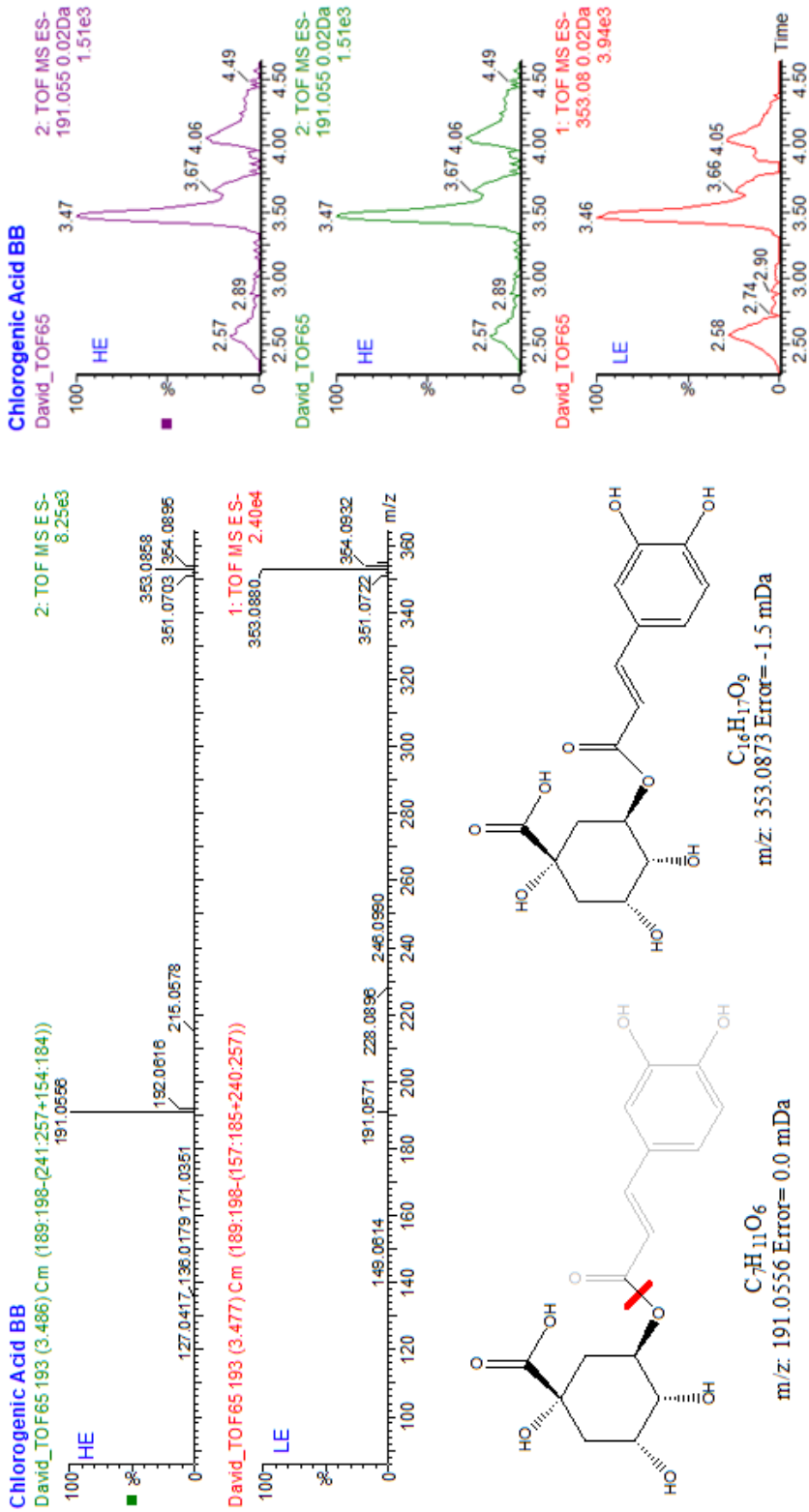
S2.8. Caffeic Acid detection and confirmation confirmation by UHPLC-QTOF MS in an example of fortified *Arabidopsis thaliana* at 500ppb. (Left) LE and HE spectrum; (right) nwXIC of the protonated molecule in LE and the fragment in HE. Confirmation of the fragment was established by their accurate mass, mass error and retention time.



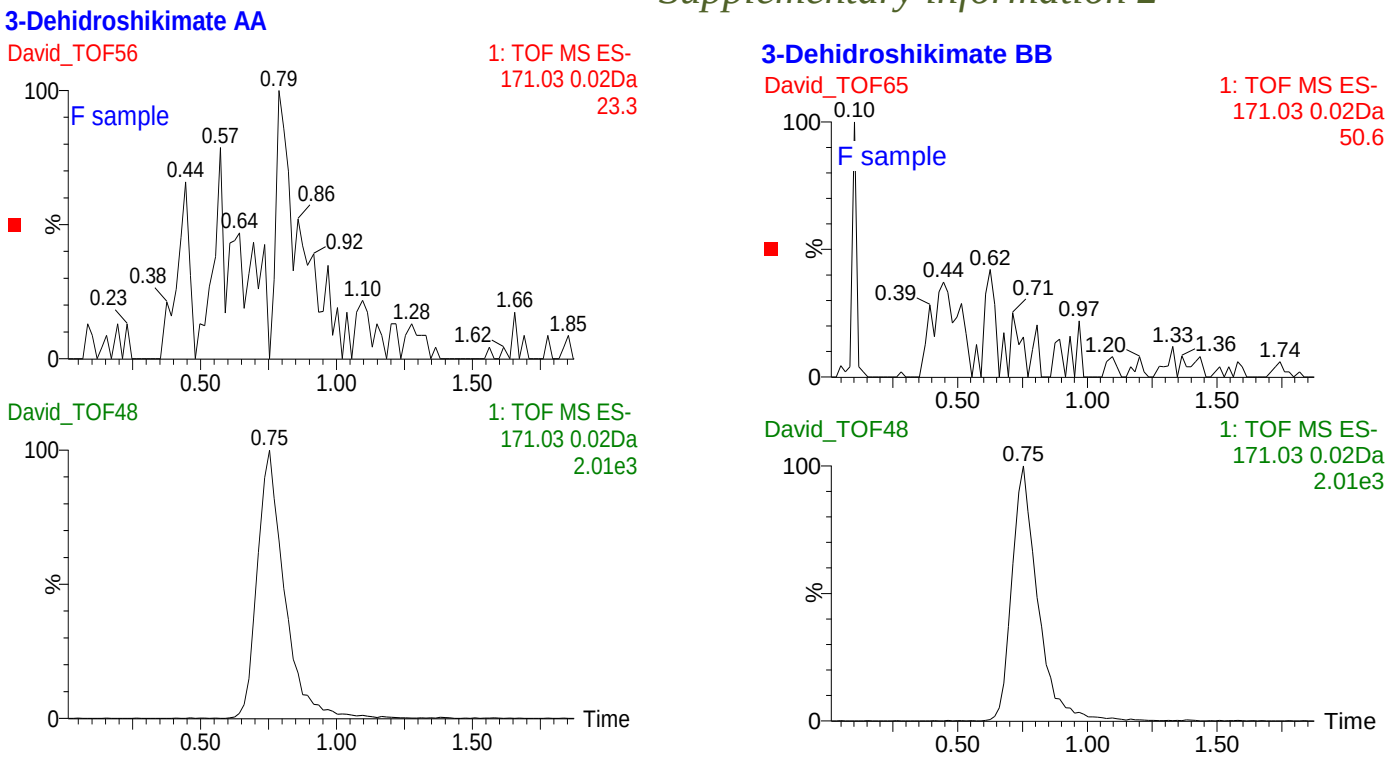
S2.9. Comparison between nwXICs for caffeic acid in the fortified plant and in a reference standard at the same level. AA is referred to *Arabidopsis thaliana* (left) and BB to *Solanum lycopersicum* (right). As it can be seen, the peak appears in both fortified samples.



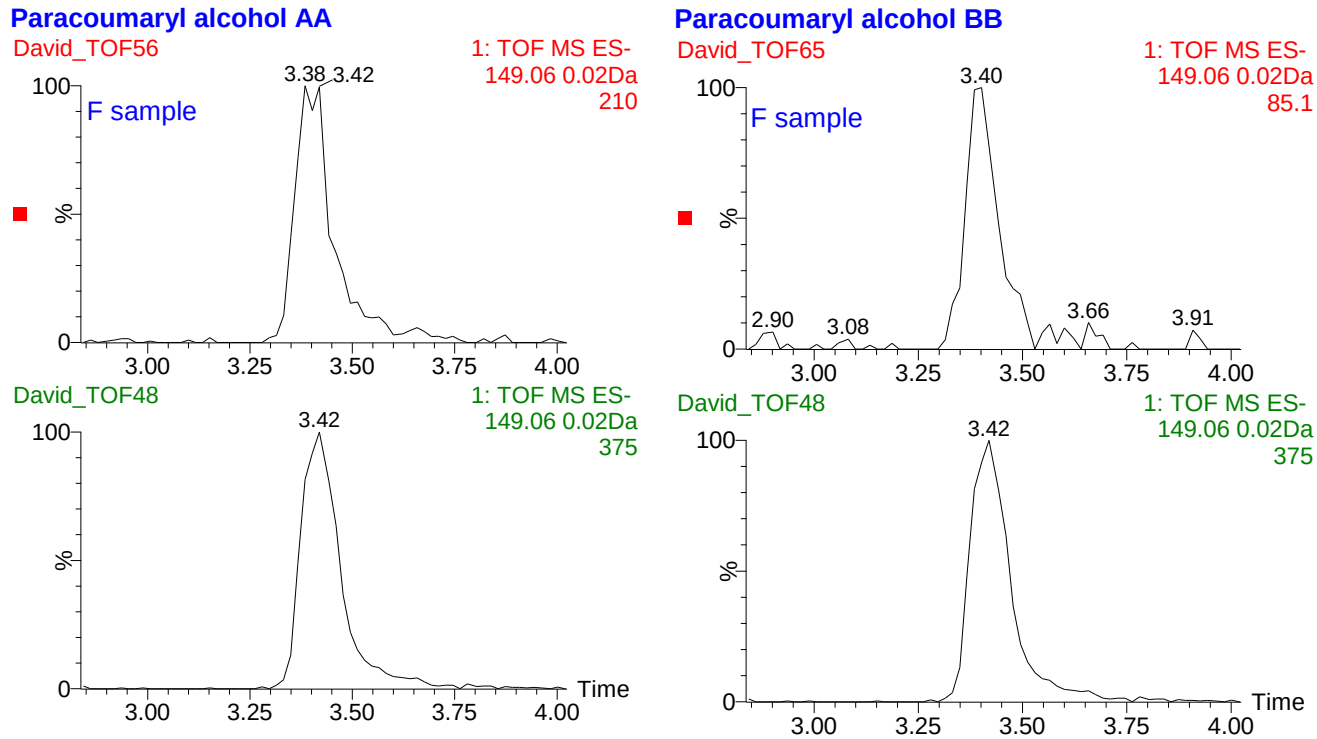
S2.10. Chlorogenic acid detection and confirmation confirmation by UHPLC-QTOF MS in an example of fortified *Arabidopsis thaliana* at 500ppb. (Left) LE and HE spectrum; (right) nwXIC of the protonated molecule in LE and the fragments in HE. Confirmation of the fragment was established by their accurate mass, mass error and retention time.



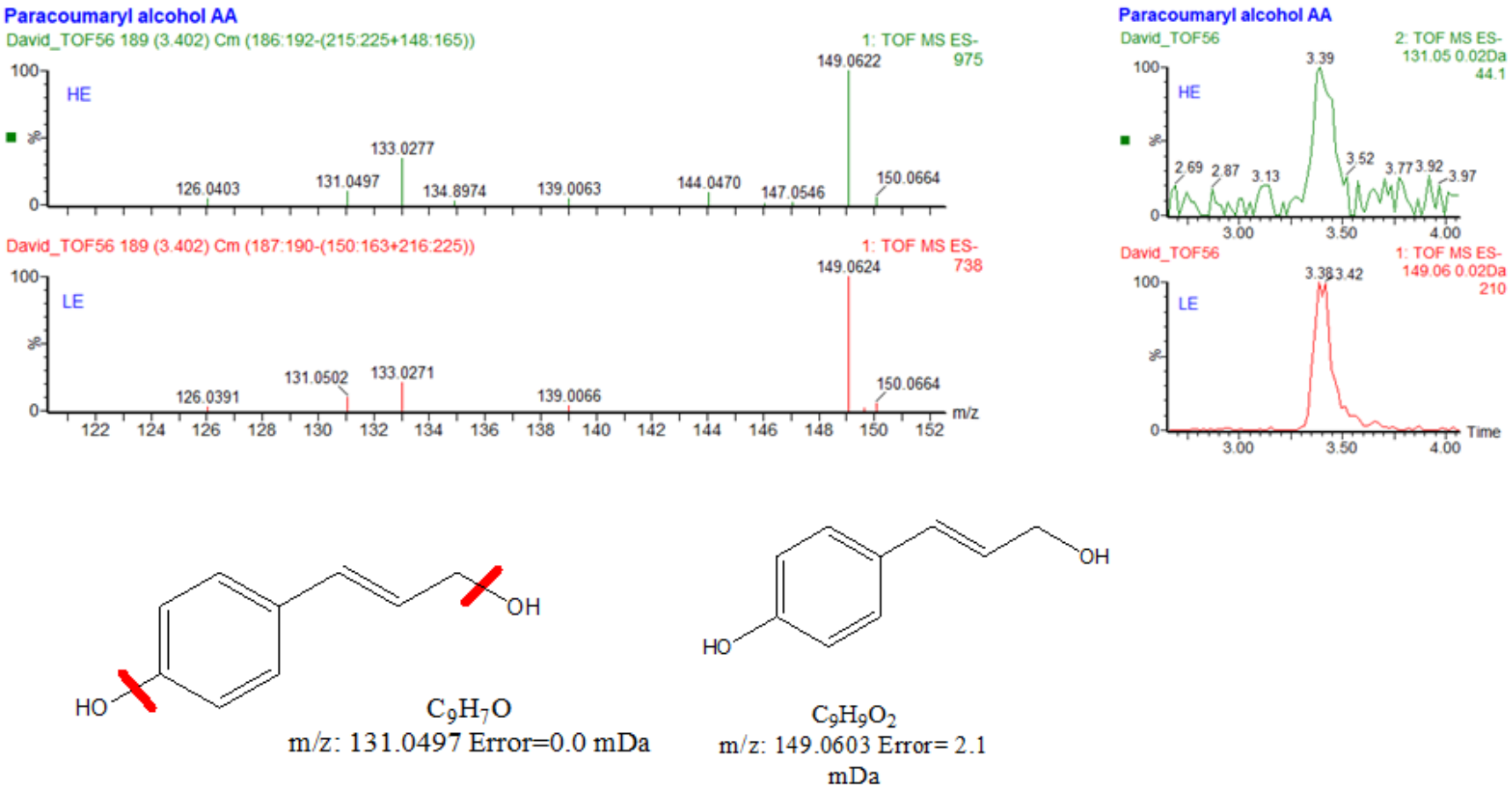
S2.11. Chlorogenic acid detection and confirmation confirmation by UHPLC-QTOF MS in an example of fortified *Solanum lycopersum* at 500ppb. (Left) LE and HE spectrum; (right) nwXIC of the protonated molecule in LE and the fragment in HE. Confirmation of the fragment was established by their accurate mass, mass error and retention time.



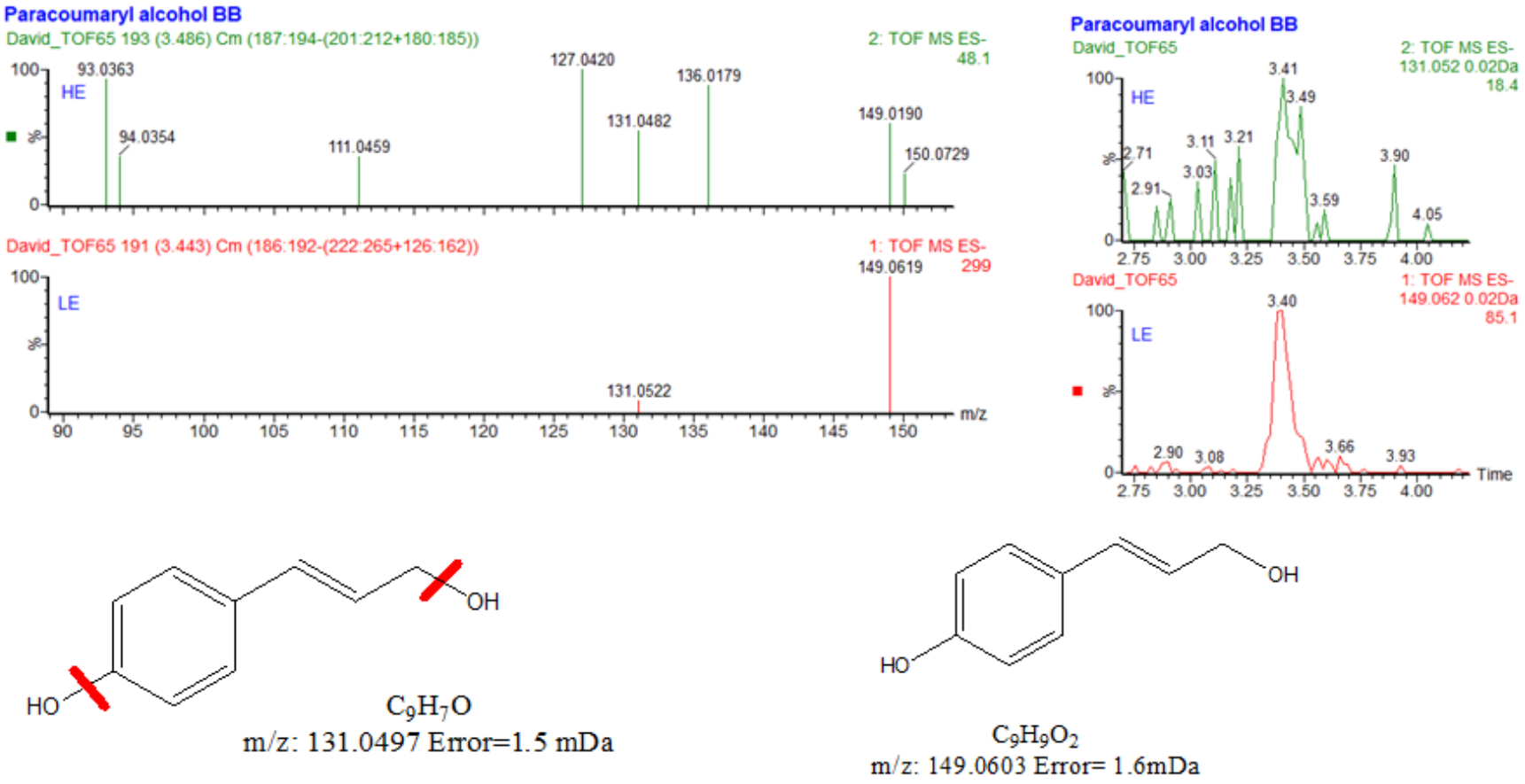
S2.12. Comparison between nwXICs for 3-dehidroshikimate in the fortified plant and in a reference standard at the same level. AA is referred to *Arabidopsis thaliana* (left) and BB to *Solanum lycopersicum* (right). As it can be seen, the peak does not appear in the fortified sample in either case.



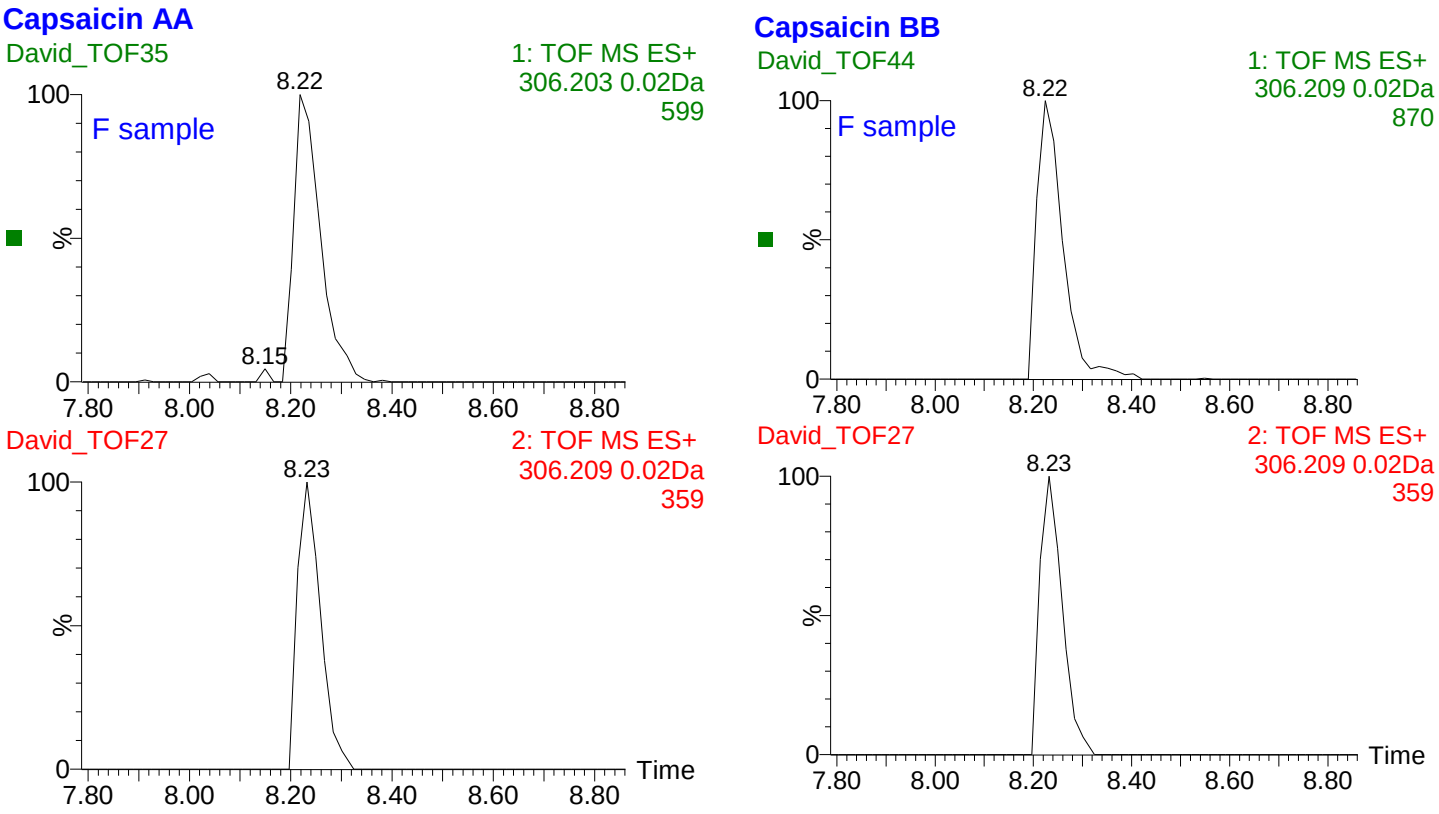
S2.13. Comparison between nwXICs for 4-coumaryl alcohol in the fortified plant and in a reference standard at the same level. AA is referred to *Arabidopsis thaliana* (left) and BB to *Solanum lycopersicum* (right). As it can be seen, the peak appears in both fortified samples.



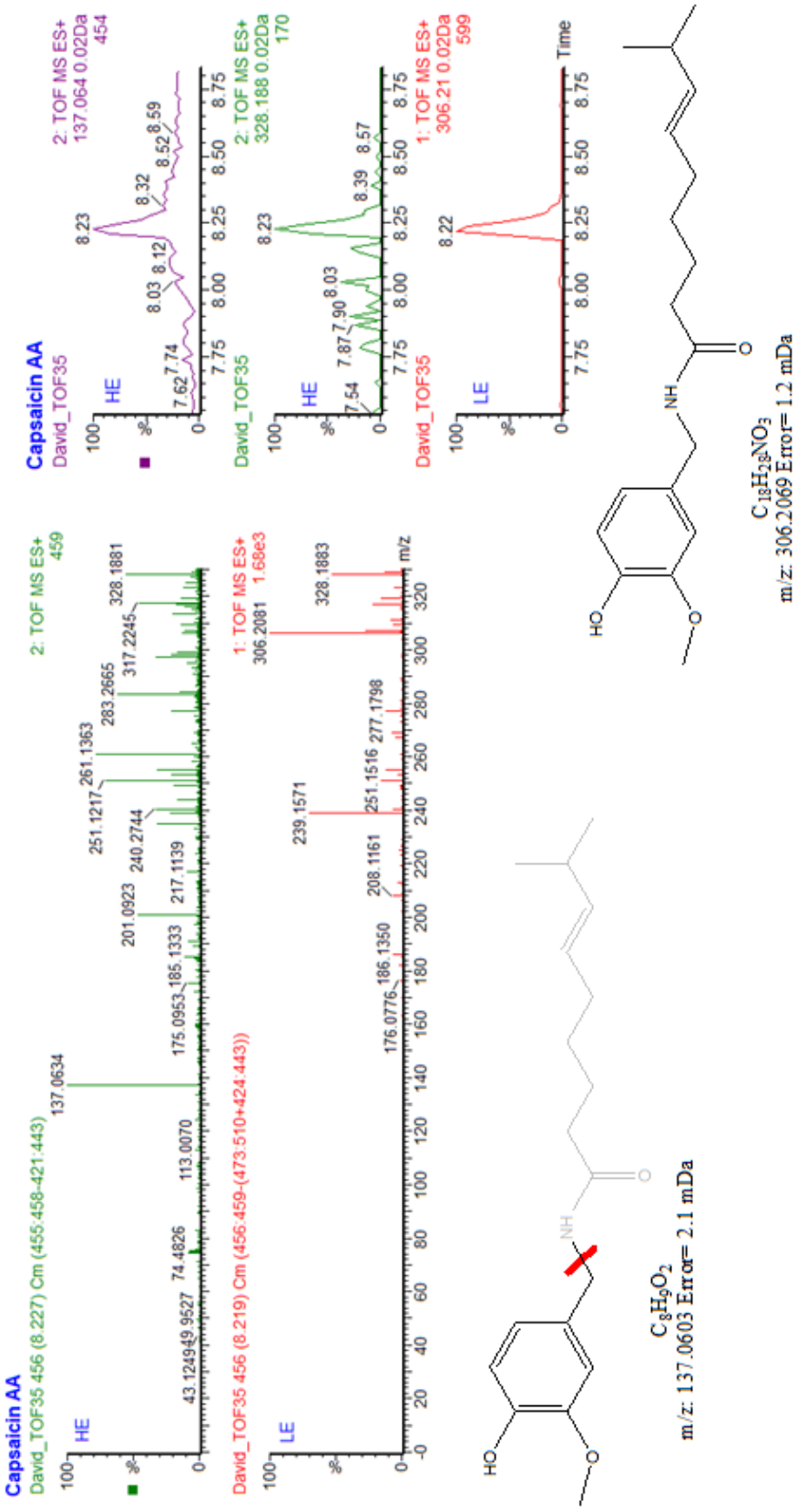
S2.14. 4-coumaryl alcohol detection and confirmation confirmation by UHPLC-QTOF MS in an example of fortified *Arabidopsis thaliana* at 500ppb. (Left) LE and HE spectrum; (right) nwXIC of the protonated molecule in LE and the fragment in HE. Confirmation of the fragment was established by their accurate mass, mass error and retention time.



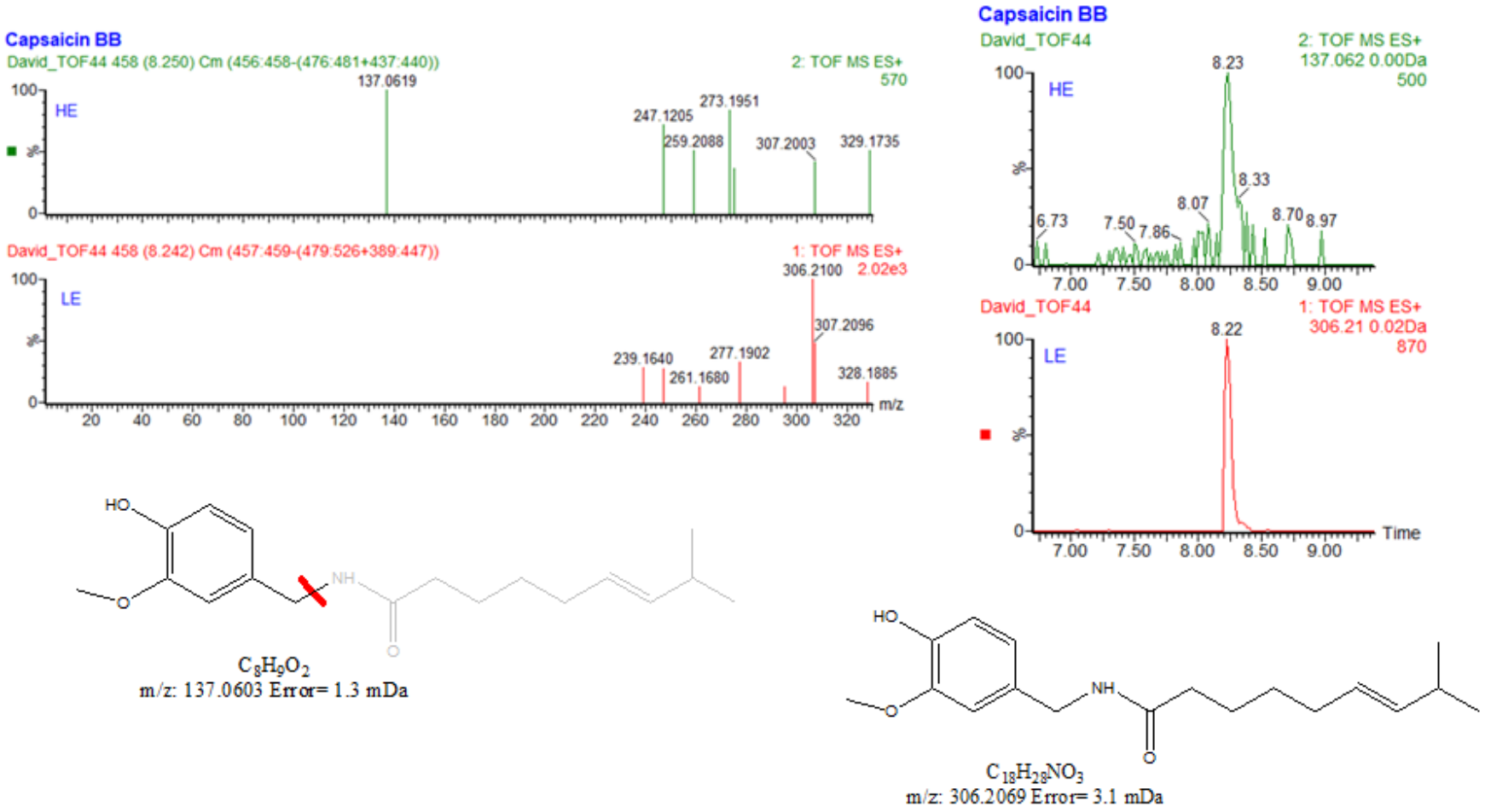
S2.15. 4-coumaryl alcohol detection and confirmation confirmation by UHPLC-QTOF MS in an example of fortified *Solanum lycopersum* at 500ppb. (Left) LE and HE spectrum; (right) nwXIC of the protonated molecule in LE and the fragment in HE. Confirmation of the fragment was established by their accurate mass, mass error and retention time.



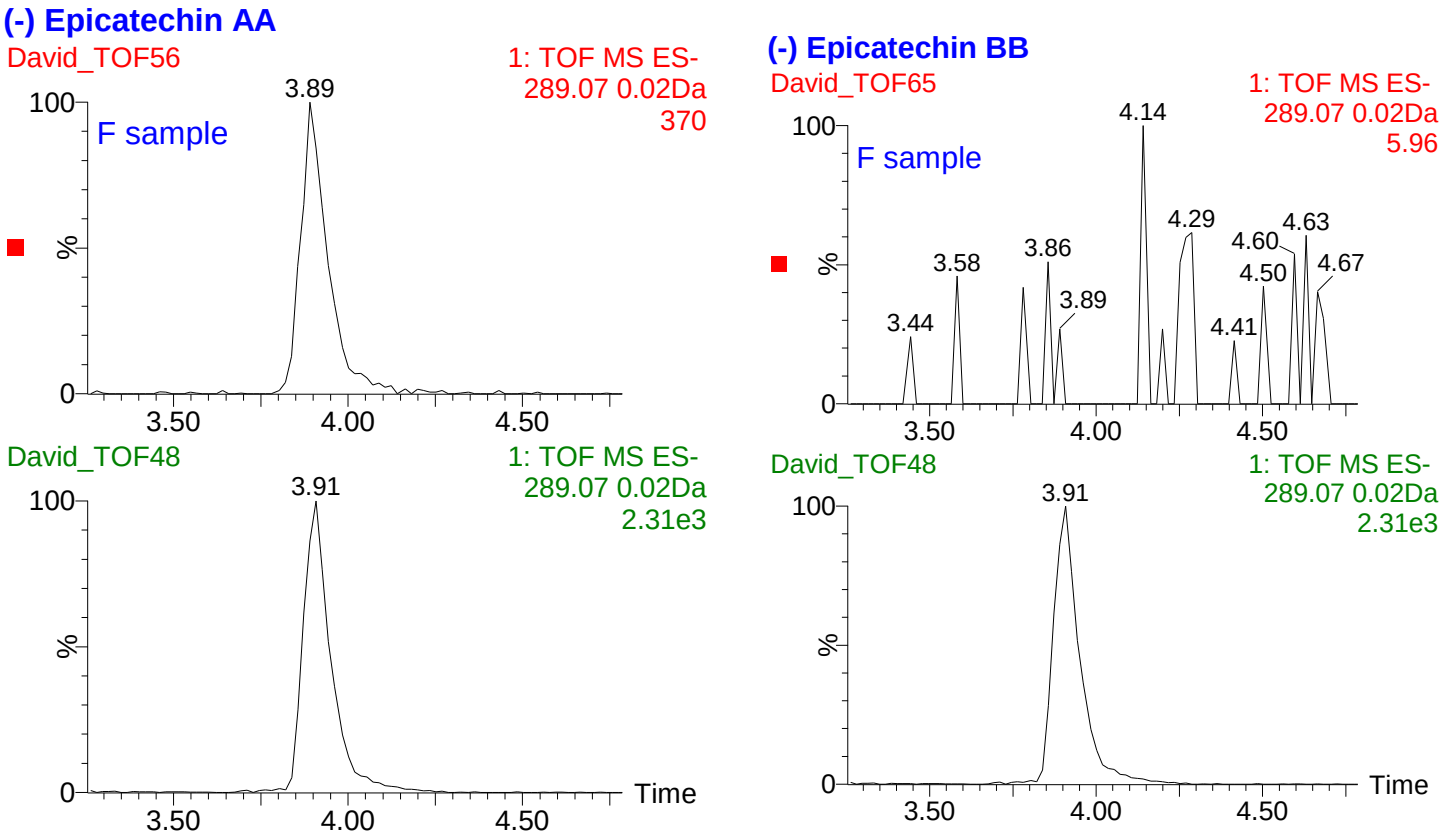
S2.16. Comparison between nwXICs for capsaicinin in the fortified plant and in a reference standard at the same level. AA is referred to *Arabidopsis thaliana* (left) and BB to *Solanum lycopersicum* (right). As it can be seen, the peak appears in both fortified samples.



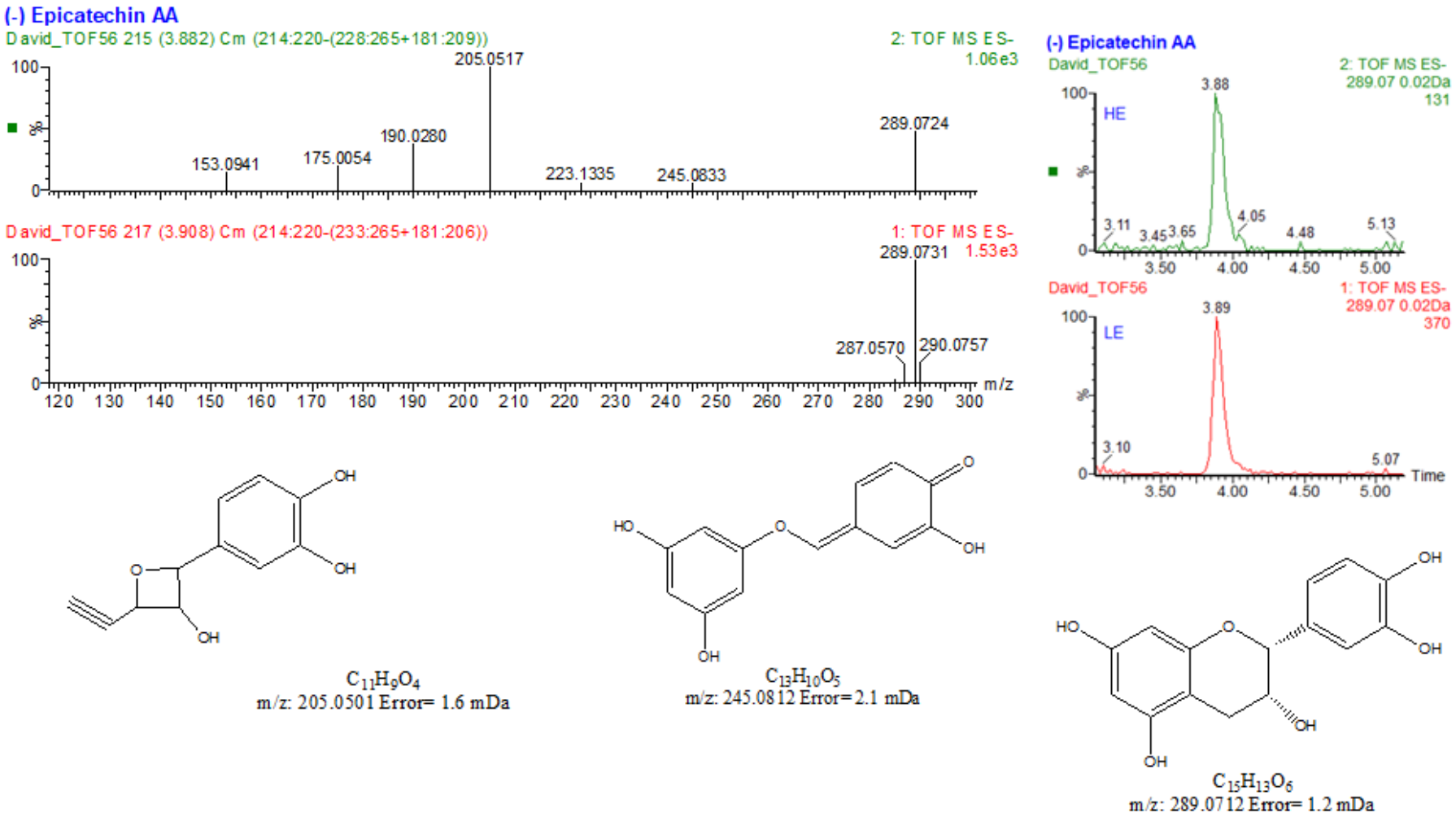
S2.17. Capsaicin detection and confirmation confirmation by UHPLC-QTOF MS in an example of fortified *Arabidopsis thaliana* at 500ppb. (Left) LE and HE spectrum; (right) nwXIC of the protonated molecule in LE and the fragment in HE. Confirmation of the fragment was established by their accurate mass, mass error and retention time. M/z equal to 328.1898 is referred to $[M+Na]^+$ adduct.



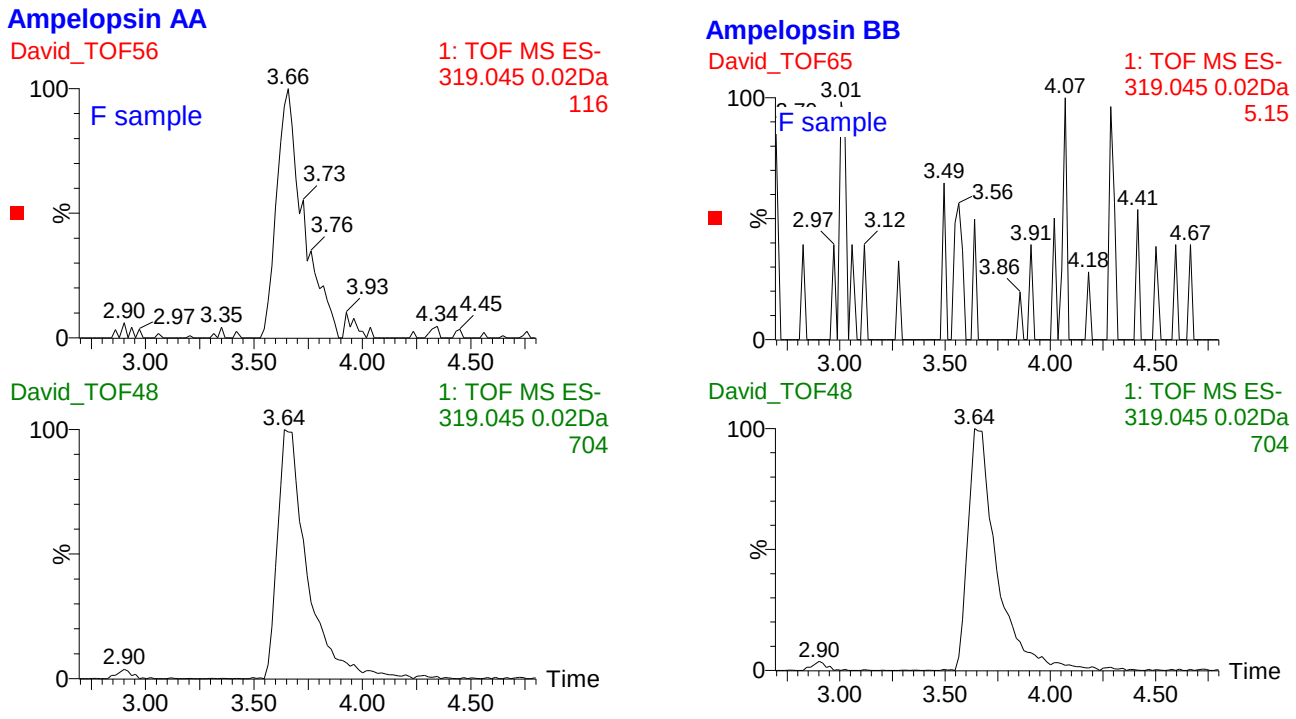
S2.18. Capsaicin detection and confirmation confirmation by UHPLC-QTOF MS in an example of fortified *Solanum lycopersum* at 500ppb. (Left) LE and HE spectrum; (right) nwXIC of the protonated molecule in LE and the fragment in HE. Confirmation of the fragment was established by their accurate mass, mass error and retention time.



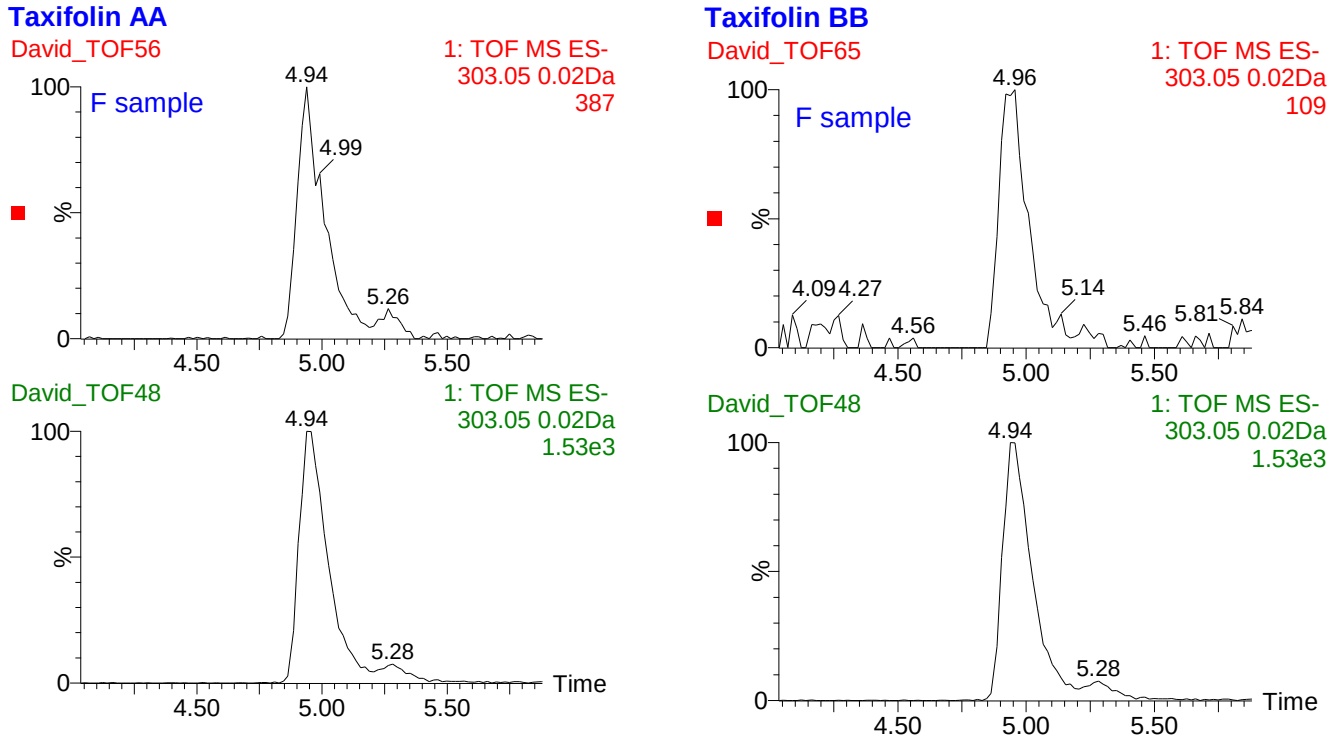
S2.19. Comparison between nwXICs for (-) epicatechin in the fortified plant and in a reference standard at the same level. AA is referred to *Arabidopsis thaliana* (left) and BB to *Solanum lycopersicum* (right). As it can be seen, the peak only appears in the fortified sample of *Arabidopsis thaliana.*



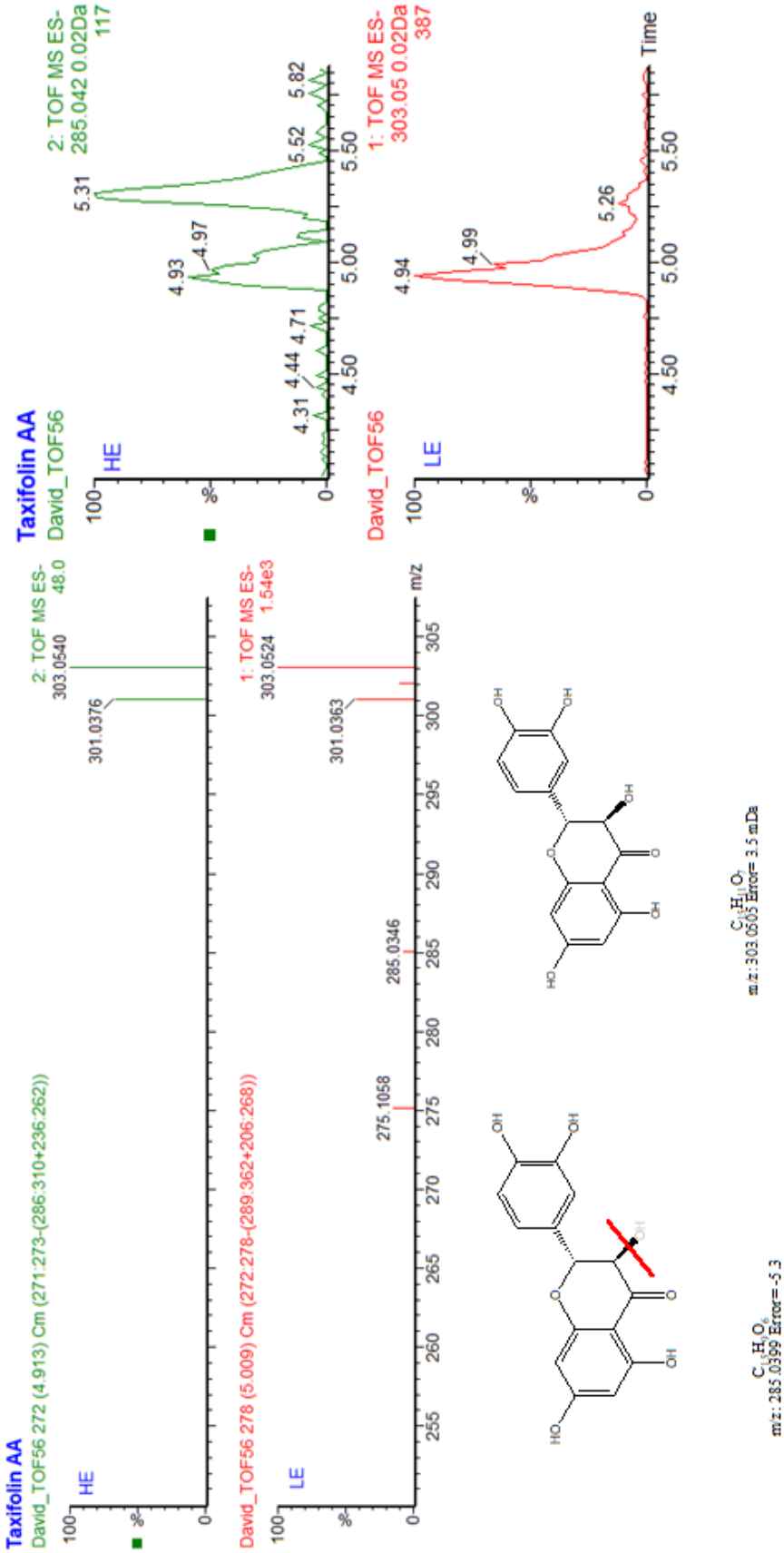
S2.20. (-) Epicatechin detection and confirmation confirmation by UHPLC-QTOF MS in an example of fortified *Arabidopsis thaliana* at 500ppb. (Left) LE and HE spectrum; (right) nwXIC of the protonated molecule in LE and main fragments in HE. Confirmation of the fragments was established by their accurate mass, mass error and retention time.



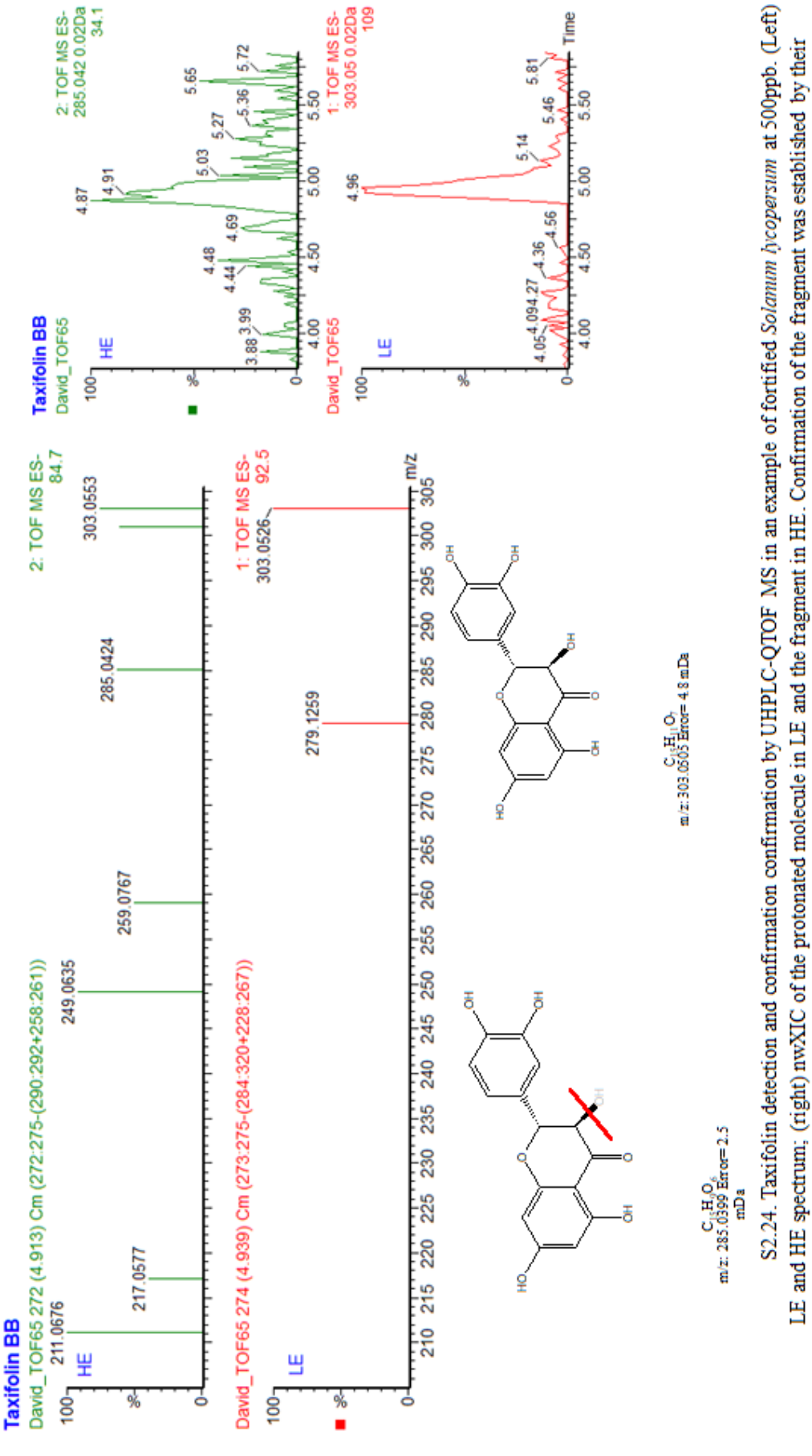
S2.21. Comparison between nwXICs for ampelopsin in the fortified plant and in a reference standard at the same level. AA is referred to *Arabidopsis thaliana* (left) and BB to *Solanum lycopersicum* (right). It sees that the peak of the sample fortified in *Arabidopsis thaliana* s coincides with the standard but the *m / z* corresponds



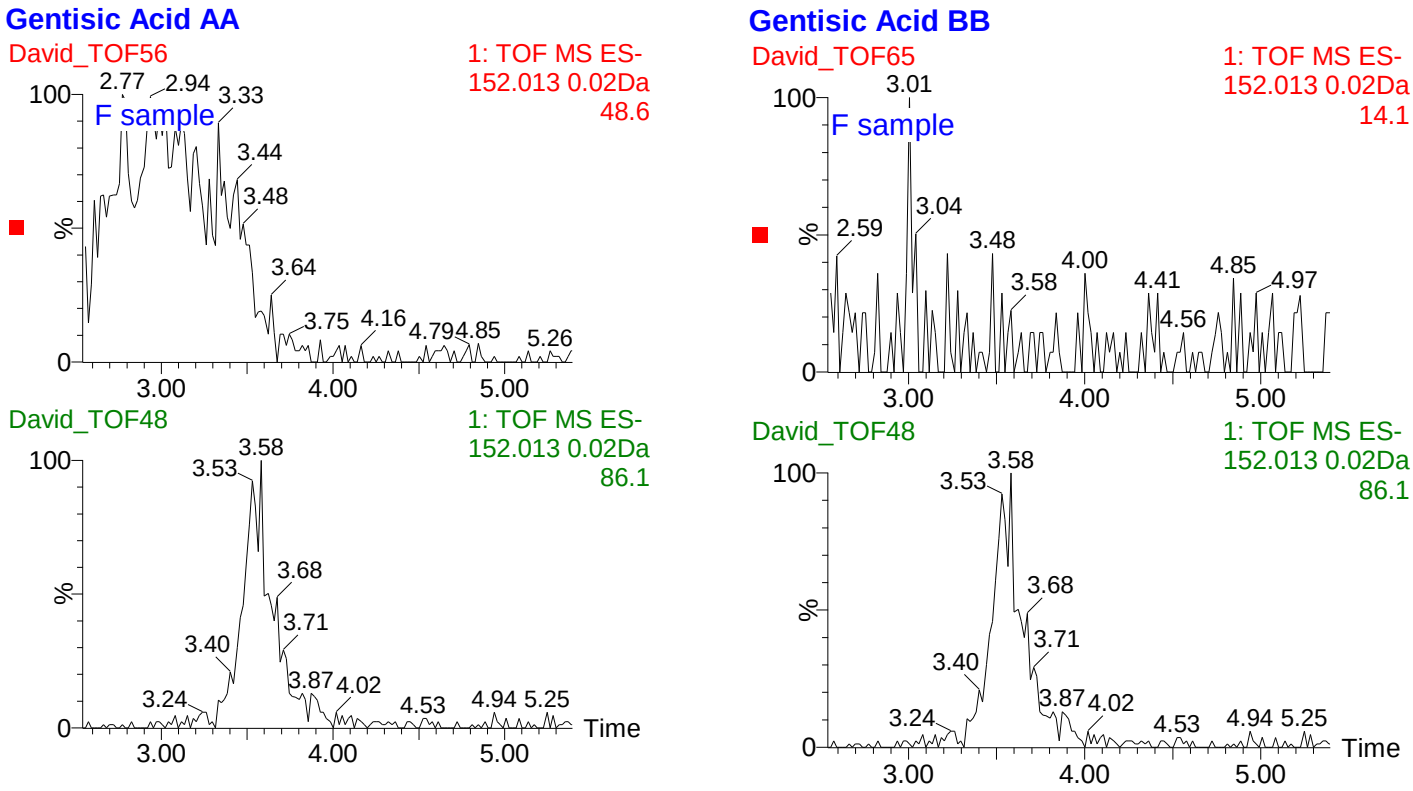
S2.22. Comparison between nwXICs for ampelopsin in the fortified plant and in a reference standard at the same level. AA is referred to *Arabidopsis thaliana* (left) and BB to *Solanum lycopersicum* (right). As it can be seen, the peak appears in both fortified samples.



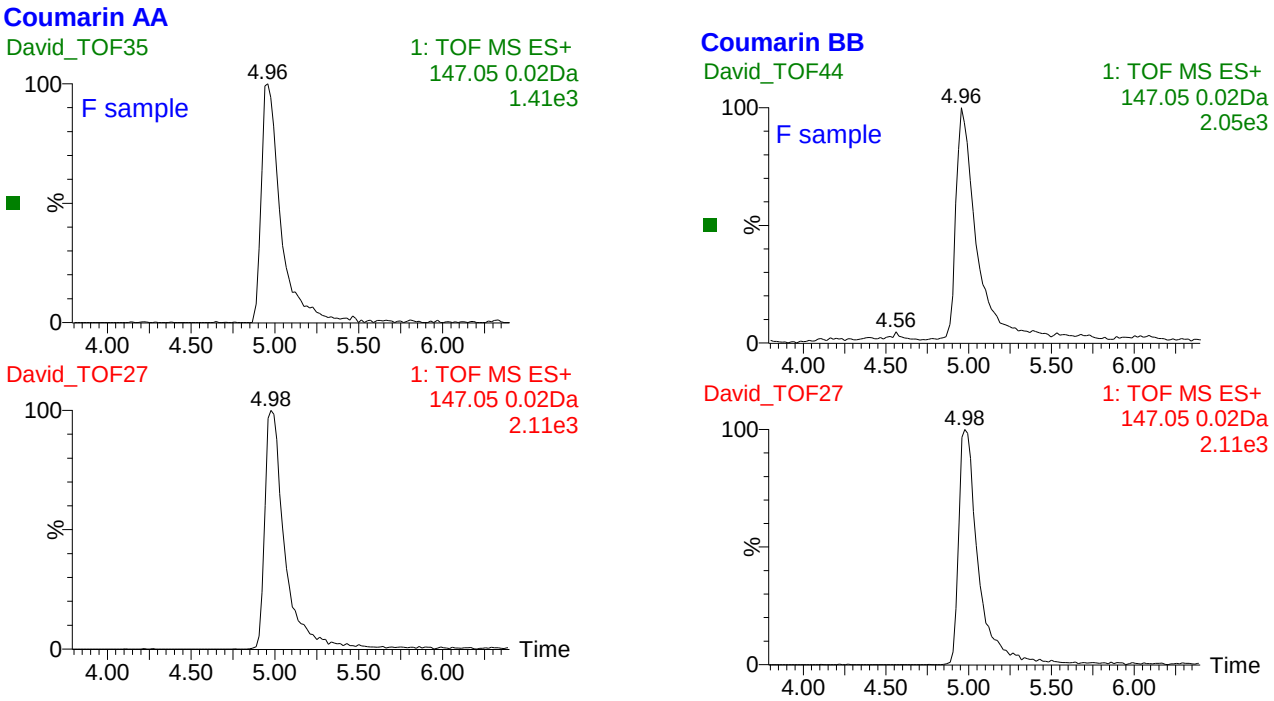
S2.23. Taxifolin detection and confirmation confirmation by UHPLC-QTOF MS in an example of fortified *Arabidopsis thaliana* at 500ppb. (Left) LE and HE spectrum; (right) nwXIC of the protonated molecule in LE and the fragments in HE. Confirmation of the fragment was established by their accurate mass, mass error and retention time.



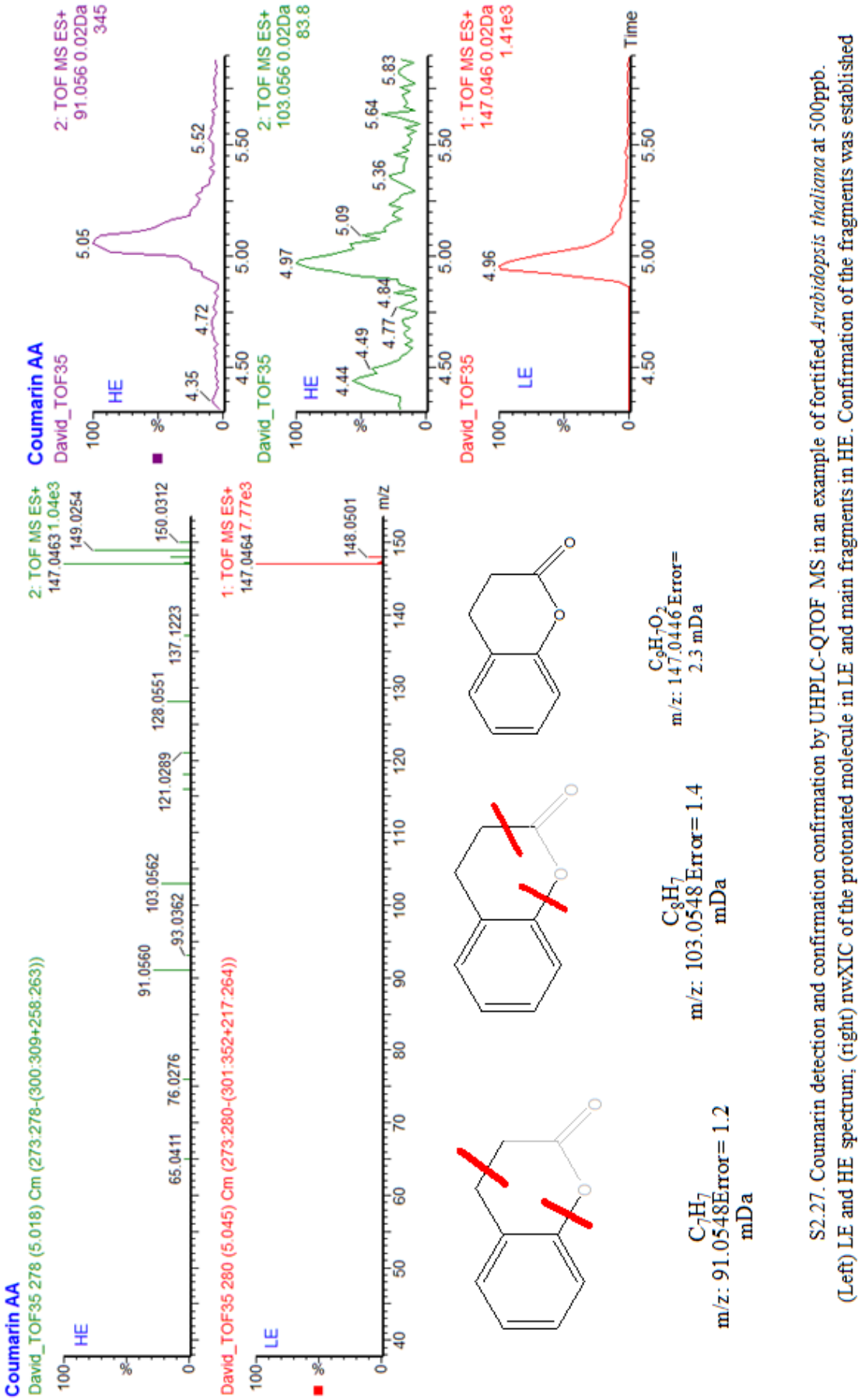
S2.24. Taxifolin detection and confirmation confirmation by UHPLC-QTOF MS in an example of fortified *Solanum lycopersum* at 500ppb. (Left) LE and HE spectrum; (right) nwXIC of the protonated molecule in LE and the fragment in HE. Confirmation of the fragment was established by their accurate mass, mass error and retention time.



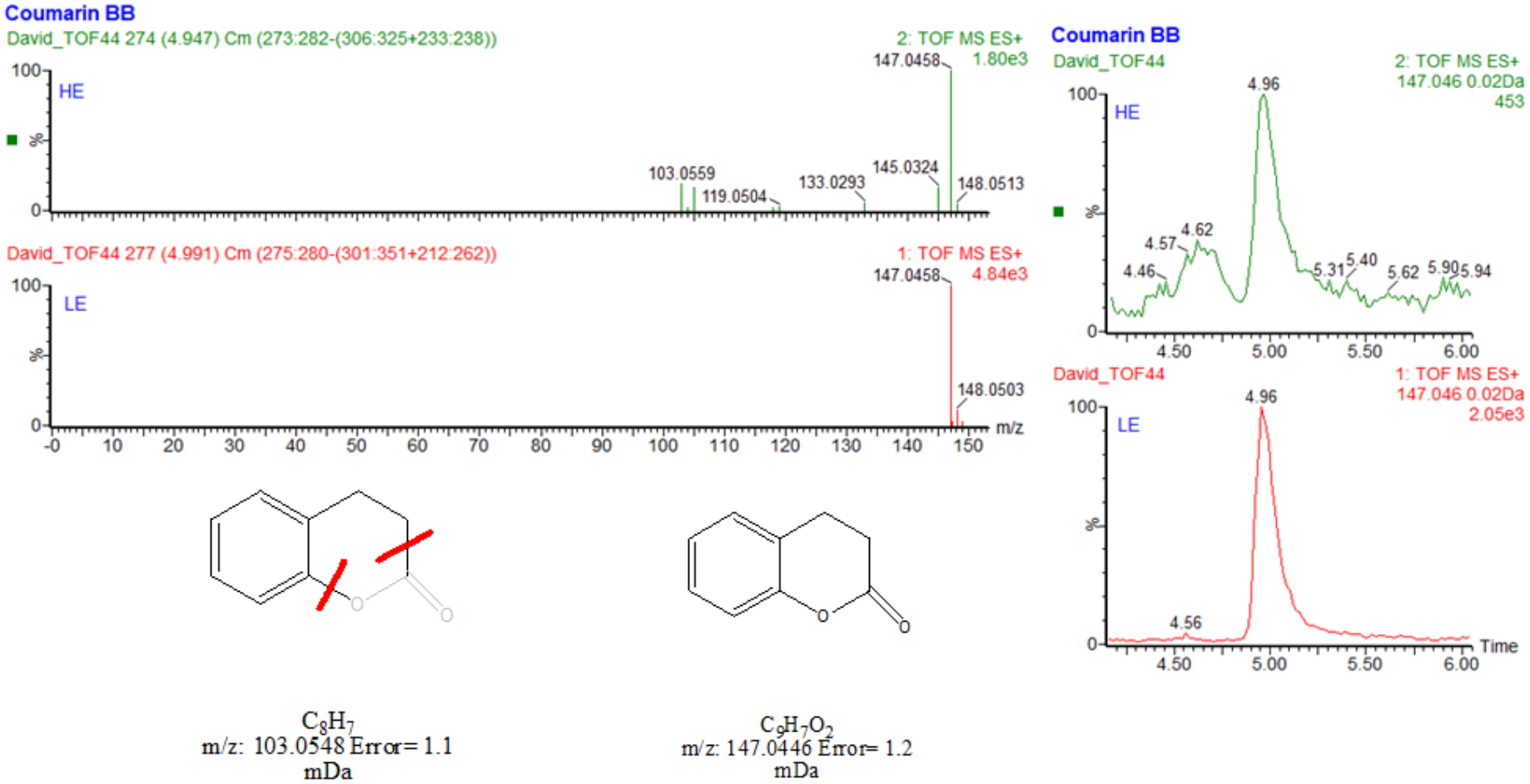
S2.25. Comparison between nwXICs for gentisic acid in the fortified plant and in a reference standard at the same level. AA is referred to *Arabidopsis thaliana* (left) and BB to *Solanum lycopersicum* (right). As it can be seen, the peak only appears in the fortified sample of *Arabidopsis thaliana*.



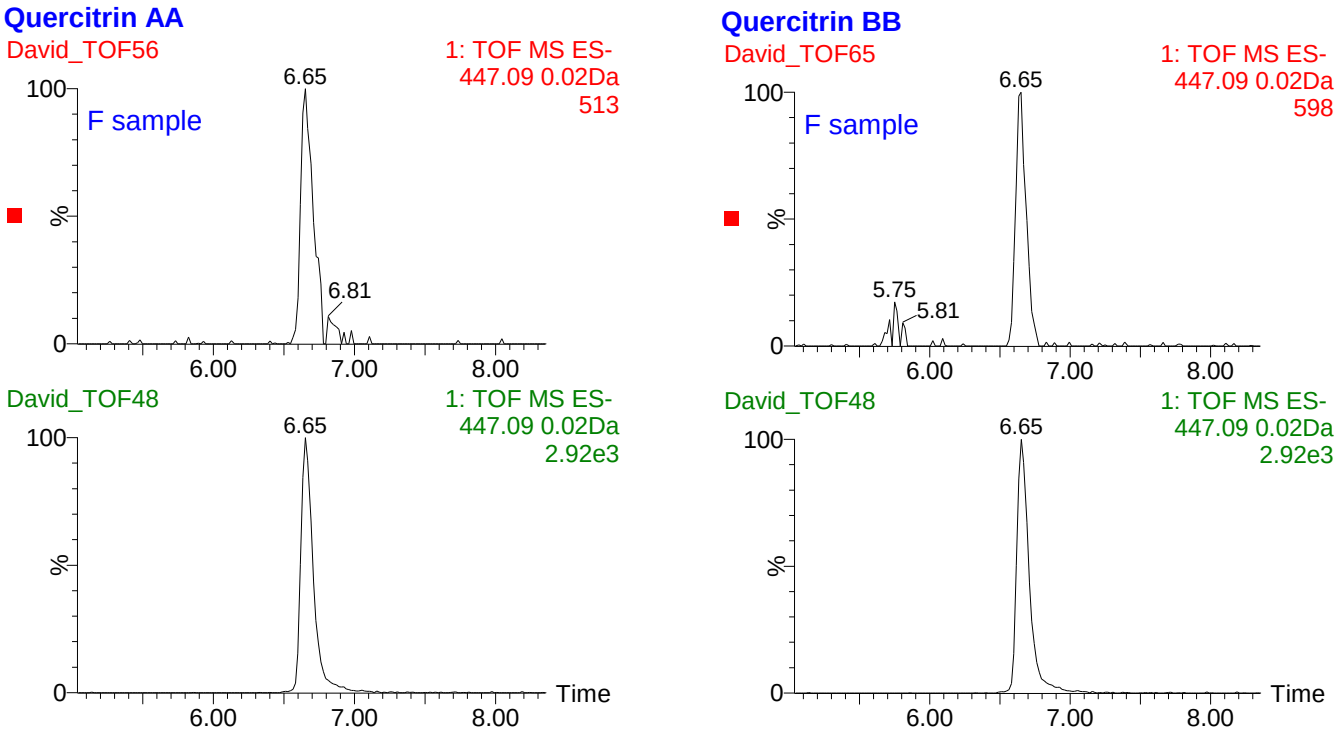
S2.26. Comparison between nwXICs for coumarin in the fortified plant and in a reference standard at the same level. AA is referred to *Arabidopsis thaliana* (left) and BB to *Solanum lycopersicum* (right). As it can be seen, the peak appears in both fortified samples.



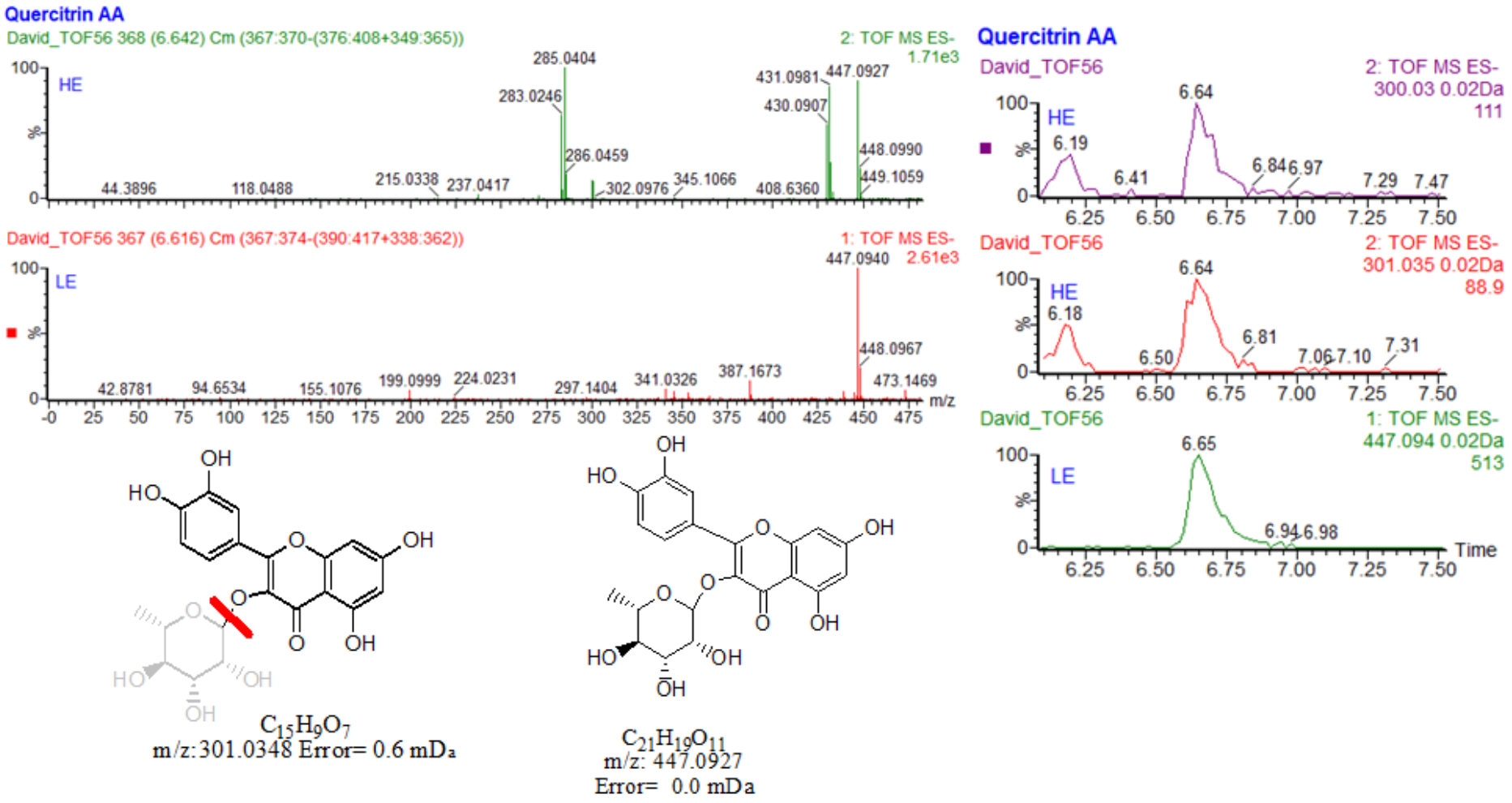
S2.27. Coumarin detection and confirmation confirmation by UHPLC-QTOF MS in an example of fortified *Arabidopsis thaliana* at 500ppb. (Left) LE and HE spectrum; (right) nwXIC of the protonated molecule in LE and main fragments in HE. Confirmation of the fragments was established by their accurate mass, mass error and retention time.



S2.28. Coumarin detection and confirmation confirmation by UHPLC-QTOF MS in an example of fortified *Solanum lycopersum* at 500ppb. (Left) LE and HE spectrum; (right) nwXIC of the protonated molecule in LE and the fragment in HE. Confirmation of the fragment was established by their accurate mass, mass error and retention time.



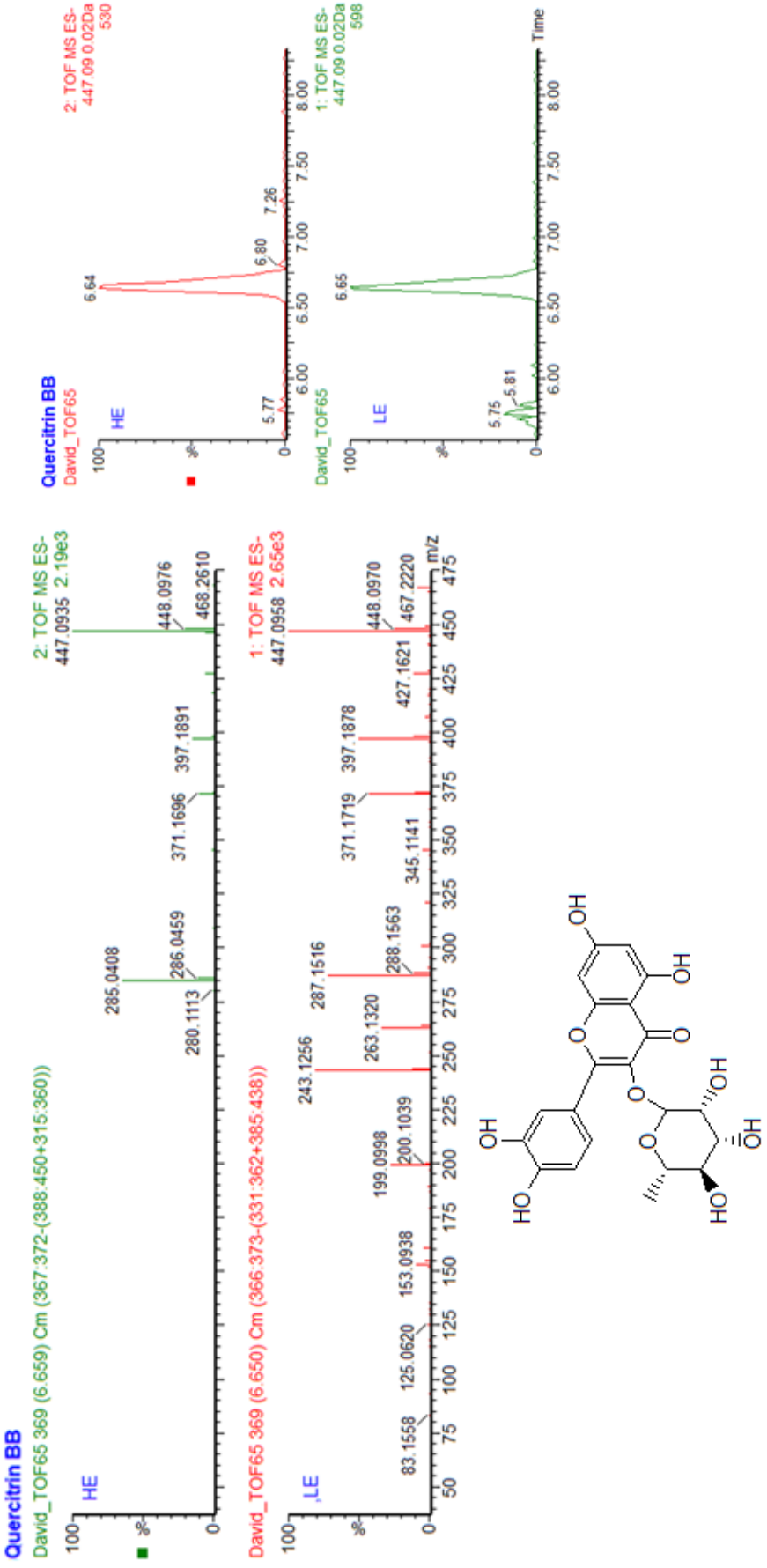
S2.29. Comparison between nwXICs for coumarin in the fortified plant and in a reference standard at the same level. AA is referred to *Arabidopsis thaliana* (left) and BB to *Solanum lycopersicum* (right). As it can be seen, the peak appears in both fortified samples.



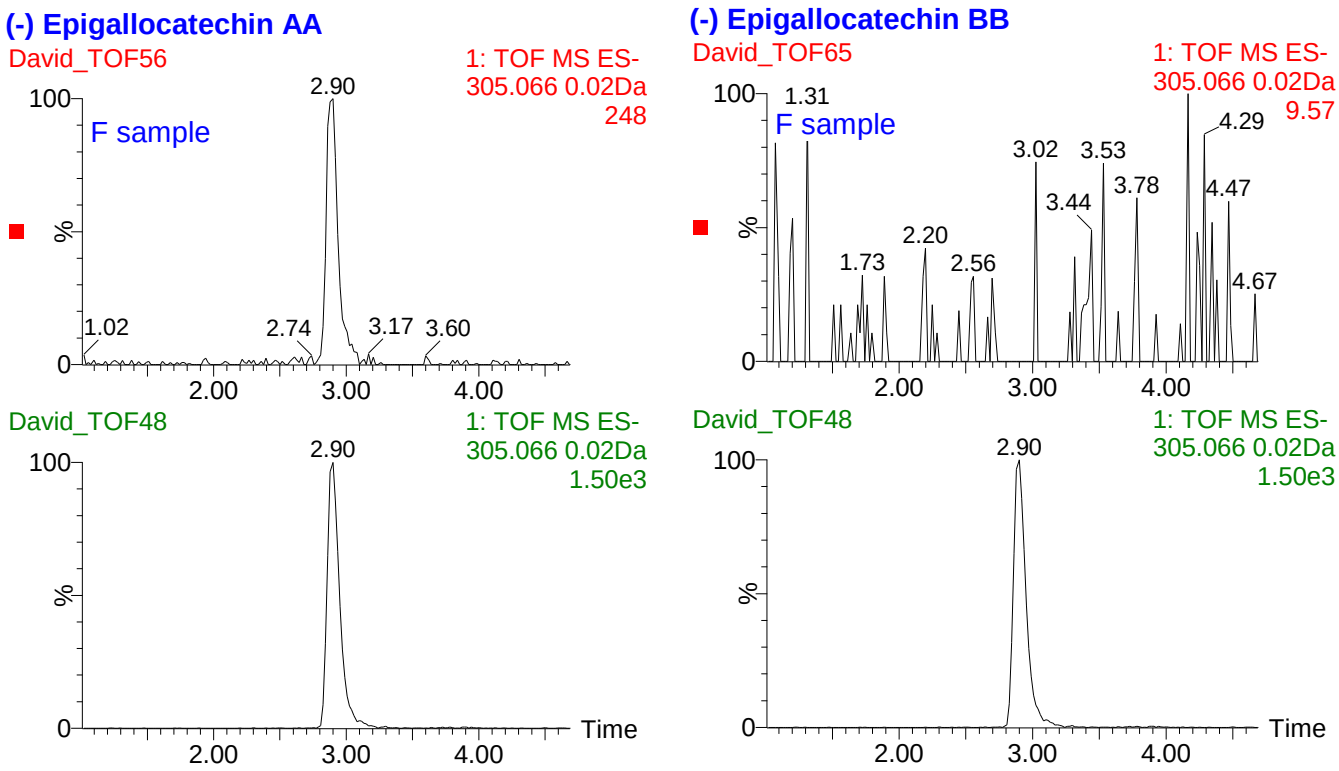
$C_{15}H_9O_7$
m/z:301.0348 Error= 0.6 mDa

$C_{21}H_{19}O_{11}$
m/z: 447.0927
Error= 0.0 mDa

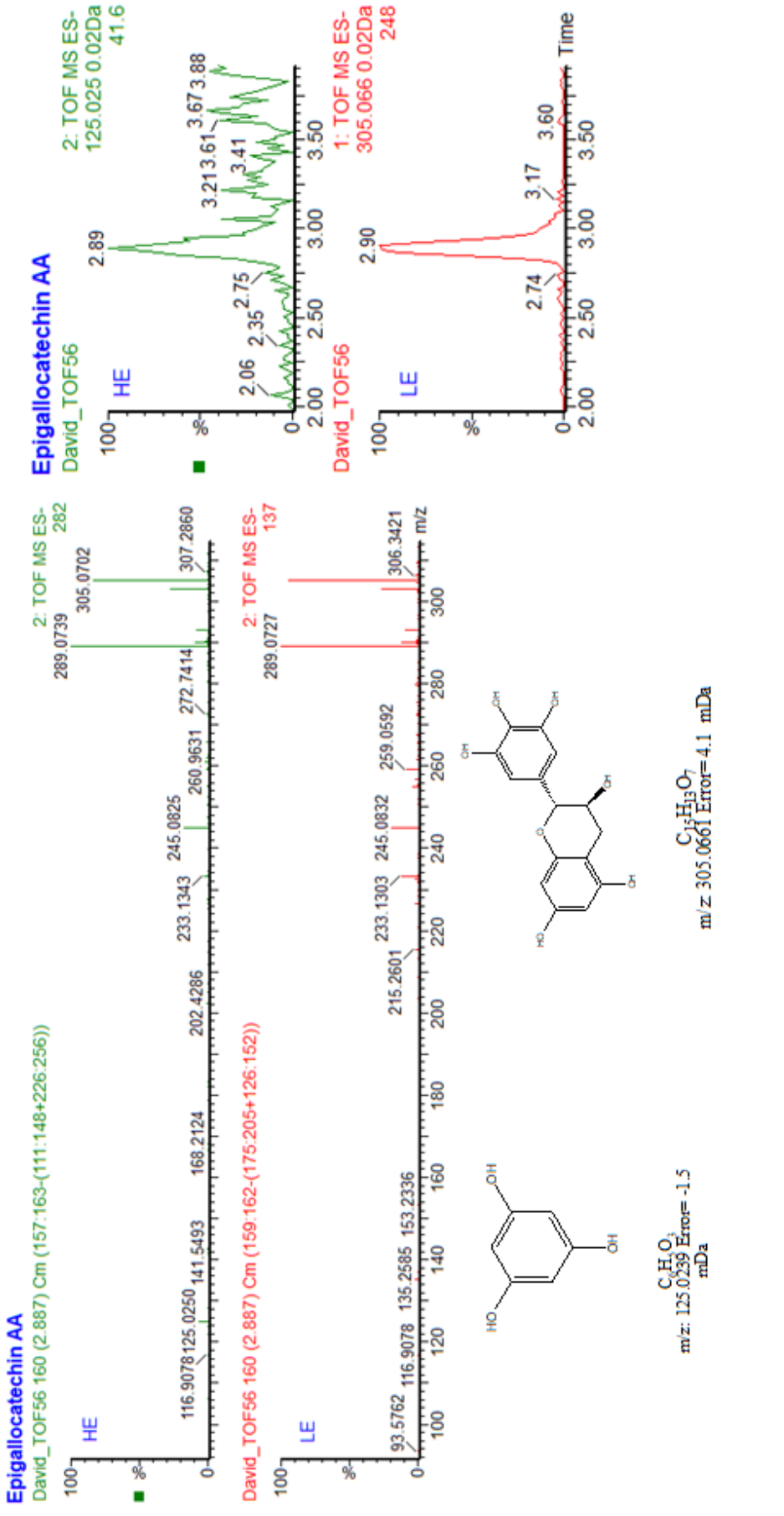
S2.30. Quercitrin detection and confirmation confirmation by UHPLC-QTOF MS in an example of fortified *Arabidopsis thaliana* at 500ppb. (Left) LE and HE spectrum; (right) nwXIC of the protonated molecule in LE and the fragment in HE. Confirmation of the fragment was established by their accurate mass, mass error and retention time.



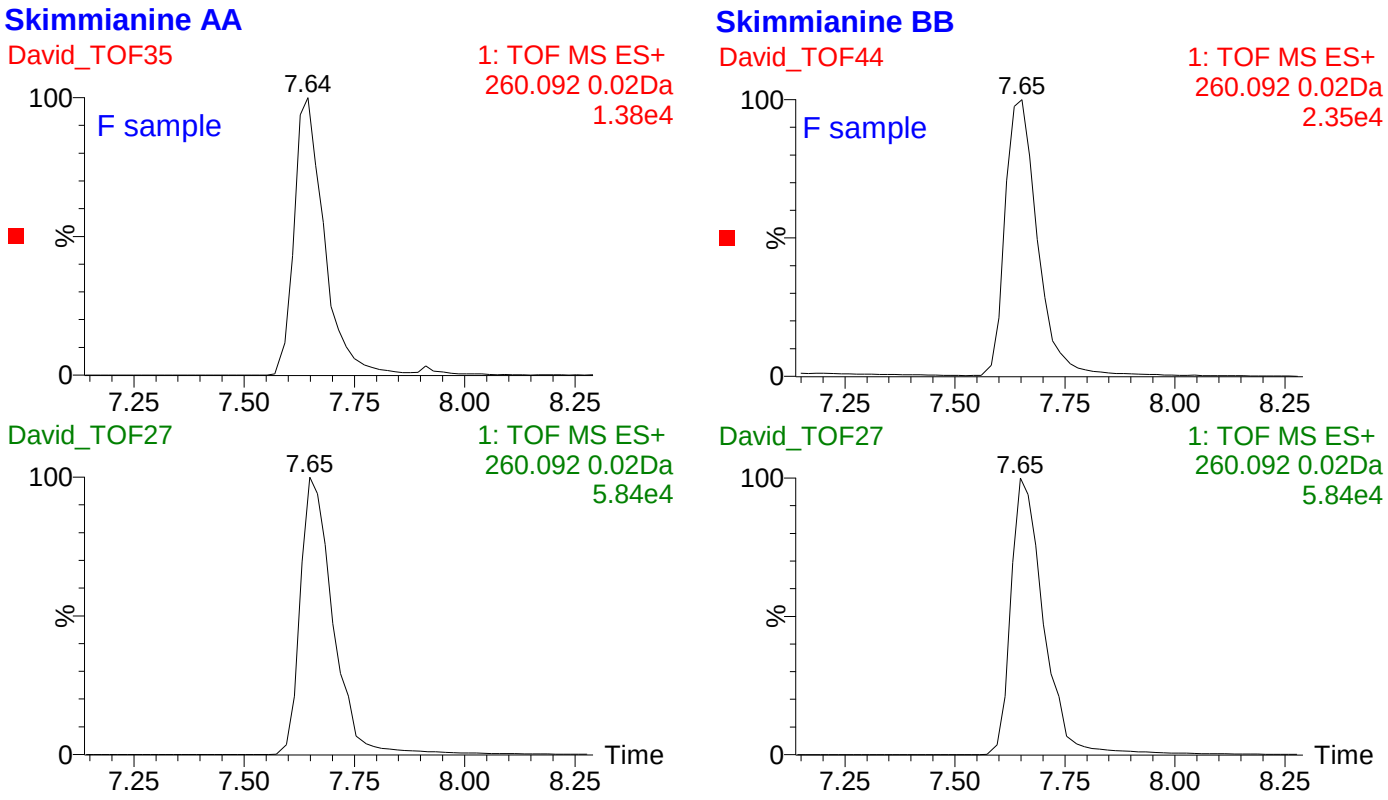
S2.31. Quercitrin detection and confirmation confirmation by UHPLC-QTOF MS in an example of fortified *Solanum lycopersum* at 500ppb. (Left) LE and HE spectrum; (right) nwXIC of the protonated molecule in LE and in HE; no fragmentation was detected.



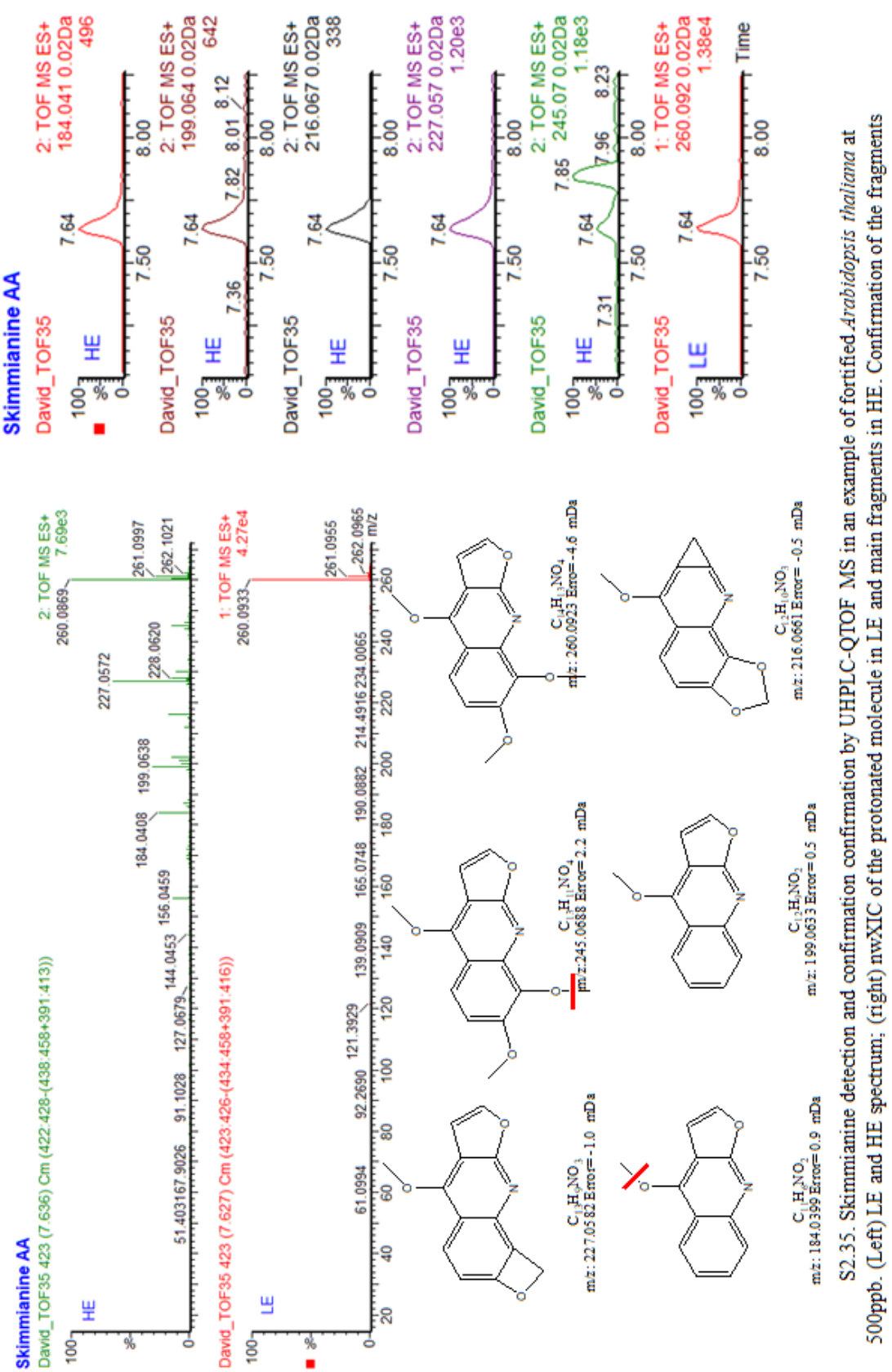
S2.32. Comparison between nwXICs for in the fortified plant and in a reference standard at the same level. AA is referred to *Arabidopsis thaliana* (left) and BB to *Solanum lycopersicum* (right). As it can be seen, the peak only appears in the fortified sample of *Arabidopsis thaliana*.



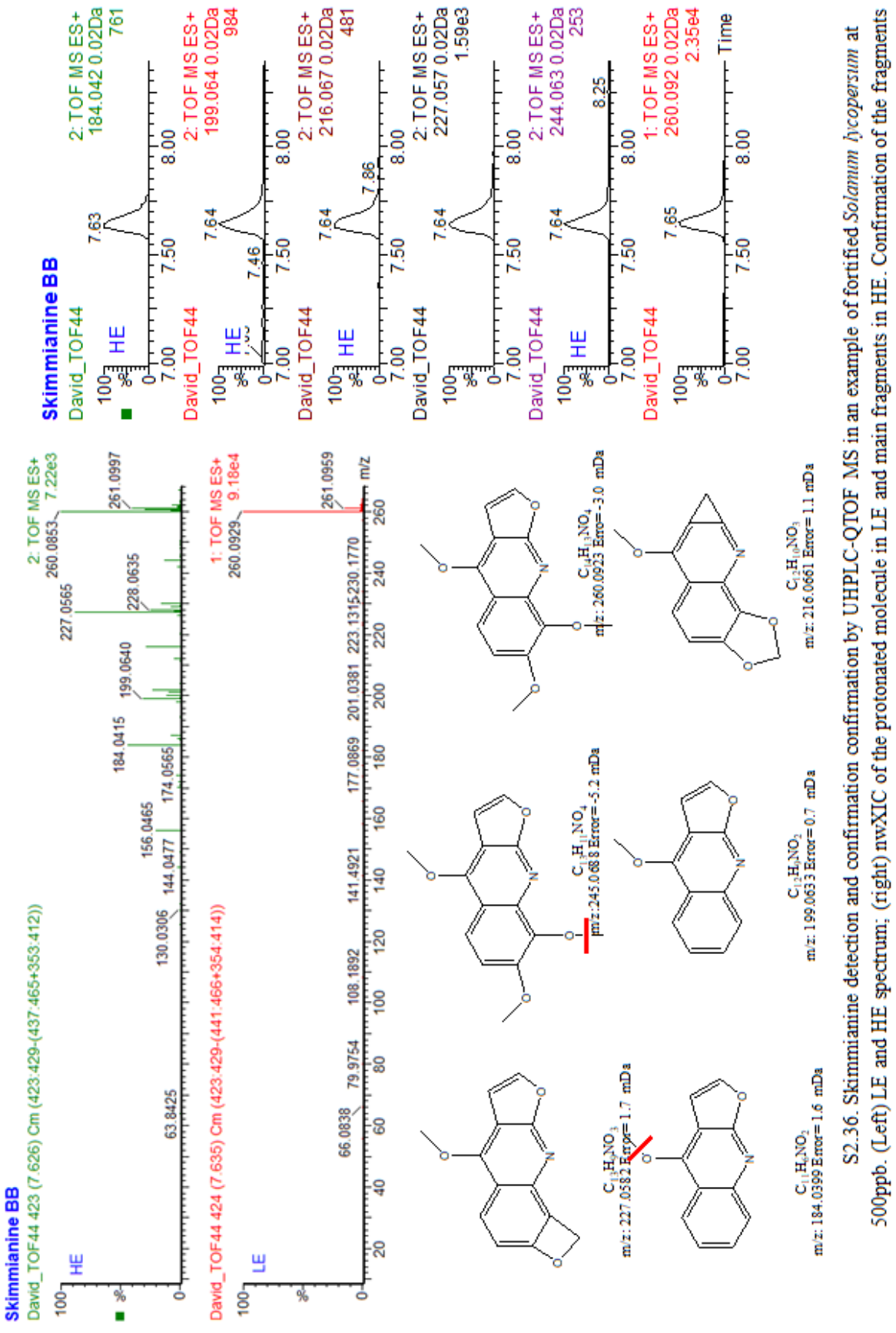
S2.33. (-) Epigallocatechin detection and confirmation confirmation by UHPLC-QTOF MS in an example of fortified *Arabidopsis thaliana* at 500ppb. (Left) LE and HE spectrum; (right) nwXIC of the protonated molecule in LE and fragment in HE. Confirmation of the fragment was established by their accurate mass, mass error and retention time



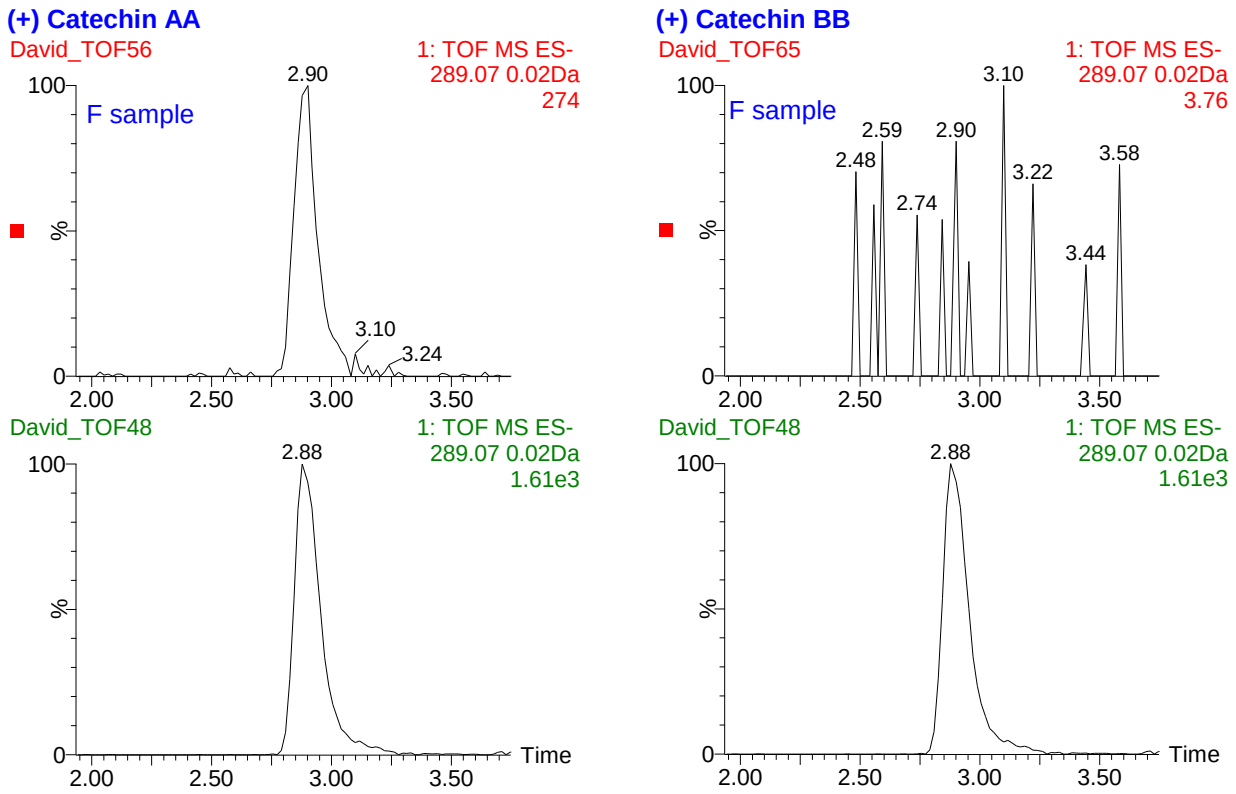
S2.34. Comparison between nwXICs for skimmianine in the fortified plant and in a reference standard at the same level. AA is referred to *Arabidopsis thaliana* (left) and BB to *Solanum lycopersicum* (right). As it can be seen, the peak appears in both fortified samples.



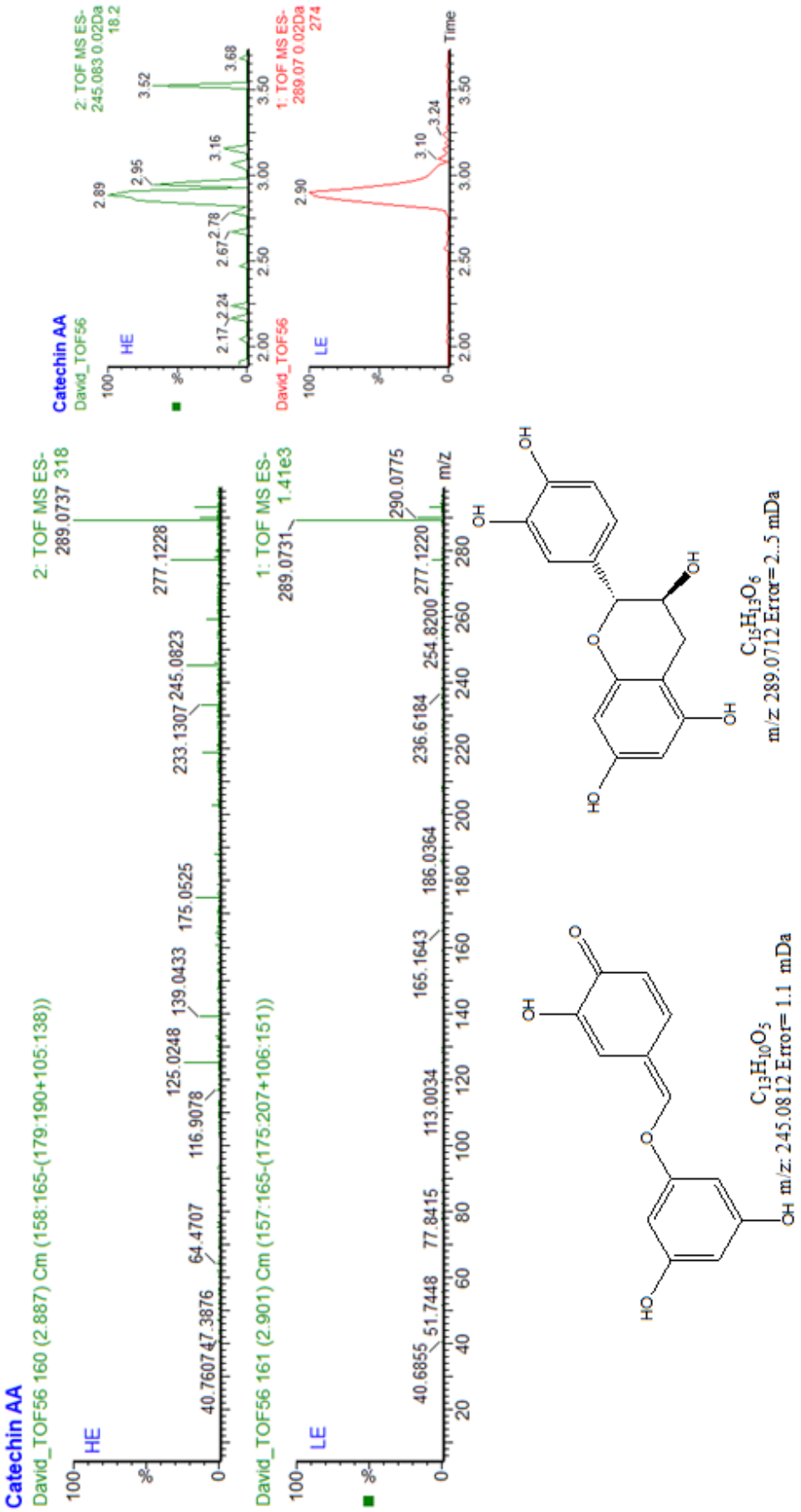
S2.35. Skimmianine detection and confirmation confirmation by UHPLC-QTOF MS in an example of fortified *Arabidopsis thaliana* at 500ppb. (Left) LE and HE spectrum; (right) nwXIC of the protonated molecule in LE and main fragments in HE. Confirmation of the fragments was established by their accurate mass, mass error and retention time.



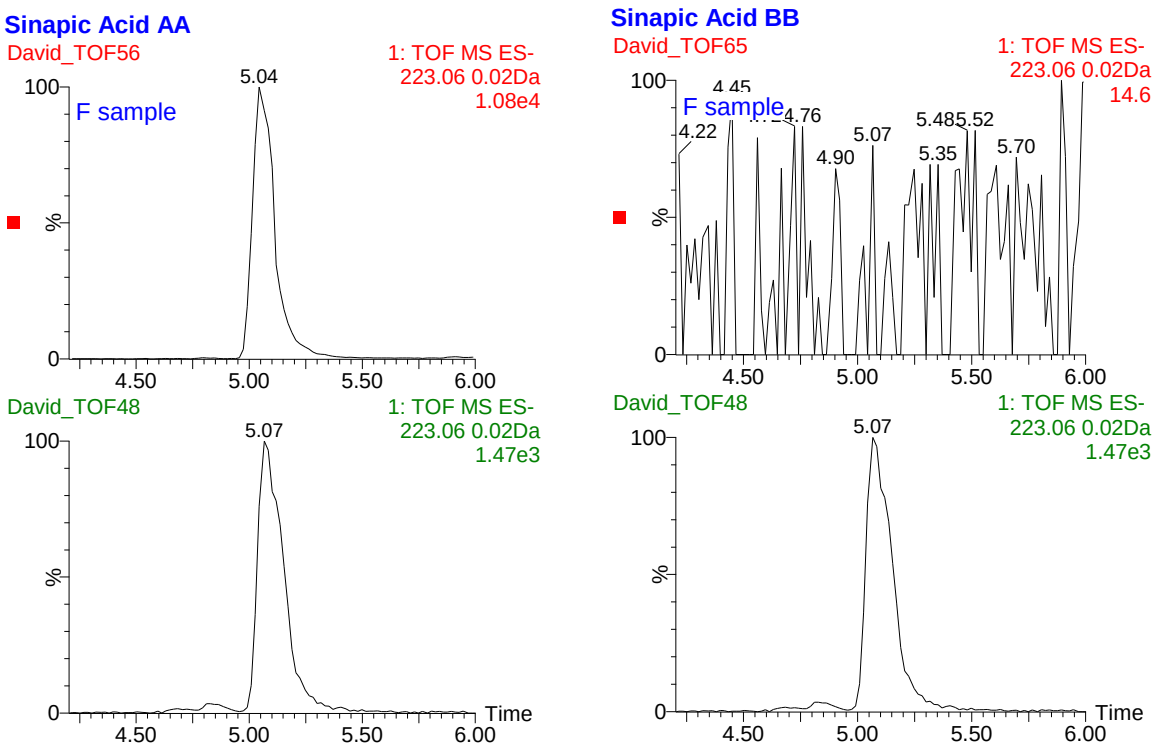
S2.36. Skimmianine detection and confirmation confirmation by UHPLC-QTOF MS in an example of fortified *Solanum lycopersum* at 500ppb. (Left) LE and HE spectrum; (right) nwXIC of the protonated molecule in LE and main fragments in HE. Confirmation of the fragments was established by their accurate mass, mass error and retention time.



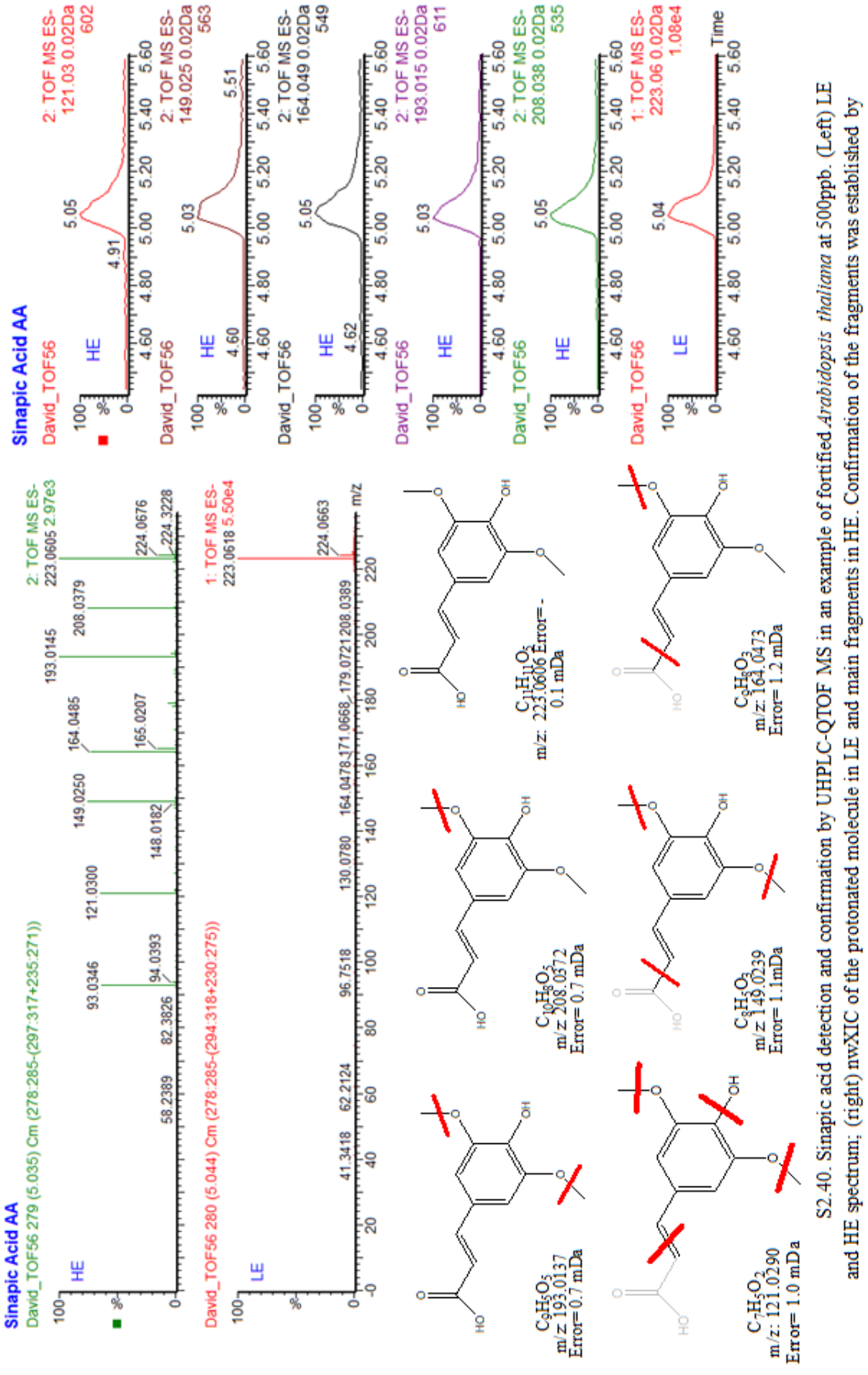
S2.37. Comparison between nwXICs for coumarin in the fortified plant and in a reference standard at the same level. AA is referred to *Arabidopsis thaliana* (left) and BB to *Solanum lycopersicum* (right). As it can be seen, the peak only appears in the fortified sample of *Arabidopsis thaliana*.



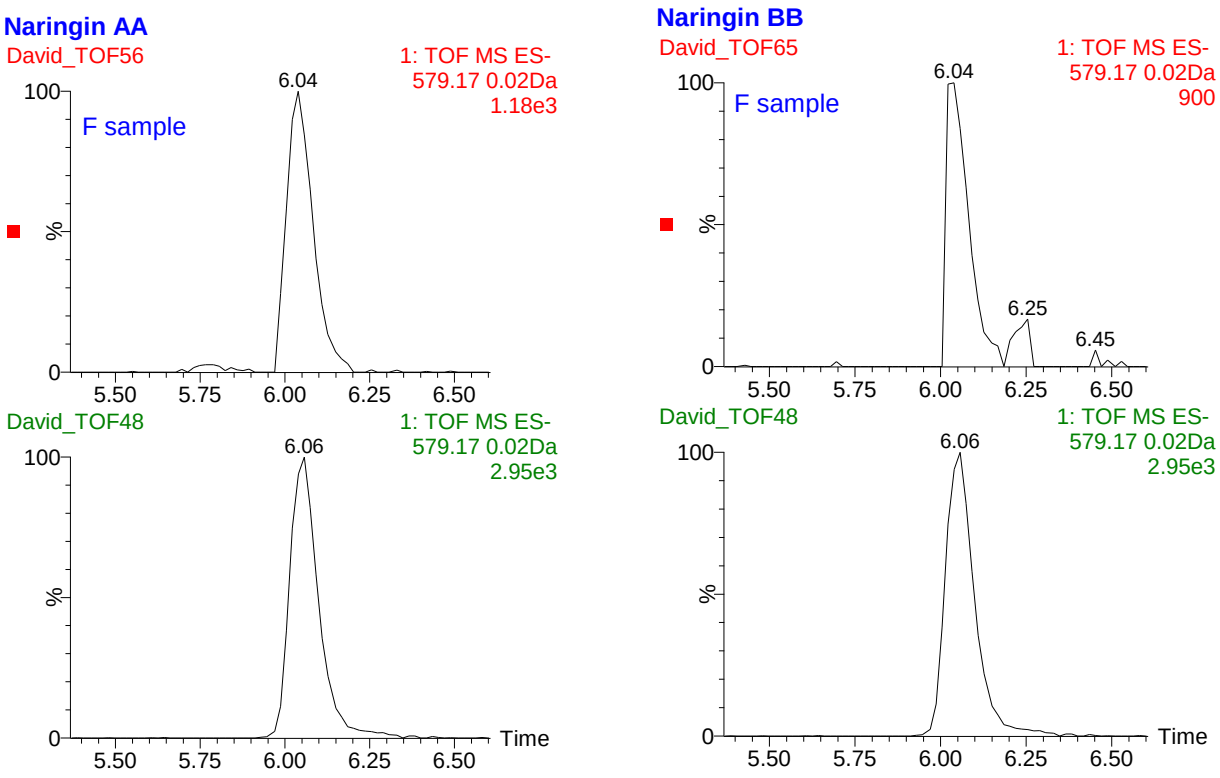
S2.38. (+) Catechin detection and confirmation confirmation by UHPLC-QTOF MS in an example of fortified *Arabidopsis thaliana* at 500ppb. (Left) LE and HE spectrum; (right) nwXIC of the protonated molecule in LE and fragment in HE. Confirmation of the fragment was established by their accurate mass, mass error and retention time.



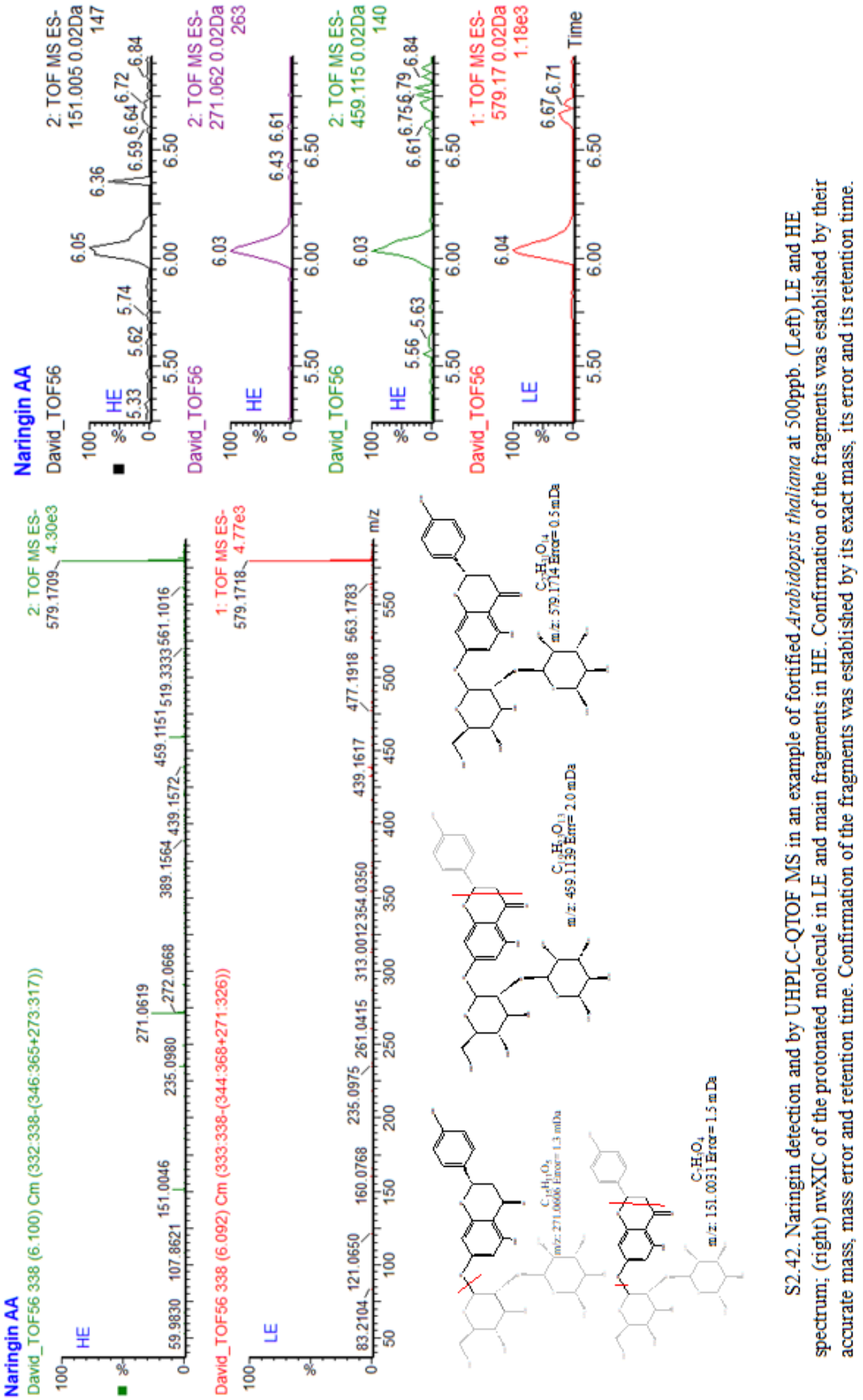
S2.39. Comparison between nwXICs for sinapic acid in the fortified plant and in a reference standard at the same level. AA is referred to *Arabidopsis thaliana* (left) and BB to *Solanum lycopersicum* (right). As it can be seen, the peak only appears in the fortified sample of *Arabidopsis thaliana*.



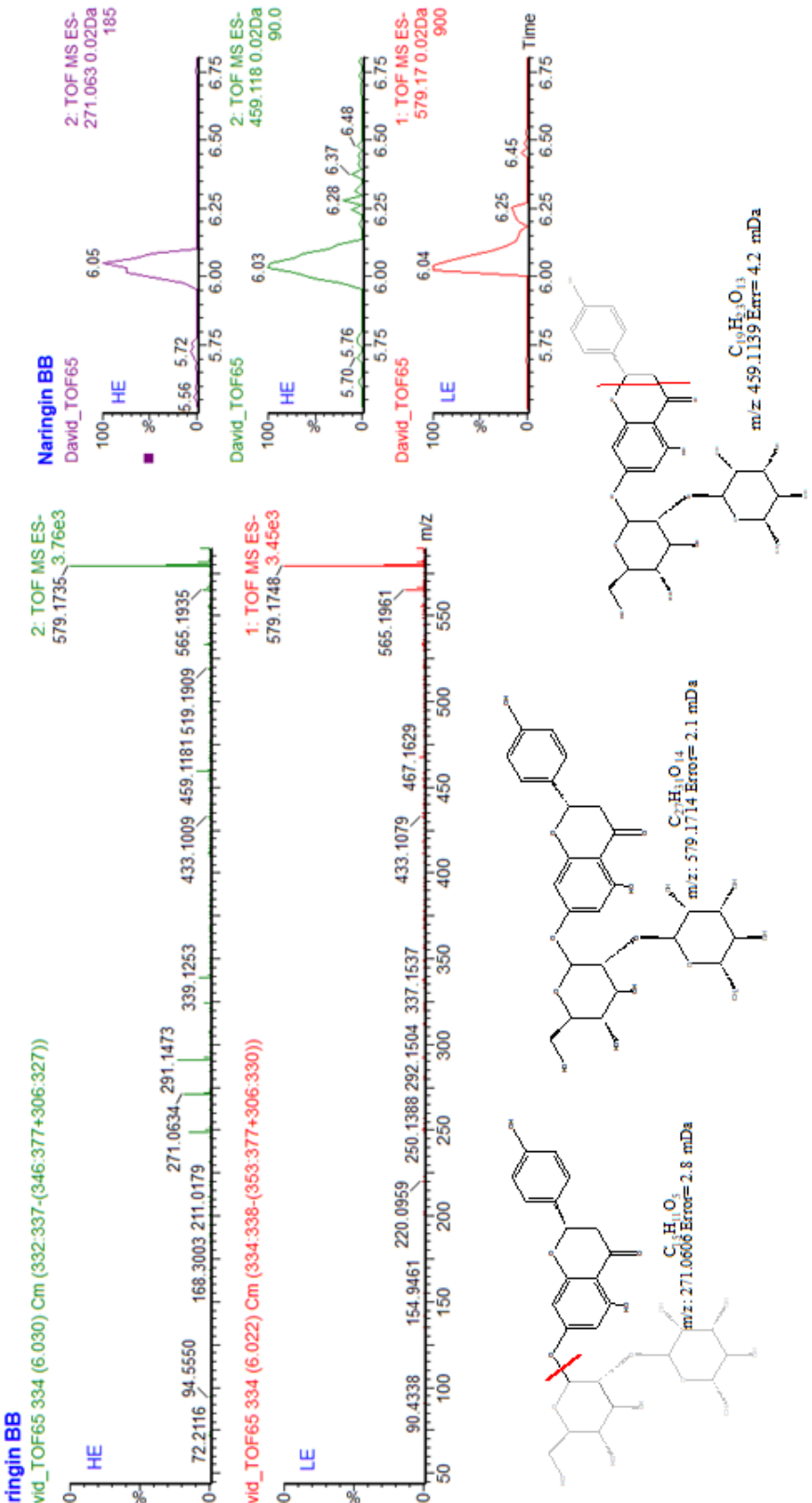
S2.40. Sinapic acid detection and confirmation by UHPLC-QTOF MS in an example of fortified *Arabidopsis thaliana* at 500ppb. (Left) LE and HE spectrum; (right) nwXIC of the protonated molecule in LE and main fragments in HE. Confirmation of the fragments was established by their accurate mass, mass error and retention time.



S.241. Comparison between nwXICs for skimmianine in the fortified plant and in a reference standard at the same level. AA is referred to *Arabidopsis thaliana* (left) and BB to *Solanum lycopersicum* (right). As it can be seen, the peak appears in both fortified samples.



S2.42. Naringin detection and by UHPLC-QTOF MS in an example of fortified *Arabidopsis thaliana* at 500ppb. (Left) LE and HE spectrum; (right) nwXIC of the protonated molecule in LE and main fragments in HE. Confirmation of the fragments was established by their accurate mass, mass error and retention time. Confirmation of the fragments was established by its exact mass, its error and its retention time.



S2.43. Naringin detection and confirmation by UHPLC-QTOF MS in an example of fortified *Solanum lycopersicum* at 500ppb. (Left) LE and HE spectrum; (right) nwXIC of the protonated molecule in LE and main fragments in HE. Confirmation of the fragments was established by their accurate mass, mass error and retention time Confirmation of the fragments was established by its exact mass, its error and its retention time.

S3. Extraction efficiency: peak area

S3.1. Peak areas of the samples fortified at 0 and 150 ppb of library compounds mixture and standardas at the same level. Peak areas were quantified by XCMS package in R.

Compound	Genotype	Fortified sample 0 ppb			Fortified sample 150 ppb			Standard 150 ppb
Neohesperidin	*Arabidopsis*	0	0	0	421	368	333	4547
Ferulic acid	*Arabidopsis*	274	203	270	346	360	386	453
Caffeic acid	*Arabidopsis*	0	0	0	829	2503	2097	4951
Chlorogenic acid	*Arabidopsis*	0	0	0	526	1089	865	3033
	Solanum	1880	1070	2580	2110	1900	2100	292
4-coumaryl alcohol	*Arabidopsis*	0	0	0	312	447	422	716
	Solanum	0	0	0	28	24	21	72
Capsaicin	*Arabidopsis*	0	0	0	114	252	247	541
	Solanum	0	0	0	694	1578	1141	4848
Epicatechin	*Arabidopsis*	0	0	0	27	937	830	4621
Taxifolin	*Arabidopsis*	0	0	0	652	1120	1062	4068
	Solanum	0	0	0	417	285	189	3627
Coumarine	*Arabidopsis*	0	0	0	428	585	589	560
	Solanum	1064	831	840	5510	7507	6788	5827
Quercitrin	*Arabidopsis*	0	0	0	895	1040	947	5441
	Solanum	105	63,6	88,3	197	187	162	556
Epigallecatechin	*Arabidopsis*	0	0	0	148	768	486	4305
Skimmianine	*Arabidopsis*	0	0	0	29593	62807	58871	110530
	Solanum	626	358	252	23746	51599	44252	110530
Catechin	*Arabidopsis*	0	0	0	148	768	486	4305
Sinapic acid	*Arabidopsis*	61065	68179	33940	80818	27641	35578	3987
Naringin	*Arabidopsis*	553	284	272	2540	2553	2306	5892
	Solanum	0	0	0	2338	1805	2439	5892

S.3.2. Peak areas of the samples fortified at 500ppb and standards at the same concentration level of library compounds mixture. Peak areas were quantified by XCMS package in R.

Compound	Genotype	Fortified sample 500 ppb			Standard 150 ppb	Standard 500 ppb
Neohesperidin	*Arabidopsis*	909	912	880	4547	11937
Ferulic acid	*Arabidopsis*	657	674	719	453	1160
Caffeic acid	*Arabidopsis*	5416	6339	6760	4951	12436
Chlorogenic acid	*Arabidopsis*	2748	3269	2604	3033	9818
	Solanum	2130	2690	2620	292	842
4-coumaryl alcohol	*Arabidopsis*	978	1207	1264	716	2247
	Solanum	50	47	46	72	226
Capsaicin	*Arabidopsis*	571	249	504	541	1510
	Solanum	3458	3187	3422	4848	13559
Epicatechin	*Arabidopsis*	1580	1897	2428	4621	11997
Taxifolin	*Arabidopsis*	2455	2786	2858	4068	11561
	Solanum	681	793	760	3627	11561
Coumarine	*Arabidopsis*	1250	944	1160	560	1430
	Solanum	14508	16289	14709	5827	14766
Quercitrin	*Arabidopsis*	2599	3118	2365	5441	15899
	Solanum	273	322	323	556	1630
Epigallecatechin	*Arabidopsis*	1433	1672	2147	4305	11563
Skimmianine	*Arabidopsis*	148716	64510	136325	110530	295854
	Solanum	73184	119130	118215	110530	295854
Catechin	*Arabidopsis*	1433	1672	2147	4305	11563
Sinapic acid	*Arabidopsis*	76557	66917	33739	3987	10821
Naringin	*Arabidopsis*	5906	6225	6561	5892	16034
	Solanum	4558	4678	4627	5892	16894