Mass Spectrometric Structural Elucidation of Sporopollenin and Lignin

by

©Abanoub Mikhael

A dissertation submitted to the School of Graduate Studies

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Department of Chemistry

Memorial University

St. John's, Newfoundland, Canada

July 2021

Abstract

Sporopollenin is a natural, highly cross-linked polymer composed of carbon, hydrogen, and oxygen, which forms the outer wall of pollen grains. Sporopollenin is highly resilient to mild chemical degradation. Because of this stability, its exact chemical structure and the biochemical pathways involved in its synthesis are not yet clear.

There is no mild degradation method capable of degrading sporopollenin without extreme modification and/or depolymerization. However, several harsh degradation methods were capable of degrading sporopollenin into a mixture of small and modified monomers that provides no clue to the sporopollenin complex native structure. In this thesis, Sporopollenin's molecular structure was demystified and unraveled without any prior degradation and/or modification (Top-down approach). This was accomplished using Secondary Ion Mass Spectrometry SIMS-TOF-MS/MS and Matrix-Assisted Laser Desorption MALDI-TOF-MS/MS. The latter mass spectrometric techniques allowed us to identify the two major building units of the sporopollenin network. The identified building units were supported by solid-state NMR studies.

Lignin oligomers are an essential source of aromatic compounds; it is composed mainly of three monomers: *p*-coumaryl (H), coniferyl (G), and sinapyl (S) alcohols which are linked covalently in different fashions. It is well known that various methods for releasing and purifying lignin from wood cellulose lead to a drastic change in lignin chemical structure. In 2015, Banoub et al. introduced a new paradigm that states that accurate structural studies should be done on lignin oligomers released from the polysaccharide without further purification or chemical transformations. This type of lignin

is called virgin released lignin (VRL). In the second part of this thesis, different MS ionization methods such as MALDI, APPI (Atmospheric Pressure Photoionization, and ESI (Electrospray ionization) coupled with tandem mass spectrometry was used to investigate the major structural features of the French Oak and Pine VRLs.

Acknowledgments

I would like to express my deep appreciation to my supervisor, Dr. Joseph H. Banoub, Adjunct Professor of Chemistry at Memorial University of Newfoundland and Head of Special Projects, Science Branch, Department of Fisheries and Oceans, Canada, for his support, advice, and guidance during my Ph.D. studies.

I also would like to thank my supervisory committee Dr. Travis Fridgen and Dr. Peter Warburton. I would like to offer my special thanks to Dr. Travis Fridgen for reviewing the drafts of the manuscripts and for his valuable advice during my Ph.D. studies.

I would like to express gratitude to both the Department of Fisheries and Oceans (DFO), St John's, and the Chemistry Department of the Memorial University of Newfoundland for providing the resources which enabled me to perform my research studies.

I would like to thank Professor Michel Delmas for providing us with the lignin VRL samples and their NMR data. Also, I would like to thank Professor Grahame Mackenzie for providing us the sporopollenin sample.

I am extremely grateful to Physical Electronics, Minnesota, USA, for the help in measuring the High-Resolution X-Ray Photoelectron Spectrometry (HR-XPS) and the TOF-SIMS-MS/MS analyses of sporopollenin.

I would like to thank Kristina Jurcic, MALDI Mass Spectrometry Facility, Western University, Ontario, for her help in measuring MALDI TOF-MS/MS of sporopollenin and oak VRL.

I would like to thank Dr. Celine Schneider, C-Cart facility, for her help in measuring the solid-state NMR of sporopollenin and her cooperation in analyzing the NMR data. Also, I would like to acknowledge Dr. Paris E. Georghiou for reviewing the sporopollenin manuscript draft and his contribution with a DFT computational model in the sporopollenin published manuscript.

I would like to express my gratitude to the School of Graduate Studies (SGS) of the Memorial University of Newfoundland for the financial support throughout my degree program.

Lastly, I would like to express thanks to my family. I am very grateful to my mother "Mona Amin", my father "Atef Fahmy" and my brothers "Michael and Mina" for their endless love, support, and encouragement. I would like to express my gratitude to my wife "Nourhan" for her love, patience, and understanding during my studies. Finally, countless thanks to my lovely daughter "Marla" whose smile has been the driving force behind my success.

Table of Contents

List of Figures

List of Tables

List of Schemes

List of Abbreviations and Symbols

SIMS Secondary Ion Mass Spectrometry

TOF Time of Flight

MALDI Matrix-Assisted Laser Desorption

NMR Nuclear Magnetic Resonance

VRL Virgin Released Lignin

MS Mass Spectrometry

APPI Atmospheric Pressure Photoionization

ESI Electrospray ionization

HR-XPS High-Resolution X-Ray Photoelectron Spectrometry

DFT Density functional theory

CIMV "La Compagnie Industrielle de la Matière Végétale"

DHB 2,5-dihydroxybenzoic acid

α-CHCA α-cyano-4-hydroxycinnamic acid

DAN 1,5-diamino naphthalene

CID collision Induced dissociation

QTOF quadrupole-time-of flight mass spectrometer

MS/MS Tandem Mass Spectrometry

KMD Kendrick Mass defect

DBE double bond equivalent

NOESY Nuclear Overhauser Effect Spectroscopy

CP-MAS Cross Polarization Magic-Angle Spinning

HSQC	heteronuclear single quantum correlation
HABA	2-(4-Hydroxyphenylazo)benzoic acid
CPMG	Carr-Purcell-Meiboom-Gill
DAG	Diacylglycerol
TAG	Triacylglycerol
LCC	lignin-carbohydrate complexes
RDA	retro Diels alder reaction
DP	declustering potential
FP	Focusing Potential
CE	Collision energy

List of Appendices

Co-authorship statement

The research presented in Chapters 2-5 of this thesis has been conducted by the principal author Abanoub Mikhael for the Doctor of Philosophy degree under the supervision of Prof. Joseph Banoub.

Chapter 2. Mikhael, A.; Jurcic, K.; Schneider, C.; Karr, D.; Fisher, G.L.; Fridgen, T.D.; Diego-Taboada, A.; Georghiou, P.E.; Mackenzie, G.; and Banoub, J.: **Demystifying and unravelling the molecular structure of the biopolymer sporopollenin**. *Rapid Commun. Mass Spectrom.* 2020; 34: e8740. https://doi.org/10.1002/rcm.8740**.** Abanoub Mikhael performed all the data interpretation and manuscript preparation except in the NMR section was in co-operation with Dr. Celine Schneider. Dr. Kristina Jurcic measured the MALDI-TOF-MS/MS of sporopollenin. Dr. David Karr measured the XPS of sporopollenin. Dr. Grogory L. Fisher measured the SIMS-TOF-MS/MS of sporopollenin. Dr. Alberto Diego-Taboada and Dr. Grahame Mackenzie provided us the sporopollenin sample. Dr. Paris Georghiou calculated the DFT sporopollenin model. The manuscript draft was edited by Dr. Paris Georghiou, Dr. Travis Fridgen, and Dr. Joseph Banoub.

Chapter 3. Mikhael, A.; Jurcic, K.; Fridgen, T.D.; Delmas, M.; Banoub, J.: **Matrix-assisted laser desorption/ionization time-of-flight/time-of-flight tandem mass spectrometry (negative ion mode) of French Oak lignin: A novel series of lignin and tricin derivatives attached to carbohydrate and shikimic acid moieties**. *Rapid Commun. Mass Spectrom*. 2020; 34: e8841. https://doi.org/10.1002/rcm.8841. Abanoub

Mikhael performed all the data interpretation and manuscript preparation. Dr. Kristina Jurcic measured the MALDI-TOF-MS/MS. Professor Michel Delmas provided us with the Oak VRL sample and its NMR data. The manuscript was edited by Dr. Travis Fridgen and Dr. Joseph Banou**b.**

Chapter 4. Mikhael, A.; Fridgen, T.D.; Delmas, M.; Banoub, J.: **Top-down lignomics analysis of the French pine lignin by atmospheric pressure photoionization quadrupole time-of-flight tandem mass spectrometry: Identification of a novel series of lignin–carbohydrate complexes**. *Rapid Commun. Mass Spectrom*. 2020; 34: e8910. https://doi.org/10.1002/rcm.8910. Abanoub Mikhael recorded the API-MS/MS data and performed all the data interpretation and manuscript preparation. Professor Michel Delmas provided us with the Pine VRL sample and its NMR data. The manuscript was edited by Dr. Travis Fridgen and Dr. Joe Banoub.

.

Chapter 5. Mikhael, A.; Fridgen, T.D.; Delmas, M.; Banoub, J.: **Top-down lignomics analysis of the French Oak lignin by atmospheric pressure photoionization and electrospray ionization quadrupole time-of-flight tandem mass spectrometry: Identification of a novel series of lignans**. *J Mass Spectrom*. 2020, e4676 https://doi.org/10.1002/jms.4676. Abanoub Mikhael performed the APPI- and ESI-MS/MS measurements, data interpretation, and manuscript preparation. Professor Michel Delmas provided us the Oak VRL sample. The manuscript was edited by Dr. Travis Fridgen and Dr. Joseph Banoub.

1. Introduction

Biopolymers are organic polymers created by plants, animals, bacteria, viruses, and fungi. The structural determination and sequencing of well-known biopolymers such as proteins, polysaccharides, lipids, and nucleic acids, usually are considered straightforward. In contrast, the structural determination and sequencing of the very intricate biopolymers, such as sporopollenin and lignin, is discussed in this thesis. It is a very complicated subject and still a matter of much debate.

Sporopollenin is a very complex biopolymer that forms the outer wall of pollen grains and functions to protect its enclosed genetic material against rain, wind, etc. There still much effort that needs to put into the elucidation of the precise structure of sporopollenin. The determination of the structure of sporopollenin has been hindered due to its extreme stability towards chemical degradation. Nevertheless, the lack of success in sporopollenin's structural characterization over the past decades has not stopped the researcher's drive to focus on its applications in different fields, especially in drug delivery. Recently, progress in developing new soft ionization mass spectrometry methods was established to address the bioanalytical challenge in analyzing these complex biopolymers. In the first part of this thesis, MS-based methods have been used to analyze sporopollenin, which helped to demystify and unravel the mystery of the complex chemical structure of sporopollenin.

In the last century, outdated gas chromatography-mass spectrometry (GC-MS) analyses in conjunction with pyrolysis were the methods of choice to investigate the

structure of sporopollenin. However, these structural studies only offered minimal information about the sporopollenin constituents.

In this thesis, the secondary ion mass spectrometry SIMS and matrix-assisted laser desorption mass spectrometry MALDI-MS supported by tandem mass spectrometry MS/MS provided in-depth details about the two major sporopollenin building blocks for the first time in literature.

The second type of biopolymer analyzed in this thesis deals with the structural investigation of lignin, the second most abundant natural biopolymer on earth. Once again, there is still much debate if lignin is a biopolymer which stand-alone with high molecular weight or is composed of short oligomers that are covalently linked in a crisscross fashion to the cellulose fibers and hemicelluloses portions of the wood.

Lignin was generally known to be constructed from three simple monomers (simple monolignols): *p*-coumaryl (H), coniferyl (G), sinapyl (S) alcohols. In the past decade, the structural complexity of lignin, was demonstrated through the identification of other monomers and moieties that were identified and incorporated in the lignin network, such as caffeyl alcohol (C), 5- hydroxyguaiacyl alcohol (F), tricin (flavonoids), and lignin-carbohydrate complexes. Also, the lignin complexity arises from the different types of lignin linkages between its monomers.

As in the case of sporopollenin, mass spectrometry using GC-MS and pyrolysis-GC-MS were used for the structural characterization of lignin over the past decades. In this thesis, we have used different soft ionization techniques such as matrix-assisted laser desorption (MALDI), atmospheric pressure photoionization (APPI), and electrospray

ionization (ESI) to investigate and sequence the structural features of the French oak and pine lignin.

In this chapter, the various research subjects entertained in this Ph.D. thesis are discussed. It is divided into three subsections describing the studied biomolecules, general basics of mass spectrometry, and its use in analyzing and sequencing sporopollenin and lignin.

1.1. Biopolymers

1.1.1. Sporopollenin

Sporopollenin is a natural, highly cross-linked polymer composed of carbon, hydrogen, and oxygen, forming the outer wall of pollen grains.[1] Sporopollenin is highly resilient to chemical degradation.[1] Because of this stability, its exact chemical structure and the biochemical pathways involved in its synthesis are not yet clear.[1]

In the last century and recently, several attempts were completed to determine the structure of sporopollenin. The work on sporopollenin began in 1931 when Zetzsche *et al.* calculated the empirical formula of *Lycopodium clavatum* sporopollenin ($C_{90}H_{142}O_{27}$).[2] Sporopollenin was suggested to be composed of varied and complex straight and branched aliphatic chains, some of which are saturated, unsaturated, and polyhydroxylated.[3]

Other suggested building blocks appeared to involve oxygenated aromatic rings and phenylpropanoid moieties.[1] The cross-linking of these straight- and branched- aliphatic chains was tentatively described as either ether cross-linking and/or carbon-carbon bonds.[3,4]

One of the most important trials to investigate the structure of sporopollenin was done by Hayatsu *et al*.[5], they found that sporopollenin was composed of a highly cross-linked network of lipids (fatty acids and/or fatty alcohols). It should be noted that Hayatsu *et al*.[5] did not propose any molecular structure for sporopollenin, and their conclusion contradicts the well-known proposal of Brooks and Shaw [6] that specified that sporopollenin was formed through the oxidative polymerization of carotenoids and/or carotenoids esters.

The most recent trial to study sporopollenin structure was done by Li *et al.* [7] by using Nuclear Magnetic Resonance (NMR) spectroscopy. This study was performed on the products resulting from the thio-acidolysis treatment of ball-milled pitch pine *Pinus rigida* sporopollenin.[7] The results showed that sporopollenin is composed mainly of aliphatic hydroxylated polyketides (alpha-pyrone ring plus hydroxylated aliphatic chain), aromatic moieties, and acetal groups as a cross-linker.[7]

Based on the data that will be presented in this thesis (Chapter 2) and the survey of the literature, we have several arguments that can be made regarding the recent structural proposal for the pitch pine *Pinus rigida* sporopollenin. [7] First, the use of high-energy ball milling of the whole pollen grain of pitch pine *Pinus rigida* could produce heat, leading to the production of aromatic moieties and definitely altering the structure of sporopollenin.[8-12] These modifications are expected to occur by ball milling as it is a well-known technique used in the synthesis and reactions of organic compounds.[8-12] Second, it is enigmatic to believe that ball milling can break spherical sporopollenin, knowing that sporopollenin can be compressed under tons of pressure then return to its original shape.[13] Third, **we** disagree

with their assignment concerning the untreated sporopollenin's carbon NMR peak at 101 ppm as an acetal group. This resonance could well be assigned to C-5 of the α-pyrone rings in the sporopollenin structure,[14] and/or the vinylic carbons in the beta diketone moieties that will be discussed later in this thesis. Fifth, the suggestion of the presence of acetal groups in the sporopollenin structure is completely unacceptable as these groups will hydrolyze readily under mildly acidic conditions.[15] This does not support the known fact of extreme chemical stability of sporopollenin under harsh conditions such as the one used to extract the studied sample herein of *Lycopodium clavatum* sporopollenin. Moreover, suppose there were acetals in sporopollenin as a cross-linker between its chains. In that case, they should have been hydrolyzed after seven days of reflux in the presence of phosphoric acid, as indicated in our experimental section (Chapter2). Also, removing these acetal crosslinkers should degrade sporopollenin to a large extent and change its properties completely, which does not happen during the extraction procedure.

1.1.1.1. Sporopollenin Extraction

In general, pollen grains (Figure 1.1) contain cytoplasmic materials (fats, genetic material, or proteins), which are enclosed by two layers and/or walls: the intine (cellulose) and the tough exine (sporopollenin). [1,16]

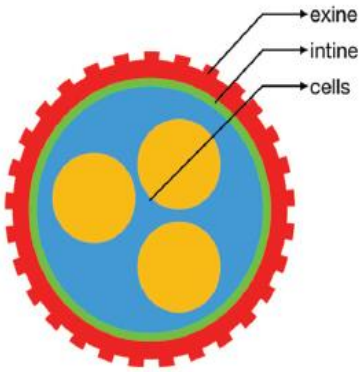


Figure 1.1. Schematic structure of a pollen grain. Reprinted with permission from *New J. Chem*. **2020**, *44*(3), 647-652

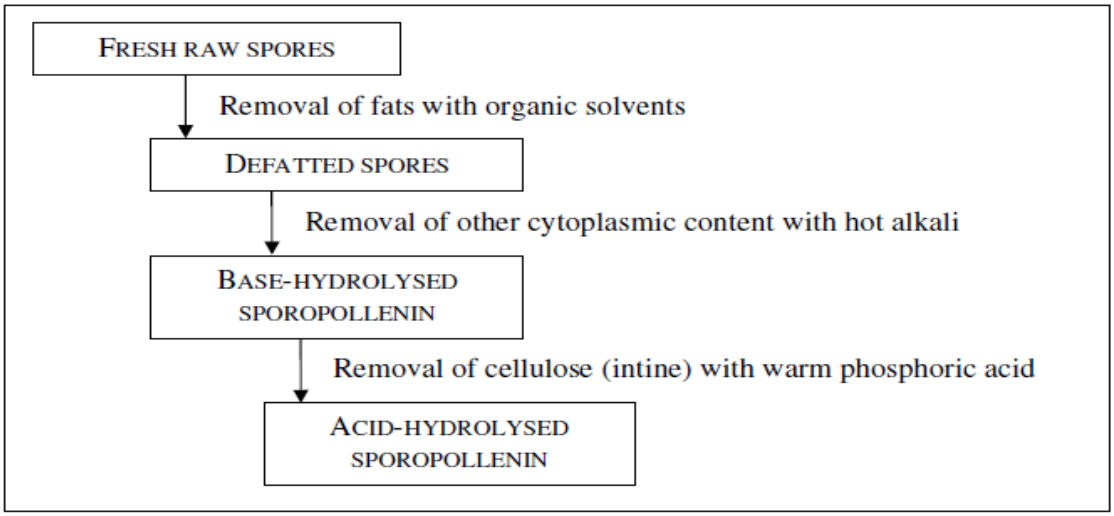


Figure 1.2. Extraction protocol for obtaining the clean, intact, unmodified *lycopodium clavatum* sporopollenin. Adapted with permission from *Physical and chemical properties of sporopollenin exine particles* (Doctoral dissertation). University of Hull, **2008**

In this thesis, we have studied the structure of *lycopodium clavatum* sporopollenin. The most preferred method used in the literature to extract the clean and unmodified *lycopodium clavatum* sporopollenin has been developed by Zetzsche *et al*. (Figures 1 and 2) [2, 17-19]This extraction method consisted of removing the cytoplasmic materials (fats, genetic material, or proteins) from pollen grains, in order to produce the intact pollen or spore wall. [2, 17-19] In this method, the removal of the cytoplasmic content was

accomplished by the treatment of pollen grains with organic solvents such as acetone plus hot alkaline solution (typically sodium or potassium hydroxide).[2, 17-19] The resulting spore wall is hollow and preserves the same morphology and/or shape of the original pollen grains. It is composed of at least the intine and the exine, representing 25% of the mass of the original pollen grain. [2, 17-19] Lastly, the resulting spore wall is devoid of any nitrogen supporting the complete removal of any proteins. [2, 17-19]

According to Zetzsche *et al*. [2, 17-19], the intine (cellulose) could be eliminated by acid hydrolysis. These acid treatments were done over several days, such as the treatment with 40% hydrochloric acid or hot 72% sulphuric acid, or hot 85% phosphoric acid. [2, 17-19]

According to Shaw *et al.*, [20], the best choice to obtain clean and cellulose-free sporopollenin is by treatment with concentrated phosphoric acid. On the other hand, the use of hydrochloric acid or sulphuric acid for cellulose removal is not recommended as it can introduce impurities in the resulting exine, such as chlorine and sulfur. [21]

It should be noted that only sulphuric acid was able to remove all the traces of the intine in contrast to phosphoric acid, which leaves traces of cellulose in the resulting exine.[22] However, concentrated phosphoric acid was selected by many researchers to remove the intine because it balances between the clearance of cellulose and the introduction of impurities like sulfur in the exines. [20, 23-26]

1.1.1.2. Sporopollenin Chemical Degradation

Due to the extreme stability and the insolubility of sporopollenin in suitable solvents, the investigation of the sporopollenin structure was a difficult task.[21] For this reason, several harsh degradation methods were used aiming to get any information about the sporopollenin chemical composition and/or structure. Some of these harsh degradations are reflux in chromic acid (Kuhn-Roth oxidation), ozonolysis, potash fusion, nitrobenzene oxidation, and treatment with aluminum iodide. [21]

All these previous degradation methods result in a mixture of small and modified monomers, which cannot provide any information about the sporopollenin native and/or original structure. [21] Additionally, this harsh degradation method can produce compounds that initially do not exist in the sporopollenin structure. For example, Gordon Shaw et al. [20] showed that aromatic compounds resulting from the potash fusion treatment of sporopollenin are artefacts that are formed due to the harsh treatment conditions. It should be noted that there is no mild degradation method that can degrade sporopollenin to fragments that can represent the sporopollenin native structure.

1.1.1.3. Sporopollenin Applications

In the last century, there have been many contradictions about the structural components of sporopollenin, but this did not disenchant many researchers who focused their efforts on the biomedical applications of sporopollenin as an excellent drug delivery system.

Scanning electron microscopy (SEM) of the empty sporopollenin displayed the porosity of the spore surface (Figure 1.3), which makes it an excellent shell for loading

materials such as drugs (Figure 1.4). [27,28] For example, sporopollenin capsules were used to deliver several drugs such as the antibiotic erythromycin, the antibiotic bacitracin, the anticancer drug oxaliplatin, ibuprofen, and paracetamol. [29-32]

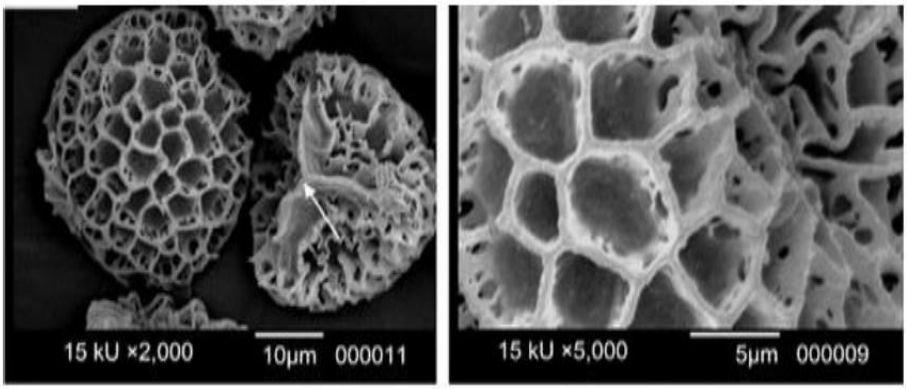


Figure 1.3. Scanning electron microscopy image of empty sporopollenin exines. Reprinted with permission from *J. Encapsul. Adsorp. Sci*. **2016**, *6*(04), 109-131

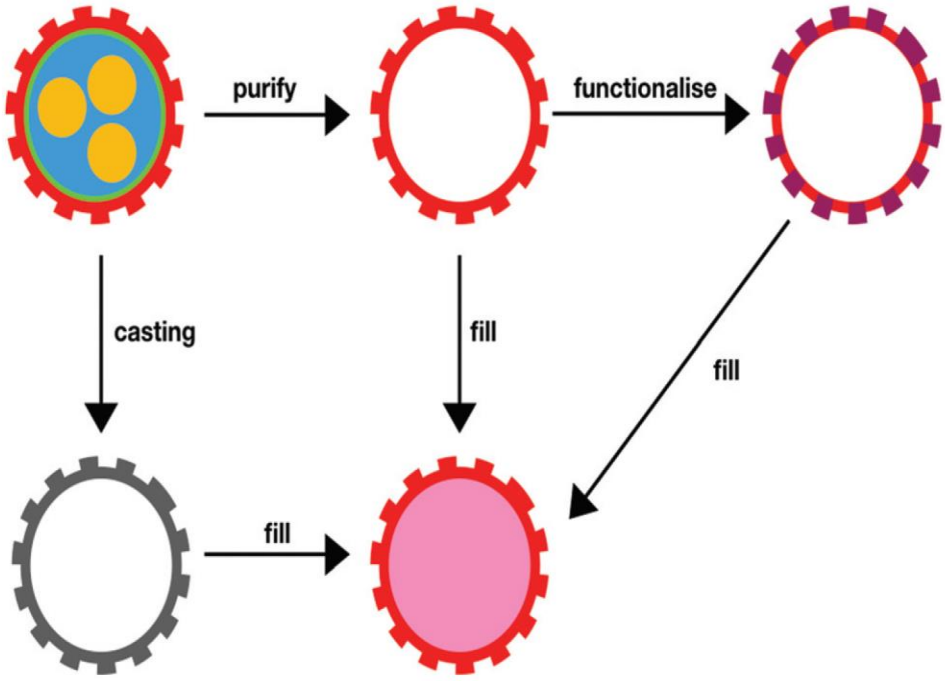


Figure 1.4. Types of sporopollenin modification. Reprinted with permission from *New J. Chem*. **2020**, *44*(3), 647-652

It goes without saying, and it is obvious that the detailed structural study of sporopollenin will help to understand better the current sporopollenin applications such as drug delivery, peptide synthesis, microreactors, and wastewater purification. [1,27,28] As well, it may lead to the discovery of new applications in the next upcoming years.

It should be remarked that the surface of sporopollenin can be cast by different inorganic materials to produce inorganic microstructures with the same toughness as the sporopollenin wall (Figure 1.4).[16] For example, sporopollenin casting can produce porous silica, calcium carbonate, and calcium phosphate sporopollenin microparticles that can be loaded with the anti-inflammatory drug ibuprofen.[33]

Also, since sporopollenin is known to contain hydroxyl groups on its outer surface, it can be functionalized to alter its surface properties, allowing it to be used for more specific applications (Figure 1.4).[16] For example, the surface of sporopollenin can be functionalized by different groups that can act as a ligand for metal removal from an aqueous environment, especially industrial wastewater. [34,35]

1.1.2. Lignin

The wood cell wall is composed mainly of cellulose, hemicellulose, and lignin (Figure 1.5).[36] Cellulose is a polymeric D-glucopyranose and the most abundant biopolymer in nature.[37] The glycan chains of cellulose (about 40 chains) are held together through hydrogen bonds to form a cellulose microfibril.[38]

The cellulose content in the two major types of wood, softwood, and hardwood, is roughly the same percentage of 45%;[39] however, hardwood has a higher percentage of hemicellulose (~25%) with respect to softwood (~17%).[39]

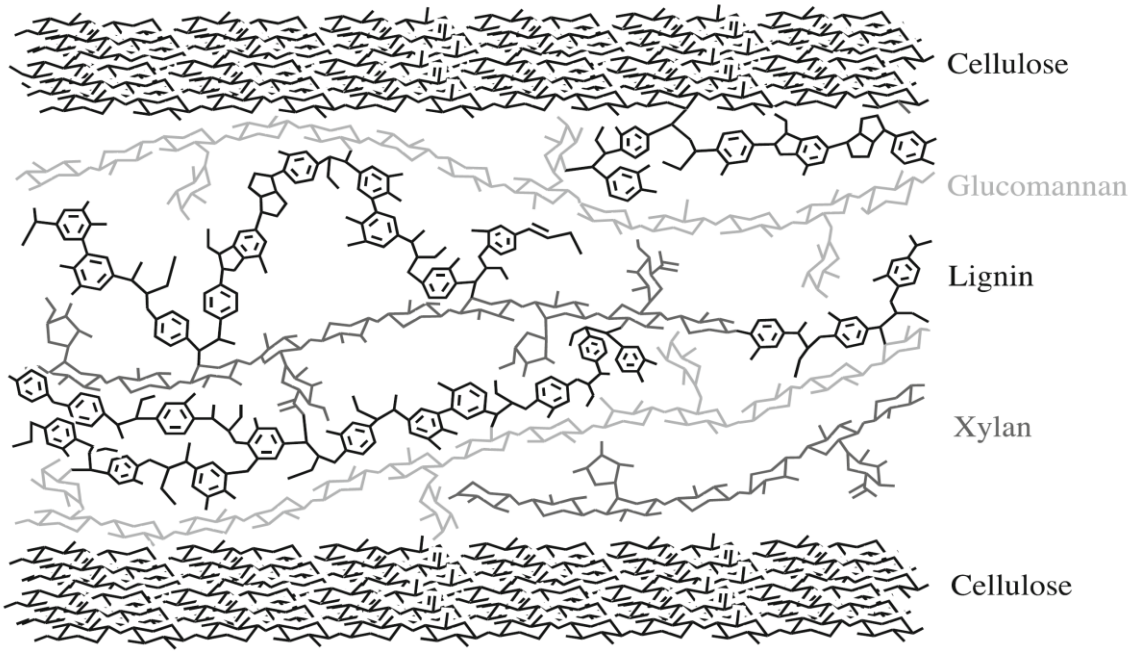


Figure 1.5. A schematically lignin polysaccharide network in wood. Reprinted with permission from *Wood Sci. Technol.* **2017**, *51*(6), 1365-1376

The hemicellulose polysaccharides (xylans, mannans, galactans) and the aromatic lignins act like cross-linkers between the cellulose fibers, which accounts for wood rigidity.[36,38] Also, lignin hydrophobicity act as a barrier to biological attacks by various aqueous enzymatic systems and, it prevents water loss from plant cell walls during water transport.[40] It should be noted that softwood has a higher percentage of lignin (~28%) with respect to hardwood (~20%).[39]

Lignin is the second most abundant biopolymer in nature; [41,42] it is the product of polymerization of three monomeric aromatic compounds (monolignols): *p*-coumaryl (H-unit), coniferyl (G-unit), sinapyl (S-unit) alcohols (Figure 1.6). [41-43] Coniferyl alcohol (G-unit) is considered the major component in softwoods, while both coniferyl and sinapyl alcohols (G and S-units) are major components in hardwoods.[39] Lastly, *p*-coumaryl alcohol (H-unit) is abundant in the grass.[39] It should be noted that some lignin units are not commonly incorporated in lignin oligomers such as caffeyl (C), 5- hydroxyguaiacyl (F), and the hypothetically proposed gallyl unit (L).[44]

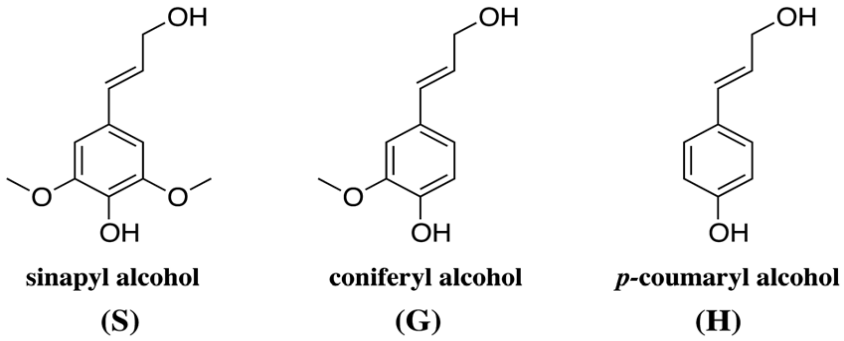


Figure 1.6. Major lignin monomers. Reprinted with permission from *Rsc Adv.* **2014**, *4*(48), 25310-25318

Lignin monomers are connected randomly through different types of linkages such as *β*-O-4, *β*-5, *β*-1, 5-5, and *β–β* linkages. The most important type of lignin monomers linkages is (β-O-4) or (8-O-4‘), representing 40-60 % of all kinds of lignin linkages. [39,40] Different types of lignin linkages are shown in Figure 1.7.

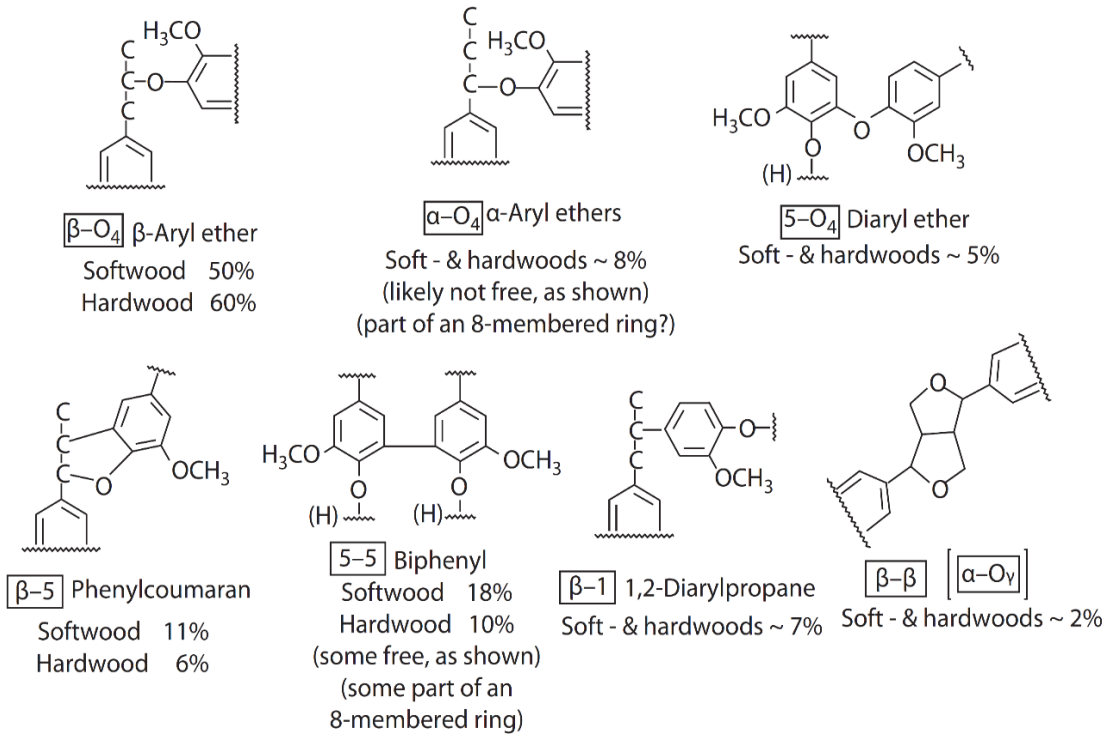
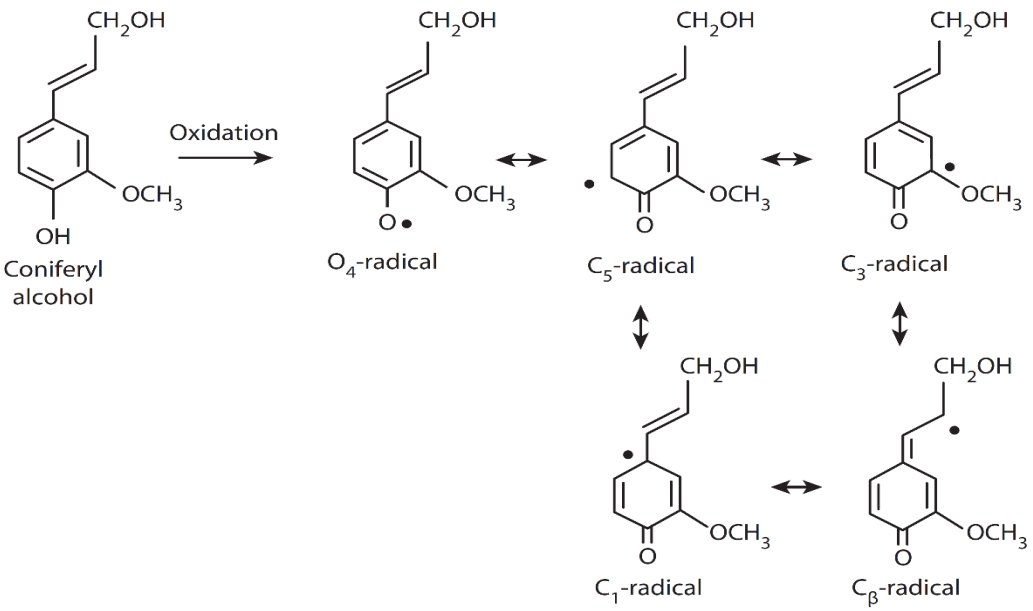
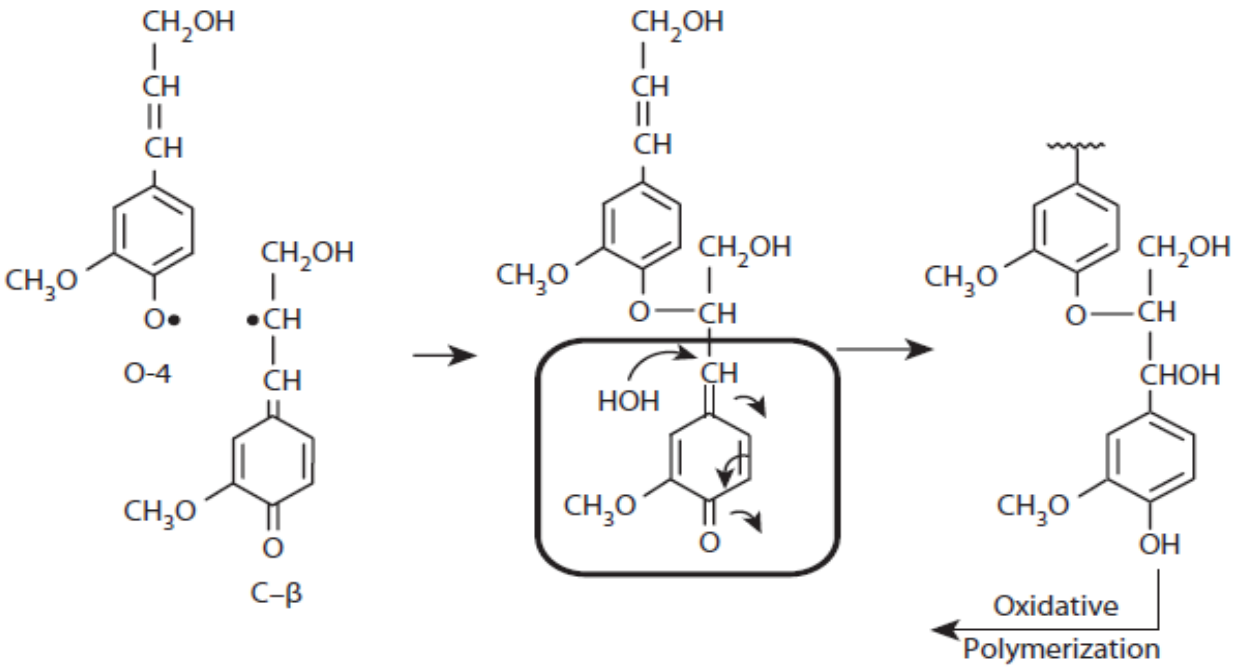


Figure 1.7. Types of lignin linkages. Reprinted with permission from Heitner, C.; Dimmel, D.; Schmidt, J.: *Lignin and lignans: advances in chemistry*, CRC press, **2016**

The various combinations of lignin monomers occur via an enzymatic oxidative radical coupling mechanism (Scheme 1.1).[39] The monolignol monomers lose hydrogen radical from their phenolic OH groups to form a phenoxy radical stabilized by resonance.[39] This resulting monolignol free radicals can then undergo radical coupling reactions to produce dilignols.[39] The formed dilignols can be coupled with more monolignols using the same polymerization mechanism to form higher lignin oligomers.[39] Due to the reactivity of phenoxide radicals and the randomization of the polymerization process, lignin lacks regularity and ordered repeating units known for other polymers such as polysaccharides and proteins.[38]



Scheme 1.1a. First step in lignin polymerization. Reprinted with permission from Heitner, C.; Dimmel, D.; Schmidt, J.: *Lignin and lignans: advances in chemistry*, CRC press, **201**

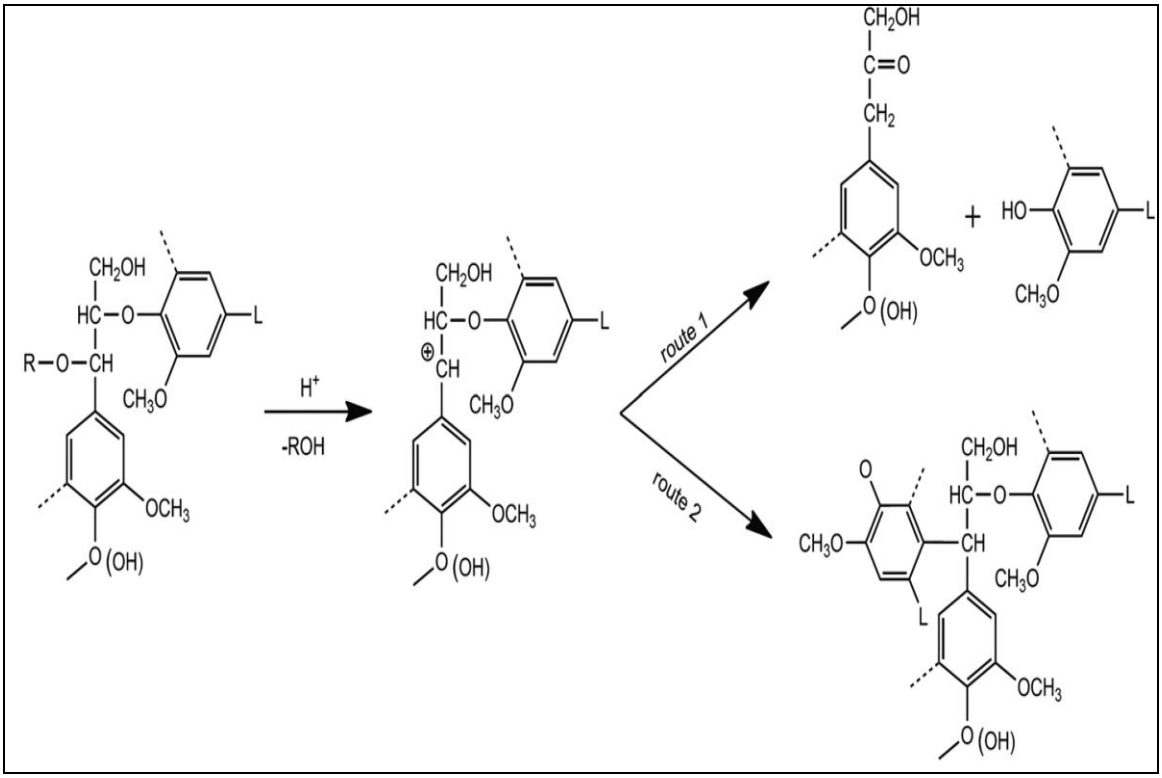


Scheme 1.1b. Formation of the most abundant lignin linkage Cβ–O4 bond formation via radical coupling. Reprinted with permission from Heitner, C.; Dimmel, D.; Schmidt, J.: *Lignin and lignans: advances in chemistry*, CRC press, **2016**

For about a century, the structure of lignin was defined as a stand-alone complex polymer composed of irregular branched units. [42,45] However, lignin is bonded with celluloses and hemicelluloses, forming the complex wood network. [32] This is supported by the association of lignin with carbohydrates forming a class of oligomers called lignin-carbohydrate complexes. [32]

1.1.2.1.Lignin Extraction

It has been established that various methods for releasing and purifying lignin from wood cellulose lead to a drastic change in lignin chemical structure.[45, 46-48] It should be noted that most of the lignin structural elucidations were completed by depolymerizing the isolated presumed lignin polymer to produce simple lignin moieties, which are easier to be studied than the native complicated lignin structure.[49] For example, lignin ester and ether bond cleavages can be achieved through acidic depolymerization; however, due to some of the resulting products' reactivity, unexpected complex products can be formed (Scheme 1.2), which tend to complicate matters further.[49]



Scheme 1.2. Unexpected complex products formed during acidic depolymerization of lignin. Reprinted with permission from *Bioresour. Technol.* **2007**, *98*(16), 3061- 3068

It is important to mention that when the lignocellulosic biomass is deconstructed into its three main components (cellulose/hemicellulose/lignin), the lignin portion is recovered from the black liquor. Usually, the resulting lignin is then subjected to several useless operations such as depolymerization, purification, and transformation.

Banoub *et al.* [38,50,51] had introduced a new paradigm that stated that originally there was no lignin existing as a huge polymer. Instead, they proposed that lignin is composed of different length oligomers that are connected to cellulose and hemicellulose in a criss-cross pattern forming the network of wood.[38] It also states that accurate structural studies should be done on the series of constituting lignin oligomers released from the glycolignin

(polysaccharide-lignin) without any further purification or chemical transformations. This type of released lignin oligomer is called virgin released lignin (VRL). [38]

The French company "La Compagnie Industrielle de la Matière Végétale" (CIMV) extracts VRL lignin from wood by using a mixture of simple organic solvents (acetic acid-formic acid-water). So, it is considered one of the most excellent extraction techniques that produce VRL lignin.[52] Some lignin extraction methods such as the Kraft, the Sulfite, the Alkaline, and the Klason process aim to obtain undegraded cellulose, but on the other hand, the resulting lignin present in the black liquor is then submitted to more chemical treatments, which without any doubts will alter and modify their original structures.[53]

For example, the Kraft and the Sulfite processes lead to the incorporation of sulfur in the lignin oligomers structure.[53] Also, the acidic (Klason process) or alkaline (NaOH) depolymerization leads to significant damage to the native lignin structure.[53] Other extraction methods such as the Bjorkman process, enzyme hydrolysis, organic solvent extraction, and ionic liquid treatment are known to produce slightly modified lignin.[53] The summary of different lignin extraction techniques is shown in Figure 1.8.

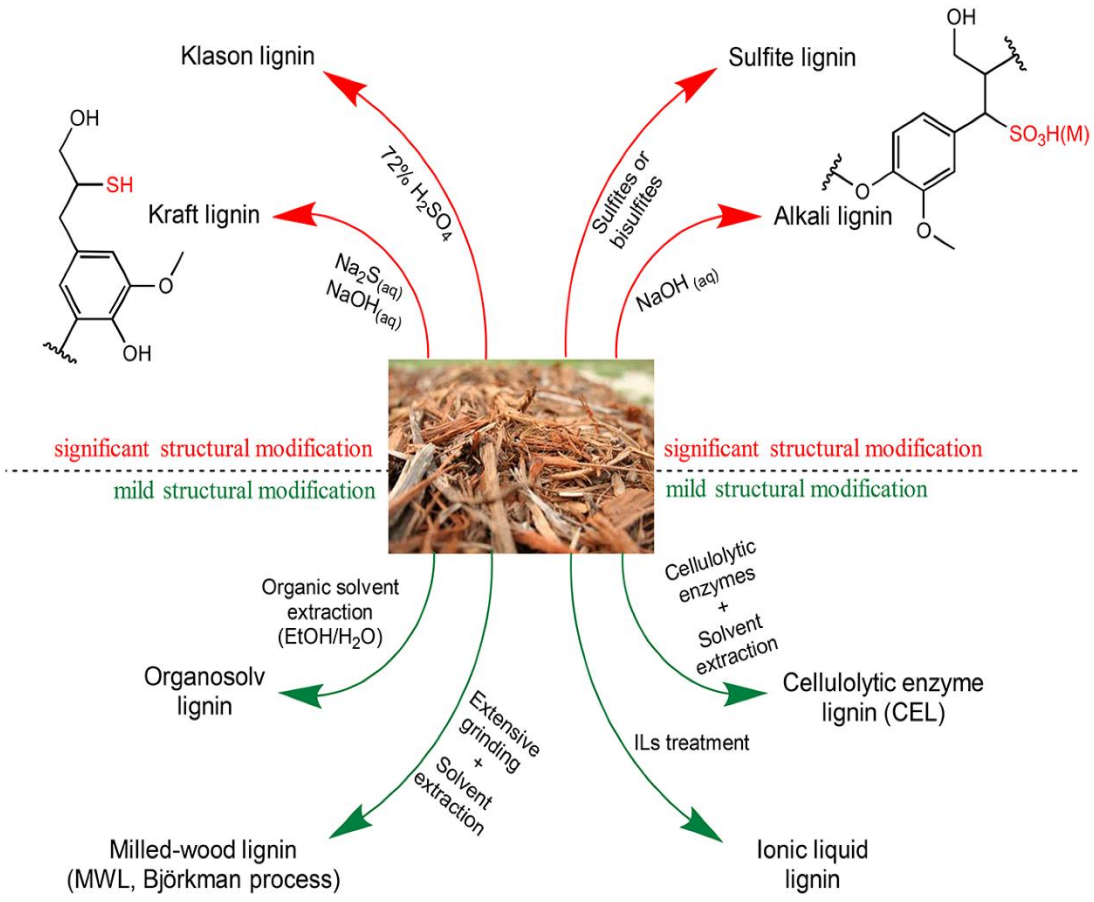


Figure 1.8. Different methods of lignin extraction. Reprinted with permission from *Chem. Rev.* **2018**, *118*(2), 614-678

1.1.2.2. Lignin Applications

The lignin component of the lignocellulose biomass is considered the most abundant renewable source composed of aromatic units in nature.[54] Over the past few decades, research on lignin valorization has grown rapidly on account of the importance of lignin in biorefineries.[55] Lignin has been used either directly or chemically modified in the production of bio-dispersant, wood panel products, emulsifiers, polyurethane foams, and automotive brakes.[55] In addition, it has been suggested that lignin depolymerization

in order to produce more aromatic compounds is an important way for the utilization of lignin.[54,56] The supposed depolymerization can be accomplished in different ways, such as acid/base-catalyzed depolymerization/hydrolysis, pyrolysis, hydrotreating, oxidation, liquid-phase reforming, gasification, and biodegradation.[55]

Different lignin conversion processes are summarized in Figure 1.9.[55] The depolymerization processes can be carried out in an oxidizing environment (in the presence of oxidants such as O_2, H_2O_2, peracetic acid, or air), a reducing environment (in the presence of reductant such as H_2), or a neutral environment. [55]

Lignin hydroprocessing (H_2 /100-350 °C) produces simple aromatic compounds such as phenols, benzene, toluene, xylene, and alkane fuels. [55]

Oxidation of lignin in a temperature range of 0−250 °C favors the production of aromatic alcohols, aldehydes, and acids.[55] Acid or base-catalyzed depolymerization reactions break the C−O or C−C linkages between lignin units to create small fragments such as phenolic monomers. [55]

Liquid-phase reforming (typically at 250−400 °C) has been used to produce hydrogen and light gases from the lignin. [55] Lastly, the gasification method produces synthesis gas (CO and H_2) from different lignin feedstocks and model compounds.[55] It should be noted that biocatalysis (using enzymes such as manganese peroxidase, lignin peroxidase, and laccase) has also been used as a process of lignin degradation in an environmentally friendly way. [55,57]

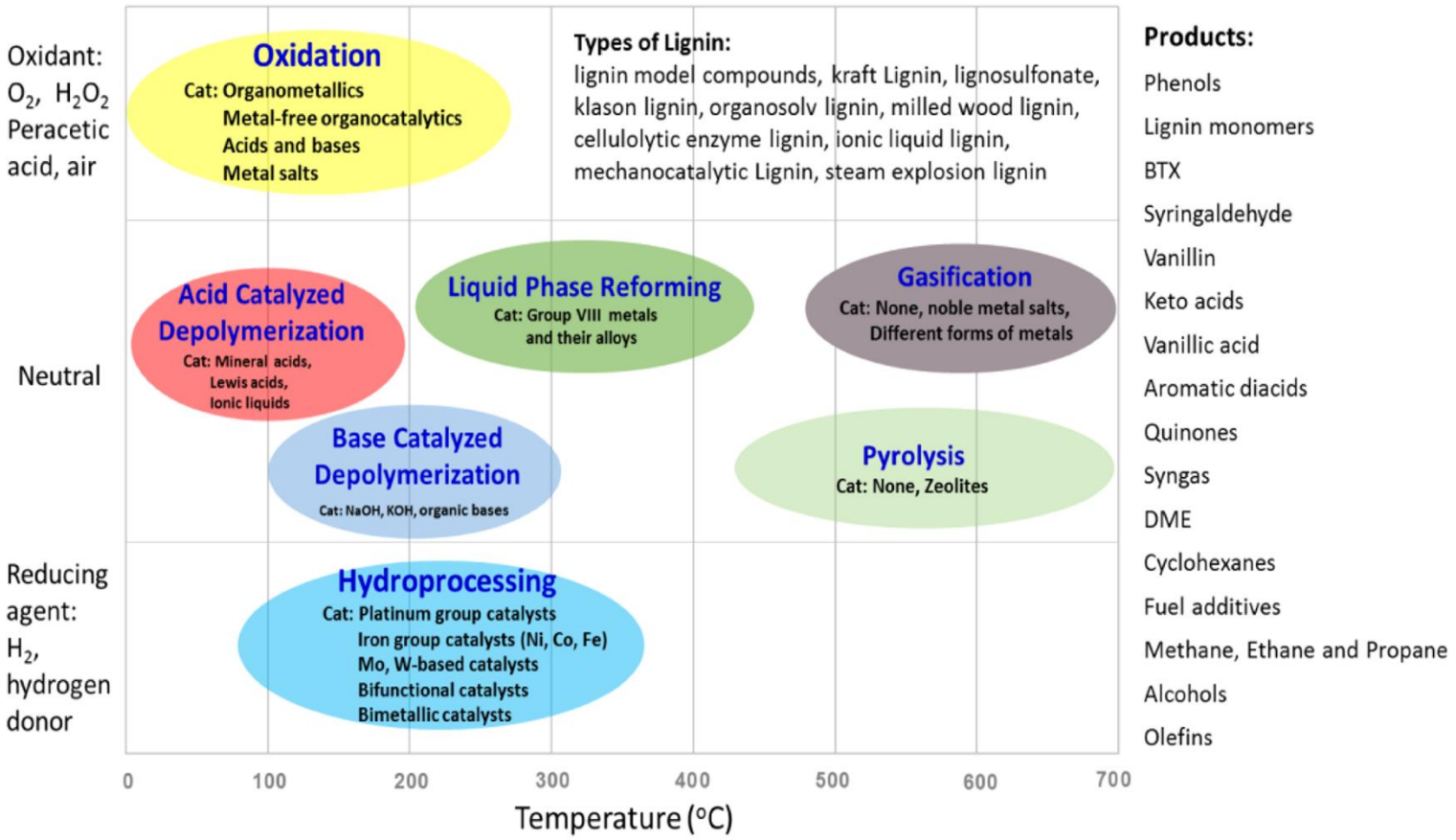


Figure 1.9. Summary of processes for conversion of lignin. Reprinted with permission from *Chem. Rev*. **2015**, *115*(21), 11559-11624

1.2. Mass Spectrometry

Mass spectrometry is an analytical technique that allows the determination of the analyte mass by detecting their mass-to-charge *(m/z)* ratio.[58] This technique can be used for the qualitative structural determination of analytes and their quantitation. [59] In the past few decades, the needs of bioanalytical chemistry were the driving force for innovation in mass spectrometry.[60] Mass spectrometry is the best option for the analysis of highly complicated mixtures produced by genomics,[61] proteomics,[62] metabolomics,[63] lipidomics,[64], and lignomics studies.[65] A conventional mass spectrometer is composed of three major sections: an ionization source, a mass analyzer, and a detector, as shown in Figure 1.10.[66]

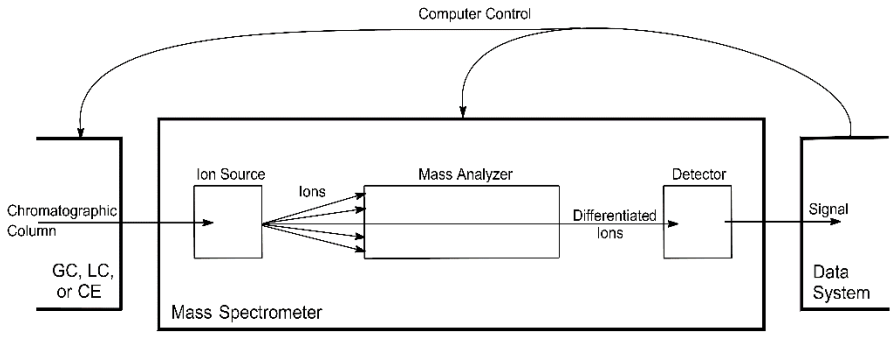


Figure 1.10. Block diagram of a chromatography/MS/computer system. Reprinted with permission from Smith, R. M.: *Understanding mass spectra: a basic approach*. John Wiley & Sons, **2004**

1.2.1. Ion sources

The main role of different ionization sources is to create charged particles of the analytes.[67] Ionization sources are classified as hard and soft ionization sources based on the quantity of internal energy that can be transferred to the molecule during the ionization process.[68] Electron Ionization (El) and Chemical Ionization (CI) are examples of hard

ionization sources that transfer a significant amount of internal energy to the analytes that cause its remarkable fragmentation.[69,70] Electrospray ionization (ESI), Atmospheric Pressure Photoionization (APPI), and Matrix-Assisted laser desorption (MALDI) ionization techniques are considered soft ionization techniques.[71-73] This is because the analyte gains a small amount of internal energy during the ionization process, which decreases the possibility of any fragmentation.[74] In general, soft ionization methods are the most widely used methods, especially for the analysis of biomolecules.[74]

1.2.1.1. Electrospray Ionization (ESI)

The operations in an electrospray ion source can be summarized as the following: a solution containing the analyte of interest and a suitable polar solvent flow slowly through a capillary tube (1-10 μl min^{-1}).[75] The flowing solution experiences a high electric field, which is created through the potential difference (2-5 kV) between the capillary tube and the counter electrode (Figure 1.11).[71] The solution emerging from the capillary tip accumulates excess charges and/or ions of the same polarity as the applied potential at its surface.[76] A cone of liquid called Taylor cone is created at the tip of the capillary tube.[71] This cone results from the balance between the coulombic repulsion forces between the excess surface charges and the surface tension of the liquid.[71] The highest amount of charges that a liquid droplet can accommodate is known as the Rayleigh limit.[77] When the Coulombic repulsion forces between the excess charges exceed the liquid surface tension, the Rayleigh limit is reached, and the cone breaks down to smaller ESI droplets.[71,74] Moreover, an inert gas (usually nitrogen) is introduced coaxially to the

capillary tube to assist in the nebulization of the spray.[74] Also, a curtain of heated inert gas is used to accelerate the drying and the desolvation process of the generated ESI droplets.[74] Finally, the dry charged analyte molecules are speeded to enter the mass analyzer for detection.[74]

A proton transfer reaction between the electrospray solution and the analyte is necessary to create the charged analyte molecules either as protonated (+ve ion mode) or deprotonated molecules (-ve ion mode).[71] However, redox reactions may take place and influence the ionization process.[78]

One of the major advantages of ESI sources is that they can be connected to high-performance liquid chromatography (HPLC) systems because of their ability to cope with flow rates up to 1000 μl min^{-1}.[79]

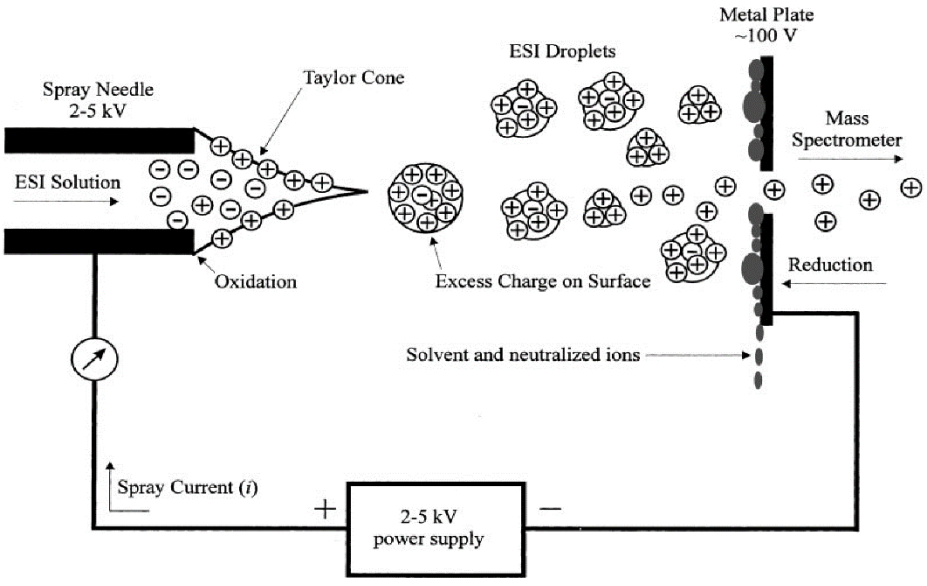


Figure 1.11. Schematic representation of the mechanism of electrospray Ionization. Reprinted with permission from *Mass Spectrom. Rev.* **2001**, *20*(6), 362-387

1.2.1.2. Atmospheric pressure photoionization (APPI)

Atmospheric pressure photoionization was designed for the ionization of very non-polar molecules that cannot be ionized easily by ESI.[80] The photoionization process of the molecules is accomplished by a UV lamp (Figure 1.12).[81] In this case, in contrast to ESI, the ionization process does not depend on the charge affinity of the targeted molecules that it is why it is capable of ionizing very non-polar molecules ($M + h\nu = M^{\bullet+} + {}^{-}e$).[82] In case of the presence of protic solvents, it can help in the formation of protonated molecules ($M^{\bullet+} + S = [M + H]^{+} + [S\text{-}H]^{\bullet}$).[81] It should be noted that the ionization energy of the analyte molecules should be less than the energy of the lamp; otherwise, a dopant like toluene or acetone is necessary to be used.[81] In case that the ionization energy of the analyte is greater than the lamp energy, photodissociation and/or radiative decay occurs instead of the ionization of the target molecules.[83] Therefore, the dopant, in this case, is needed to be ionized first by the lamp, then it can transfer its charge to the targeted molecules (indirect ionization).[83]

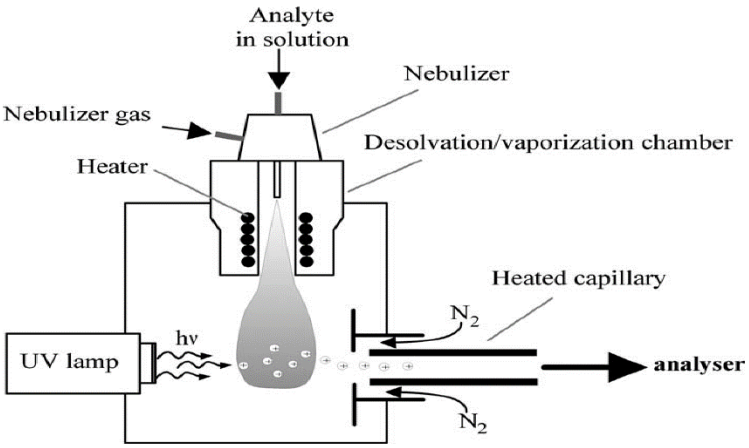


Figure 1.12. Atmospheric Pressure Photoionization (APPI) source. Adapted with permission from De Hoffmann, E.; Stroobant, V.: *Mass Spectrometry: Principles and Applications*, 3rd ed. John Wiley and Sons, Chichester, **2007**

1.2.1.3. Matrix-Assisted Laser Desorption Ionization (MALDI)

One of the most widely used ionization sources is the matrix-assisted laser desorption ionization, which is coupled commonly with a TOF mass analyzer for the analysis of different biomolecules such as oligonucleotides,[84] oligosaccharides,[85] proteins,[86] glycoproteins[87], lipids[88] and lignin.[89]

MALDI has many advantages, which are the high sensitivity to low concentration analytes, detection of high molecular weight biomolecules, and unaffected by salts.[90] Also, the cause of a small amount of fragmentation to large biomolecules allowed the use of MALDI for their direct analysis on tissue slices (MALDI imaging).[90]

The first step in performing the MALDI analysis is to mix the matrix solution with the analyte, followed by spotting the resulting mixture on a MALDI plate.[91] This latter method is the most commonly used spotting method and is called the dried droplet method.[92] However, there are other spotting techniques such as quick and dirty, fast evaporation, sandwich, seed-film layer, two-layer, and spin coating.[93-95] The next step is to leave the spotted mixture on the MALDI plate at room temperature to allow the evaporation of the solvent (drying) and the co-crystallization of the analyte with the matrix.[94] It should be indicated that there are different types of matrices that were used to analyze different classes of biomolecules.[74] For example, 2,5-dihydroxybenzoic acid (DHB) was used for oligosaccharides, peptides, and proteins,[96] sinapic acid was used for proteins, α-cyano-4-hydroxycinnamic acid (α-CHCA) for a mixture of peptides,[97,98] and 1,5-diamino naphthalene (DAN) was used for lipids.[99] The dried spot on the MALDI plate

is then irradiated by a laser beam (usually a nitrogen laser at 337 nm), causing the desorption and the ionization of the matrix and the analyte (Figure 1.13).[81]

The absorption of the laser energy (hv) by the matrix molecules leads to their excitation (MH + hv = [MH]*.[100] The excited matrix molecules transfer a portion of their energy to the analyte.[100] As a result of the latter process, desorption of the analyte and matrix molecules as gas-phase clusters occurs.[100] The collision of the matrix molecules with the analyte molecule in the formed gas-phase clusters leads to the ionization of the analytes through a proton transfer reaction.[101]

Lastly, It should be noted that there are MALDI sources that are able to operate under atmospheric pressure.[102] Also, laser-induced acoustic desorption (LIAD) and desorption/ionization on silicon (AP-DIOS) were developed as atmospheric pressure sources and without the need to use a matrix.[103,104] Also, liquid chromatography was successfully coupled with MALDI ionization sources, which is an efficient tool in proteome analysis.[105]

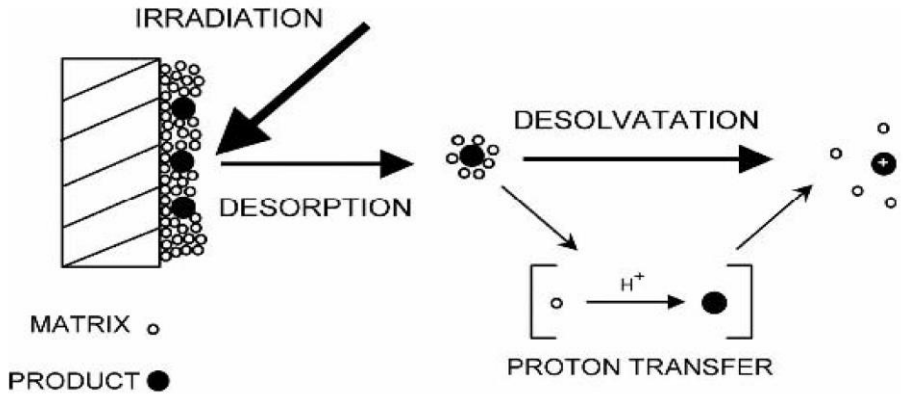


Figure 1.13. Diagram of the principle of MALDI. Reprinted with permission from De Hoffmann, E.; Stroobant, V.: *Mass Spectrometry: Principles and Applications*, 3rd ed. John Wiley and Sons, Chichester, **2007**

1.2.1.4. Secondary Ion Mass Spectrometry (SIMS)

SIMS is a mass spectrometric technique in which the solid surface of interest is hit by an energetic, focused ion beam (~0.1 and 50 keV).[106] This energetic beam of ions is called the primary ion beam. Examples of these beams are O^-, O_2^+, Cs^+, Ga^+, In^+, Au^{n+}, Bi^{n+}, SF_5^+, C_{60}^+, C_{84}^+, $C_{24}H_{12}^+$, Ne^+, Ar^{n+}, Xe^+, etc. [106] This primary ion beam leads to the *sputtering* of secondary ions (analyte) from the irradiated area, which is then directed to a mass spectrometer for analysis (Figure 1.14).[106] The total ion mass spectra represent the chemical composition of the solid surface.[107] Also, the created total ion image is considered as the topography of the studied material.[107] Moreover, the image of a specific ion showed its spatial distribution on the characterized surface.[107]

SIMS is considered a solid-state mass spectrometry, which has many applications in different fields such as material science, earth science, and bioscience.[106] For example, SIMS has been used to investigate heterogeneous surfaces such as the lignocellulosic biomass.[108] In this case, the spatial image showed the distribution of cellulose, hemicellulose, and lignin in the biomass.[108]

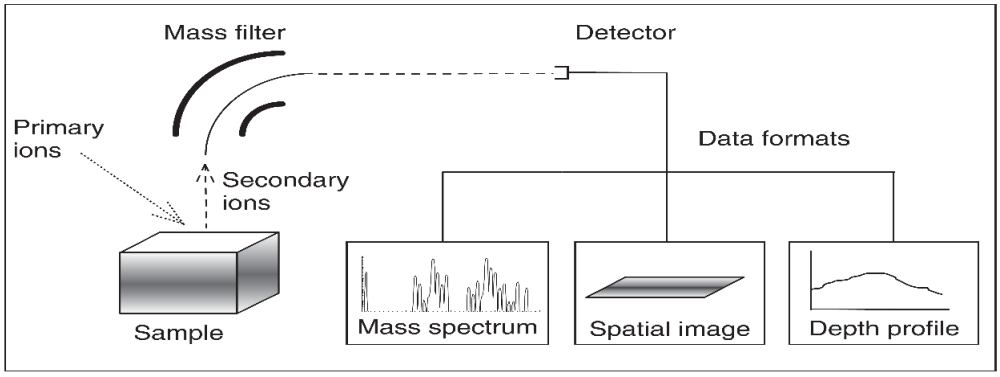


Figure 1.14. Pictorial illustration of the main components found in a SIMS instrument. Reprinted with permission from Van der Heide, P.: *Secondary ion mass spectrometry: an introduction to principles and practices*. John Wiley & Sons, **2014**

1.2.2. Mass Analyzers

The main task of different mass analyzers is to separate different ions according to their mass to charge (*m/z*) values.[81] Two major types of mass analyzers were developed over the past decades. [74] The first type is the scanning analyzers, which scan ions within a specific mass range, such as the quadrupole and the time-of-flight (TOF) mass analyzers.[74] The second type is the trapping mass analyzers in which all ions can be trapped simultaneously, such as the quadrupole ion trap, the orbitrap, and the Fourier transform ion cyclotron resonance (FTICR) mass analyzers. [74] In this type of mass analyzer, a specific ion can be selected for performing more experiments such as collision Induced dissociation CID (fragmentation of the ion through the collision with a neutral gas). [74]

Furthermore, hybrid mass spectrometers were also developed by combining two or more mass analyzers in one instrument, such as the quadrupole-time-of-flight (QTOF) mass spectrometer, which is extensively used in proteomics. [74] The resolution, sensitivity, mass range, and mass accuracy are the main characteristics that differentiate one mass analyzer from others. [74]

1.2.2.1. Quadrupole Mass Filter

The linear quadrupole mass filter is one of the most used mass analyzers during the past decades.[109] Separation techniques such as liquid chromatography (LC),[110] gas chromatography (GC),[111] and capillary electrophoresis (CE) are frequently coupled with

the quadrupole mass analyzer.[112] However, it is considered a low-resolution mass analyzer that provides at least unit resolution for small molecules analysis (<2000 Da).[113]

The quadrupole is composed of four parallel metal rods (Figure 1.15A).[113] A direct current (DC) and a radiofrequency (RF) potentials are applied to the four rods to create two positively charged rods (on the x-direction) and two negatively charged rods (on the y-direction), as shown in Figure 1.15B. [113] By varying the values of the DC and the RF potentials, only one ion of a certain *m/z* value (M_1) will have a stable ion trajectory in the quadrupole and will be allowed to reach the detector, while the ions with a different *m/z* value (M_2, M_3) will have unstable ion trajectory and get neutralized through the collision with the charged rods (Figure 1.15A).[113] The applied electric field on the rods is calculated by the following formula: $\Phi_o = \pm (U \pm V\cos \omega t)$, where Φ_o represents the voltage applied to the rods, U is the DC voltage, V is the RF voltage amplitude, and ω is the frequency.[113] To describe briefly how the quadrupole mass filter separates ions, we will consider that there are positively charged ions at the intersection point of the X-axis and the Y-axis. [114] For example, if we consider that the two positively charged rods on the x-axis are experiencing DC voltage only, the positive ions will be focused on the center away from these rods due to the repulsion forces. [114] If the AC current is applied to these rods, the positive RF cycle will have no effect on the positive ions in the center. [114]

In contrast, the negative RF cycle will make the positive ions move slightly from the center towards the positive rods. [114] In this case, light ions (low *m/z*) will move faster towards the rods with respect to heavy ions (high *m/z*), causing a separation between the ions according to their *m/z* values. [114] Similarly, if we consider that the two negatively

charged rods on the y-axis are experiencing negative DC potential only, all positive ions in the center will be attracted to the rods and get neutralized.[114] However, if the AC current is applied, the positive RF cycle will push the ions away from the rods towards the center. [114] In this case, lighter ions will move faster towards the center with respect to heavier ions causing the separation between ions according to their *m/z* values. [114] The ions that will pass through the mass analyzer should have a stable ion trajectory in both directions xz and yz. i.e., the ions that will not hit the rods and get neutralized. [113,114]

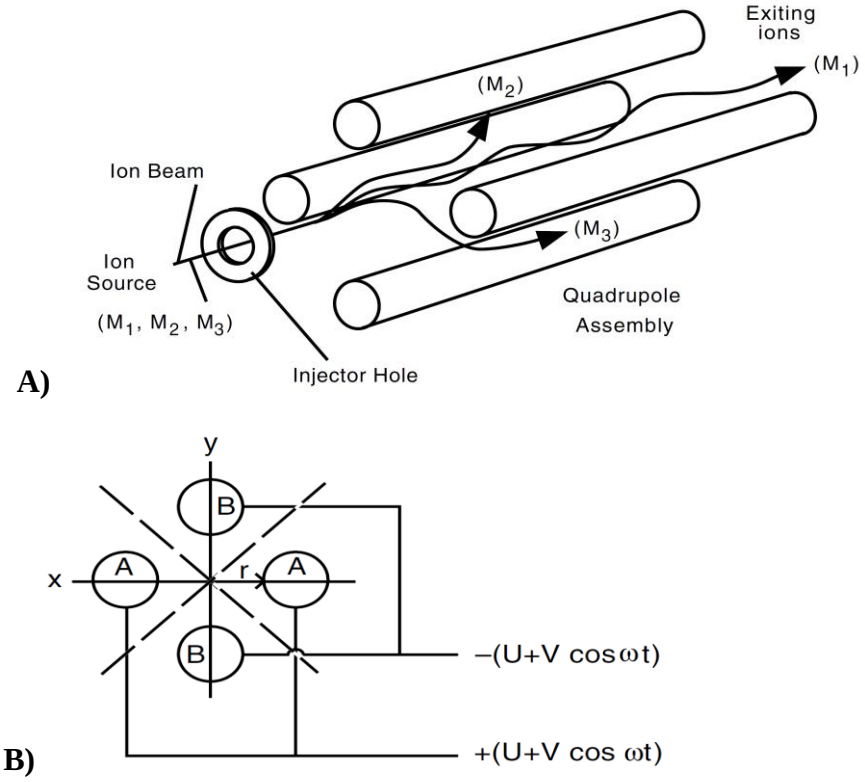


Figure 1.15. Principle of Quadrupole mass analyzer. Adapted with permission from Herbert, C. G.; Johnstone, R. A.: *Mass spectrometry basics*. CRC press, **2002**

A) a beam of three ions (M_1, M_2, and M_3) enters the quadrupole mass analyzer, M_1 was allowed to exit the analyzer while M_2 and M_3 collide with the rods and get neutralized. Planes that experience zero electric field are indicated in dotted lines. The rods on the x- and y-axes are shown with the polarity of the applied voltage (+/-). The direction of ion motion (the z-axis) is perpendicular to the plane of the paper.

1.2.2.2. Time of Flight (TOF) mass analyzer

The TOF analyzer is suitable to be coupled with pulsed ionization sources, such as laser desorption ionization.[81] Its principle of operation depends on allowing the ions with different m/z values and with the same kinetic energy to drift in a field-free region (Figure 1.16).[113] Each ion has a kinetic energy K.E = $zV = \frac{1}{2} mv^2$, where z is the charge, V is the accelerating potential, m is the mass of the ion, and v is the velocity of the ion. [113] From the last equation, we can deduce that $v = (2zV/m)^{1/2}$, which indicates that the velocity of the ion is inversely proportional to its mass. [113] The velocity of the ion can be calculated by dividing the length of the drift tube (L) by the time that the ion takes to traverse this tube. [113]

The TOF analyzer resolution is limited due to the inhomogeneity in the kinetic energy of the ion beam; however, a reflectron can be used to improve the TOF analyzer resolution. [113] The reflectron lens electrostatic field can control the velocity of the ions to make all isomass ions that differ in their K.E arrives at the detector at the same time.[113] After the ions move through the drift tube, they will experience the electrostatic field of the reflector lens placed at the end of the drift tube. [113] The ion's velocity will decrease gradually in the reflectron region until they stop, then the ions are accelerated again to move in the opposite direction in a second drift tube. [113] It should be noted that faster ions of a specific *m/z* value will spend more time in the reflectron region than the slower ions before being reflected to flow in the second drift tube. [113] Thus, retarding faster ions more than slower ions will allow the slower ions to catch up with the faster ones causing all the isomass ions to have a constant K.E. [113]

The TOF analyzers are widely used in MALDI-TOF mass spectrometers, which are used in many applications, such as the analysis of biomolecules, synthetic polymers, and polymer biomolecule conjugates.[115]

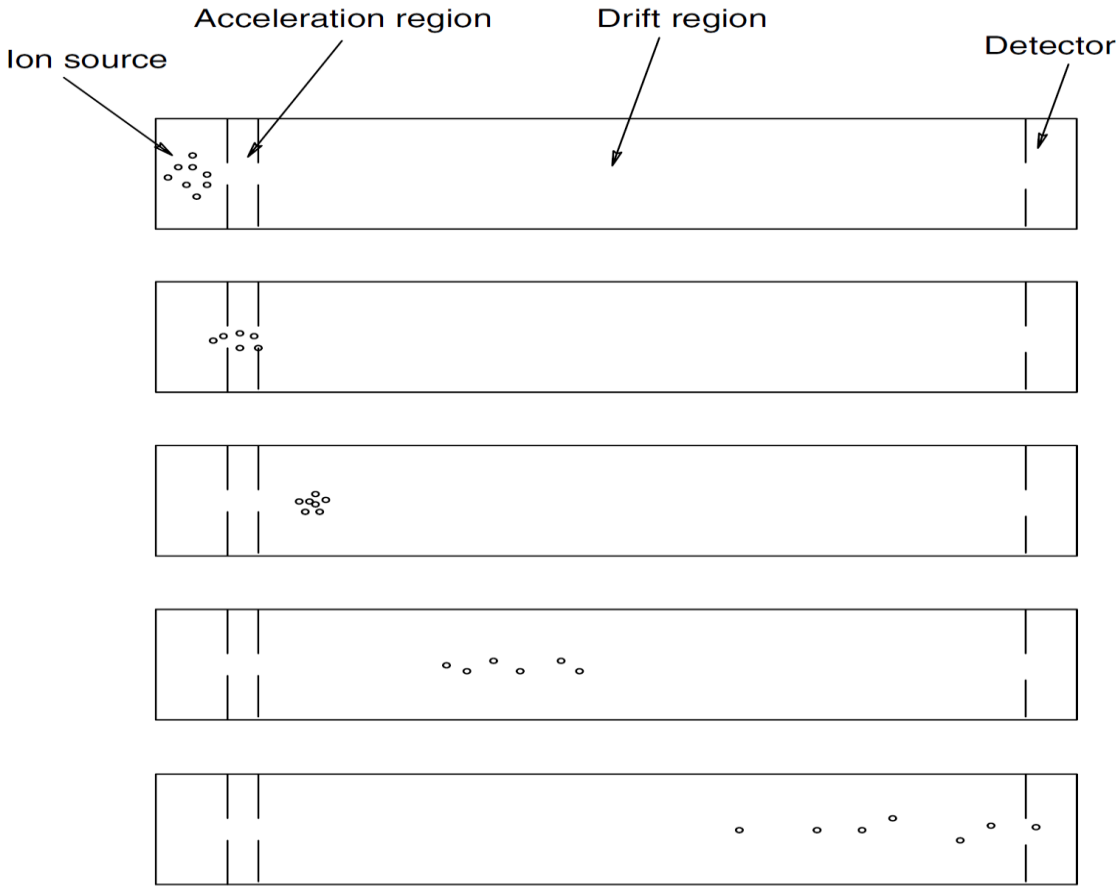


Figure 1.16. Principles of the time-of-flight mass analyzer. Reprinted with permission from Herbert, C. G.; Johnstone, R. A.: *Mass spectrometry basics*. CRC press, **2002**

1.2.2.3. Quadrupole-Time of Flight (Q-TOF) hybrid mass spectrometer

There are several types of hybrid mass spectrometers, which are composed of more than one mass analyzer. [113] The combination of more than one analyzer together is usually giving more advantages than those obtained from using each analyzer separately. [113] For example, the Q-TOF mass spectrometer contains a quadrupole mass filter in conjunction

with a time of flight mass analyzer. [113] Figure 1.17 showed a schematic diagram of an orthogonal Q/TOF instrument. [113] The quadrupole can operate in two different modes, narrow or wideband pass-mode to select specific ions, while the TOF analyzer measures the final mass spectrum. [113] In the narrow mode, the quadrupole selects specific ions to pass, while, in the wide mode, all ions are allowed to pass to the TOF analyzer. [113] The hybrid QTOF instruments can be used with different ionization sources such as ESI,[116] APPI,[117] and MALDI.[118] Also, this hybrid instrument is characterized by its high sensitivity and can detect masses up to 10,000 Da with full sensitivity.[113]

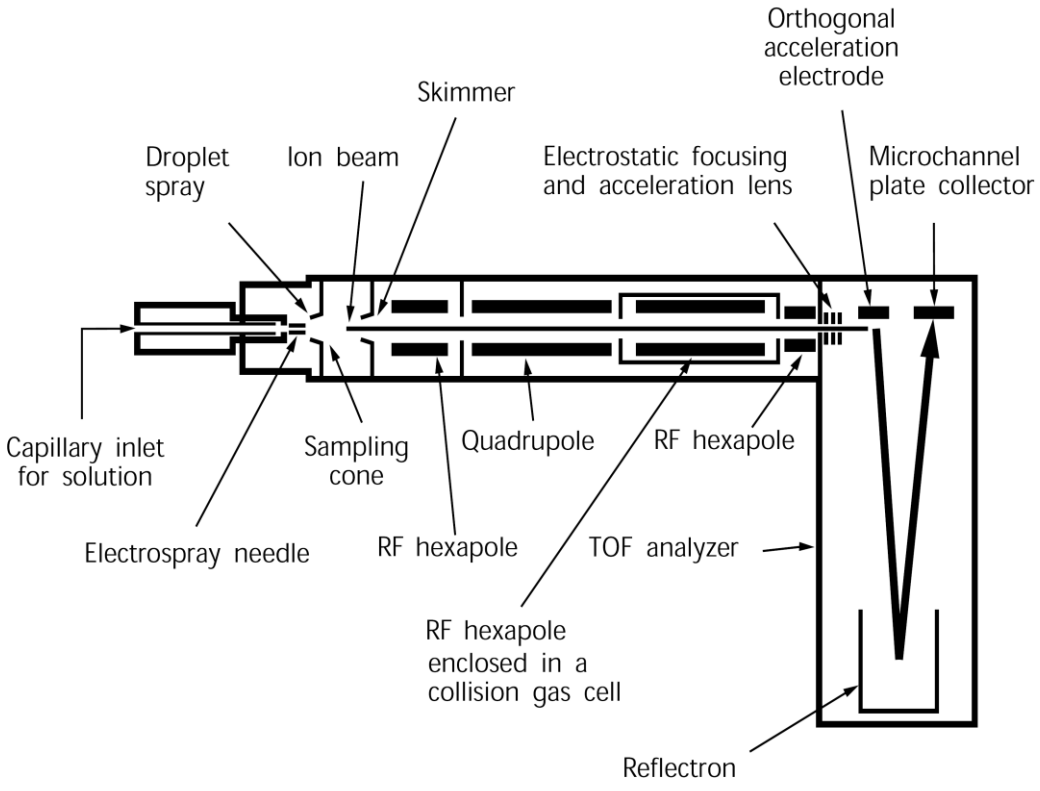


Figure 1.17. Schematic diagram of an orthogonal Q/TOF instrument. Reprinted with permission from Herbert, C. G.; Johnstone, R. A.: *Mass spectrometry basics*. CRC press, **2002**

1.2.2.4. Ion Traps

Ions separation in ion traps depends on trapping ions for a certain time.[74] The trapping of ions facilitates ion/molecule reaction studies and performing sequential tandem MS experiments.[74]

1.2.2.4.1. Quadrupole Ion Trap (QIT)

The quadrupole ion trap (QIT) is considered a three-dimensional quadrupole mass filter.[74] It is composed of three electrodes, a central ring electrode (Doughnut shaped), and two endcap electrodes (Figure 1.18).[66] The upper endcap electrode is used to allow ions into the QIT, while the lower one is used to escape the ions from the QIT for detection. [66] In addition, the upper endcap has a filament that emits electrons to ionize analytes in case if they are not ionized before entering the QIT. [66]

Like the quadrupole mass filter, both DC current and RF are used to confine the ions between the three electrodes.[66] The trapped ions have two types of movements or oscillations, a radial movement with respect to the ring electrode and an axial movement with respect to the two endcaps.[81,119] The ion motion in both directions is independent.[81,119] The ions should have a stable trajectory in both directions to be trapped successfully,[81,119] i.e., ions do not hit the ring electrode or exit from the endcap.

It should be noted that focusing the trapped ions in the center of the trap enhances their stability. This is assisted by inert gas (usually 10^{-2}–10^{-3} torr of He), which slows the motion of the ions through collisions (collisional focusing).[66] The ejection of ions for detection is accomplished by increasing the applied RF potential to the endcaps. [66] The

ions are expelled according to their *m/z* values, so lighter ions are expelled first as they have a higher frequency of axial oscillations with respect to heavier ions. [66]

In general, the quadrupole ion trap has several advantages, such as high sensitivity and good resolution.[120] However, its mass accuracy is low. [120]

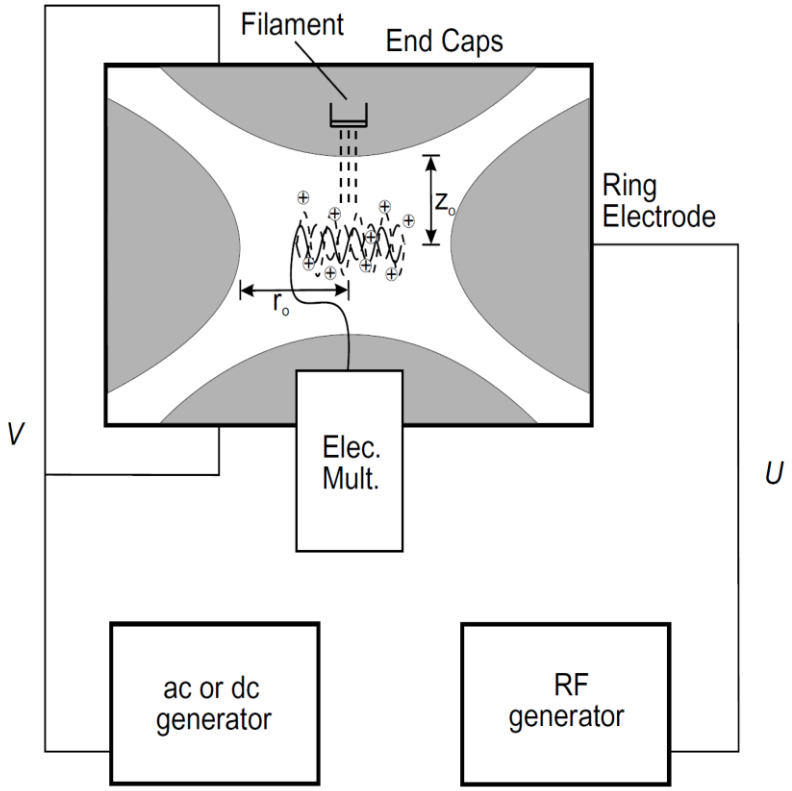


Figure 1.18. Quadrupole ion trap (QIT) mass spectrometer. Reprinted with permission from Smith, R. M.: *Understanding mass spectra: a basic approach*. John Wiley & Sons, **2004**

1.2.2.4.2. Fourier Transform Ion Cyclotron (FTICR) Mass Analyzer

In FTICR mass analyzer, the ions are analyzed and detected in a cell (the heart of the analyzer), which is placed in a uniform strong magnetic field.[121] The FTICR cell is commonly comprised of a couple of trapping plates, a couple of excitation plates, and a

couple of detection plates (Figure 1.19).[121] The ions that enter the cell are trapped by applying a small potential to the trapping plates. [121] Due to the applied magnetic field, each trapped ion will precess with a specific ion cyclotron frequency $\nu_{c,}$ according to its *m/z* value. [121] In general, the cyclotron frequency is inversely proportional to the ion *m/z* value. [121] A radiofrequency rf is applied to the excitation plates to excite the ions to a higher cyclotron radius in order to be detected at the detection plates. [121] When the ions are close to the detection plates, they are detected as an image current in the form of a time-domain signal. [121] This latter signal is transformed into a mass spectrum by using a mathematical technique called is the Fourier transform. [121]

The main advantages of the FTICR are its high resolution, high mass accuracy, and efficient ion trapping.[122] Also, FTICR-MS has been used for the analysis of complex samples such as proteins, glycoproteins, and oligonucleotides.[123-125]

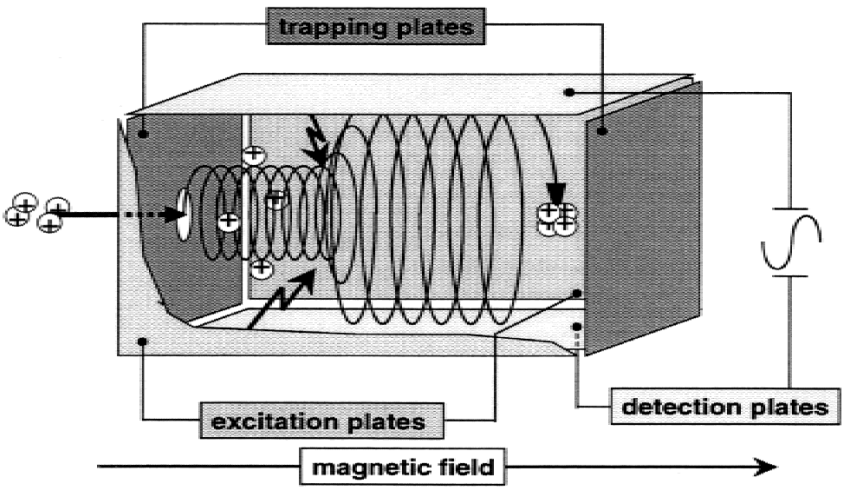


Figure 1.19. Scheme of a cubic analyzer cell. The cyclotron motion and the excitation process are shown schematically. Reprinted with permission from *Biotechnol. Bioeng.* **2000**, *71*(2), 149-161

1.2.2.4.3. Orbitrap Mass Analyzer

The increase in the complexity of analytical mixtures over the last two decades leads to the development of a new mass analyzer called the Orbitrap.[126] This mass analyzer has the excellent advantages of all previous mass analyzers, such as the image current detection from the FTICR, trapping of ions between electrodes from an ion trap, and the last thing, the use of electrostatic fields from time of flight TOF.[126] In addition, the Orbitrap not only has better advantages on the previous analyzers, but it also overcomes the disadvantages of some of its older relatives, such as the need for a superconducting magnet in FTICR, space charge limitations in the ion trap, and the dynamic range of the TOF mass analyzer.[126]

The Orbitrap mass analyzer operation depends on the orbital trapping theory. [127,128] This theory was introduced and explained in 1923 by Kingdon by using a device consisting of a charged wire surrounded by a metal can. [127,128] The electrostatic field established by this wire causes ions that possess sufficiently high velocity to survive and orbit the wire rather than directly colliding with it. [127,128]

In the new millennium, Alexander Makarov designed the first orbitrap mass analyzer by replacing the thin wire used by Kingdon with a spindle-shaped electrode that is enclosed by two symmetrical outer electrodes separated by a dielectric material (Figure 1.20).[128,129] The inner electrode traps the ions radially by only using the electrostatic attractions that are compensated by the centrifugal force (high velocity of ions). [128,129] Whereas the outer electrodes have two main roles. The first role is to confine the ions axially by the establishment of a trapping field. [128,129] The second task is to act as the receiver plates that

produce image current detection as a time-domain signal, which is converted by the Fourier transform equation to a mass spectrum. [128,129]

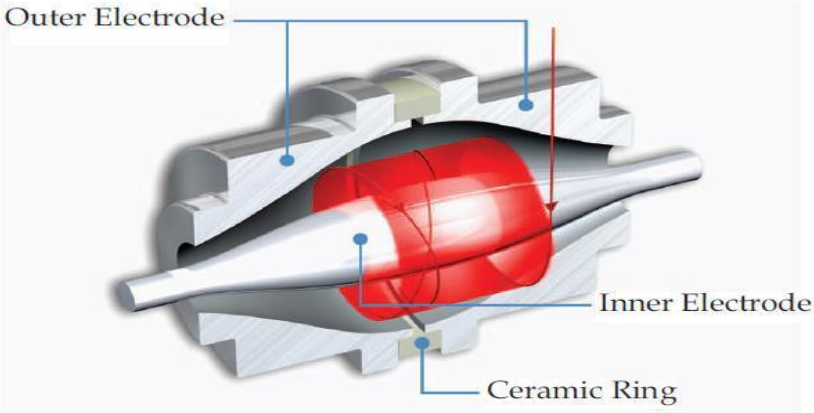


Figure 1.20. A general schematic of an orbitrap mass analyzer. Adapted with permission from High Resolution Mass Spectrometry Using FTICR and Orbitrap Instruments. In Salih, S.M. (Ed.), *Fourier Transform - Materials Analysis*, InTech, Rijeka, Croatia, **2012**.

1.2.3. Ions in the Gas Phase

1.2.3.1. Gas-phase Ions Formation and Their Fragmentation

The formation of gas-phase ions is occurring in the ionization source.[74] It was observed that during the ionization process, the energy gained by the ionized molecule plays a role in the formation of in-source fragment ions.[74]

1.2.3.2. Tandem Mass Spectrometry

The principle of tandem mass spectrometry (MS/MS) is based on the isolation of a precursor ion, which is then activated to fragment into product ions plus neutral fragments.[81] Over the years, tandem mass spectrometry has been used comprehensively to characterize the structure of unknown compounds.[130]

Tandem MS instruments can perform in two ways, either tandem in space or tandem in time. [81] In tandem in space, at least two analyzers and a collision cell are necessary. [81] The first mass analyzer is used to choose the precursor ion of interest, which is then allowed to enter the collision cell (Figure 1.21). [81] Inside the collision cell, the selected precursor ion experiences a collision with a neutral gas, which results in the fragmentation of the precursor ion to its characteristic product ions. [81] This latter process is called collision-induced dissociation CID. [81] The resulting product ions are then allowed to enter a second mass analyzer in order to be separated and detected.[81,131] The triple quadrupole mass spectrometer QqQ (lower case q represents the collision cell) and hybrid instruments such as the QqTOF are common instruments used for tandem in space MS.[81] In the case of tandem MS in time (Figure 1.21), all events that were described for tandem in space MS (isolation, fragmentation, and detection) occur sequentially in one region and/or one analyzer. [81] Ion trap mass analyzers such as QIT, FTICR, and Orbitrap are typical tandem in time mass analyzers. [81]

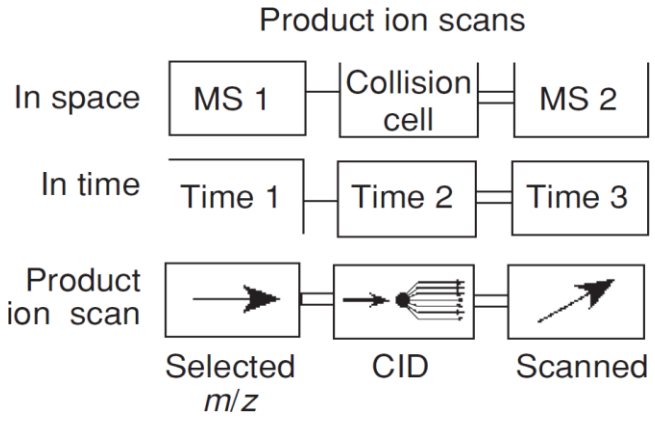


Figure 1.21. Comparison of a product ion scan performed by a space-based and a time-based instrument. Reprinted with permission from De Hoffmann, E.; Stroobant, V.: *Mass Spectrometry: Principles and Applications*, 3rd ed. John Wiley and Sons, Chichester, **2007**

1.2.4. Kendrick mass defect plots

A typical Kendrick mass defect (KMD) plot uses CH_2 (methylene group) as the base unit to classify and sort structurally related ions in a complex spectrum, such as the one of crude oil (Figure 1.22).[132] The KMD plot is a plot between the Kendrick mass defect versus the nominal Kendrick mass.[132] The nominal Kendrick mass can be calculated from the following formula: Kendrick *m/z* = IUPAC *m/z* x (14.00000/14.01565), while the Kendrick mass defect = nominal Kendrick *m/z* - exact Kendrick *m/z*. [132] All ions that differ in the number of CH_2 only will have the same defect and align on the same horizontal line in the KMD plot (Figure 1.22). [132] Also, the ions with a different mass defect that are aligned on the same vertical line differ in the double bond equivalent (DBE) or the number of heteroatoms. i.e., the ion with higher DBE or number of heteroatoms will have a higher mass defect. [132]

It should be noted that in KMD plots, other units instead of CH_2 can be used according to the studied class of compounds, and in this case, the modified Kendrick *m/z* = IUPAC *m/z* x (nominal *m/z* of a base unit/exact *m/z* of base unit).[133] Moreover, 2D-KMD plots can use the mass defect with respect to two different base units in order to relate two structural features together in one plot.

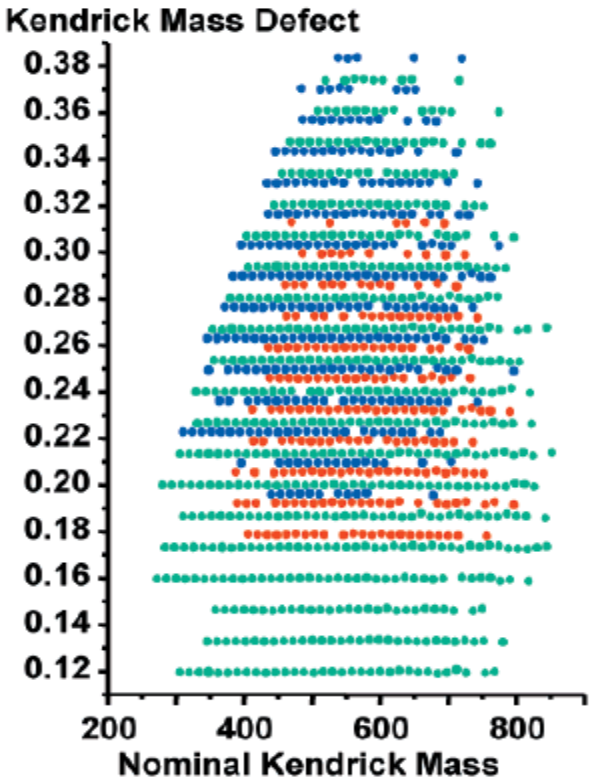


Figure 1.22. KMD plot of a crude oil sample using CH2 as a base unit. Adapted with permission from *Anal. Chem.* **2001**, *73* (19), 4676-4681

1.3. Mass Spectrometry Analysis of Biomolecules

1.3.1. Mass Spectrometry Analysis of Sporopollenin

Gas Chromatography-Mass spectrometry GC-MS was one of the most used mass spectrometric techniques to investigate sporopollenin structure.[134-136] GC-MS studies that were done on sporopollenin samples after performing an ozonolysis reaction prove the presence of unsaturated fatty acids through the detection of different types of dicarboxylic acids (Table 1.1). [134] The identification of varying length dicarboxylic acids (C_6-C_9) indicates the variety and presumed location of the double bonds.[134]

Table 1.1. Main compounds identified using GC-MS after Oxidation of different sporopollenin samples. Reprinted with permission from *Sex. Plant Reprod.* **1999**, *12*(3), 171-178

		P. pinaster	*B. alba*	*A. elatior*	*C. annuum*
α,ω-dicarboxylic acids (%)					
HOOC-(CH_2)n-COOH	C_6	28.8	–	63.2	39.9
	C_7	–	–	–	22.1
	C_8	–	–	–	16.7
	C_9	–	–	–	9.8
Fatty acids (%)					
CH_3-(CH_2)n-COOH	n=13	20.3	–	–	–
	n=14	12.1	13.8	–	–
	n=16	10.3	–	4.8	–
n-alkanes (%)					
CH_3-(CH_2)n-CH_3		6.7 (n=26)	50.7 (n=24–28)	–	–
Non-identified major compounds (%)		–	22.8 (M^+ 340)	–	–
Other minor compounds (%)		21.8	12.7	32.0	11.5

Moreover, some trials using pyrolysis GC-MS revealed *p*-coumaric acid and/ or aromatic compounds and considered them the major building blocks of sporopollenin (Figure 1.23).[135,136] It is well known that pyrolysis can change the structure of the original material, especially those with long aliphatic chains, which can aromatize and produce aromatic compounds that were not present or exist initially.[137] Bernard *et al.*[138] studied the thermal degradation of sporopollenin and showed that this occurred mainly in two main stages. The first one below 500 °C, where sporopollenin undergoes dehydrogenation and deoxygenation simultaneously. [138] Whereas the second stage is above 500 °C, where aromaticity increases dramatically, this was verified using infrared spectroscopy, which showed the increase in the intensity of the C-H aromatic stretching band with increasing temperature (Figure 1.24).[138]

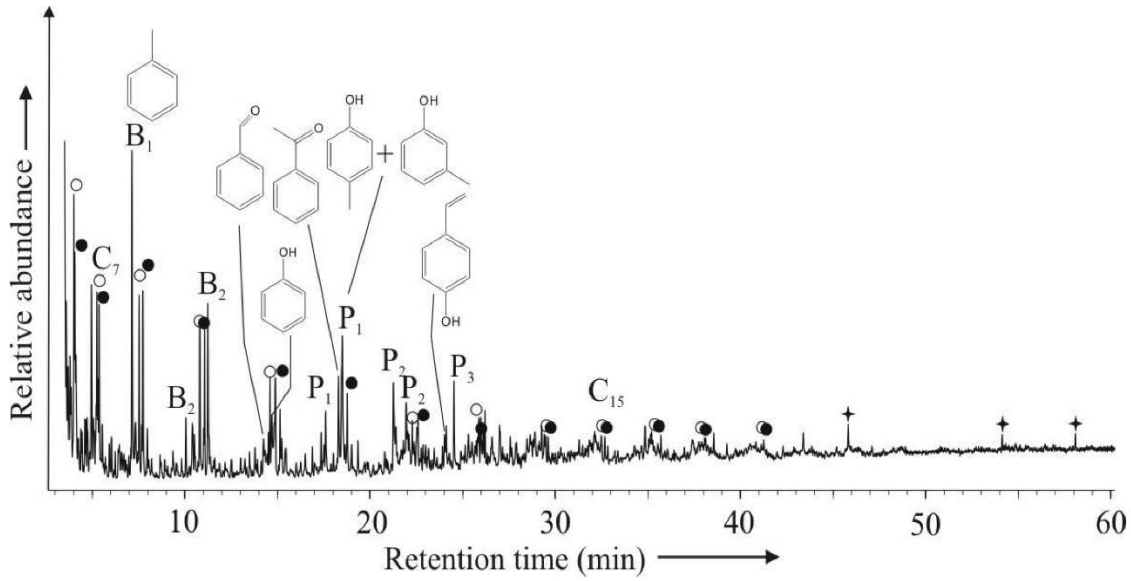


Figure 1.23. Total ion chromatogram resulting from the curie point pyrolysis-GC-MS of Creataceous megaspore *Dijkstraiporites helios.* Reprinted with permission from *The molecular composition of sporopollenin from fossil megaspores as revealed by micro-FTIR and pyrolysis-GC-MS.* 23rd International Meeting on Organic Geochemistry, **2007.**

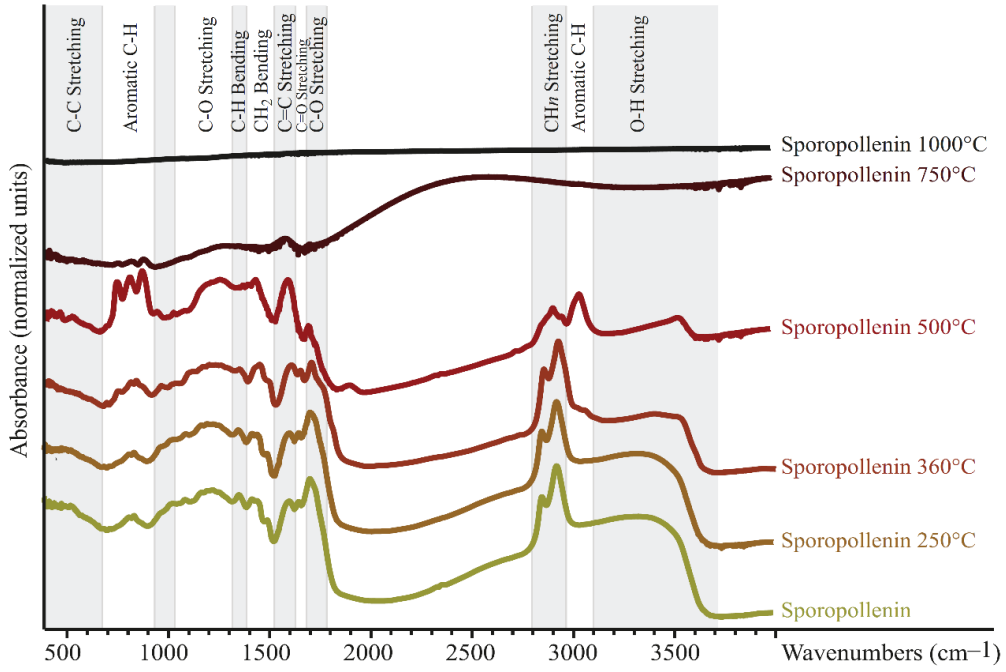


Figure 1.24. IR spectra of sporopollenin residue at different thermal degradation stages. Reprinted with permission from *Heliyon.* **2015**, *1*(2), 1-28

Another fact from Bernard's study that deoxygenation occurs in the first stage was mainly by the loss of CO_2, which supports Hayatsu *et al.* proposal of fatty acids network.[5, 138] This study proves without any doubt that high-temperature processes such as pyrolysis and/or high energy ball milling can produce a series of aromatic compounds, which most probably lead to misleading and incorrect structural interpretations.

GC-MS was not the only used mass spectrometric technique to investigate the structure of sporopollenin over the years; Moore *et al*. [139] used MALDI-TOF-MS (positive and negative modes) to investigate *lycopodium clavatum* sporopollenin (Figure 1.25). The ions were found in the range of *m/z* 4000-1700.[139] The authors were trying to prove that the simple repeating unit of sporopollenin was composed of coumaric acid; however, the differences between the ion masses did not show a correlation between them, which support the complexity of the sporopollenin structure.[139]

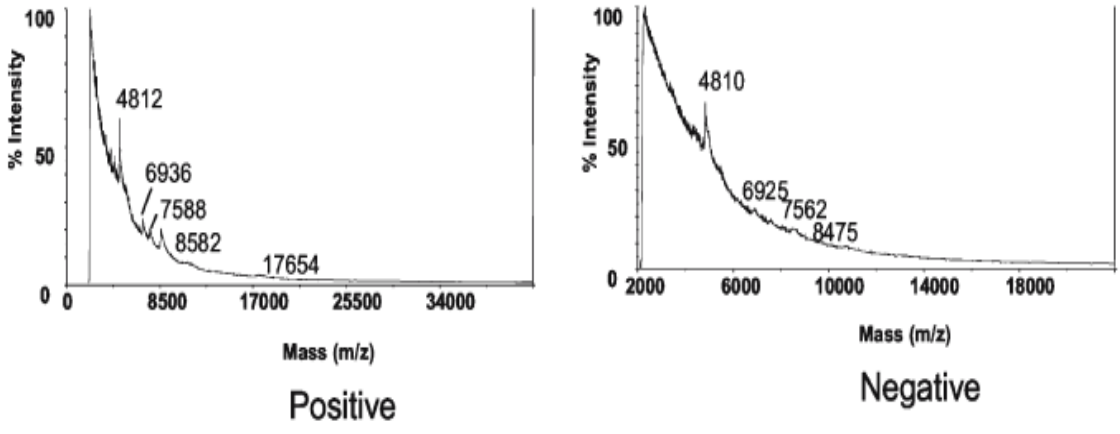


Figure 1.25. MALDI MS of *Lycopodium clavatum* sporopollenin. Reprinted with permission from *Protoplasma*. **2006**, *228*(1-3), 151-157

1.3.1.1. Mass spectrometry for investigating the sporopollenin biosynthesis

Besides chemical and analytical investigations to discover the mystery of the sporopollenin envelop, there was a tremendous push to study and propose the biosynthetic pathways of its formation. [140,141] Several biosynthetic studies showed that the polyhydroxylated tetraketide molecule was one of the essential sporopollenin monomers (Figure 1.26).[140,141] This polyhydroxylated tetraketide arises from the condensation of fatty acyl coenzyme A with malonyl coenzyme A in the presence of polyketide synthetase enzyme (PKS), as shown in Scheme 1.3, followed by a hydroxylation step using a cytochrome enzyme (CYP).[140,141] Finally, the resulting polyhydroxylated ketide is reduced by the action of tetraketide alpha pyrone reductase enzyme (TKPR). [140,141] It should be noted that the fatty acyl Co-A substrates are formed by the action of Acyl Co-A synthetase (ACO) on fatty acids.[140,141] The formation of polyhydroxylated tetraketide was supported by Sung Soo Kim *et al.* [140] through the incubation of different types of fatty acyl Co-A with malonyl Co-A in the presence of PKS enzyme that is identified to be involved in the biosynthesis of sporopollenin. The resulting products from this incubation process were investigated by Liquid Chromatography LC-ESI-MS/MS (Figure 1.27). [140]

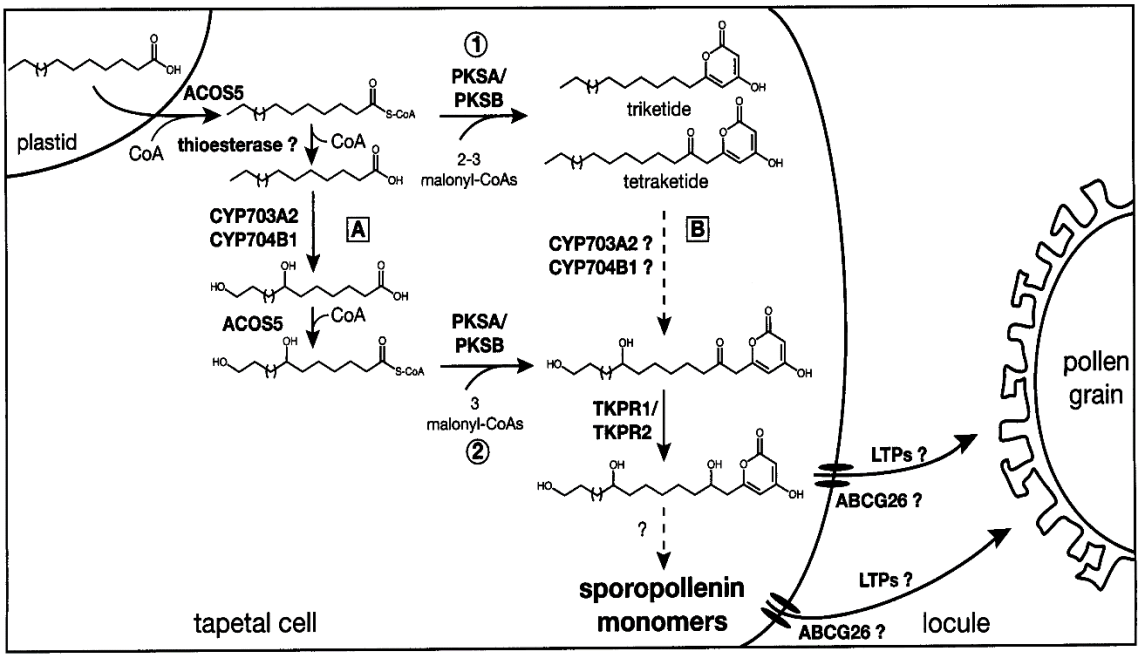
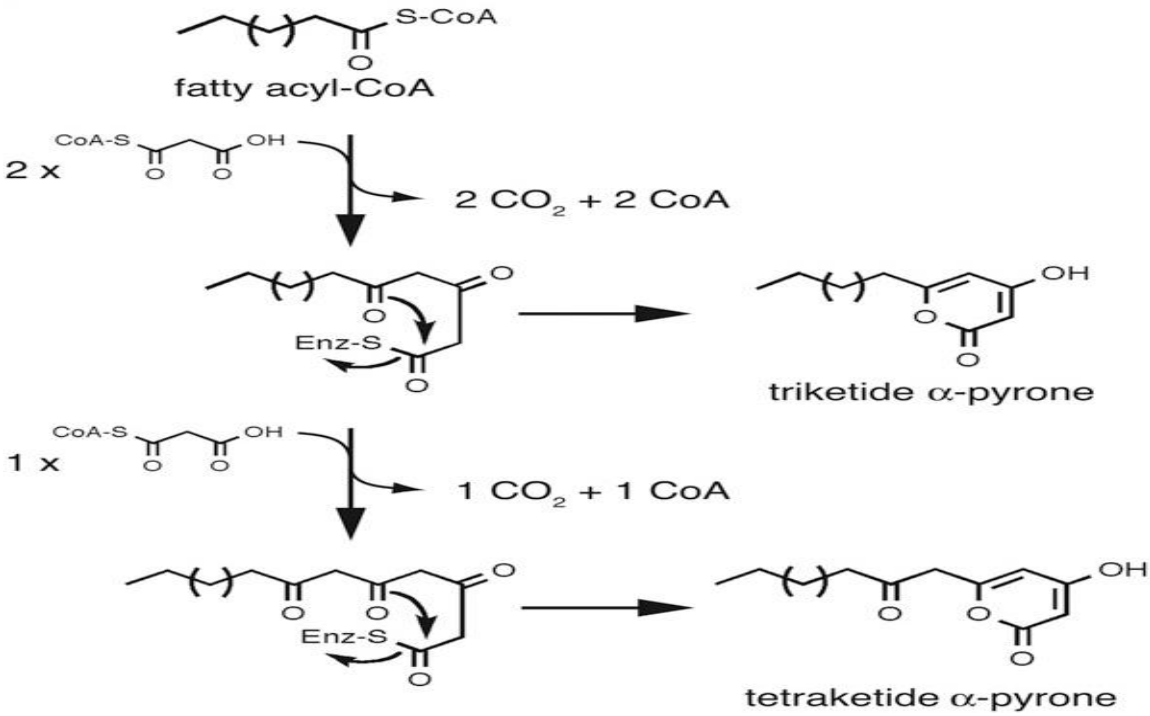


Figure 1.26. Proposed biosynthesis of sporopollenin. Reprinted with permission from *Plant Cell.* **2010**. *22*(12), 4067-4083



Scheme 1.3. Putative mechanism of fatty acyl-CoA condensation with malonyl- CoA catalyzed by PKSs. Adapted with permission from *Plant Cell*. **2010**, *22*(12), 4045-4066

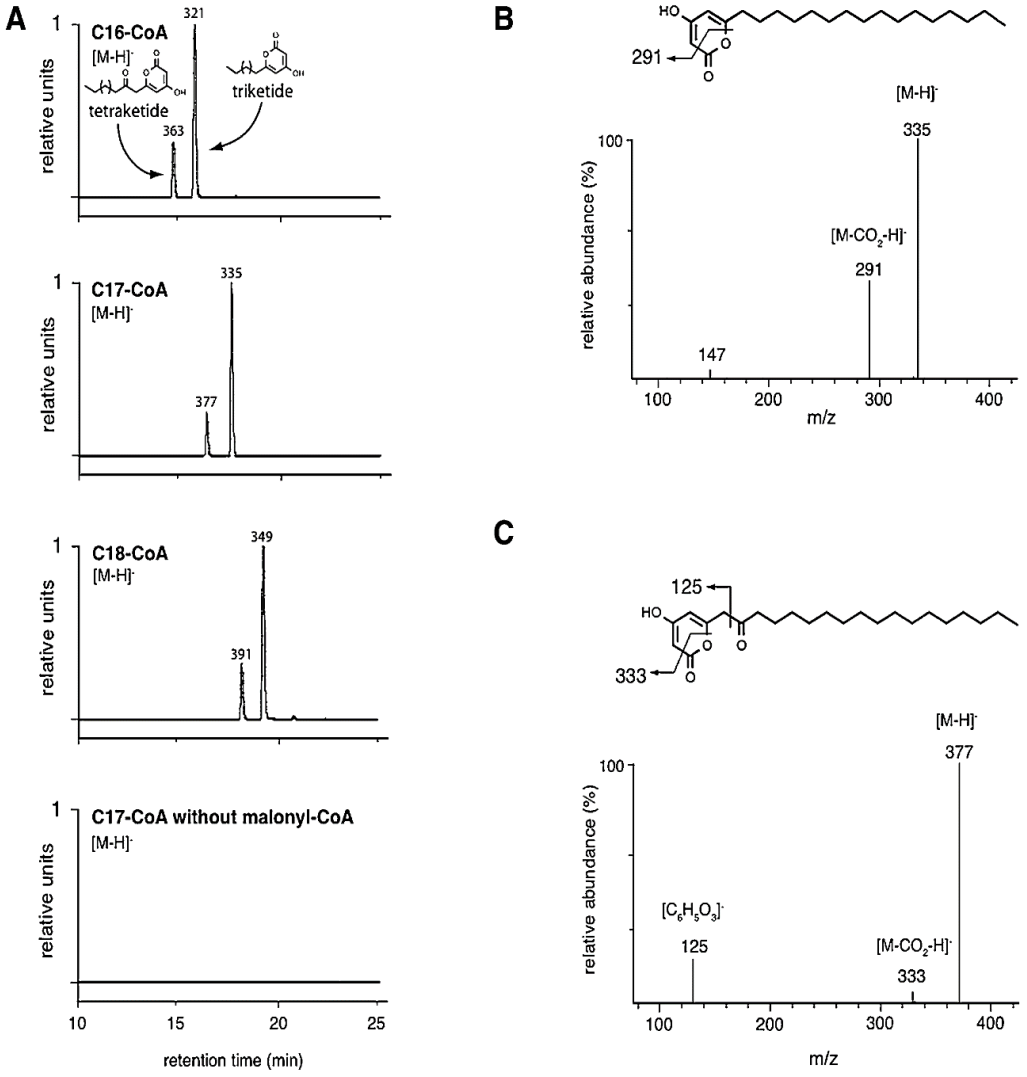


Figure 1.27. Detection of PKS Reaction Products by LC-MS/MS. Adapted with permission from *Plant Cell*. **2010**, *22*(12), 4045-4066. Reaction products were separated by reverse-phase UPLC and investigated by negative ion mode ESI- MS/MS.
(A) The substrates C16-CoA, C17-CoA, or C18-CoA was used as starter substrate as indicated, and cognate tri- and tetraketide reaction products were detected. Lastly, a C17-CoA substrate was incubated without malonyl-CoA to act as a negative control.
(B) Fragmentation mechanism of the $[M-H]^-$ ion of the triketide a-pyrone obtained from C17-CoA incubation with malonyl-CoAas substrate.
(C) Fragmentation mechanism of the $[M-H]^-$ ion of the tetraketide α-pyrone reaction product resulted from C17-CoA incubation with malonyl CoA.

1.3.2. Lignin Mass Spectrometry

1.3.2.1. Pyrolysis-gas chromatography-mass spectrometry (Py-GC-MS)

Pyrolysis causes the degradation of lignin samples to small fragments and/or monomers which can be separated by gas chromatography and then identified by MS. [142] The thermal degradation of lignin occurs through the cleavage of C-O and C-C linkages between monomers to produce relatively simple phenols.[143] Lignin monomers such as *p*-hydroxyphenyl, guaiacyl, and syringyl structures with other characteristic structures have been identified by this technique.[144,145] Since the thermal degradation does not affect aromatic methoxy groups of G and S units, it is possible to identify the H: G:S ratio of the analyzed lignin sample. [144,145] However, this ratio provides minimal information about the structure of the complex lignin mixture.

1.3.2.2. Soft ionization MS methods

Mass spectrometry (MS) is one of the essential tools to determine the structure of lignin oligomers. [146] Soft ionization methods like Electrospray Ionization (ESI), [146-149] Atmospheric Pressure Photoionization (APPI), [50, 150-153] and Matrix-Assisted Laser Desorption/Ionization (MALDI) had been used extensively for the sequencing of lignin oligomers.[89,154-156]

For example, APPI-MS was used to identify fifty-seven wheat straw lignin oligomers.[50] Also, ESI-MS was used for the Eucalyptus globulus dioxane lignin (EDL) structural elucidation.[157] This study was done on the low molecular weight fraction of EDL separated by using size exclusion chromatography (Figure 1.28).[157] EDL ESI-MS (

negative ion mode) measured by a Quadrupole-Time of Flight Spectrometer (Q-TOF) showed a series of lignin oligomers *inter alia* at *m/z* 643 (trimer) and 869 (tetramer), 1065 (pentamer), and 1291 (hexamer).[157] These series of lignin oligomers differ from each other by 226 Da (S-unit) and 196 Da (G-unit). [157] Low energy-collision induced dissociation CID-tandem mass spectrometry MS/MS was used to confirm the structures of these lignin oligomers. [157] For example, the precursor ion at *m/z* 643 was assigned to exist as lignin trimer S(8-O-4') S(8-8'')S, and the precursor ion at *m/z* 839 was assigned to exist as a lignin tetramer with two possible structures S(8-O-4') S(8-O-4'')S(8-8''')S or S(8-O-4') S(8-8'')S(8-O-8''') based on their low-energy CID-fragmentation pathway (Figure 1.29).

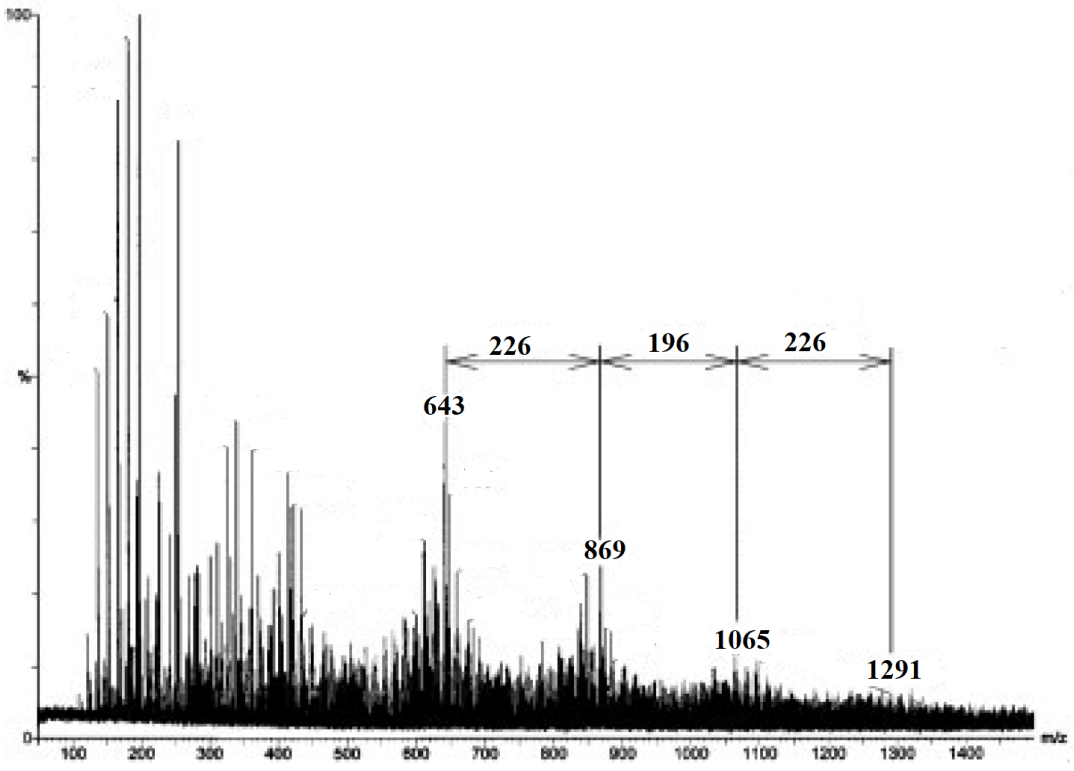


Figure 1.28. Negative ion mode ESI-MS of the low molecular weight EDL fraction. Reprinted with permission from *Macromol. Biosci*. **2003**, *3*, 339-343

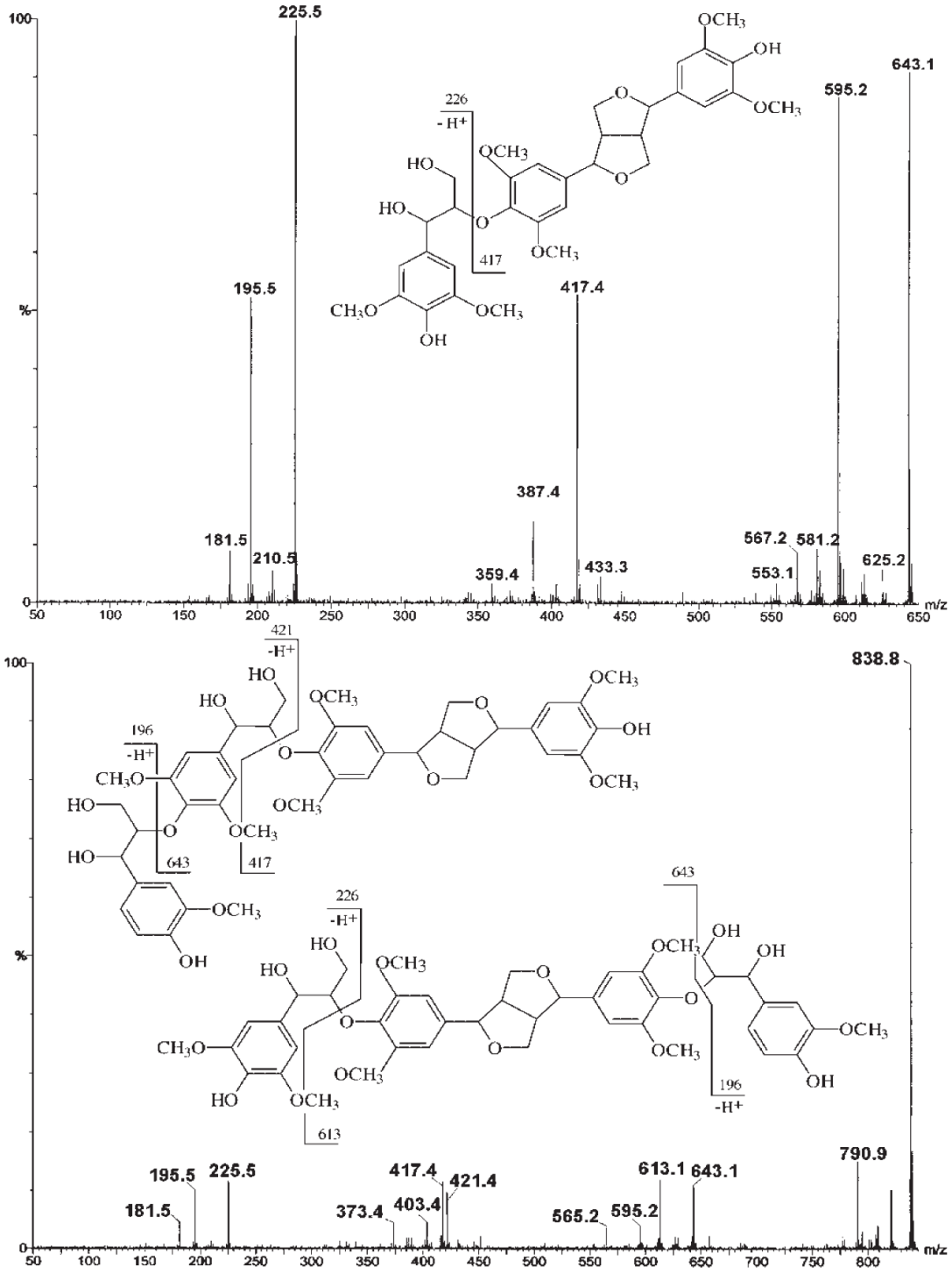


Figure 1.29. MS/MS spectra of lignin oligomers with *m/z* 643 (top image) and *m/z* 839 (bottom image) and correspondent inferred structures. Reprinted with permission from *Macromol. Biosci*. **2003**, *3*, 339-343

Moreover, Morreel and coworkers studied the gas-phase fragmentations of different synthetic deprotonated lignin oligomers containing the main linkage such as β–O–4′, β–5′ and β–β′ using negative ion mode ESI-MS/MS (Figure 1.30).[158] This latter study aimed to differentiate between different lignin linkages through their characteristic neutral losses and/or fragmentation pathway.[158] For example, The product ion scan of the deprotonated β-aryl ether [G(t8-O- 4′)G] anion loses one water molecule (-18 Da) to yield the product ion $[M\text{-}H\text{-}H_2O]^-$. [158] This last product ion loses formaldehyde (-30 Da) to give the product ion $[M\text{-}H\text{-}H_2O\text{-}CH_2O]^-$. [158] Unlike the [G(t8-O- 4′)G] anion, the product ion scan of the deprotonated phenylcoumaran [G(8–5′)G] anion does not show the consecutive loss of H_2O and CH_2O (-48 Da). [158] However, it showed the loss of either water molecule (-18 Da) or formaldehyde (-30 Da) to afford the product ions assigned as $[M\text{-}H\text{-}H_2O]^-$ and $[M\text{-}H\text{-}CH_2O]^-$, respectively. [158] Lastly, the product ion scan of the deprotonated [G(8–8′)G] resinol did not show water loss due to the lack of a primary alcohol group; however, the formaldehyde loss was observed to yield the product ion $[M\text{-}H\text{-}CH_2O]^-$. [158]

Besides using different soft ionization over the years to characterize complex lignin samples, some recent work showed the high complexity of lignin samples.[159,160] For example, the use of different matrices in the MALDI-MS studies of a specific lignin sample gives different spectrum and lignin composition based on the used matrices.[159] Moreover, the use of different ionization techniques such as APCI, APPI, and ESI gives different information about the lignin sample under investigation.[160]

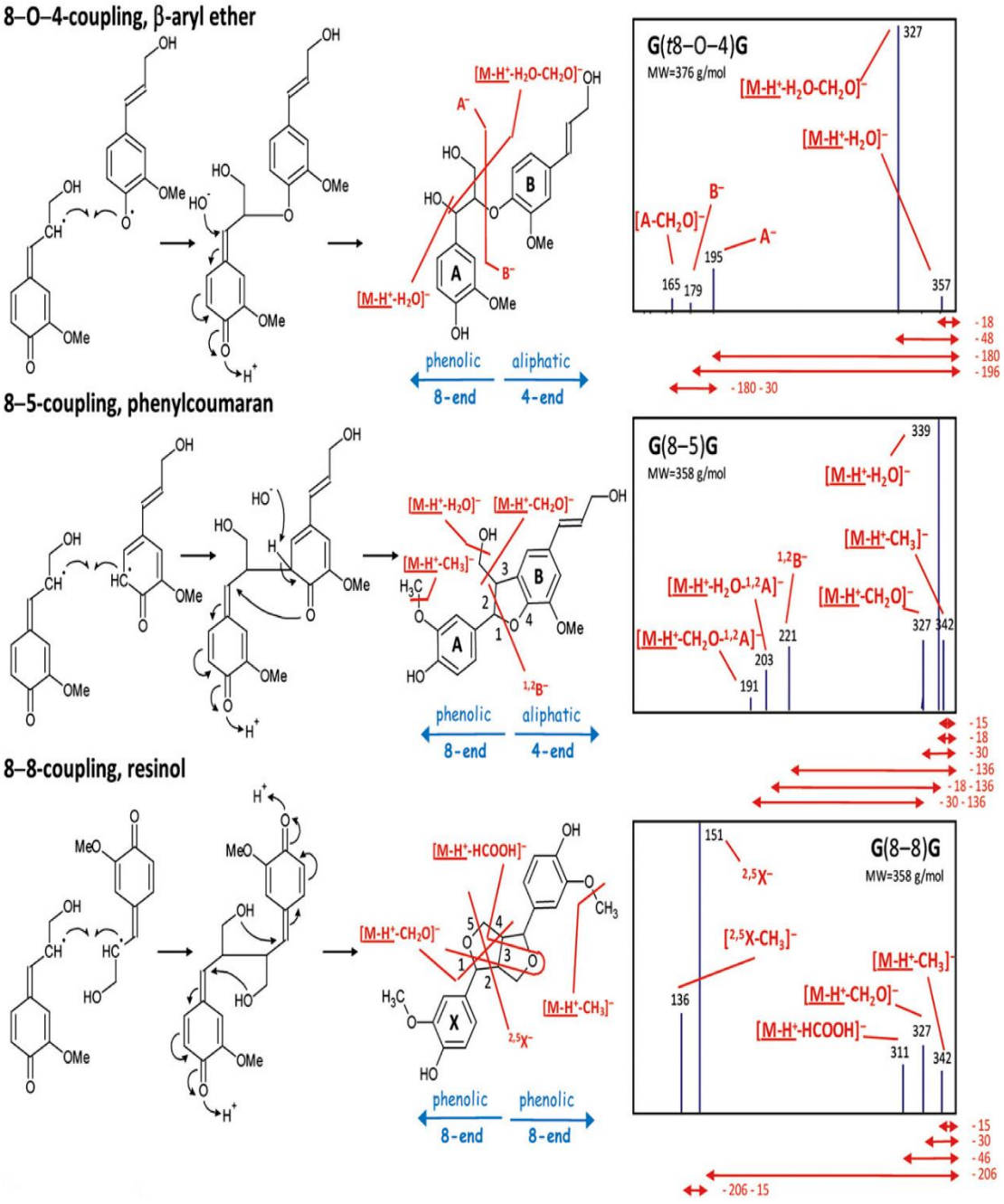


Figure 1.30. The product ion scan and the fragmentation mechanism of the deprotonated lignin oligomers [G(t8-O- 4′)G], [G(8–5′)G] and[G(8–8′)G]. Adapted with permission from *Plant Physiol.* **2010**, *153*(4), 1464-1478

1.3.2.3. KMD plots in lignin analysis

Due to the complexity of the lignin samples, several studies aimed to get general information using Kendrick mass defect (KMD) plots instead of studying each individual ion in the complex spectrum. [133, 160, 161] For example, the 2D KMD plot shown in Figure 1.31 is composed of the mass defect with respect to the unit [OCH_2] on the x-axis (methoxylation) and the mass defect with respect to the phenol core [C_6H_4O].[133]

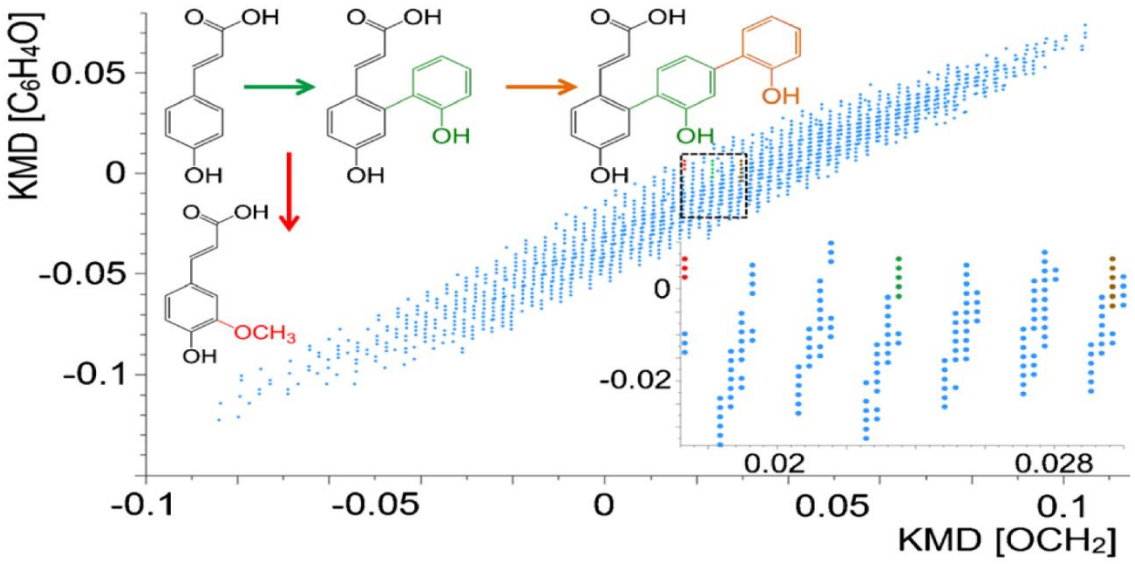


Figure 1.31. 2D KMD plot for a lignin sample after decomposition. A part of the plot is enlarged to show the relation between three different species (red, green, brown). The proposed core structures of these compounds are shown. Adapted with permission from *Anal. Bioanal. Chem.* **2016**, *408*(18), 4835-4843

To clarify the usefulness of this plot, the magnified part of this plot will be explained. In this magnified part, there are different color data points (red, green, and brown). [133] Based on the CID-MS/MS analysis, the first red data point (*m/z* 163.0389) was assigned as coumaric acid (zero aromatic methoxy group). [133] This last red point and the

two following red points are aligned on the same vertical (same OCH_2 mass defect), and their masses differ by 30 Da.[133] This means that the next two red points can be assigned as ferulic acid (one aromatic methoxy group) and sinapic acid (two aromatic methoxy groups), respectively.[133] Likewise, the first red point aligned on the same horizontal line with the first green and first brown data points (same C_6H_4O mass defect). [133] This indicated that the three data points have the same core structure (coumaric acid) but differed in the number of phenolic groups, as demonstrated in Figure 1.31. [133]

1.3.2.4. Mass accuracy and lignin molecular formulas assignment

The identification of unknown analytes with high confidence requires highly accurate mass measurement. [162] However, in highly complex mixtures, the assignment of molecular formulas based on mass accuracy solely is a challenging task.[162] It has been reported that even mass accuracy of less than 1 ppm is not sufficient to exclude enough formulas that can be assigned to ions composed of C, H, N, O, P, and S. [163]

It has been established that the ranking of possible formulas according to their isotopic pattern fitting percentage can greatly assist mass accuracy in reducing the number of candidates that can be assigned to each analyte. [163,164] For example, the reduction in the number of possible formulas at different mass accuracies due to the impact of isotopic abundance accuracy is shown in Table 1.2. [163] Interestingly, a mass spectrometer with 3 ppm mass accuracy and 2% error on isotopic abundances is more efficient than a hypothetical mass spectrometer capable of 0.1 ppm mass accuracy (Table 1.2). [163]

Table 1.2. The effect of mass and isotopic abundance accuracy on decreasing the number of possible assigned formulas. Reprinted with permission from *BMC bioinform*. **2006**, *7*(1), 1-10.

	without isotope abundance information					2% isotopic abundance accuracy	5% isotopic abundance accuracy
molecular mass [Da]	10 ppm	5 ppm	3 ppm	1 ppm	0.1 ppm	3 ppm	5 ppm
150	2	1	1	1	1	1	1
200	3	2	2	1	1	1	1
300	24	11	7	2	1	1	6
400	78	37	23	7	1	2	13
500	266	115	64	21	2	3	33
600	505	257	155	50	5	4	36
700	1046	538	321	108	10	10	97
800	1964	973	599	200	20	13	111
900	3447	1712	1045	345	32	18	196

Based on the nature of the analyzed complex sample, several filters can be utilized to reduce the number of possible formulas for each analyte. For example, the search for lignin oligomers can be restricted to formulas composed of C, H, and O only. [165,166] Also, the expected C/O ratio and the degree of unsaturation (double bond equivalent DBE) range for different lignin oligomers can be used as a filter in the search for possible lignin formulas. [165,166] For example, lignin monomers have DBE $\geq$ 4 and C/O ratio $\geq$ 2.5, while lignin dimers have DBE $\geq$ 8 and C/O $\geq$ 2.5.[165,166]

It should be noted that the previous restrictions, along with mass accuracy and isotopic distribution fitting percentage, can help in reducing the number of formulas that can be assigned to each unknown lignin ion. However, due to the extreme complexity of the lignin mixture, overlapping isotopic distribution patterns and centroid mass shifts are very possible, making these restrictions insufficient to assign formulas.[167]

In this thesis, we have used medium resolution mass spectrometers such as APPI/ESI-QqTOF and MALDI-TOF/TOF to analyze the French pine and oak virgin released lignin (Chapter 3-5). These latter instruments are known to have a medium mass accuracy range of approximately ± 10-50 ppm. [168] Consequently, mass accuracy and/or isotopic distribution are insufficient, and tandem mass spectrometry (MS/MS) was essential to find the most likely formula and/or structure for the selected precursor lignin ions.

It should be noted that the selected lignin precursor ion could be a mixture of isobars and/or isomers which significantly complicates the resulting product ion scans. It has been reported by Qi et al. that within 0.2 *m/z* units of the lignin complex MS spectrum, there are twenty ions and/or assigned features (Figure 1.32).[160] It is known that a quadrupole mass analyzer can select precursor ions with a minimum isolation window of ± 0.5 *m/z* units. [160] Therefore, if we assume that a quadrupole mass analyzer is used to isolate one of these twenty ions for MS/MS, all of them will be selected simultaneously. In addition, it is expected that each ion could be a mixture of isomers which can complicate matters further and produce very complex product ion scans. In this thesis, most of the selected lignin precursor ions were proposed to be a mixture of isomers and/or isobars to facilitate the interpretation of their complex and/or messy product ion scans (Chapters 4 and 5).

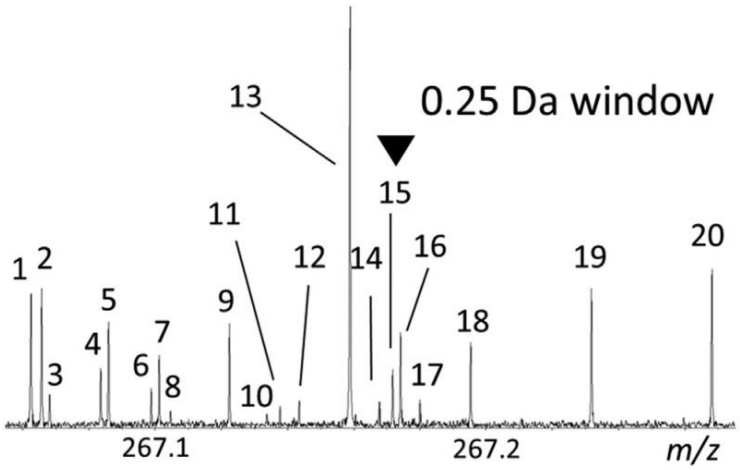


Figure 1.32. Assigned features in a 0.2 *m/z* unit of a complex lignin spectrum. Reprinted with permission from *Sci. Total Environ.* **2020**, *713*, 136573.

1.4. Conclusion

It is evident that the structural elucidation of complex biopolymers and/or biomolecules such as sporopollenin and lignin is still a matter of debate. This debate arises from the difficulty in extracting these natural materials, especially lignin, in their native forms. i.e., without any structural modification and/or degradation.

In the case of sporopollenin, its extreme stability allows it to be extracted in its native form; however, this stability creates a challenge in its structural elucidation. Due to this stability, there is no mild extraction method that can degrade sporopollenin mildly to simplify its structural elucidation. In this thesis, the structural elucidation of sporopollenin without any degradation or modification (top-down approach) and its challenges are discussed. In Chapter 2, the sporopollenin structure enigma was solved using different mass spectrometric techniques combined with nuclear magnetic resonance (NMR) and high-resolution x-ray photoelectron spectroscopy (HR-XPS) studies.

In Chapters (3-5), the complexity and structural diversity of the French pine and oak virgin released lignins was revealed using different ionization techniques (MALDI, APPI, and ESI). The analyzed virgin released lignins are considered to be an excellent representative of the native lignin structure. For this reason, several novel lignin structures were identified, and important conclusions were drawn from the analysis of these complex virgin released lignins.

References

[1] Mackenzie, G.; Boa, A. N.; Diego-Toboada, A.; Atkin, S. L.; Sathyapalan, T.: Sporopollenin, the Least known yet toughest natural biopolymer. *Front. Mater.* **2015**, *2*, 1-5.

[2] Zetzsche, F.; and Kälin, O.: Untersuchungen über die membrane der sporen und pollen v.4. Zur autoxidation der sporopollenine. *Helv. Chim. Acta*. **1931**, *14*(1), 517–519.

[3] Ahlers, F.; Bubert, H.; Steuernagel, S.; Wiermann, R.: The nature of oxygen in sporopollenin from the pollen of Typha angustifolia L. *Z. Naturforsch. C*. **2000**, *55*(3-4), 129-136.

[4] Domínguez, E.; Mercado, J. A.; Quesada, M. A.; Heredia, A.: Isolation of intact pollen exine using anhydrous hydrogen fluoride. *Grana*. **1998**, *37*(2), 93–96.

[5] Hayatsu, R.; Botto, R. E.; McBeth, R. L.; Scott, R. G.; Winans, R. E.: Chemical alteration of a biological polymer" sporopollenin" during coalification: origin, formation, and transformation of the coal maceral sporinite. *Energy Fuels*. **1988**, *2*(6), 843-847.

[6] Brooks, J.; Shaw, G.: Chemical structure of the exine of pollen walls and a new function for carotenoids in nature. *Nature*. **1968**, *219*(5153), 532-533.

[7] Li, F. S.; Phyo, P.; Jacobowitz, J.; Hong, M.; Weng, J. K.: The molecular structure of plant sporopollenin. *Nat. plants*. **2019**, *5*(1), 41-46.

[8] Stolle, A.; Szuppa, T.; Leonhardt S. E.; Ondruschka, B.: Ball milling in organic synthesis: solutions and challenges. *Chem. Soc. Rev.* **2011**, *40*(5), 2317-2329.

[9] Stolle, A.; Brindaban, R. (Eds): *Ball milling towards green synthesis: applications, projects, challenges;* Royal Society of Chemistry, Cambridge, UK, **2014**.

[10] Wang, G-W: Mechanochemical organic synthesis. *Chem. Soc. Rev*. **2013**, *42*(18), 7668-7700.

[11] Icli, B.; Christinat, N.; Tonnemann, J.; Schuttler, C.; Scopelitti, R.; Severin, K.: Synthesis of molecular nanostructures by multicomponent condensation reactions in a ball mill. *J. Am. Chem. Soc.* **2009**, *131*(9), 3154-3155.

[12] Declerck, V.; Nun, P.; Martinez, J.; Lamaty, F.: Solvent-free synthesis of peptides. *Angew. Chem. Int. Ed.* **2009**, *48*(49), 9318-9321.

[13] Application for the approval of sporopollenin exine capsules from *lycopodium clavatum* spores Under Regulation (EC) No 258/97 of the European Parliament and of the Council of 27th January **1997** Concerning Novel Foods and Novel Food Ingredients.
[14] Zhang, H.; Saurav, K.; Yu, Z.; Mándi, A.; Kurtán, T.; Li, J.; et al.: α-Pyrones with diverse hydroxy substitutions from three marine-derived Nocardiopsis Strains. *J. Nat. Prod.* **2016**, *79*(6), 1610-1618.
[15] Ouellette, R. J.; Rawn, J. D.: *Organic chemistry: structure, mechanism, and synthesis.* Elsevier, **2014**.
[16] Pomelli, C. S.; D'Andrea, F.; Mezzetta, A.; Guazzelli, L.: Exploiting pollen and sporopollenin for the sustainable production of microstructures. *New J. Chem*. **2020**, *44*(3), 647-652.

[17] Zetzsche, F.; Huggler, K.: Untersuchungen über die Membran der Sporen und Pollen. I.1. Lycopodium clavatum L. *Justus Liebigs Ann. Chem*. **1928**, *461*, 89-108.

[18] Zetzsche, F.; Kalt, P.; Lietchi, J.; Ziegler, E.: Zur Konstitution des Lycopodium-Sporonins, des Tasmanins und des Lange-Sporonins. XI. Mitteilung über die Membran der Sporen und Pollen. *J. Prakt. Chem.* **1937**, *148*, 267-286.

[19] Zetzsche, F.; Vicari, H.: Untersuchungen über die Membran der Sporen und Pollen III.2. Picea orientalis, Pinus silvestris L., Corylus Avellana L. *Helv. Chim. Acta*. **1931**, *14*, 62-67.

[20] Shaw, G.: The chemistry of sporopollenin. In Brooks, G. et al. (Eds.), *Sporopollenin* (pp. 305-348), Academic Press, London and New York, **1971**.
[21] Barrier, S.: *Physical and chemical properties of sporopollenin exine particles* (Doctoral dissertation). University of Hull, **2008**.
[22] Shaw, G.; Apperley, D. C.: ^{13}C-NMR spectra of Lycopodium clavatum sporopollenin and oxidatively polymerised β-carotene. *Grana*. **1996**, *35*(2), 125-127.
[23] Kettley, S. J.: *Novel derivatives of sporopollenin for potential applications in solid phase organic synthesis and drug delivery* (Doctoral dissertation). University of Hull, UK, **2001**.
[24] Boasman, A. J.: *Investigation into the amination and thiolation of sporopollenin*, (Doctoral dissertation). University of Hull, UK, **2003**.
[25] Brooks, J.; Shaw, G.: Recent developments in the chemistry, biochemistry, geochemistry and post-tetrad ontogeny of sporopollenins derived from pollen and spore exines. In Heslop-Harrison, J. (Ed.), *Pollen: development and physiology* (pp. 99-114). Butterworths, London, **1971**.
[26] W. J.; Schneider, D. M.; Labovitz, J.; Opella, S. J. High Resolution Solid State ^{13}C NMR Spectroscopy of Sporopollenins from Different Plant Taxa. *Plant Physiol*. **1988**, *86*(1), 134-136.
[27] Diego-Taboada, A.; Maillet, L.; Banoub, J. H.; Lorch, M.; Rigby, A. S.; Boa, A. N.; et al.: Protein free microcapsules obtained from plant spores as a model for drug delivery: ibuprofen encapsulation, release and taste masking. *J. Mater. Chem. B*. **2013**, *1*(5), 707-713.

[28] Dyab, A. K.; Abdallah, E. M.; Ahmed, S. A.; Rabee, M. M.: Fabrication and characterisation of novel natural Lycopodium clavatum sporopollenin microcapsules loaded in-situ with nano-magnetic humic acid-metal complexes. *J. Encapsul. Adsorp. Sci*. **2016**, *6*(04), 109-131.
[29] Dyab, A. K.; Mohamed, M. A.; Meligi, N. M.; Mohamed, S. K.: Encapsulation of erythromycin and bacitracin antibiotics into natural sporopollenin microcapsules: antibacterial, cytotoxicity, in vitro and in vivo release studies for enhanced bioavailability. *RSC adv*. **2018**, *8*(58), 33432-33444.
[30] Mujtaba, M.; Yilmaz, B. A.; Cansaran-Duman, D.; Akyuz, L.; Yangın, S.; Kaya, M.; et al.: Newly isolated sporopollenin microcages from Cedrus libani and Pinus nigra as carrier for Oxaliplatin; xCELLigence RTCA-based release assay. *Polym. Bull*. **2021**, 1-22.
[31] Alshehri, S. M.; Al-Lohedan, H. A.; Chaudhary, A. A.; Al-Farraj, E.; Alhokbany, N.; Issa, Z.; et al.: Delivery of ibuprofen by natural macroporous sporopollenin exine capsules extracted from Phoenix dactylifera L. *Eur. J. Pharm. Sci*. **2016**, *88*, 158-165.
[32] Mujtaba, M.; Sargin, I.; Akyuz, L.; Ceter, T.; & Kaya, M.: Newly isolated sporopollenin microcages from Platanus orientalis pollens as a vehicle for controlled drug delivery. *Mater. Sci. Eng. C*. **2017**, *77*, 263-270.
[33] Hall, S. R.; Bolger, H.; Mann, S.: Morphosynthesis of complex inorganic forms using pollen grain templates. *Chem. Comm.* **2003**, *22*, 2784-2785.
[34] Ahmad, N. F.; Kamboh, M. A.; Nodeh, H. R.; Abd Halim, S. N. B.; Mohamad, S.: Synthesis of piperazine functionalized magnetic sporopollenin: a new organic-inorganic hybrid material for the removal of lead (II) and arsenic (III) from aqueous solution. *Environ. Sci. Pollut. Res.* **2017**, *24*(27), 21846-21858.
[35] Hassan, A. M.; Ibrahim, W. A. W.; Bakar, M. B.; Sanagi, M. M.; Sutirman, Z. A.; Nodeh, H. R.; et al.: New effective 3-aminopropyltrimethoxysilane functionalized magnetic sporopollenin-based silica coated graphene oxide adsorbent for removal of Pb (II) from aqueous environment. *J. Environ. Manage*. **2020**, *253*, 1-11.
[36] Aminzadeh, S.; Zhang, L.; Henriksson, G.: A possible explanation for the structural inhomogeneity of lignin in LCC networks. *Wood Sci. Technol.* **2017**, *51*(6), 1365-1376.
[37] Kamide, K.: *Cellulose and cellulose derivatives*. Elsevier, **2005**.
[38] Banoub, J.; Delmas Jr, G. H.; Joly, N.; Mackenzie, G.; Cachet, N.; Benjelloun-Mlayah, B.; et al.: A critique on the structural analysis of lignins and application of novel tandem mass spectrometric strategies to determine lignin sequencing. *J. Mass Spectrom.* **2015**, *50*(1), 5-48.
[39] Heitner, C.; Dimmel, D.; Schmidt, J.: *Lignin and lignans: advances in chemistry*, CRC press, **2016**.
[40] Stevanovic, T.: Chemical Composition and Properties of Wood. In Belgacem, M. N.; Pizzi, A. (Eds.), *Lignocellulosic Fibers and Wood Handbook: Renewable Materials for Today's Environment* (pp 49-110). Scrivener Publishing, Salem, Massachusetts, USA, **2016**.
[41] Vanholme, R.; Demedts, B.; Morreel, K.; Ralph, J.; Boerjan, W.: Lignin biosynthesis and structure. *Plant physiol.* **2010**, *153*(3), 895-905.
[42] Doherty, W. O.; Mousavioun, P.; Fellows, C. M.: Value-adding to cellulosic ethanol: Lignin polymers .*Ind. Crop. Prod.* **2011**, *33*(2), 259-276.

[43] Strassberger, Z.; Tanase, S.; Rothenberg, G.: The pros and cons of lignin valorisation in an integrated biorefinery. *Rsc Adv.* **2014**, *4*(48), 25310-25318.
[44] Vanholme, R.; Morreel, K.; Darrah, C.; Oyarce, P.; Grabber, J. H.; Ralph, J.; et al.: Metabolic engineering of novel lignin in biomass crops. *New Phytol.* **2012**, *196*(4), 978-1000.
[45] Glasser, W.G.: Lignin. In Casey J. P. (Ed.), *Pulp and Paper. Chemistry and Chemical Technology* (pp. 39–111), *Vol.1,* 3rd ed. John Wiley and Sons, New York, **1980**.
[46] Lu, F.; Ralph, J.: Derivatization followed by reductive cleavage (DFRC Method), a new method for lignin analysis: protocol for analysis of DFRC monomers. *J. Agric. Food Chem*. **1997**, *45*(7), 2590-2592.
[47] Guerra, A.; Filpponen, I.; Lucia, L.A.; Saquing, C.; Baumberger, S.; Argyropoulos, D.S.: Toward a better understanding of the lignin isolation process from wood. *J. Agric. Food Chem*. **2006**, *54*(16), 5939-5947.
[48] Perlack, R. D.; Wright, L. L.; Turhollow, A. F.; Graham, R. L.; Stokes, B. J.; Erbach, D. C.: *Biomass as a feedstock for a bioenergy and bioproducts industry: the technical feasibility of a billion-ton annual supply*. Oak Ridge National Laboratory, Tennessee, **2005**.
[49] Li, J.; Henriksson, G.; Gellerstedt, G.: Lignin depolymerization/repolymerization and its critical role for delignification of aspen wood by steam explosion. *Bioresour. Technol.* **2007**, *98*(16), 3061- 3068.
[50] Banoub, J.; Benjelloun-Mlayah, B.; Ziarelli, F.; Joly, N.; Delmas, M.: Elucidation of the complex molecular structure of wheat straw lignin polymer by atmospheric pressure photoionization quadrupole time-of-flight tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **2007**, *21*, 2867-2888.
[51] Banoub, J.H.; Delmas, M.: Structural elucidation of the wheat straw lignin polymer by atmospheric pressure chemical ionization tandem mass spectrometry and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *J. Mass Spectrom.* **2003**,*38*(8):900-903.
[52] Lam, H.Q.; Le Bigot, Y.; Delmas, M.; Avignon, G.: A new procedure for the destructuring of vegetable matter at atmospheric pressure by a catalyst/solvent system of formic acid/acetic acid. Applied to the pulping of triticale straw. *Ind. Crops Prod*. **2001**, *14*, 139-144.
[53] Sun, Z.; Fridrich, B.; de Santi, A.; Elangovan, S.; Barta, K.: Bright side of lignin depolymerization: toward new platform chemicals. *Chem. Rev.* **2018**, *118*(2), 614-678.
[54] Holladay, J. E.; White, J. F.; Bozell, J. J.; Johnson, D.: *Top value-added chemicals from biomass-Volume II—Results of screening for potential candidates from biorefinery lignin* (No. PNNL-16983). Pacific Northwest National Lab. (PNNL), Richland, WA, United States, **2007**.
[55] Li, C.; Zhao, X.; Wang, A.; Huber, G. W.; Zhang, T.: Catalytic transformation of lignin for the production of chemicals and fuels. *Chem. Rev*. **2015**, *115*(21), 11559-11624.
[56] Gasser, C. A.; Hommes, G.; Schäffer, A.; Corvini, P. F. X.: Multicatalysis reactions: new prospects and challenges of biotechnology to valorize lignin. *Appl. Microbiol. Biotechnol*. **2012**, *95*, 1115−1134.

[57] Ahmad, M.; Taylor, C. R.; Pink, D.; Burton, K.; Eastwood, D.; Bending, G. D.; et al.: Development of novel assays for lignin degradation: comparative analysis of bacterial and fungal lignin degraders. *Mol. BioSyst*. **2010**, *6*, 815−821.
[58] De Hoffmann, E.: *Mass spectrometry*. Kirk-Othmer Encyclopedia of Chemical Technology, **2000**.
[59] Aceña, J.; Stampachiacchiere, S.; Pérez, S.; Barceló, D.: Advances in liquid chromatography–high-resolution mass spectrometry for quantitative and qualitative environmental analysis. *Anal. Bioanal. Chem.* **2015**, *407*(21), 6289-6299.
[60] Hu, Q.; Noll, R.G.; Li, H.; Makarov, A.; Hardman, M.; Cooks, R.G.: The Orbitrap: a new mass spectrometer, *J. Mass Spectrom*. **2005**, *40*, 430–443.
[61] Yates III, J. R.: Mass spectrometry: from genomics to proteomics. *Trends Genet.* **2000**, *16*(1), 5-8.
[62] Aebersold, R.; Mann, M.: Mass spectrometry-based proteomics. *Nature*. **2003**, *422*(6928), 198-207.
[63] Dettmer, K.; Aronov, P. A.; Hammock, B. D.: Mass spectrometry-based metabolomics. *Mass spectrom. Rev.* **2007**, *26*(1), 51-78.
[64] Blanksby, S. J.; Mitchell, T. W.: Advances in mass spectrometry for lipidomics. *Ann. Rev. Anal. Chem.* **2010**, *3*, 433-465.
[65] Morreel, K.; Kim, H.; Lu, F.; Dima, O.; Akiyama, T.; Vanholme, R.; et al.: Mass spectrometry-based fragmentation as an identification tool in lignomics. *Anal. Chem.* **2010**, *82*(19), 8095-8105.
[66] Smith, R. M.: *Understanding mass spectra: a basic approach*. John Wiley & Sons, **2004**.
[67] Van Berkel, G. J.: An overview of some recent developments in ionization methods for mass spectrometry. *Eur. J. Mass Spectrom.* **2003**, *9*(6), 539-562.
[68] Portolés, T.; Pitarch, E.; López, F. J.; Hernández, F.; Niessen, W. M.: Use of soft and hard ionization techniques for elucidation of unknown compounds by gas chromatography/time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **2011**, *25*(11), 1589-1599.
[69] Mirsaleh-Kohan, N.; Robertson, W. D.; Compton, R. N.: Electron ionization time-of-flight mass spectrometry: Historical review and current applications. *Mass Spectrom. Rev.* **2008**, *27*(3), 237-285.
[70] Harrison, A. G.: *Chemical ionization mass spectrometry*. CRC press, **1992**.
[71] Cech, N. B.; Enke, C. G.: Practical implications of some recent studies in electrospray ionization fundamentals. *Mass Spectrom. Rev*. **2001**, *20*(6), 362-387.
[72] Raffaelli, A.; Saba, A.: Atmospheric pressure photoionization mass spectrometry. *Mass Spectrom. Rev*. **2003**, *22*(5), 318-331.
[73] Cohen, L. H.; Gusev, A. I.: Small molecule analysis by MALDI mass spectrometry. *Anal. Bioanal. Chem.* **2002**, *373*(7), 571-586.
[74] Jahouh, F.: *Oligonucleotides sequencing and glycation sites determination of neoglycoconjugate vaccines using tandem mass spectrometry* (Doctoral dissertation). Memorial University of Newfoundland, **2012**.

[75] Ho, C. S.; Lam, C. W. K.; Chan, M. H. M.; Cheung, R. C. K.; Law, L. K.; Lit, L. C. W.; et al.: Electrospray ionisation mass spectrometry: principles and clinical applications. *Clin. Biochem. Rev.* **2003**, *24*(1), 3.
[76] Cole, R. B.: Some tenets pertaining to electrospray ionization mass spectrometry. *J. Mass Spectrom.* **2000**, *35*(7), 763-772.
[77] Rayleigh, L.: XX. On the equilibrium of liquid conducting masses charged with electricity. *Lond. Edinb. Dubl. Phil. Mag. J. Sci.* **1882**, *14*(87), 184-186.
[78] Abonnenc, M.; Qiao, L.; Liu, B.; Girault, H. H.: Electrochemical aspects of electrospray and laser desorption/ionization for mass spectrometry. *Ann. Rev. Anal. Chem.* **2010**, *3*, 231-254.
[79] Kim, H. Y.; Salem, N.: Application of thermospray high-performance liquid chromatography/mass spectrometry for the determination of phospholipids and related compounds. *Anal. Chem.* **1987**, *59*(5), 722-726.
[80] Awad, H.; Khamis, M. M.; El-Aneed, A.: Mass spectrometry, review of the basics: ionization. *Appl. Spectrosc. Rev.* **2015**, *50*(2), 158-175.
[81] De Hoffmann, E.; Stroobant, V.: *Mass Spectrometry: Principles and Applications*, 3rd ed. John Wiley and Sons, Chichester, **2007**.
[82] Takino, M.; Daishima, S.; Nakahara, T.: Determination of chloramphenicol residues in fish meats by liquid chromatography–atmospheric pressure photoionization mass spectrometry. *J. Chromatgr. A.* **2003**, *1011*(1-2), 67-75.
[83] Tubaro, M.; Marotta, E.; Seraglia, R.; Traldi, P.: Atmospheric pressure photoionization mechanisms. 2. The case of benzene and toluene. *Rapid Commun. Mass Spectrom.* **2003**, *17*(21), 2423-2429.
[84] Koomen, J. M.; Ruotolo, B. T.; Gillig, K. J.; McLean, J. A.; Russell, D. H.; Kang, M.; et al.: Oligonucleotide analysis with MALDI–ion-mobility–TOFMS. *Anal. Bioanal. Chem.* **2002**, *373*(7), 612-617.
[85] Mechref, Y.; Novotny, M. V.; Krishnan, C.: Structural characterization of oligosaccharides using MALDI-TOF/TOF tandem mass spectrometry. *Anal. Chem.* **2003**, *75*(18), 4895-4903.
[86] Ryan, D. J.; Spraggins, J. M.; Caprioli, R. M.: Protein identification strategies in MALDI imaging mass spectrometry: a brief review. *Curr. Opin. Chem. Biol.* **2019**, *48*, 64-72.
[87] Drake, R. R.; West, C. A.; Mehta, A. S.; Angel, P. M.: Maldi mass spectrometry imaging of n-linked glycans in tissues. In Yamaguchi, Y.; Kato, K. (Eds.), *Glycobiophysics* (pp. 59-76). Springer, Singapore, **2018**.
[88] Leopold, J.; Popkova, Y.; Engel, K. M.; Schiller, J.: Recent developments of useful MALDI matrices for the mass spectrometric characterization of lipids. *Biomolecules*. **2018**, *8*(4), 173.
[89] Kosyakov, D. S.; Anikeenko, E. A.; Ul'yanovskii, N. V.; Khoroshev, O. Y.; Shavrina, I. S.; Gorbova, N. S.: Ionic liquid matrices for MALDI mass spectrometry of lignin. *Anal. Bioanal. Chem.* **2018**, *410*(28), 7429-7439.
[90] Walch, A.; Rauser, S.; Deininger, S.O.; Höfler, H.: MALDI imaging mass spectrometry for direct tissue analysis: a new frontier for molecular histology. *Histochem. Cell Biol*. **2008**, *130*(3), 421-434.

[91] Armstrong, D. W.; Zhang, L. K.; He, L.; Gross, M. L.: Ionic liquids as matrixes for matrix-assisted laser desorption/ionization mass spectrometry. *Anal. Chem.* **2001**, *73*(15), 3679-3686.
[92] Meier, M. A.; Schubert, U. S.: Evaluation of a new multiple-layer spotting technique for matrix-assisted laser desorption/ionization time-of-flight mass spectrometry of synthetic polymers. *Rapid Commun. Mass Spectrom.* **2003**, *17*(7), 713-716.
[93] Vorm, O.; Roepstorff, P.; Mann, M.: Improved resolution and very high sensitivity in MALDI TOF of matrix surfaces made by fast evaporation. *Anal. Chem.* **1994**, *66*(19), 3281-3287.
[94] Gaertner, A. L.; Chow, N. L.; Fryksdale, B. G.; Jedrzejewski, P.; Miller, B. S.; Paech, S.; et al.: Increasing Throughput and Data Quality for Proteomics. In Kamp, R. M.; Calvete, J. J.; Choli-Papadopoulou, T. (Eds.). *Methods in Proteome and Protein Analysis* (pp. 371-397). Springer, Berlin, Heidelberg, **2004**.
[95] Perera, I. K.; Perkins, J.; Kantartzoglou, S.: Spin-coated samples for high resolution matrix-assisted laser desorption/ionization time-of-flight mass spectrometry of large proteins. *Rapid Commun. Mass Spectrom.* **1995**, *9*(2), 180-187.
[96] Karas, M.; Baher, U.; Ingedoh, A.; Nordhoff, E.; Stahl, B.; Strupat, K.; Hillenkamp. F.: Principles and Applications of Matrix-assisted UV-laser Desorption/Ionization Mass Spectrometry. *Anal. Chem. Acta.* **1990**, *241*(2), 175-185.
[97] Beavis, R. C.; Chait. B. T.: Cinnamic acid derivatives as matrices for ultraviolet laser desorption mass spectrometry of proteins. *Rapid Commun. Mass Spectrom.* **1989**, *3*(12),432-435.
[98] Beavis, R. C.; Chait. B. T.: α-Cyano-4-hydroxycinnamic acid as a matrix for matrixassisted
laser desorption mass spectrometry. *Org. Mass Spectrom.* **1992**, *27*(2), 156-158.
[99] Thomas, A.; Charbonneau, J. L.; Fournaise, E.; Chaurand, P.: Sublimation of new matrix candidates for high spatial resolution imaging mass spectrometry of lipids: enhanced information in both positive and negative polarities after 1, 5-diaminonapthalene deposition. *Anal. Chem.* **2012**, *84*(4), 2048-2054.
[100] Knochenmuss, R.; Zenobi, R.: MALDI ionization: the role of in-plume processes. *Chem. Rev.* **2003**, *103*(2), 441-452.
[101] Busch, K. L.: Desorption ionization mass spectrometry. *J. Mass Spectrom*. **1995**, *30*, 233-240.
[102] Laiko, V. V.; Baldwin, M. A.; Burlingame, A. L.: Atmospheric pressure matrix-assisted laser desorption/ionization mass spectrometry. *Anal. Chem.* **2000**, *72*(4), 652-657.
[103] Laiko, V. V.; Taranenko, N. I.; Berkout, V. D.; Musselman, B. D.; Doroshenko, V. M.: Atmospheric pressure laser desorption/ionization on porous silicon. *Rapid Commun. Mass Spectrom.* **2002**, *16*(18), 1737-1742.
[104] Peng, W. P.; Yang, Y. C.; Kang, M. W.; Tzeng, Y. K.; Nie, Z.; Chang, H. C.; et al.: Laser-induced acoustic desorption mass spectrometry of single bioparticles. *Angew. Chem. Int. Ed.* **2006**, *45*(9), 1423-1426.
[105] Mueller, D. R.; Voshol, H.; Waldt, A.; Wiedmann, B.; Van Oostrum, J.: LC-MALDI MS and MS/MS—an efficient tool in proteome analysis. In Bertrand, E; Faupel, M. (Eds.), *Subcellular Proteomics* (pp. 355-380). Springer, Dordrecht, **2007**.

[106] Van der Heide, P.: *Secondary ion mass spectrometry: an introduction to principles and practices*. John Wiley & Sons, **2014**.
[107] Mou, H. Y.; Wu, S.; Fardim, P.: Applications of ToF-SIMS in surface chemistry analysis of lignocellulosic biomass: A review. *BioResources*. **2016**, *11*(2), 5581-5599.
[108] Tolbert, A.; Ragauskas, A. J.: Advances in understanding the surface chemistry of lignocellulosic biomass via time-of-flight secondary ion mass spectrometry. *Energy Sci. Eng.* **2017**, *5*(1), 5-20.
[109] Dawson, P. H. (Ed.): *Quadrupole mass spectrometry and its applications*. Elsevier, **2013**.
[110] Yang, L.; Mann, T. D.; Little, D.; Wu, N.; Clement, R. P.; Rudewicz, P. J.: Evaluation of a four-channel multiplexed electrospray triple quadrupole mass spectrometer for the simultaneous validation of LC/MS/MS methods in four different preclinical matrixes. *Anal. Chem.* **2001**, *73*(8), 1740-1747.
[111] Fiehn, O.; Kopka, J.; Trethewey, R. N.; Willmitzer, L.: Identification of uncommon plant metabolites based on calculation of elemental compositions using gas chromatography and quadrupole mass spectrometry. *Anal. Chem.* **2000**, *72*(15), 3573-3580.
[112] Schmitt-Kopplin, P.; Frommberger, M.: Capillary electrophoresis–mass spectrometry: 15 years of developments and applications. *Electrophoresis*. **2003**, *24*(22-23), 3837-3867.
[113] Herbert, C. G.; Johnstone, R. A.: *Mass spectrometry basics*. CRC press, **2002**.
[114] Henchman, M.; Steel, C.: Understanding the quadrupole mass filter through computer simulation. *J. Chem. Educ.* **1998**, *75*(8), 1049-1054.
[115] Pusch, W.; Kostrzewa, M.: Application of MALDI-TOF mass spectrometry in screening and diagnostic research. *Curr. Pharm. Des.* **2005**, *11*(20), 2577-2591.
[116] Oberacher, H.; Pitterl, F.: On the use of ESI-QqTOF-MS/MS for the comparative sequencing of nucleic acids. *Biopol*ymers. **2009**, *91*(6), 401-409.
[117] Tachon, N.; Jahouh, F.; Delmas, M.; Banoub, J. H.: Structural determination by atmospheric pressure photoionization tandem mass spectrometry of some compounds isolated from the SARA fractions obtained from bitumen. *Rapid Commun. Mass Spectrom.* **2011**, *25*(18), 2657-2671.
[118] Wattenberg, A.; Organ, A. J.; Schneider, K.; Tyldesley, R.; Bordoli, R.; Bateman, R. H.: Sequence dependent fragmentation of peptides generated by MALDI quadrupole time-of-flight (MALDI Q-TOF) mass spectrometry and its implications for protein identification. *J. Am. Soc. Mass Spectrom.* **2002**, *13*(7), 772-783.
[119] March, R. E.: Quadrupole ion trap mass spectrometer. In Meyers, R.A. (Ed.), *Encyclopedia of Analytical Chemistry: Applications, Theory and Instrumentation*. John Wiley and Sons, Ltd, **2012**.
[120] Wong, P. S.; Cooks, R.G.: Ion trap mass spectrometry. *Curr. Sep.* **1997**, *16*, 85-92.
[121] Schmid, D. G.; Grosche, P.; Bandel, H.; Jung, G.: FTICR-mass spectrometry for high-resolution analysis in combinatorial chemistry. *Biotechnol. Bioeng.* **2000**, *71*(2), 149-161.
[122] Feng, X.; Siegel, M. M.: FTICR-MS applications for the structure determination of natural products. *Anal. Bioanal. Chem.* **2007**, *389*(5), 1341-1363.

[123] Lee, S. W.; Berger, S. J.; Martinović, S.; Paša-Tolić, L.; Anderson, G. A.; Shen, Y.; et al.: Direct mass spectrometric analysis of intact proteins of the yeast large ribosomal subunit using capillary LC/FTICR. *Proc. Natl. Acad. Sci*. **2002**, *99*(9), 5942-5947.

[124] Peterman, S. M.; Mulholland, J. J.: A novel approach for identification and characterization of glycoproteins using a hybrid linear ion trap/FT-ICR mass spectrometer. *J. Am. Soc. Mass Spectrom.* **2006**, *17*(2), 168-179.

[125] Muddiman, D. C.; Wunschel, D. S.; Liu, C.; Paša-Tolić, L.; Fox, K. F.; Fox, A.; et al.: Characterization of PCR products from bacilli using electrospray ionization FTICR mass spectrometry. *Anal. Chem.* **1996**. *68*(21), 3705-3712.

[126] Zubarev, R. A.; Makarov, A.: Orbitrap Mass Spectrometry. *Anal. Chem.* **2013**, *85*, 5288−5296.

[127] Kingdon, K. H.: A method for the neutralization of electron space charge by positive ionization at very low gas pressures. *Phys. Rev.* **1923**, *21*, 408–18.

[128] Madeira, P. J. A.; Alves, P.A.; Borges, C. M.: High Resolution Mass Spectrometry Using FTICR and Orbitrap Instruments. In Salih, S.M. (Ed.), *Fourier Transform - Materials Analysis*, InTech, Rijeka, Croatia, **2012**.

[129] Eliuk, S.; Makarov, A.: Evolution of Orbitrap Mass Spectrometry Instrumentation. *Annu. Rev. Anal. Chem.* **2015**, *8*, 61–80.

[130] Griffiths, W. J.; JONSSON, A. P.; LIU, S.; RAI, D. K.; WANG, Y.: Electrospray and tandem mass spectrometry in biochemistry. *Biochem. J.* **2001**, *355*(3), 545-561.

[131] Li, J.; Assmann, S. M.: Mass spectrometry. An essential tool in proteome analysis. *Plant physiol*. **2000**, *123*(3), 807-810.

[132] Hughey, C. A.; Hendrickson, C. L.; Rodgers, R. P.; Marshall, A. G.; Qian, K.: Kendrick Mass Defect Spectrum: A Compact Visual Analysis for Ultrahigh-Resolution Broadband Mass Spectra. *Anal. Chem.* **2001**, *73* (19), 4676-4681

[133] Qi, Y.; Hempelmann, R.; Volmer, D. A.: Two-dimensional mass defect matrix plots for mapping genealogical links in mixtures of lignin depolymerisation products. *Anal. Bioanal. Chem.* **2016**, *408*(18), 4835-4843.

[134] Domínguez, E.; Mercado, J. A.; Quesada, M. A.; Heredia, A.: Pollen sporopollenin: degradation and structural elucidation. *Sex. Plant Reprod.* **1999**, *12*(3), 171-178.

[135] Wehling, K.; Niester, C.; Boon, J. J.; Willemse, M. T. M.; Wiermann, R.: p-Coumaric acid-a monomer in the sporopollenin skeleton. *Planta.* **1989**, *179*(3), 376-380.

[136] Dutta, S.; Hartkopf-Fröder, C.; Wilkes, H.; Greenwood, P.; Littke, R.; Mann, U.: *The molecular composition of sporopollenin from fossil megaspores as revealed by micro-FTIR and pyrolysis-GC-MS.* 23rd International Meeting on Organic Geochemistry, P88-MO. Devon, UK Torquay, **2007**.

[137] Sáiz-Jiménez, C.: Production of alkylbenzenes and alkylnaphthalenes upon pyrolysis of unsaturated fatty acids. *Naturwissenschaften*. **1994**, *81*(10), 451-453.

[138] Bernard, S.; Benzerara, K.; Beyssac, O.; Balan, E.; Brown Jr, G. E.: Evolution of the macromolecular structure of sporopollenin during thermal degradation. *Heliyon*. **2015**, *1*(2), 1-28.

[139] Moore, S. E. M.; Hemsley, A. R.; French, A. N.; Dudley, E.; Newton, R. P.: New insights from MALDI-ToF MS, NMR, and GC-MS: mass spectrometry techniques applied to palynology. *Protoplasma*. **2006**, *228*(1-3), 151-157.
[140] Kim, S. S.; Grienenberger, E.; Lallemand, B.; Colpitts, C. C.; Kim, S. Y.; de Azevedo Souza, C.; et al.: LAP6/POLYKETIDE SYNTHASE A and LAP5/POLYKETIDE SYNTHASE B encode hydroxyalkyl α-pyrone synthases required for pollen development and sporopollenin biosynthesis in Arabidopsis thaliana. *Plant Cell*. **2010**, *22*(12), 4045-4066.
[141] Grienenberger, E.; Kim, S. S.; Lallemand, B.; Geoffroy, P.; Heintz, D.; de Azevedo Souza, C.; et al.: Analysis of TETRAKETIDE α-PYRONE REDUCTASE function in Arabidopsis thaliana reveals a previously unknown, but conserved, biochemical pathway in sporopollenin monomer biosynthesis. *Plant Cell.* **2010**. *22*(12), 4067-4083.
[142] Bocchini, P., Galletti, G. C.; Camarero, S.; Martinez, A. T.: Absolute quantitation of lignin pyrolysis products using an internal standard. *J. Chromatogr. A.* **1997**, *773*(1-2), 227-232.
[143] Ralph, J.; Hatfield, R.D.: Pyrolysis-GC-MS characterization of forage materials. *J. Agr. Food Chem.***1991**, *39*(8), 1426-1437.
[144] Rodrigues, J.; Meier, D.; Faix, O.; Pereira, H.: Determination of tree to tree variation in syringyl/guaiacyl ratio of Eucalyptus globulus wood lignin by analytical pyrolysis. *J. Anal. Appl. Pyrolysis*. **1999**, *48*(2), 121-128.
[145] Sonoda, T.; Ona, T.; Yokoi, H.; Ishida, Y.; Ohtani, H.; Tsuge, S.: Quantitative analysis of detailed lignin monomer composition by pyrolysis-gas chromatography combined with preliminary acetylation of the samples. *Anal. Chem.* **2001**, *73*(22), 5429-5435.
[146] Sheng, H.; Tang, W.; Gao, J.; Riedeman, J. S.; Li, G.; Jarrell, T. M.; et al.: (−) ESI/CAD MS n Procedure for Sequencing Lignin Oligomers Based on a Study of Synthetic Model Compounds with β-O-4 and 5-5 Linkages. *Anal. Chem.* **2017**, *89*(24), 13089-13096.
[147] Andrianova, A.A.; DiProspero, T.; Geib, C.; Smoliakova, I. P.; Kozliak, E. I.; Kubátová, A.: Electrospray ionization with high-resolution mass spectrometry as a tool for lignomics: lignin mass spectrum deconvolution. *J. Am. Soc. Mass Spectrom.* **2018**, *29*(5), 1044-1059.
[148] Prothmann, J.; Spégel, P.; Sandahl, M.; Turner, C.: Identification of lignin oligomers in Kraft lignin using ultra-high-performance liquid chromatography/high-resolution multiple-stage tandem mass spectrometry (UHPLC/HRMS n). *Anal. Bioanal. Chem.* **2018**, *410*(29), 7803-7814.
[149] Prothmann, J.; Li, K.; Hulteberg, C.; Spégel, P.; Sandahl, M.; Turner, C.: Nontargeted Analysis Strategy for the Identification of Phenolic Compounds in Complex Technical Lignin Samples. *ChemSusChem*. **2020**, *13*,1-9.
[150] Kosyakov, D. S.; Ipatova, E. V.; Krutov, S. M.; Ul'yanovskii, N. V.; Pikovskoi, I. I.: Study of products of the alkaline decomposition of hydrolysis lignin by atmospheric pressure photoionization high-resolution mass spectrometry. *J. Anal. Chem.* **2017**, *72*(14), 1396-1403.

[151] Pikovskoi, I. I.; Kosyakov, D. S.; Shavrina, I. S.; Ul'yanovskii, N. V.: Study of Nettle (Urtica dióica) Lignin by Atmospheric Pressure Photoionization Orbitrap Mass Spectrometry. *J. Anal. Chem.* **2019**, *74*(14), 1412-1420.
[152] Kosyakov, D. S.; Ul'yanovskii, N. V.; Anikeenko, E. A.; Gorbova, N. S.: Negative ion mode atmospheric pressure ionization methods in lignin mass spectrometry: A comparative study. *Rapid Commun. Mass Spectrom.* **2016**, *30*(19), 2099-2108.
[153] Kubátová, A.; Andrianova, A. A.; Hatton, J.; Kozliak, E. I.: Atmospheric pressure ionization-mass spectrometry as a tool for structural characterization of lignin. *Rapid Commun. Mass Spectrom.* **2020**, *34*, 1-10.
[154] Terrell, E.; Carré, V.; Dufour, A.; Aubriet, F.; Le Brech, Y.; Garcia-Pérez, M.: Contributions to Lignomics: Stochastic Generation of Oligomeric Lignin Structures for Interpretation of MALDI–FT-ICR-MS Results. *ChemSusChem*. **2020**, doi:10.1002/cssc.202000239.
[155] Bowman, A. S.; Asare, S. O.; Lynn, B. C.: Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry analysis for characterization of lignin oligomers using cationization techniques and 2, 5-dihydroxyacetophenone (DHAP) matrix. *Rapid Commun. Mass Spectrom.* **2019**, *33*(8), 811-819.
[156] Kosyakov, D. S.; Ul'yanovskii, N. V.; Sorokina, E. A.; Gorbova, N. S.: Optimization of sample preparation conditions in the study of lignin by MALDI mass spectrometry. *J. Anal. Chem.* **2014**, *69*(14), 1344-1350.
[157] Evtuguin, D.V.; Amado, F.M.L.: Application of Electrospray Ionization Mass Spectrometry to the Elucidation of the Primary Structure of Lignin. *Macromol. Biosci*. **2003**, *3*, 339-343.
[158] Morreel, K.; Dima, O.; Kim, H.; Lu, F.; Niculaes, C.; Vanholme, R.; et al.: Mass spectrometry-based sequencing of lignin oligomers. *Plant Physiol.* **2010**, *153*(4), 1464-1478.
[159] Qi, Y.; Volmer, D. A.: Chemical diversity of lignin degradation products revealed by matrix-optimized MALDI mass spectrometry. *Anal. Bioanal. Chem.* **2019**, *411*(23), 6031-6037.
[160] Qi, Y.; Fu, P.; Li, S.; Ma, C.; Liu, C.; Volmer, D. A.: Assessment of molecular diversity of lignin products by various ionization techniques and high-resolution mass spectrometry. *Sci. Total Environ.* **2020**, *713*, 136573.
[161] Qi, Y.; Hempelmann, R.; Volmer, D. A.: Shedding light on the structures of lignin compounds: photo-oxidation under artificial UV light and characterization by high resolution mass spectrometry. *Anal. Bioanal. Chem.* **2016**, *408*(28), 8203-8210.
[162] Tolmachev, A.V.; Monroe, M.E.; Jaitly, N.; Petyuk, V.A.; Adkins, J.N.; Smith, R.D.: Mass measurement accuracy in analyses of highly complex mixtures based upon multidimensional recalibration. *Anal. Chem*. **2006**, *78*(24), 8374-8385.
[163] Kind, T.; Fiehn, O.: Metabolomic database annotations via query of elemental compositions: mass accuracy is insufficient even at less than 1 ppm. *BMC bioinform.* **2006**, *7*(1), 1-10.
[164] Kind, T.; Fiehn, O.: Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC bioinform.* **2007**, *8*(1), 1-20.

[165] Jarrell, T.M.; Marcum, C.L.; Sheng, H.; Owen, B.C.; O'Lenick, C.J.; Maraun, H.; et al.: Characterization of organosolv switchgrass lignin by using high performance liquid chromatography/high resolution tandem mass spectrometry using hydroxide-doped negative-ion mode electrospray ionization. *Green Chem*. **2014**, *16*(5), 2713-2727.
[166] Prothmann, J.; Spégel, P.; Sandahl, M.; Turner, C.: Identification of lignin oligomers in Kraft lignin using ultra-high-performance liquid chromatography/high-resolution multiple-stage tandem mass spectrometry (UHPLC/HRMS^n). *Anal. Bioanal. Chem*. **2018**, *410*(29), 7803-7814.
[167] Sleno, L.; Volmer, D.A.; Marshall, A.G.: Assigning product ions from complex MS/MS spectra: the importance of mass uncertainty and resolving power. *J. Am. Soc. Mass Spectrom.* **2005**, *16*(2), 183-198.
[168] Schaeffer-Reiss, C. (2008). A brief summary of the different types of mass spectrometers used in proteomics. In *Functional Proteomics* (pp. 3-16). Humana Press.

Chapter 2

Demystifying and Unravelling the Molecular Structure of the Biopolymer Sporopollenin

This chapter is reproduced with permission from Mikhael, A.; Jurcic, K.; Schneider, C.; Karr, D.; Fisher, G.L.; Fridgen, T.D.; Diego-Taboada, A.; Georghiou, P.E.; Mackenzie, G.; and Banoub, J.: **Demystifying and unravelling the molecular structure of the biopolymer sporopollenin**. *Rapid Commun. Mass Spectrom.* 2020; 34: e8740. https://doi.org/10.1002/rcm.8740

2.1. Abstract

We report the unsolved molecular structure of the complex biopolymer sporopollenin exine extracted from *Lycopodium clavatum* pollen grains. The compiled MS and NMR analyses showed that sporopollenin contained poly(hydroxyacid) dendrimer-like networks with glycerol as a core unit, which accounted for the sporopollenin empirical formula. In addition, these analyses showed that the hydroxy acid monomers forming this network contained a β-diketone moiety. Moreover, MALDI-TOF-MS and MS/MS allowed us to identify a unique macrocyclic oligomeric unit composed of polyhydroxylated tetraketide-like monomers. Lastly, High-resolution X-ray photoelectron spectroscopy showed the absence of aromaticity in sporopollenin.

2.2. Introduction

The shells of plant spores are natural microcapsules, which have evolved and have grown to protect the reproductive pollen of plants from air and light.[1] The shells are formed from two attached layers; the outer shell is called the sporopollenin (exine), which is mainly lipophilic, and the inner shell (intine) is mainly composed of cellulose.[1] Sporopollenin has been described as being "one of the most exceptionally resistant materials known in the organic world."[2] Its composition is defined to be a highly cross-linked polymer, which is composed of carbon, hydrogen, and oxygen atoms.[2] It is highly resilient to chemical degradation, as a result of which its exact chemical structure and biochemical pathways involved in its synthesis are not yet clear.[2]

Sporopollenin has been thought to be composed of varied and complex straight and branched aliphatic chains, some of which are saturated, unsaturated, and polyhydroxylated.[3] Other suggested building blocks appear to involve oxygenated aromatic rings and phenylpropanoid moieties.[2] The cross-linking of these straight- and branched- aliphatic chains was tentatively described as being due to ether and/or carbon-carbon bond formations. [3,4] Gordon Shaw *et al.* [5], one of the earlier pioneers studying the structure of sporopollenin, concluded that the UV properties of sporopollenin were attributed to their similarities to polycarotene. Also, it was proposed by Hayatsu *et al.* [6] that sporopollenin may be composed of a highly cross-linked network of lipids, including fatty acids and/or alcohols, thus contradicting the proposal that sporopollenin exine was a carotenoid derivative.

The main attempts to investigate the structure of sporopollenin can be summarized into three major hypotheses: (a) That the sporopollenin exine can be constructed completely by an aliphatic biopolymer, or (b) that it can be completely built up as an aromatic biopolymer, and finally, (c) that it exists as a mixture of aliphatic and aromatic biopolymer. [7-9]

It is very important to point out that pyrolysis GC-mass spectrometric analysis of different sporopollenins involving heating under inert atmosphere (pyrolysis), coupled with electron ionization mass spectrometry, has consistently yielded *p*-coumaric acid and/or aromatic compounds that apparently, represented the major building block of sporopollenin.[10,11] However, it is well known that pyrolysis can change the structure of the original analyte, especially when such materials contain long aliphatic chains, which can aromatize as reported before for unsaturated fatty acid. [12] Thus, aromatic compounds that originally did not exist can be produced during pyrolysis GC-MS analysis, and even linear saturated polymers such as polyethylene can produce aromatics.[13] Based on those outdated and older studies contingent on pyrolysis GC-EI-MS, there is still a belief that sporopollenin contains aromatic compounds identified either as *p*-coumaric acid and/or, to a lesser extent, ferulic acid.[10, 11, 14, 15].

Bernard *et al.* [16] studied the thermal degradation of sporopollenin and identified that this occurred in two main stages. The first one occurs below 500 ºC, where sporopollenin undergoes simultaneous dehydrogenation and deoxygenation. The second stage at temperatures above 500 ºC is where aromatic products are formed. This study showed that heating substances at high temperatures and/or pyrolysis can produce aromatic

compounds that lead to blatantly misleading errors in the structure elucidation of sporopollenin. Another main detail from the Bernard study [16] is that deoxygenation occurs mainly by the loss of CO_2, supporting the hypothesis of Hayatsu *et al*.[6] who proposed that sporopollenin may be composed of a lipid network of fatty acids.

Recently, the group of Li *et al.*[15] proposed a general molecular structure of the pine sporopollenin using high-energy ball-milling and a newly developed thioacidolysis degradative method together with state-of-the-art solid-state NMR techniques. These authors indicated that the so-called degraded sporopollenin exine derived from the pine pollen grains was primarily composed of aliphatic hydroxylated polyketides, aromatic moieties, and acetal groups as cross-linkers. Despite the fact ball milling has been used extensively for the synthesis of new compounds,[17] it is also well known that ball milling of complex natural materials such as lignin and pollen grains can also lead to alteration of their original structure and may produce new compounds that originally were not part of the original structures.[18] For example, it has been shown that milled wood lignin (MWL) contains more hydroxyl groups than its precursor(s) due to the extensive depolymerization during the ball milling process.[18] The presence of more hydroxyl groups indicated that homolytic bond cleavage occurred between lignin monomers, which in turn produced reactive radicals that can lead to the formation of different new compounds.[18]

The biosynthesis studies aiming to discover the exact molecular structure of sporopollenin have revealed an inkling about the major constituents of sporopollenin. Some studies reveal that polyhydroxylated tetraketides composed of α-pyrone rings with hydroxylated aliphatic chains are important sporopollenin monomers.[19,20] It has been

proposed that these polyhydroxylated ketides, along with fatty alcohols and or fatty acids, could form the sporopollenin exine building blocks [21]. Also, it has been established that sporopollenin absorbs UV radiation in the range of 280-315 nm, which is typically the range of UV absorption of alpha pyrone rings. [22,23]

In this work, we present a different point of view concerning the molecular structure of the sporopollenin exine of *Lycopodium clavatum,* which contrasts with the work of Li *et al.*[15] on the structure of the so-called degraded pine sporopollenin exine. We present herein; a top-down structural elucidation accomplished on *Lycopodium clavatum* sporopollenin.

By using high energy collision TOF-SIMS-MS/MS and MALDI-TOF/TOF-CID-MS/MS, we identified the two major building units in the *Lycopodium clavatum* sporopollenin molecular structure. The first one is the poly(hydroxyacid) dendrimer-like network with glycerol as a core unit containing β-diketone moieties in their chains. This network represents the *Lycopodium clavatum* sporopollenin empirical formula.

The second building unit is the rigid macrocyclic backbone composed of polyhydroxylated tetraketide-like monomers, upon which the poly(hydroxyacid) network is constructed, forming the sporopollenin biopolymer. In addition, 1D- and 2D-solid-state ^{1}H- and ^{13}C-NMR experiments were consistent with the main diagnostic resonances of all H- and C-atoms constituting the proposed structure of sporopollenin exine. Finally, X-ray photoelectron spectroscopy indicates the absence of aromaticity in the *Lycopodium clavatum* sporopollenin.

2.3. Experimental

2.3.1. Sporopollenin extraction

The sporopollenin exine used in this manuscript was provided by Prof. Mackenzie's group from Hull University, UK. It was prepared as described: *L. clavatum L.* (club moss) spores were purchased from Tibrewala International (Nepal), acetone from Aldrich UK, and potassium hydroxide, ethanol, orthophosphoric acid, hydrochloric acid, and sodium hydroxide from Fisher Scientific UK Ltd. Sporopollenin exine capsules (SECs) were extracted from *L. clavatum L.* spores as follows. Spores (300 g) were stirred in acetone (900 mL) under reflux for 4 h, filtered, and dried overnight in open air. They were stirred under reflux for 12 h in an aqueous solution of potassium hydroxide (54 g in 900 mL), the solution being renewed after 6 h, filtered, washed with hot water (5 x 300 mL) and hot ethanol (5 x 300 mL), and dried overnight in open air. The particles were stirred under reflux for 7 days in orthophosphoric acid (900 mL), filtered, washed with water (5 x 300 mL), acetone, 2 mol/L hydrochloric acid, 2 mol/L sodium hydroxide (each 300 mL), water (5 x300 mL), acetone and ethanol (each 300 mL), and dried at 60 °C until constant weight (90 g which account for mass loss of 70% of the total mass of spores). Elemental analysis of the sporopollenin (g/100 g) was carbon 68.90, hydrogen 7.90, nitrogen 0.00, as determined on a Fisons Instruments Carlo Erba EA 100 C H N S analyzer.[24]

2.3.2. High-Resolution X-ray Photoelectron Spectrometry (HR-XPS).

XPS surface analysis was carried out using a PHI *VersaProbe* III instrument (Physical Electronics, Minnesota, USA) equipped with a monochromated Al K_α x-ray

source (hν = 1486.6 eV) and dual-beam charge neutralization comprised of low energy electrons (≤ 25 eV) and low energy Ar^+ ions (≤ 10 eV). The power of the x-ray beam was set to 24.6 W with a 100 μm beam diameter, and all analyses were performed with a 45^o take-off angle. The hemispherical analyzer was set to a pass energy of 224 eV for survey scans, and 26.0 eV for high resolution scans with 1.0 eV/step and 0.1 eV/step, respectively; the energy scale was calibrated with reference to the C 1s peak at a binding energy (BE) of 284.8 eV (C–C, C–H). Peak fittings were performed using a Shirley-type background and a Gaussian-Lorentzian peak fit function with 85% Gaussian line shape. The high-resolution spectra were smoothed using a 3-point Savitzky Golay function. The chemical species corresponding to each binding energy were attributed using standard materials and the PHI XPS Handbook; quantification was accomplished using Scofield sensitivity factors. Data acquisition was accomplished using PHI SmartSoft-XPS software, and data processing was performed using PHI MultiPak (Physical Electronics, Minnesota, USA) software.

2.3.3. TOF-SIMS and High-Energy CID-MS/MS

TOF-SIMS tandem MS imaging analysis was performed using a PHI nanoTOF II Parallel Imaging MS/MS instrument (Physical Electronics, Minnesota, USA). A detailed description of this TOF-TOF instrument has been reported previously, including the attained spatial resolution, monoisotopic precursor selection, and kilo electron volt collision-induced dissociation (keV-CID) of the selected precursor ions to generate the tandem MS product ion spectra.[25-27] The elements and qualities of the tandem MS spectra for composition analysis and structure elucidation have been further elaborated in other reports.[28,29] In the present study, all the spectra and images were recorded with

electrodynamically bunched pulses of a 30 keV Bi3+ primary ion beam of which the DC current was measured to be ≈ 9 nA. The ion beam was operated in the HR2 mode to achieve < 500 nm lateral resolution at high mass resolving power and without the use of delayed extraction. The field-of-view of each analytical area was either 400 µm × 400 µm or 200 µm × 200 µm divided by 256 × 256 image pixels. TOF-SIMS tandem MS imaging data were acquired from the samples, held nominally at room temperature, over a range of *m/z* 0-2,000. The samples were prepared by placing spore powder onto double-sided tape, which was lightly tapped to remove excess powder, and then blown with clean and dry nitrogen gas to remove loosely bound spores. While potential surface contamination moieties were observed, no pre-sputter was used. The ion fluence for each tandem MS analysis at *m/z* 253, 255, and 281 (≤ 2.50 × 1012 Bi3+/cm2) was below the static limit of analysis. While the cumulative dose in this analysis area exceeded the static limit, there was no observed degradation of the precursor ion signal. The ion fluence for each tandem MS analysis at *m/z* 575 and 603 was 1.41 × 1013 Bi3+/cm2 and 1.99 × 1013 Bi3+/cm2, respectively. Degradation of the precursor ion signals was not observed over the course of both analyses at the same location of the sample. During analysis, low energy electrons (≤ 25 eV) and low energy Ar+ ions (≤ 10 eV) were applied for charge compensation. Data acquisition was accomplished using PHI SmartSoft-TOF software, and data processing was performed using PHI TOF-DR (Physical Electronics, MN) software.

2.3.4. MALDI-TOF-MS and High-Energy CID (1KeV) MS/MS.

Prior to MALDI mass spectrometric analysis, the Sporopollenin sample was washed three times using LCMS water (Millipore Sigma) by mixing at 1800 rpm for 45 minutes, followed by the removal of the supernatant. MALDI matrices α–cyano–4–hydroxycinnamic acid (CHCA), 2-(4-Hydroxyphenylazo)benzoic acid (HABA), and 1,5-Diaminonaphthalene (DAN) (all obtained from Sigma-Aldrich) were prepared as 10 g/L solutions in ethanol: acetonitrile at 1:1 ratio with or without 0.1% trifluoroacetic acid (TFA). Each matrix was further mixed with water at 8:1 ratio by volume (matrix: water) to reduce the spreading of the matrix on the hydrophobic plate surface and thus to obtain a thicker matrix layer. The sandwich method was used to spot the matrix and sample onto the MALDI plate. Here, a layer of the matrix was deposited first (0.5uL) followed by the layer of the sample (0.5uL). Lastly, the second layer of the matrix (0.5uL) was deposited. Mass Spectrometric data were obtained using an AB Sciex 5800 MALDI TOF/TOF System (Framingham, MA, USA). Data acquisition and data processing were respectively done using a TOF TOF Series Explorer and Data Explorer (both from AB Sciex). The 5800 TOF/TOF instrument is equipped with a 349 nm Nd: YLF OptiBeam On-Axis laser, and the laser pulse rate was 400 Hz. Reflectron positive and negative modes were used for MS acquisitions. Reflectron and MSMS modes were externally calibrated at 50 ppm mass tolerance. Each MS mass spectrum was collected as a sum of 1000 laser shots, while MSMS mass spectra were obtained as a sum of 1500 shots.

Taking advantage of the in-source decay, additional MSMS spectra, described as *quasi*-MS^3 spectra (Section 2.1), were acquired for certain product ions identified in

MS/MS. Here, product ions identified in the initial MS/MS, were selected as precursor ions and their *quasi*-MS^3 were acquired, to produce the maximal fragmentation coverage allowed by the mass spectrometer which was used.

2.3.5. Solid-State ^{1}H- and ^{13}C-NMR; ^{1}H-^{1}H NOESY 2D and 2D ^{1}H-^{13}C HSQC

The NMR spectra were obtained at 298 K on a Bruker Avance II 600 spectrometer, equipped with an SB Bruker 3.2mm MAS triple-tuned probe operating at 600.33MHz for ^{1}H and 150.97MHz for ^{13}C. Chemical shifts were referenced to tetramethylsilane (TMS) using adamantane as an intermediate standard for ^{13}C. Spinning rates are indicated within the figures. ^{1}H spectra were recorded with a regular 90 pulse (zg) as well as with CPMG filter to separate the crystal-like structure from the amorphous signal. $^{13}C\{^{1}H\}$ cross-polarization (CPMAS) spectra were collected with a Hartmann-Hahn match at 62.5 kHz and 100 kHz ^{1}H decoupling, with a contact time of 2ms, a recycle delay of 2s and 15k scans. To obtain quantitative data within a reasonable time frame, a multiCP pulse sequence was used,[30] with t_z=0.5s and p5=200ms, ns=15k. ^{1}H-^{1}H NOESY 2D spectra were recorded for various mixing time (t_{mix}=10, 50, 100, and 200ms) at a low spinning rate v_r=5kHz with ns=8 scans. 2D ^{1}H-^{13}C HSQC were recorded at v_r=23kHz (sw=152ppm) for J=125Hz and J=170Hz with ns=32 scans. Processing, peaks deconvolution, and integration, as well as spectrum prediction, were all performed using MestReNova software (Santiago de Compostela, Spain).

2.4. Results and Discussion

The *Lycopodium clavatum* sporopollenin used in this investigation was extracted by the classical method of Zetzsche and Kälin. [31] using hot acetone, potassium hydroxide, and phosphoric acid. The hot acetone is used to remove lipid content on the *L. clavatum* spore surface and within the cytoplasm, the potassium hydroxide is used to hydrolyse and remove nitrogenous components, whilest the role of the phosphoric acid is to degrade and remove polysaccharide materials, thus leaving only the sporopollenin remaining.[31] This series of sequential treatments are known as a successful method to extract the protein-free hollow intact and clean *lycopodium clavatum* sporopollenin exine microcapsule. [32] It should be noted that these resulting clean and hollow sporopollenin microcapsule exines have been used previously for the encapsulation of a wide variety of compounds, including both polar (e.g. drugs, dyes, proteins, carbohydrates, and oligonucleotides) and non-polar products (e.g. oils and waxes). [33]

Based on the knowledge that sporopollenin is extremely stable and insoluble in any solvent, we were confronted with a difficult choice for MS analysis. Considering the insolubility of sporopollenin, ESI and APCI techniques were excluded. For this reason, we decided to use the more energetic TOF-SIMS secondary ions to bombard sporopollenin, aiming to discover any ions characteristic for the sporopollenin exine.

2.4.1. Identification of the poly(hydroxy acid) network building unit representing the empirical formula of *lycopodium clavatum* sporopollenin exine.

2.4.1.1. TOF-SIMS and CID MS/MS

To target the sporopollenin wall the energetic secondary ions of TOF-SIMS were employed which result in the formation of characteristic ions in both the negative and the positive ion mode TOF-SIMS. Some ions were selected for high energy (Kev) CID-MS/MS which resulted in breaking every C-C bond of the precursor ions thus providing structurally relevant information.[25] The TOF-SIMS tandem MS imaging (+ ion mode) of the sporopollenin showed the presence of ions at *m/z* 575 and 603. These ions are characteristic for diacylglycerol (DAG) derivatives and were assigned as $[C_{16}H_{25}O_4C_3H_4C_{14}H_{21}O_4+H]^+$ and $[C_{16}H_{25}O_4C_3H_4C_{16}H_{25}O_4+H]^+$, respectively (Figure A1.1a and Figure A1.2).

The product ion scans of the precursor ions at *m/z* 575 and 603 show the formation of the characteristic acylium ions at *m/z* 237 and 265 assigned as $[C_{14}H_{21}O_3]^+$ and $[C_{16}H_{25}O_3]^+$, respectively (Figures A1.4 and Scheme A1.1 and A1.2). Each one of the resulting acylium product ions eliminated only one water molecule, indicating that one oxygen is present on their chains as a hydroxyl group, while the remaining oxygen exists most probably as a keto group that cannot lead to the loss of a second molecule of water. The carboxylate anions corresponding to these acylium ions were detected in the negative ion mode TOF-SIMS-MS at *m/z* 253 and 281 assigned as $[C_{14}H_{22}O_4\text{-}H]^-$ and $[C_{16}H_{26}O_4\text{-}H]^-$, respectively (Figure A1.1b and Figure A1.3). The product ion scan of these carboxylate ions showed the loss of one water molecule confirming the presence of a

hydroxyl group in the chain of these acids, as described above (Figures A1.5 and Scheme A1.3).

A proposed fragmentation mechanism for these carboxylate anions at *m/z* 253 and 281 is shown in Scheme A1.3. Overall, the fragmentation mechanism shown in Scheme A1.1 suggests that these carboxylic acids contain a β diketone moiety, in which, one oxygen exists in its enol form (which is lost as a water molecule), whereas the other one exists in a keto form (which cannot be lost as a water molecule). This structural pattern is therefore supportive for the presence of a β-diketone structure (Figure A1.6 and A1.7), as it facilitates the formation of an intramolecular hydrogen bond that gives extra stability to these types of structures.[34] This hydrogen bonding resembles the intramolecular hydrogen bond seen with the β-diketone moiety in tetrahydrocurcumin (THC) molecules.[35] Also, the loss of one water molecule in the product ion scans of the precursor ions $[C_{14}H_{22}O_4\text{-}H]^-$ and $[C_{16}H_{26}O_4\text{-}H]^-$ at *m/z* 253 and 281 resemble the fragmentation pattern reported for the diketone THC.[35] Lastly, images showing the chemical distribution of the ions at *m/z* 253 and 281 (-ve ion mode), and ions at *m/z* 575 and 603 (+ve ion mode) are shown in Figure A1.8.

It should be noted that these identified fatty acids contain an extra terminal hydroxyl group in their original structures, which allowed them to be attached together through ester bond formation producing the poly(hydroxyacid) network of sporopollenin exine . This fact was deduced, after performing the MALDI-TOF-MS (+ ion mode using CHCA as a matrix) of sporopollenin, where an ion at *m/z* 1643.9948 was identified and assigned as $[C_{89}H_{142}O_{27}+H]^+$. This ion confirms that each of the fatty acids forming the

poly(hydroxyacid) network contain three oxygens in their chains, where two oxygen atoms are present in the β-diketone moiety and one oxygen atom is present as a terminal hydroxyl group) (see section 3.1.3, Figure 2.2, Figure 2.3 and Table 2.1). The proposed chemical structures of all other ions formed in the positive or negative ion mode TOF-SIMS-MS are shown in Figure A1.2 and Figure A1.3).

2.4.1.2. MALDI-TOF-MS (+ ion mode) using DAN as a matrix and CID-MS/MS

After revealing the presence of fatty acids using TOF-SIMS, the extracted sporopollenin exine was analyzed by MALDI-TOF-MS (+ ion mode) using DAN as a matrix. The use of DAN as a matrix was chosen to enhance the discovery of poly(hydroxyacid) moieties as it is an excellent matrix for the identification of lipids.[36] The MALDI spectrum of sporopollenin using DAN as a matrix in the range of *m/z* 1300-2000 showed a complex series of ions that were very closely related, indicating the presence of a heterogeneous mixture (Figure A1.9).

Furthermore, when measuring the Kendrick mass defect plot, we noticed the formation of a bundle of ions that were strictly related to each other's (Figure A1.10). [37] This Kendrick mass defect plot displayed a series of ions that, in general, varied in the number of methylene groups (14 Da). This is demonstrated in the expanded part of the MALDI MS in the range of *m/z* 1300-1500 Da (Figure A1.9).

Based on the chemical structures of the fatty acid and /or DAG ions that were identified using the TOF-SIMS-MS/MS, we proposed the chemical structures for some selected ions in the Sporopollenin MALDI-MS spectrum using DAN as a matrix.

As an example, we assigned the radical cation at *m/z* 1965.1278 as $[C_{106}H_{164}O_{33}]^{+\bullet}$, which contained two triglycerides (six C14 fatty acids plus two glycerol units). These two triglycerides are connected by an extra spacer (C16 fatty acid) (Figure 2.1). This latter ion at *m/z* 1965.1278 has a structure that appears like a branching unit in triacylglycerol (TAG) dendrimers.[38] Also, it should be noted that the formula of this ion at *m/z* 1965.1278 fits very well with the empirical formula $[C_{90}H_{142}O_{27}]$ of *Lycopodium clavatum* sporopollenin as indicated by the C/H, C/O and H/O ratios shown in Table A1.1. As another example, we have assigned the ion at *m/z* 1441.8037 as $[C_{78}H_{120}O_{24}+H]^{+}$, and its molecular structure is supported by its proposed CID-MS/MS fragmentation pathways, as shown in Figure A1.12 and Scheme 2.1.

In order to propose a chemical formula with reasonable mass accuracy for this ion, we proposed that this ion contained one alpha-pyrone ring plus a hydroxylated chain in its chemical structure (Scheme 2.1). The identification of the alpha pyrone ring with a hydroxylated aliphatic chain in the sporopollenin structure will be discussed later in section 2.4.2.

Also, we proposed the presence of the pyrone ring and the hydroxylated chain in other selected ions in the MALDI-MS of sporopollenin using DAN as a matrix such as *m/z* 1302.7322, *m/z* 1315.7599, *m/z* 1328.7941, *m/z* 1343.7914, *m/z* 1357.7900, *m/z* 1371.7622, *m/z* 1412.8228, *m/z* 1427.7810, *m/z* 1498.8486, *m/z* 1511.8441 for the purpose to have a good mass accuracy within 25 ppm. (Figure A1.11 and Table A1.1).

Needless to say, further MS/MS studies are needed to confirm the assigned chemical formulae and/or chemical structures of these ions. Lastly, it should be noted the

MALDI-MS of sporopollenin using DAN showed a mixture of protonated ions and radical cation as DAN is known to produce radical cations of the target analyte. [39]

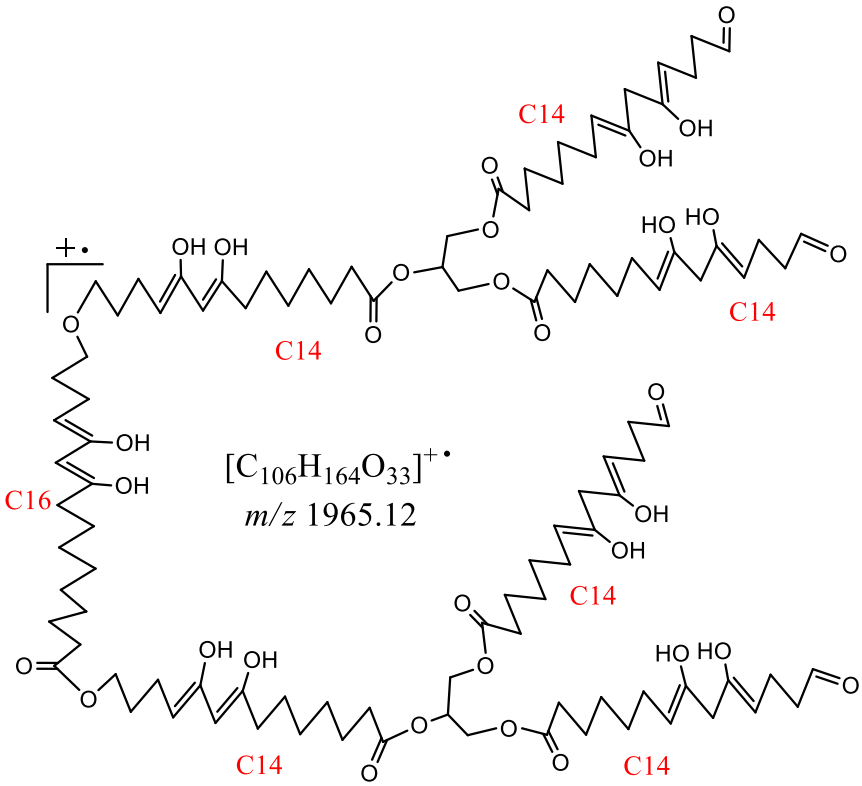
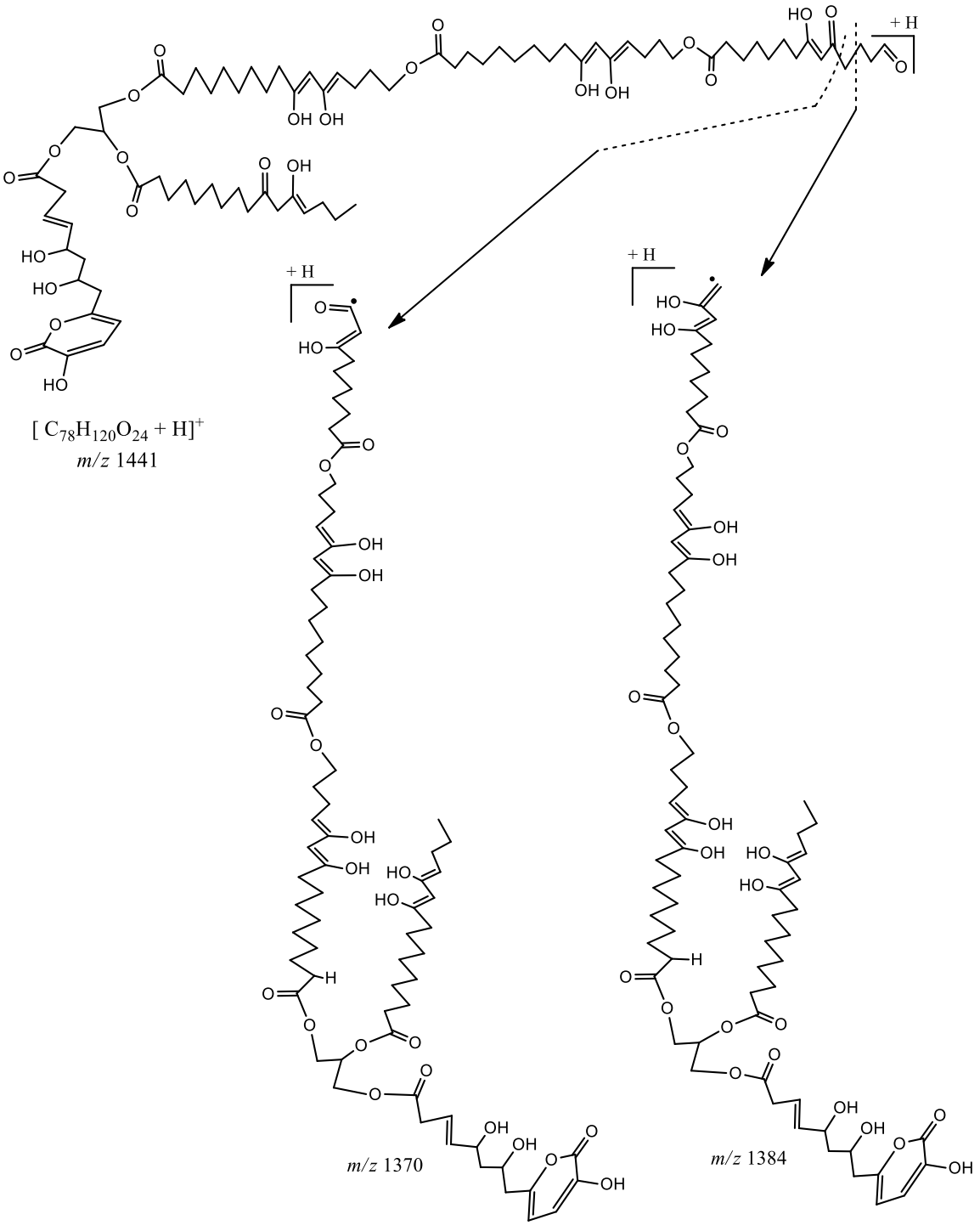
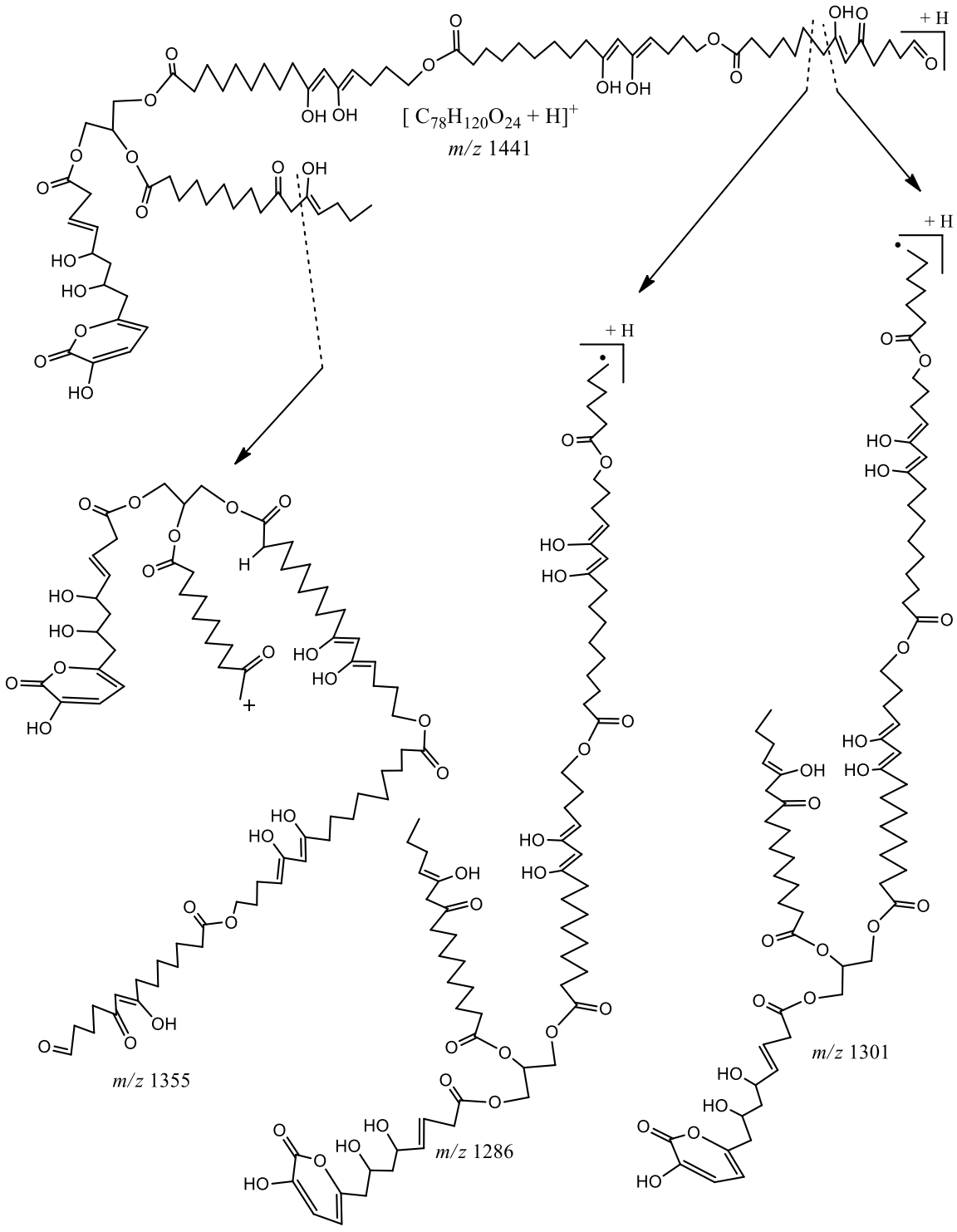


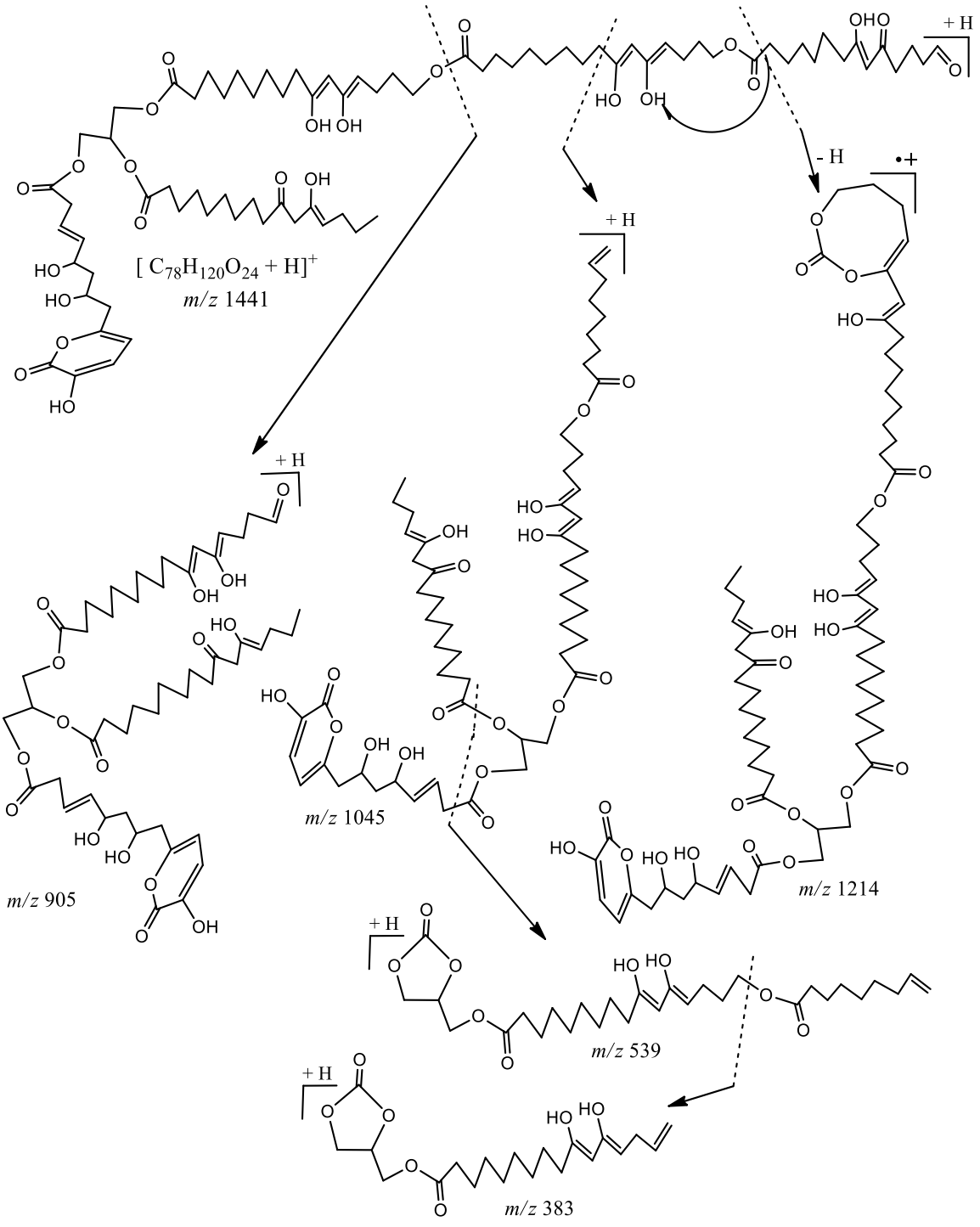
Figure 2.1. The proposed structure of the poly(hydroxyacid) with glycerol as a core unit at *m/z* 1965 identified in MALDI-TOF-MS using DAN as a matrix



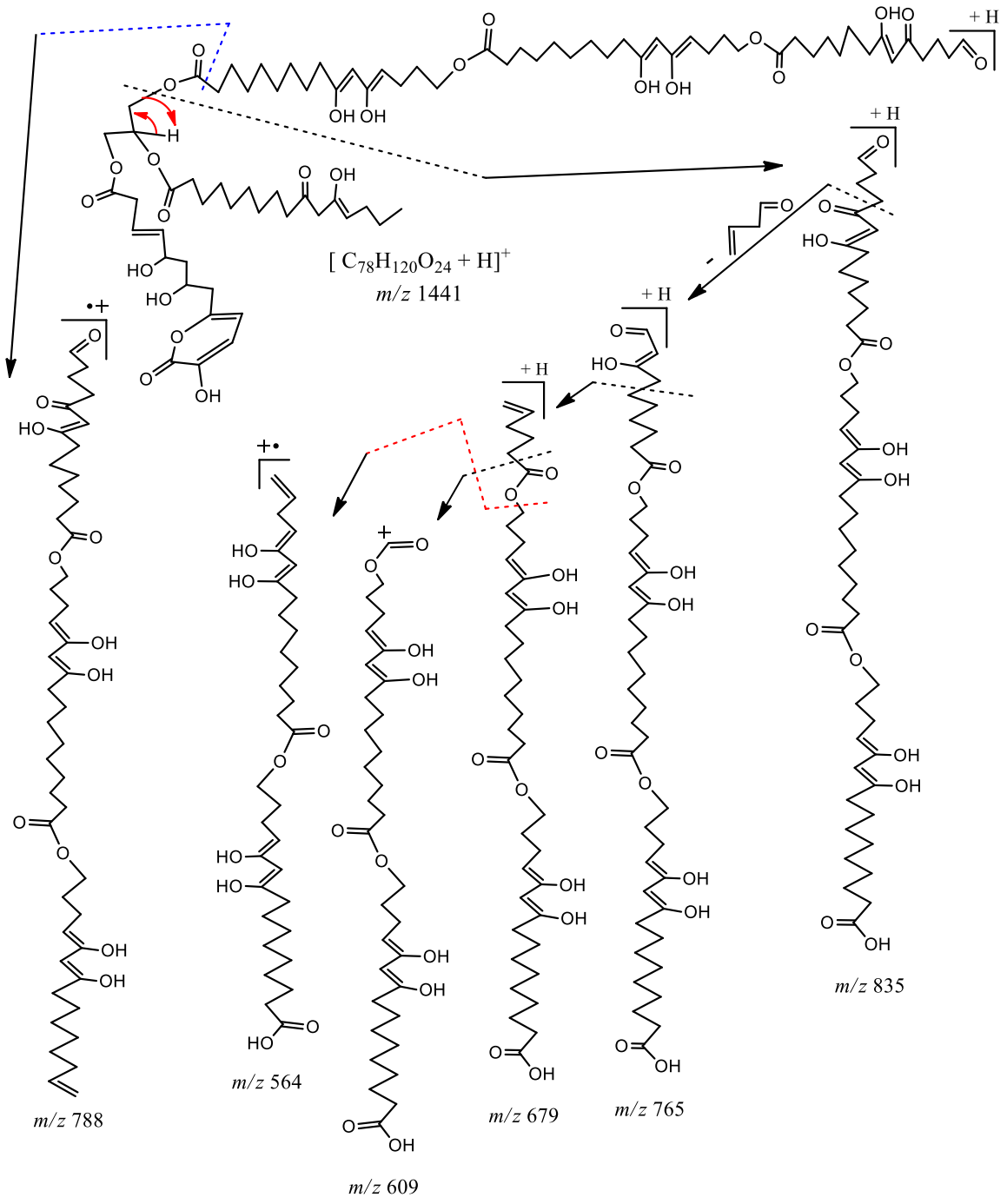
Scheme 2.1a. Fragmentation mechanism for the product ion scan of the precursor ion at *m/z* 1441.93 identified in the sporopollenin exine MALDI-TOF-MS/MS using DAN as a matrix.



Scheme 2.1b. Fragmentation mechanism for the product ion scan of the precursor ion at *m/z* 1441.93 identified in sporopollenin exine MALDI-TOF-MS/MS using DAN as a matrix



Scheme 2.1c. Fragmentation mechanism for the product ion scan of the precursor ion at *m/z* 1441.93 identified in sporopollenin exine MALDI-TOF-MS/MS using DAN as a matrix



Scheme 2.1d. Fragmentation mechanism for the product ion scan of the precursor ion at *m/z* 1441.93 identified in sporopollenin exine MALDI-TOF-MS/MS using DAN as a matrix

2.4.1.3. MALDI-TOF-MS (+ ion mode) using CHCA as a matrix

The sporopollenin MALDI-TOF-MS using CHCA as a matrix showed the presence of a heterogeneous mixture of higher molecular weight poly(hydroxyacid) ions than those obtained using DAN as the matrix (Figure 2.2). However, it should be noted that the mass region around *m/z* 100-1300 was mainly composed of CHCA matrix peaks,[40] whereas the region from *m/z* 1500 to 2500 Da was composed of significant analyte ions.

In the expanded region of the spectra from *m/z* 1500 to 1700, a series of ions that differ by C_2H_4 (2 x CH_2) in their structures (Figure **2.2** and Table 2.1) can be seen. The most important one is the ion at *m/z* 1643.9948 (previously described in the TOF-SIMS Section 3.1.1), which was assigned as $[C_{89}H_{142}O_{27}+H]^+$ .This ion is composed of a diglyceride containing six fatty acids (five C_{14} and one C_{16}) plus the glycerol moiety (Figure 2.3 and Table 2.1). This ion was identified with good accuracy (+ 8.0 error ppm), and its chemical formula fits the experimental isotopic distribution pattern with 71.2%. This assignment corresponds very well with the empirical formula of *Lycopodium clavatum* sporopollenin ($C_{90}H_{142}O_{27}$) calculated by Zetzsche *et al.* [31] and whose work was the most useful information that helped us in our own interpretation of all of the data reported in this contribution.

It is also interesting to mention that the region in the range of *m/z* 2300 to 2500 showed a series of ions which differed by C_2H_4 units in their structure. As an example from this region, the ion at *m/z* 2428.4433 (TAG) was assigned as $[C_{133}H_{207}O_{39}]^+$, which consists of nine fatty acids (seven C_{14} and two C_{16} fatty acids) plus glycerol as a core unit (Figure 2.3 and Table **2.**1).

The chemical formulas of other selected ions in the MALDI-TOF- MS of sporopollenin using CHCA as a matrix are shown in Table 2.1. It should be noted that, all ions chemical formula's C/H, C/O and H/O ratios are very close to the C/H, C/O and H/O ratios of the sporopollenin empirical formula (Table2.1).

The proposed structures and/or chemical formulas assigned to the ions in the MALDI-TOF-MS of sporopollenin using CHCA as a matrix (Figure 2.2 and Table 2.1) provide further support for the presence of poly(hydroxyacid) networks with glycerol as a core unit and importantly, the ion at *m/z* 1643.9948 reveal the structure of the smallest unit (empirical formula) of *Lycopodium clavatum* sporopollenin.

With this new finding, clear evidence was finally obtained to support and characterize the ions which corresponded to the empirical formula $C_{90}H_{142}O_{27}$ of sporopollenin extracted from *Lycopodium clavatum*. Lastly, It should be noted that the ions in the MALDI-MS of sporopollenin using CHCA as a matrix showed a variety of protonated ions, molecular ions and deprotonated ions [M]+ or $[M-H]^+$ (Table 2.1), which is not surprising based on the complexity of the sporopollenin structure. [41]

To summarize Sections 1.1, 1.2 , and 1.3, a series of branched poly(hydroxyacids) with glycerol as a core unit has been identified by using TOF-SIMS- and MALDI-TOF/TOF-MS/MS. The poly(hydroxyacids) are composed mainly of various combinations of C16 and C14 hydroxyacids, each having three oxygen atoms on their chains. One oxygen atom exist as a terminal hydroxyl group, and the other two oxygen atoms are presented in the β-diketone moiety of the fatty acids monomers. These β-diketone moieties could account for some of the antioxidant activity of sporopollenin exine. [2,42] The

structures of these branched poly(hydroxyacid) provide the first-ever report in the literature, showing the existence of a natural "dendrimer"-like molecule.

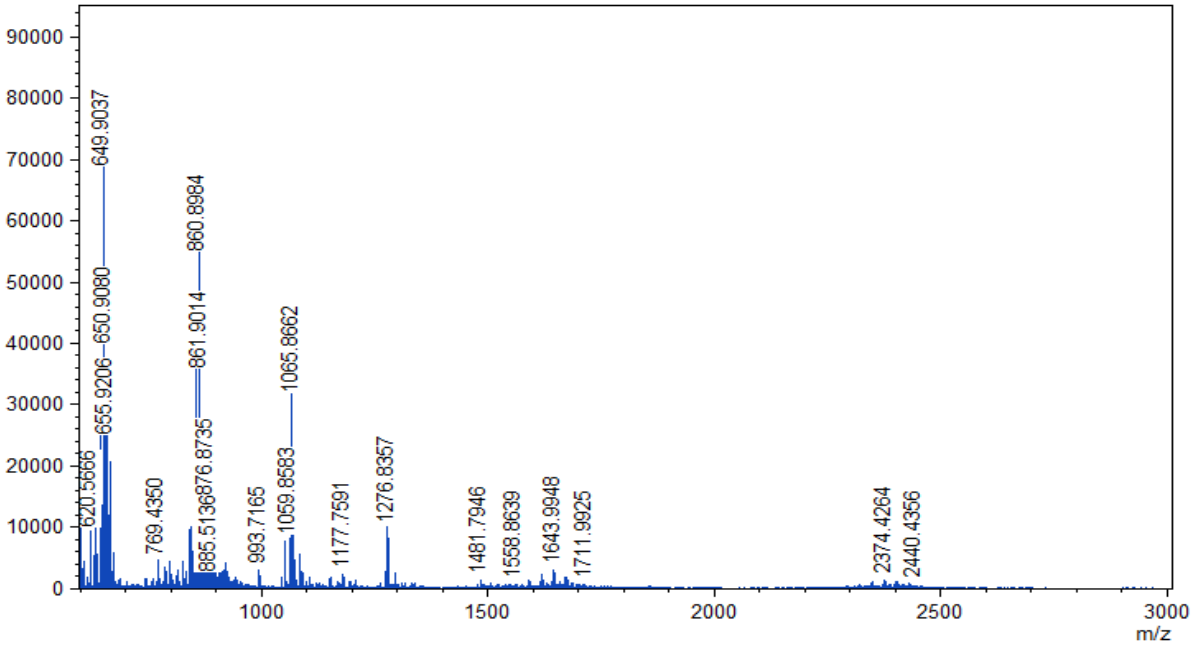


Figure 2.2a. MALDI-TOF-MS of sporopollenin exine using CHCA as a matrix

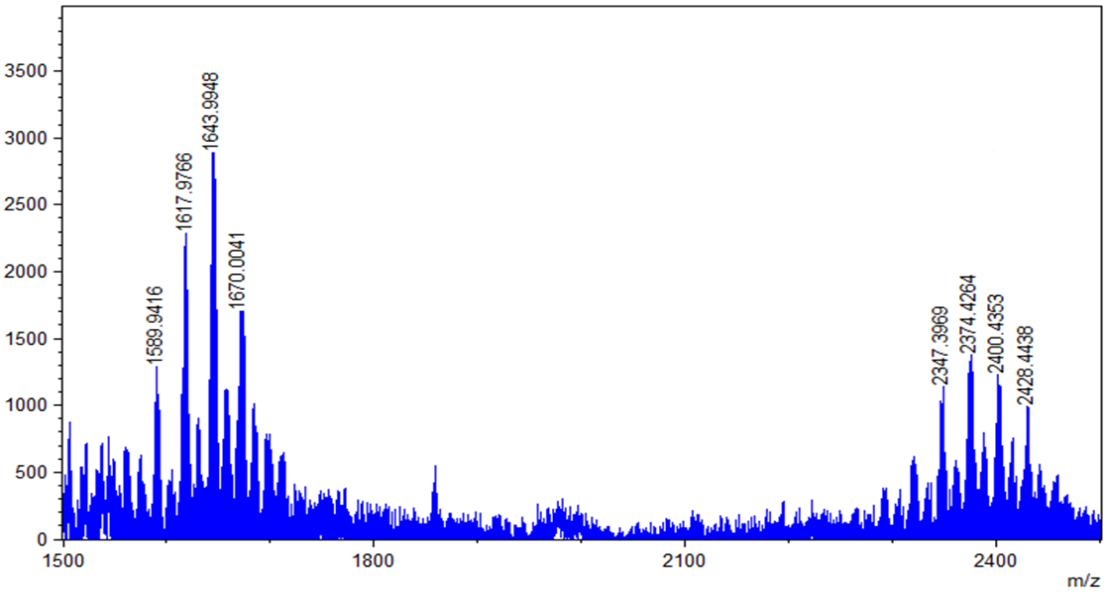


Figure 2.2b. Expansion of the sporopollenin exine MALDI-TOF-MS in the high mass region at *m/z* 1500-3000.

Table 2.1. Proposed chemical formulas for some selected ions in the MALDI-TOF-MS using CHCA as a matrix. All ions showed close fits with the C/H (0.63), C/O (3.33) and H/O (5.26) ratios of the empirical formula of *Lycopodium clavatum* sporopollenin.

Experimental *m/z*	**Assigned chemical formula**	**Error ppm**	**C/H (0.63)**	**C/O (3.33)**	**H/O (5.26)**
1589.9416	$[C_{85}H_{136}O_{27} + H]^+$	4.3	0.62	3.15	5.04
1617.9766	$[C_{87}H_{140}O_{27} + H]^+$	6.5	0.62	3.22	5.18
1643.9948	$[C_{89}H_{142}O_{27} + H]^+$	8.0	0.63	3.30	5.26
1670.0041	$[C_{91}H_{145}O_{27}]^{+\bullet}$	4.1	0.63	3.37	5.37
2347.3969	$[C_{127}H_{197}O_{39} + H]^+$	19.5	0.64	3.26	5.05
2374.4264	$[C_{129}H_{202}O_{39} - H]^+$	21.8	0.64	3.31	5.15
2400.4353	$[C_{131}H_{204}O_{39} - H]^+$	18.8	0.64	3.36	5.20
2428.4438	$[C_{133}H_{208}O_{39} - H]^+$	9.2	0.64	3.41	3.41

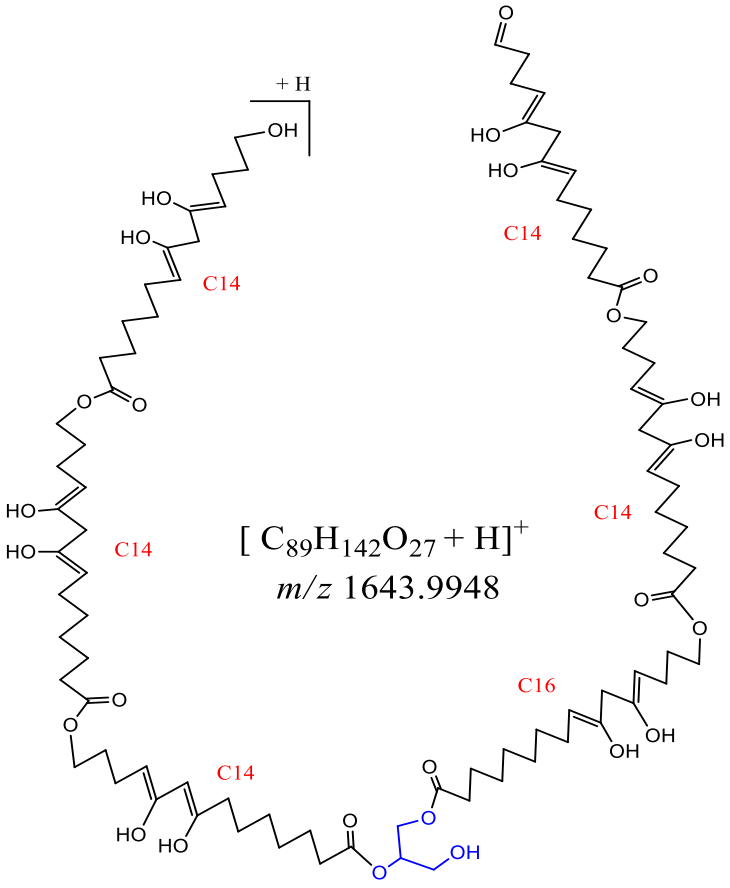


Figure 2.3a. The proposed structure of *m/z* 1643.9948 identified in sporopollenin MALDI-TOF-MS using CHCA as a matrix

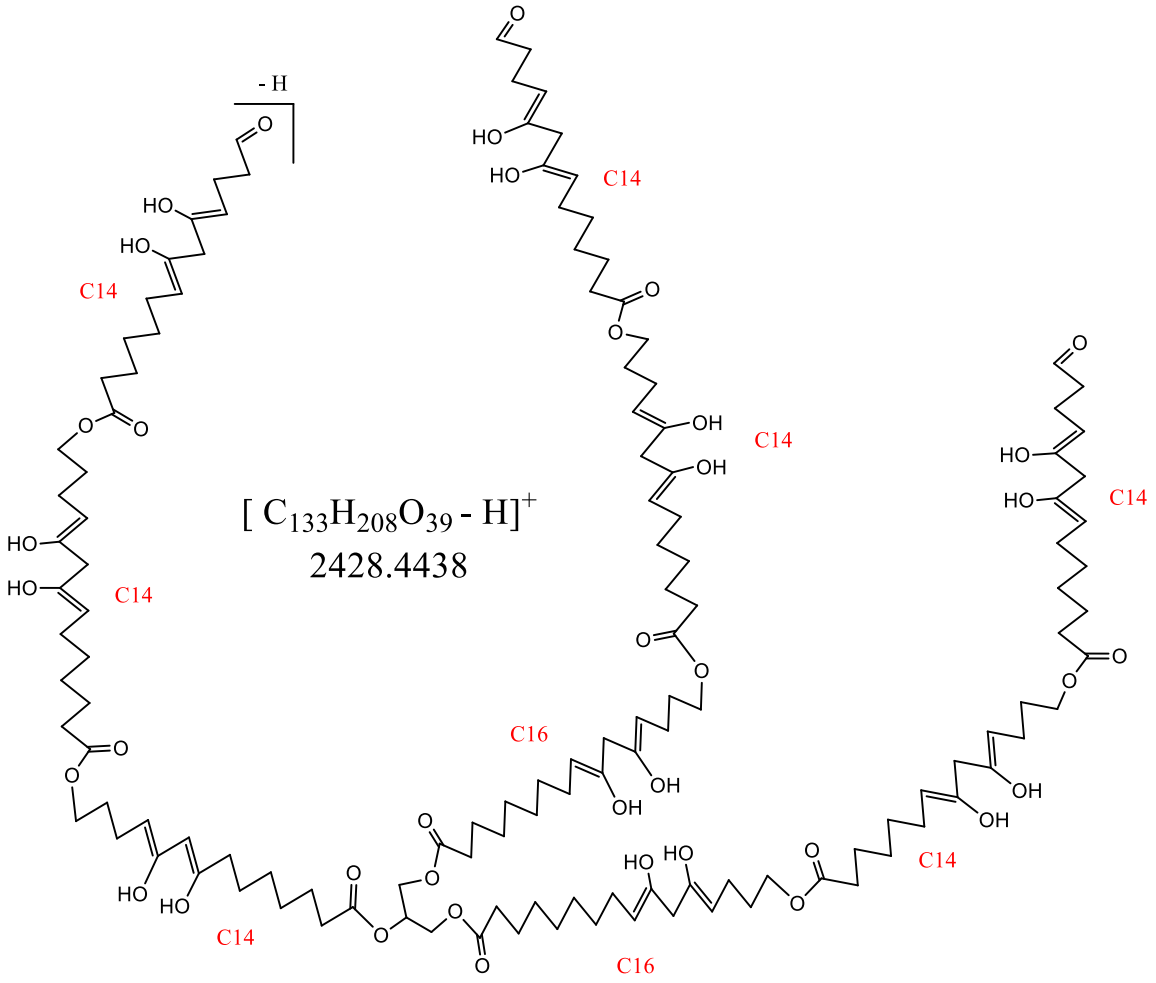


Figure 2.3b. The proposed structure of *m/z* 2428.4438 identified in sporopollenin MALDI- TOF-MS using CHCA as a matrix

2.4.2. Identification of the polyhydroxylated tetraketide-like repeating unit in the sporopollenin exine structure.

2.4.2.1. MALDI-TOF-MS (+ion mode) using HABA as a matrix and CID-MS/MS

The MALDI-TOF-MS (+ ion mode) of the sporopollenin exine using HABA as a matrix showed a different MS pattern (a polymer with a 280 Da repeating unit) than the one observed with either DAN or CHCA. In contrast with these two matrices, the MALDI-MS with the HABA matrix gave new structural information about sporopollenin and afforded a series of sodiated molecular ions containing both Na and K in their molecular formulae. These ions at *m/z* 1983.4810, 1703.4147, 1423.3455 and 1143.2730 were identified as $[C_{94}H_{91}KNa_2O_{41}+Na]^+$, $[C_{81}H_{79}KNa_2O_{34}+Na]^+$, $[C_{68}H_{67}KNa_2O_{27}+Na]^+$and $[C_{55}H_{55}KNa_2O_{20}+Na]^+$, respectively (Figure 2.4 and Table 2.2). These four sodiated molecular ions differed from each other by the 280 Da repeating unit which is assigned as a polyhydroxylatedtetraketide-like derivative with chemical formula $C_{13}H_{12}O_7$. This series of ions contain carboxylates salts in their structures, which were probably formed during the KOH step in the extraction process of sporopollenin from *Lycopodium clavatum* pollen grains in addition to the washing step at the end using NaOH as indicated in the experimental section. [43]

During the KOH treatment, proteins were removed from the pollen grains, and the ester bonds in the sporopollenin outer surface became partially hydrolyzed to carboxylic acid salts, thereby increasing the hydrophilicity of the sporopollenin exine. [43] It should also be noted that these particularly small K^+ or Na^+ metal ions could easily enter the massive network of spherical sporopollenin via multi-directional nano-diameter sized channels.[2]

As well, the step employed after the base (KOH) hydrolysis is an acid hydrolysis using phosphoric acid, thus although the phosphoric acid neutralizes the potassium carboxylate resulting from the first step, potassium is still present, as shown in the data presented in this section. Furthermore, another series with almost the same chemical formulae, except containing Na^+ instead of K^+ (-16Da), afforded ions identified at *m/z* 1687.4452, 1407.3719 and 1127.2996 and assigned as $[C_{81}H_{79}Na_3O_{34}+Na]^+$, $[C_{68}H_{67}Na_3O_{27}+Na]^+$ and $[C_{55}H_{55}Na_3O_{20}+Na]^+$ (Figure 2.4 and Table 2.2).

Unexpectedly, the high-energy KeV CID-MS/MS of these precursor ions at *m/z* 1983, 1703, 1423 and 1143 showed a base peak formed by the loss of the repeating unit of 280 Da (i.e., *m/z* 1983 produces *m/z* 1703; *m/z* 1703 produces *m/z* 1423; *m/z* 1423 produces *m/z* 1143 and finally *m/z* 1143 produces *m/z* 863).

For example, the product ion scan of the precursor ion at *m/z* 1983 showed the sequential loss of 280 Da units (*m/z* 1983→ *m/z* 1703→ *m/z* 1423→ *m/z* 1143→ *m/z* 863) (Figures 2.**5**, and Scheme 2.2). Moreover, this latter product ion at *m/z* 863 fragments to yield the product ion at *m/z* 287. This indicates that the product ion at *m/z* 863 is composed of three attached units of 287 Da (3 x 287 Da + 2H= 863). From this MS/MS pattern, we deduced that the precursor ion at *m/z* 1983 is in general composed of four units of 280 Da and three units of 287 Da. Also, the product ion scan of the precursor ion at *m/z* 1983 is initiated by loss of -CH=CH-CH_2-CH=CH- (-66 Da), which is consistent with the presence of unsaturated fatty acids. Moreover, we noticed that the product ion scan of the precursor ion at *m/z* 1983 is not initiated by the loss of 44 Da thus supporting the absence of free carboxylic acid groups (i.e. CO_2 present in rings or as salts). Other CID-MS/MS

fragmentation pathways that support the proposed structure of the precursor ion at *m/z* 1983 are shown in Scheme A1.4.

As further support for our proposed structure for the precursor ion at *m/z* 1983 and its CID-MS/MS fragmentation patterns presented in Scheme 2.2 and Scheme A1.4, the *quasi*-MS^3 spectrum of the product ion at *m/z* 863 was measured.[44-46] (Figure A1.13).

The *quasi*-MS^3 spectrum of the product ion at *m/z* 863, showed the presence of both Na and K atoms in the chemical composition of this product ion. The presence of both Na and K atoms in this ion, created a challenge in proposing a structure for this ion at *m/z* 863 (which is composed of three units of *m/z* 287 Da and /or it fragments in to three ions of *m/z* 287); (Figures 2.5).

After examining a series of rational potential structures, we have chosen the most reasonable structure for *m/z* 863 that fit with both the exact masses of this polymeric series in the MALDI-MS using HABA as a matrix, and the MS/MS of the original precursor ion at *m/z* $1983^{\cdot}$. Consequently, the product ion at *m/z* 863 was assigned as the two following isomeric ions $[C_{42}H_{43}KNa_2O_{13}+Na]^+$ and/ or $[C_{42}H_{43}Na_3O_{13}+K]^+$. These latter ions were composed of three tetraenoic C14 carboxylic acid derivatives. Furthermore, the *quas*i-MS^3 of the product ion at *m/z* 863 showed the loss of carbon dioxide to give the $[M+Na-CO_2]^+$ product ion at *m/z* 819 and the formation of its corresponding $[M+H-CO_2]^+$ at *m/z* 797 (Figure A1.13). This *quasi*-MS^3 fragmentation pattern supports the presence of the free carboxylic acid group in the structure of *m/z* 863, which was presumably formed after the cleavage of the cyclic oligomer at *m/z* 1983. Other fragmentation patterns that support the structure of the product ion at *m/z* 863 are shown in Scheme 2.3. It is of course possible

that the branched carboxylate salts in the precursor ion at *m/z* 1983 or its product ion at *m/z* 863 may have derived from alpha pyrone rings (cyclic esters or lactones) that were hydrolyzed during the extraction process. As well, the presence of the double bonds may originally exist as shown in the structure of the product ion at *m/z* 863, or could have been formed during the extraction process by dehydration reactions of a hydroxylated chain

It is notable that the molecular structures of this series of ions obtained in the MALDI-MS using HABA as a matrix were identified with good mass accuracy (< 25ppm) as shown in Table 2.2. Furthermore, the theoretical isotopic distribution of the chemical formula assigned to the highest intensity peak in the MS at *m/z* 1127 was calculated and showed excellent fitting with the experimental isotopic distribution (ca. 90%, as shown in Figure A1.14).

In summary, the MALDI-TOF/TOF-MS/MS using HABA as a matrix showed the presence of polyhydroxylated tetraketide-like repeating units (α-pyrone ring plus hydroxylated aliphatic chain) that are linked together to form a macrocyclic rigid backbone. These macrocyclic oligomers and/or its related polymers are rich in hydroxyl groups to which the lipid network of the poly(hydroxyacid) can be ancored via ether linkages. Consequently, these macrocyclic oligomers and/or its related polymers appear to act like the building block upon which the poly(hydroxyacid) network can be built, forming a scaffold of the spherical sporopollenin. It is also possible that the sporopollenin wall consists of alternating layers of macrocyclic backbone composed of the polyhydroylated tetraketide-like monomers and the poly(hydroxyacid) network.

Table 2.2. The chemical formulas of the ions identified in the sporopollenin MS using HABA as a matrix

Experimental *m/z*	**Assigned chemical formula**	**Error ppm**
1983.4810	$[C_{94}H_{91}KNa_2O_{41}+Na]^+$	22.3
1703.4147	$[C_{81}H_{79}KNa_2O_{34}+Na]^+$	21.3
1687.4452	$[C_{81}H_{79}Na_3O_{34}+Na]^+$	24.2
1423.3455	$[C_{68}H_{67}KNa_2O_{27}+Na]^+$	17.9
1407.3719	$[C_{68}H_{67}Na_3O_{27}+Na]^+$	18.4
1143.2730	$[C_{55}H_{55}KNa_2O_{20}+Na]^+$	9.9
1127.2996	$[C_{55}H_{55}Na_3O_{20}+Na]^+$	10.5
1105.3091	$[C_{55}H_{55}Na_3O_{20}+H]^+$	2.9

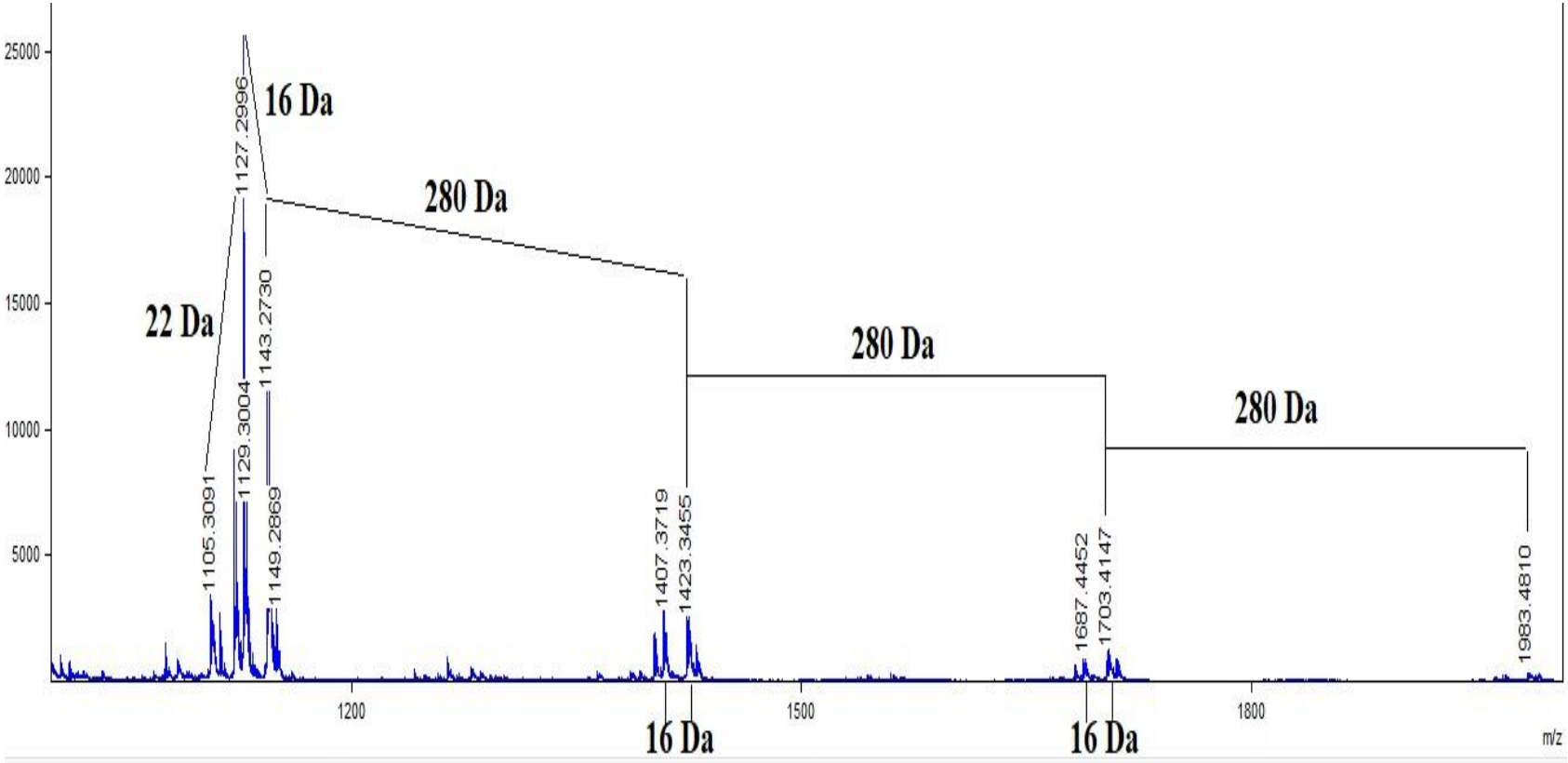


Figure 2.4. MALDI-TOF-MS (+ ion mode) using HABA as a matrix and displaying the presence of a repeating unit of 280 Da

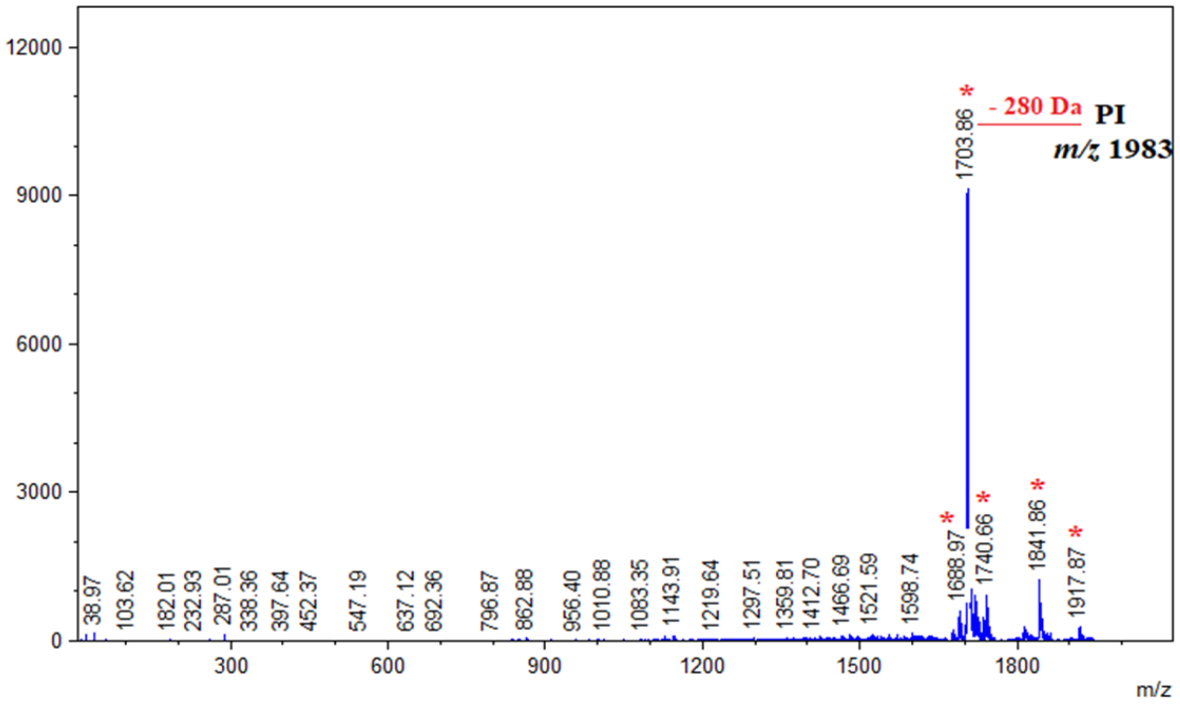


Figure 2.5a. Product ion scan of the precursor ion at *m/z* 1983 showing a loss of 280 Da to produce the base peak at *m/z* 1703

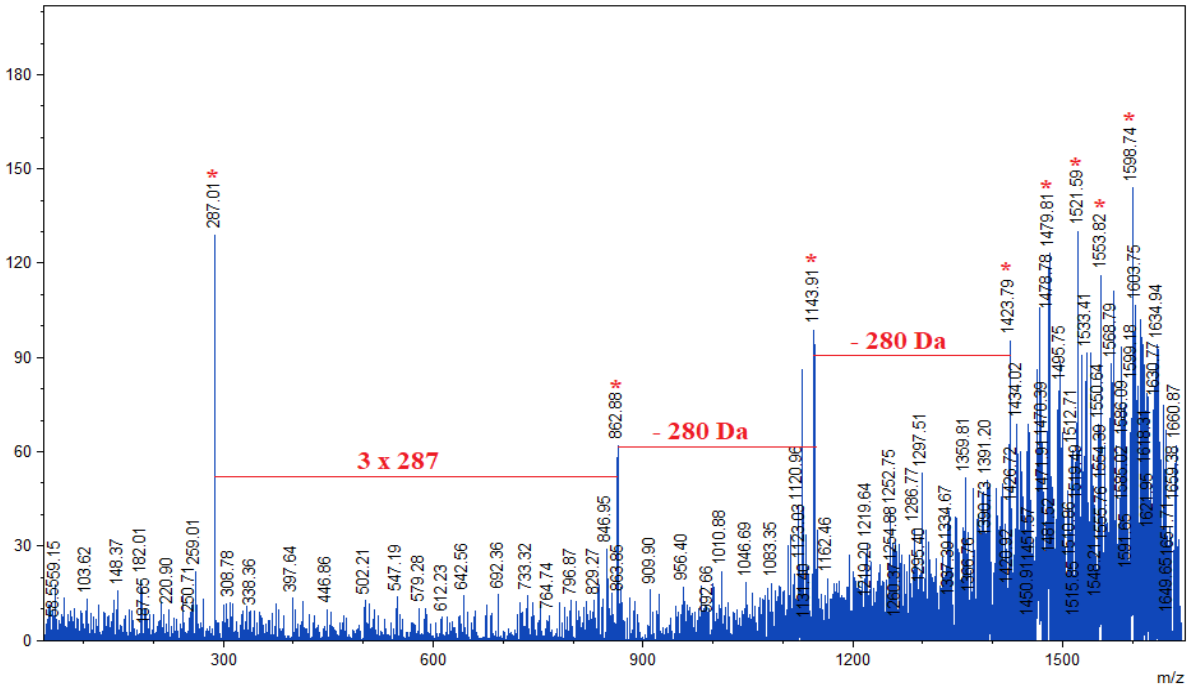
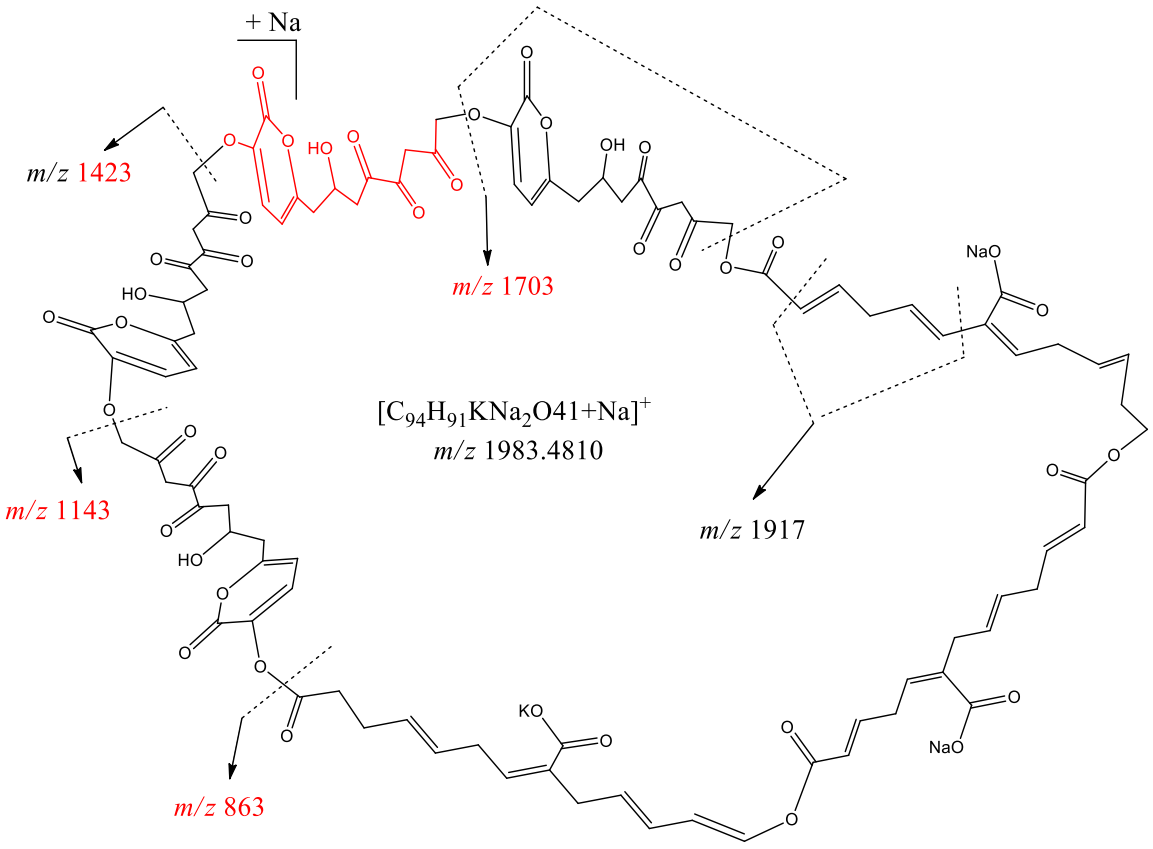
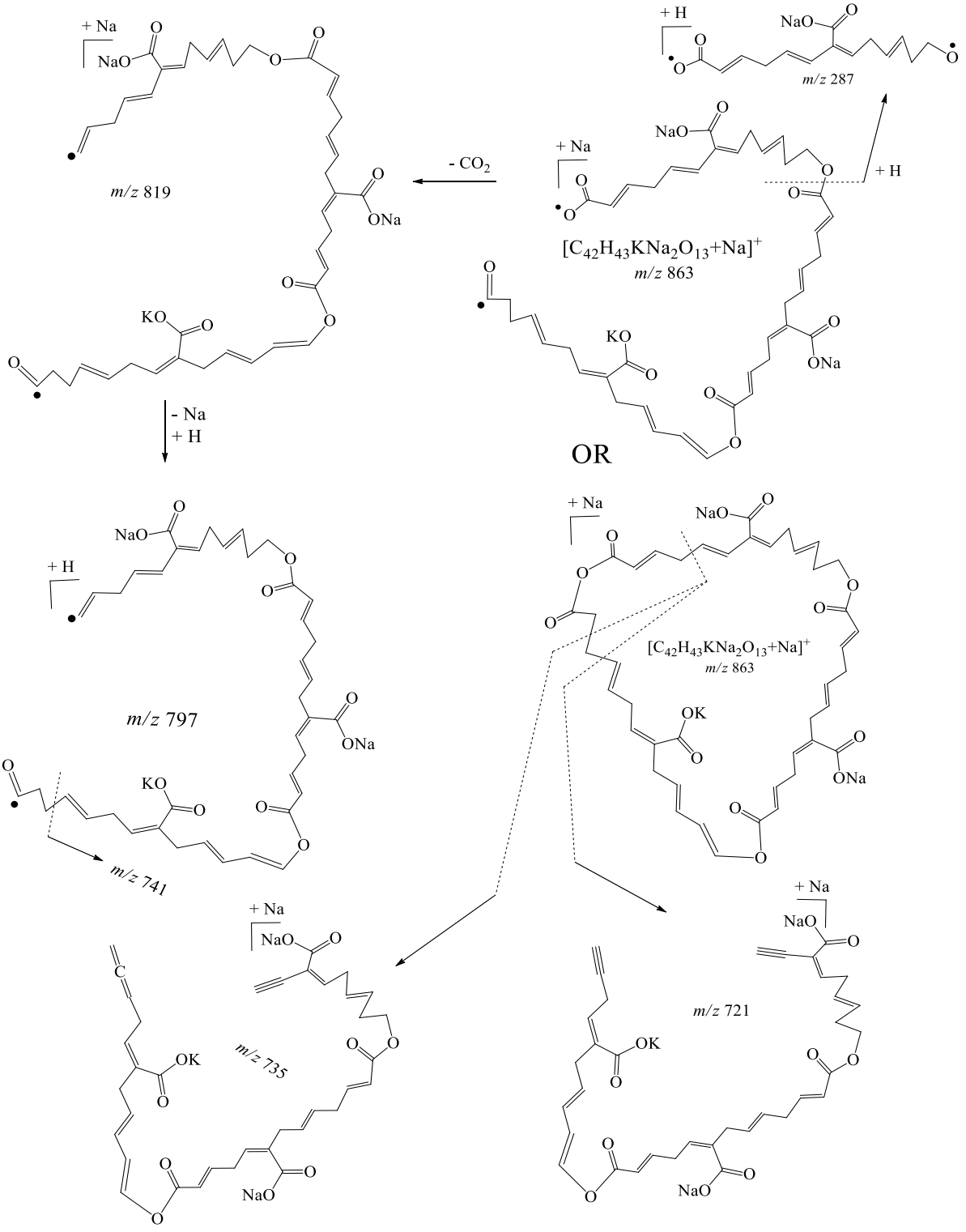


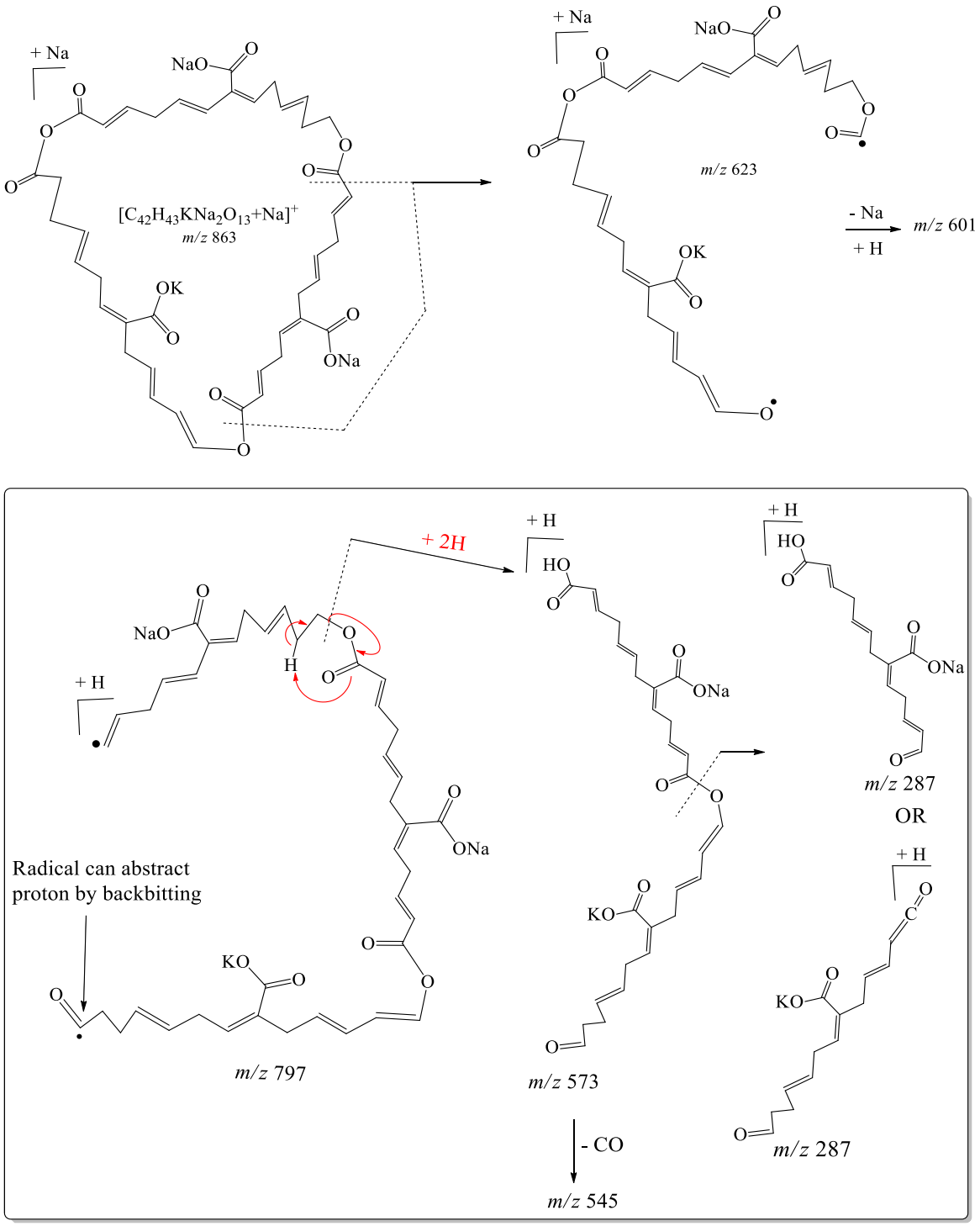
Figure 2.5b. Expansion of the low mass region in the product ion scan of the precursor ion at *m/z* 1983



Scheme 2.2. Fragmentation pattern of the precursor ion at *m/z* 1983. The red part of the molecule represents the repeating unit of 280 Da.



Scheme 2.3a. *Quasi*-MS^3 fragmentation patterns of the product ion at *m/z* 863.



Scheme 2.3b. *Quasi*-MS^3 fragmentation patterns of the product ion at *m/z* 863. Note: Intramolecular hydrogen transfers in some fragments can occur by backbiting, [47], and McLafferty rearrangement from the alcohol side, as indicated by red arrows. [48,49]

It should be noted that although many MALDI matrices have been introduced since the development of the MALDI ionization technique, there has been a continuous effort to improve our understanding of the analyte/matrix interaction and ionization mechanisms. Similarly, there has been tremendous efforts dedicated to improve the quality of MS data, based on matrix selectivity/specificity for an analyte. [50,51] It is well known that CHCA is a common MALDI matrix used for analytes such as peptides and proteins, whereas DAN matrix has been extensively employed for lipid analysis. [52,53] On the other hand, HABA has been suggested as an efficient matrix for analyte types such as glycans, steroids, glycoproteins, glycolipids, and synthetic polymers.[36,54,55] Therefore, the selection of the most suitable matrix for a particular analyte, is often a "trial and error' process.[56] Based on the complexity of the sporopollenin exine structure, it was expected that different matrices could provide different information's about the sporopollenin structure. For this reason, it was not surprising that both CHCA and DAN matrices were efficient in the ionization of the lipid network of sporopollenin and showing a series of masses that are in general 14 and/or 28 Da apart, whereas HABA was efficient in ionizing the oligomeric series of the macrocyclic unit in sporopollenin and producing masses that are in general 280 Da apart.

Lastly, it should be noted that every characteristic ion in the MALDI-TOF-MS of the sporopollenin exine using different matrices, may be composed of a mixture of isobars with different chemical composition and/or adduct type (M+ H or Na or K). However, the presence of Na and/or K was more notable in the MALDI-TOF-MS and the CID-MS/MS studies of the sporopollenin exine using HABA as a matrix. This may be because that the

major peaks in the MS using HABA was 280 Da apart from each other's unlike the complex spectrum in DAN and/or CHCA where peaks are closely related to each other and their isotopic distribution are overlapped.

2.4.3. Verification of the proposed structure of the sporopollenin exine using solid- state NMR

2.4.3.1. Sporopollenin ^{1}H- and ^{13}C-NMR, 2D ^{1}H-^{1}H NOESY, Rotor-synchronized ^{13}C{^{1}H} HSQC, and ^{13}C{^{1}H} multi CP-MAS NMR

It should be noted that the solid-state ^{13}C-NMR study of the sporopollenin exine appears to be similar to that obtained by Li *et al.* for the pine sporopollenin.[15] The standard ^{1}H spectrum of sporopollenin shows both sharp and broad peaks indicative of a mixture "liquid-like" or a more crystalline phase (green line) and a more amorphous phase (pink line) (Figure A1.15a). Using a CPMG filter, we can easily remove the amorphous phase and observe only the liquid-like structure (Figure A1.15B). The "liquid-like phase indicates the presence of a mobile sub-structure, such as long fatty acid chains.

After an extensive literature search, we were fortunate to find a great similarity between the ^{1}H-NMR spectrum of the "liquid-like" phase of our sporopollenin with that of the ^{1}H-NMR spectrum of whole seeds of *Lesquerella lyrate* which contain lipids (Figures A1.16).[57] The two spectra are almost identical; nonetheless, in sporopollenin, two important peaks were absent. The first peak is at 3.5 ppm characteristic for hydroxy fatty acid (-CHOH-) indicating that OH in the fatty acid chains of sporopollenin is present as a substituent on a central double bond (vinylic), and the second peak at 4.8 ppm characteristic for TAG Estolides (branched ester bond). This supports the proposal that the

hydroxycarboxylic acids in the sporopollenin network are connected linearly. The amorphous part appears to indicate the presence of a more rigid sub-structure containing alpha-pyrone rings (broad peak at 6.79 ppm) and hydroxylated chains (broad peaks at 3.79 and 0.98 ppm). These assignments support the presence of polyhydroxylated tetraketide-like monomer composed of alpha-pyrone ring and a hydroxylated aliphatic chain. Based on the absence of Estolides (branched esters), we proposed that the hydroxyl end groups of the poly(hydroxyacid) network are attached to the macrocyclic backbone through ether bonds instead of the branched ester bonds.

The 2D ^{1}H-^{1}H NOESY solid state experiments indicated the presence of ten cross-signals corresponding to different identified functional groups (Figures A1.17a-d). This allowed us to get an idea of the proximity of the proton groups.

Whereas the HSQC (Figure A1.18) indicates the ^{13}C correlation, at least with the most populated groups revealed by 1D ^{1}H-NMR and 2D ^{1}H-^{1}H NOESY. The results are listed in Table A1.2.

The ^{13}C{^{1}H} CPMAS NMR confirms the mixture of crystal-like and amorphous characteristics of the sample. Using the multi-CP pulse sequence, we can obtain quantitative data within a reasonable timeframe. Deconvolution and integration of the fitted peaks indicate the relative percentage of each group. The results are presented in Figures 2.6a-c, Table 2.3, and Table A1.3.

Using the building block obtained from TOF-SIMS and MALDI-MS, we were able to propose a model that represents the empirical formula of *Lycopodium clavatum*

sporopollenin and fitted with the quantitative data obtained from C-13 NMR to a large extent (Figures 2.6a-c, and Table 2.3).

It should be noted that the weak peak at ~55 ppm is assigned to the methylene group between two ketone groups (β-diketone supported by TOF-SIMS-MS/MS), which was left unassigned by Li *et al.* [15] This 55 ppm carbon can result from the β-diketone moiety which exists in the enol-enol form This carbon supports the possibility of the β-diketone moiety to exist in different forms in the sporopollenin network (i.e. keto-keto, keto-enol and enol-enol). Please note, that the presence of the β-diketone moiety has been explained, one more time, in the summary of the manuscript in the form of questions and answers in note SI-2.

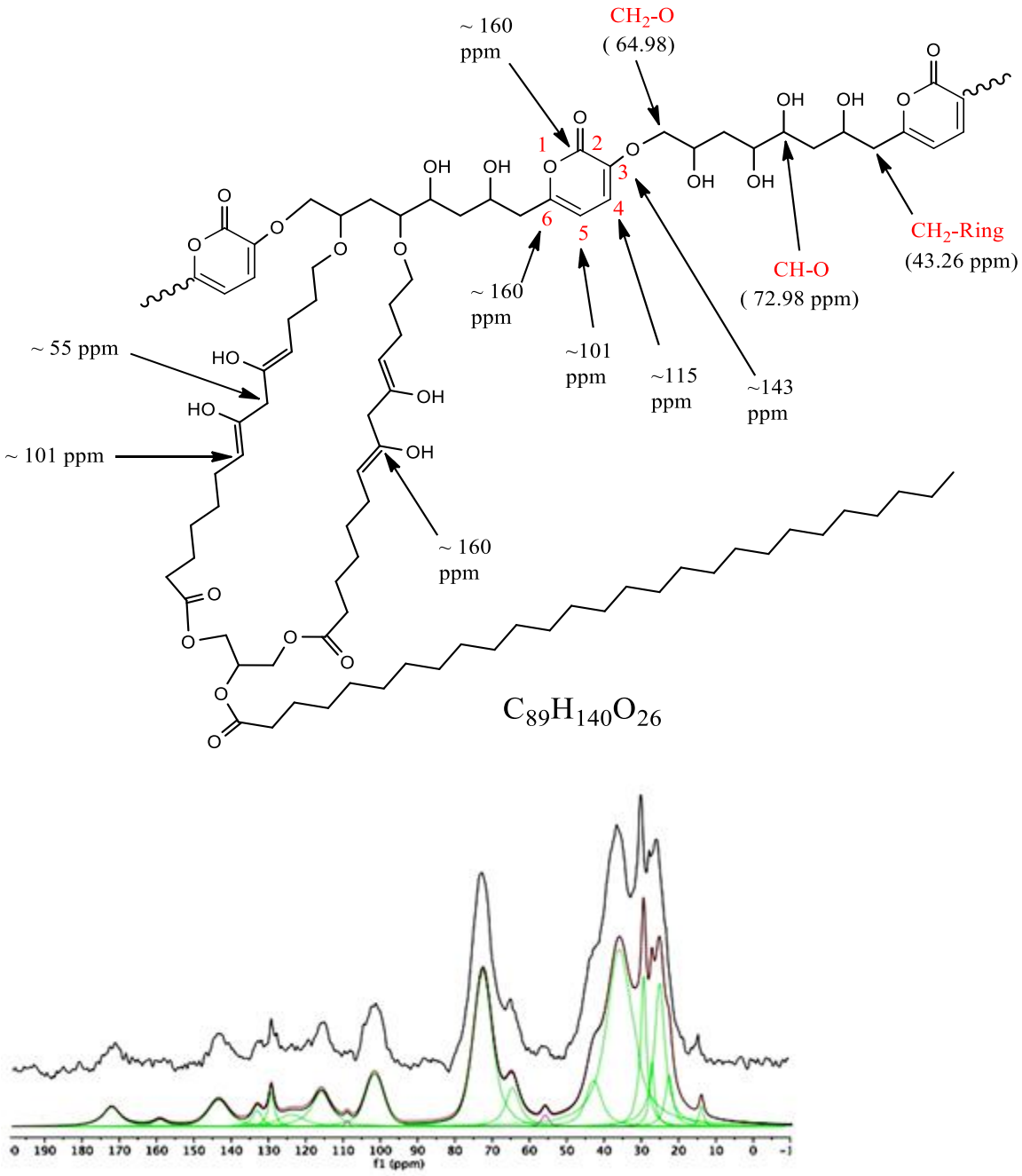


Figure 2.6. *Top:* Empirical formula model; *bottom*: experimental $^{13}C\{^{1}H\}$ multi-CP MAS and deconvoluted NMR spectra.

Table 2.3. Distribution of carbon atoms in the structure according to the deconvolution of the experimental spectrum and the *Lycopodium clavatum* empirical formula model designed according to moieties identified using MS.

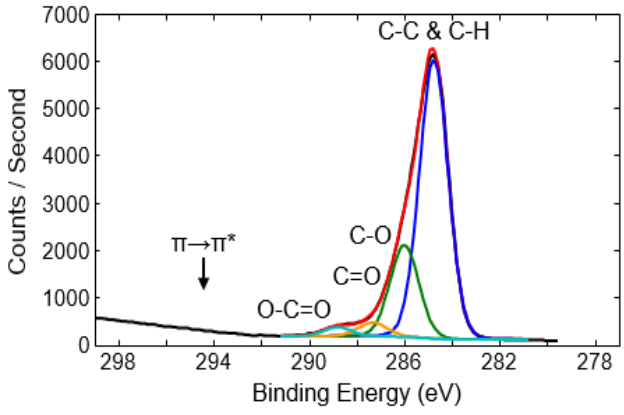
% C	Integration ppm range	Deconvolution	Model
CH_3 (Chain)	0-20	1	1
CH_2 (Chain or ring)	20-50	56	50
CO -CH_2-CO (in the enol-enol form)	50-60	2	2
CH_2-O or CH-O (Chain)	60-90	19	20
Ring C_5 or Chain C=C-O	90-110	6	6
Ring C_4 or chain C=C-CH_2-C=C	110-130	8	6
C=C (Chain)	130-140	2	2
Ring C_3	140-150	3	3
=C-OH or Ring C_2 or Ring C_6	150-165	1	5
COOH(Chain)	165-185	2	5

2.4.4. Verification of the total absence of aromatics in the sporopollenin exine

2.4.4.1. High-Resolution X-ray Photoelectron Spectroscopy (HR-XPS)

The HR-XPS C(1s) surface analysis of sporopollenin (Figure 2.**7**) showed the presence of the main types of linked carbons atoms in the *Lycopodium clavatum* sporopollenin together with their relative percentages in the sample, based upon the areas calculated under each peak. Most importantly, it showed the complete absence of any satellite peak at the higher binding energy that results from the π-π^* transition. It should be noted that the absence of this "shake-up line" is consistent with the lack of aromaticity in the sporopollenin. [58]

We would like to reiterate that the solid-state ^{1}H NMR spectra of *Lycopodium clavatum* sporopollenin exine was composed of two overlapping spectra which account for the presence of two different structural components exactly like what we have deduced by using the MALDI-TOF-MS and MS/MS analyses. The sharp region in the ^{1}H-NMR spectrum represents the poly(hydroxyacid) network, whereas the broad region containing a peak at 6.79 ppm may have indicated and made us assume in the beginning, that sporopollenin may contain aromatic components**.** However, high-resolution X-ray photoelectron spectroscopy (XPS) showed the complete absence of aromaticity in sporopollenin. Therefore the broad proton NMR spectrum more likely is supportive for the presence of α-pyrone rings in the tetraketide components of sporopollenin. It should be noted that α-pyrone rings possess very weak aromatic character, as they can undergo electrophilic additions, ring-opening and Diels-Alder reactions, which are not typical reaction for a stable aromatic compound.[59]



C 1s functionality	peak BE (eV)	rel. abundance (%)
C-C & C-H	284.8	71
C-O	286.0	23
C=O	287.3	4
O-C=O	288.0	2
π→π*	N/A	0

Figure 2.7. *Lycopodium clavatum* sporopollenin X-ray photoelectron spectroscopy of carbon (1s).

2.4.5. Sporopollenin Model

In an attempt to correlate our obtained data to the scanning electron microscopic images (SEM) images of different sporopollenin exines, we have used the identified building blocks in this manuscript to build a general model for the formation of the sporopollenin exine that appears consistent with its various network SEM images.

For example, the *passiflora sp. (Passifloraceae) sporopollenin* SEM image (Figure A1.19) showed a macrocyclic substructure with a cross-linked network built on it.[60] Although *passiflora sp.* is a different type of sporopollenin than that of *Lycopodium clavatum*, its SEM image is consistent with the building units identified in this manuscript (*i.e* the macrocyclic rigid backbone and dendrimeric-like network) and helped us to visualize how these units may be linked together as shown in our hypothetical proposed model (Figure A1.20)

2.4.6. DFT computational model using *Gaussian09* of a hypothetical macrocyclic oligomer

A DFT computational model using *Gaussian09* [61] of a hypothetical macrocyclic oligomer based upon the structure shown in Scheme 2.2 but in which the α-pyrone rings are intact and are not ring-opened by the conditions which were used is shown in Figure 2.8. The structure was optimized using the B3LYP/6-31G basis set. The macrocycle clearly shown the "bowl"-shape of the possible macrocyclic segment in the sporopollenin exine (Figures A1.19 and A1.20).

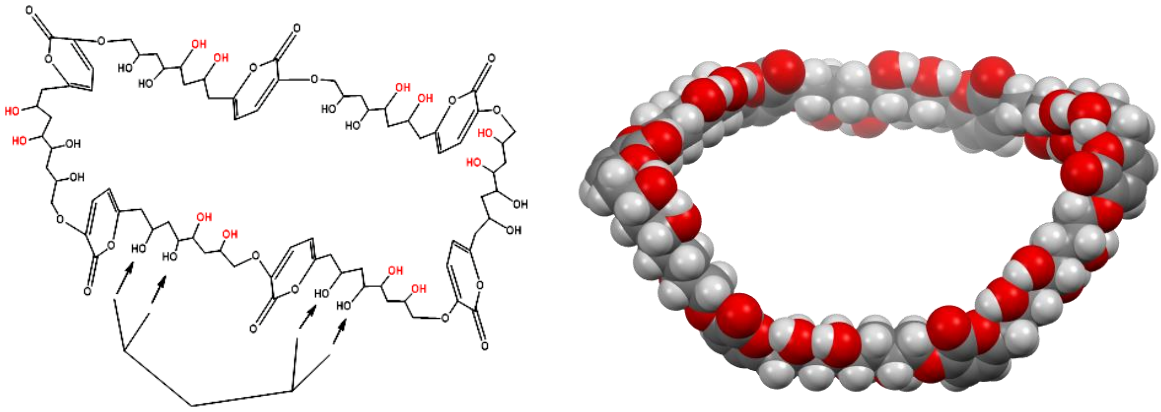


Figure 2.8. *Left*: Hypothetical structure based upon Scheme 2.2 showing an example of the dendrimer linking points, and *right:* B3LYP/6-31G DFT-optimized computed molecular stucture.

2.5. Conclusion

Although there have been much controversy and uncertainty about the structural constituents and the molecular identity of the biopolymer sporopollenin, many scientists around the world have focused mostly on its biomedical applications, since the sporopollenin exine could exist as a spherical dendrimer, a typical type of microcapsule used for drug delivery applications.[62] Until recently, the sporopollenin exine was described as being a highly resistant biopolymer, which was thought to be composed of aromatics, phenolics, and long-chain aliphatic acids. [63, 64]

In this manuscript, we have proven the total absence of aromaticity in the sporopollenin exine using XPS analysis. We also have shown, for the first time, the presence of two main sporopollenin building units using solid-state ^{1}H- and ^{13}C-NMR, 2D^{1}H-^{1}H, and 2D ^{13}C-^{1}H NMR experiments, SIMS-TOF-MS, MALDI-TOF-MS, and

CID-MS/MS. These analyses indicated the presence of a macrocyclic backbone composed of polyhydroxylated tetraketides-like monomers that represent the main macrocyclic and rigid backbone of the sporopollenin biopolymer. This macrocyclic backbone can be covalently attached by ether linkages to the poly(hydroxyacid) chain network(s) to form the sporopollenin biopolymer.

As a consequence of the SIMS-TOF-MS and KeV CID-MS/MS analyses the chemical structural features present on the outermost surfaces of the sporopollenin exine could be discerned through the discovery of ions characteristic for diacylglycerol (DAG). This led to MALDI-TOF-MS and CID-TOF/TOF-MS/MS analyses that allowed us to decipher each constituent of the sporopollenin repeating units.

In this context, it should be noted that the use of different matrices was of great assiatance. The use of 1,5-diaminonaphthalene (DAN) and α-cyano-4-hydroxycinnamic acid (CHCA) allowed us to obtain a complete characterization of the complex network of the poly(hydroxyacids). As well, the use of the 2-(4-hydroxyphenylazo)benzoic acid (HABA) also allowed us to establish the presence of the macrocyclic polymer composed of the polyhydroxylated tetraketide-like monomers (rigid backbone).

As well, a "bowl'-shaped DFT geometry-optimized structure of a hypothetical structure based upon the experimental obersvations made could be computed which is consistent with our structural conclusions.

Furthermore, we can state that specifically, the sporopollenin exine of *Lycopodium clavatum* does not contain any aromatics and bears no resemblance to lignin. It should be noted that in 1966, Gordon Shaw, one of the earliest pioneers in the studies of

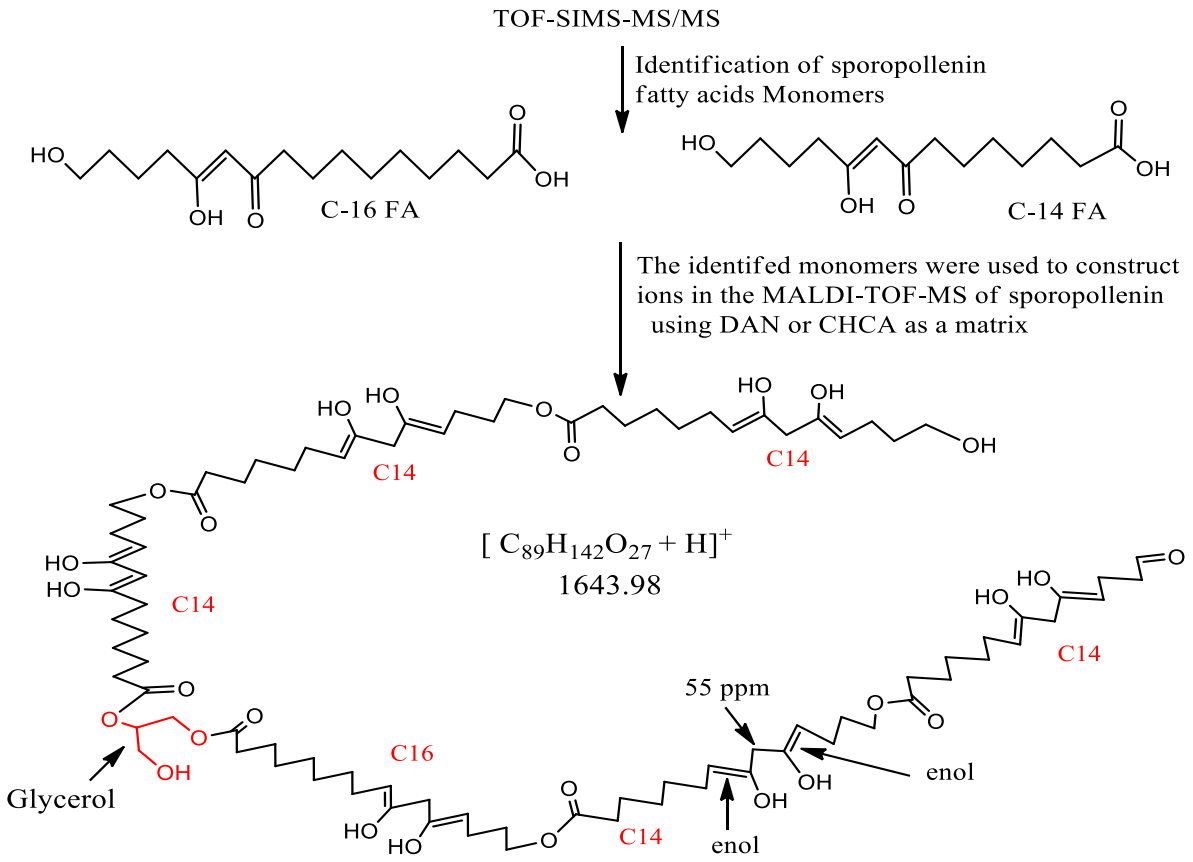
sporopollenin, withdrew his proposal that sporopollenin exine contained lignin since it did not give any positive test for lignins.[65]

Based on our data presented here, we can develop a new and experimentally well-proven opinion that sporopollenin exine is composed of aliphatic biopolymer with *pseudo* aromatic α-pyrone rings presented in the polyhydroxylated tetaketide-like component of sporopollenin.

In this manuscript, therefore, we were able to reveal the novel two major builidng units of the spherical sporpollenin exine and the summary of this well laborious and complicated study is presented in Figure 2.9.

Nevertheless, the exact biosynthesis of sporopollenin is now left to the expert biologists, who hopefully could decipher the exact biosynthesis of sporopollenin, especially since the regulation genes of sporopollenin biosynthesis have recently been discovered.[66]

Our future work will focus on using different matrices in the MALDI-TOF-MS/MS in the positive or the negative ion mode, which may reveal further diagnostic structural details of the *Lycopodium clavatum* sporopollenin. Moreover, other sporopollenin species will be the subject of our future studies using the same characterization techniques presented here. Lastly, we aim to perform some computational modeling studies to build a complete sporopollenin network using the units identified in this manuscript and to study the stability of this network and the possible interactions between their chains such as intramolecular hydrogen bonding and other van der Waals interactions.



1) The diagnostic ions in the MALDI-TOF-MS were identified with good ppm error and their C/H, C/O and H/O ratios fitted with the empirical formula of *Lycopodium clavatum* sporopollenin exine

2) This allowed us to identify the presence of the poly(hydroxyacid) network containing glycerol as a core unit, which represents the first builidng block of the sporopollenin exine.

3) A unique peak at 55 ppm in the ^{13}C-NMR is characterisitic to the betadiketone moiety identified using TOF-SIMS-MS/MS (enol-enol form of the betadiketone moiety)

4) The structure of this building unit of the sporopollenin exine, is sustained by the sharp ^{1}H-NMR component of the whole sporopollenin, which supports the presence of a liquid-like or crystalline network of sporopollenin.

5) The sharp H-1 NMR spectrum was almost identical to the proton NMR of Triacylglycerol estolides of *Lesquerella lyrate*, but in sporopollenin, two important peaks at 3.5 (CH-OH) and 4.8 ppm (Estolides) were absent. This supports the prescence of betadiketone moiety (No CH-OH) in the fatty acid monomers of sporopollenin and that the hydroxy carboxylic acid monomers are connected lineraly (No Estolides).

Figure 2.9a. Summary of the identification of the poly(hydroxyacid) network of sporopollenin

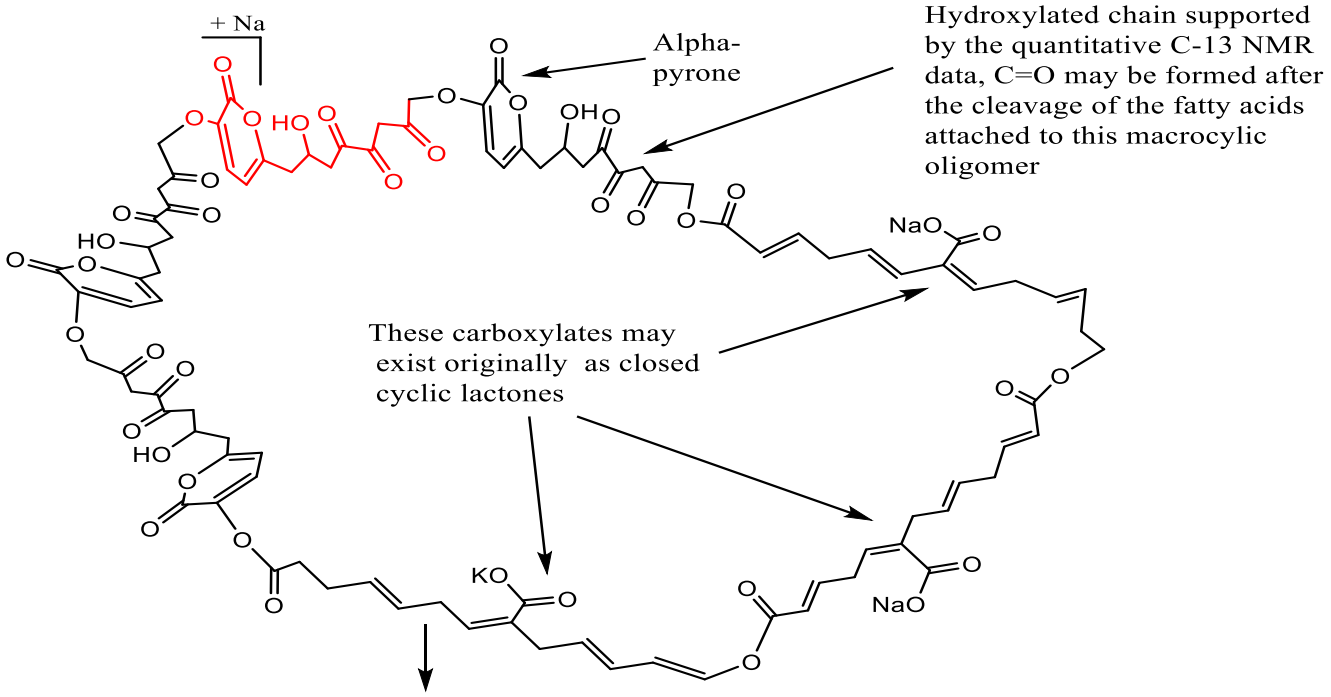
MALDI-TOF-MS/MS of sporopollenin
using HABA as a matrix

↓

Macrocyclic oligomeric seies with 280 Da repeating unit

↓

Example: *m/z* 1983 (this ion can be represented by different double bonds postions)



1) From this study, we deduced the prescence of the second building unit in the sporopollenin network which is a macrocyclic rigid backbone composed of alpha pyrone rings and hydroxylated chains

2) The prescence of this rigid backbone is supported by the broad ^{1}H-NMR component in the whole H-1 NMR of sporopollenin. This support the rigidity of this building unit and showing that its movement is restricted unlike the liquid-like chains of the poly(hydroxy acid) network.

3) Although, alpha-pyrone rings have weak aromatic character (No aromaticity in high resolution XPS), its protons appears in the broad ^{1}H-NMR at 6.79 ppm. The broad peaks at 0.98 and 3.79 represent the hydroxylated chains in the macrocyclic

4) The main components of this macrocyclic backbone is supported by the ^{13}C-NMR.

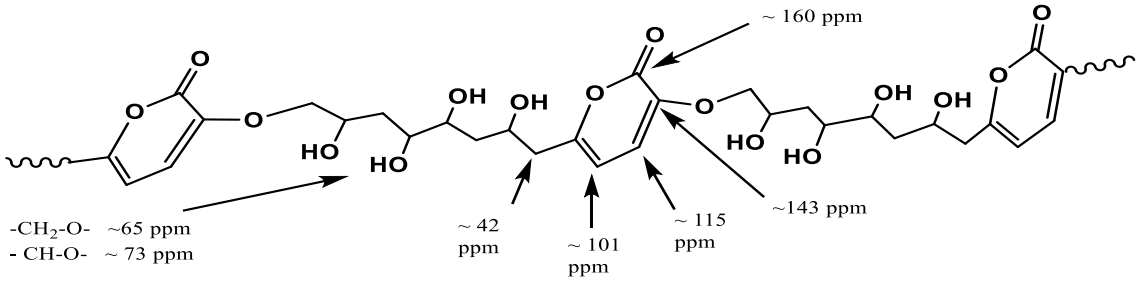


Figure 2.9b. Summary of the identification of the macrocyclic backbone of sporopollenin

References

[1] Mackenzie, G.; Beckett, S.; Atkin, S.; Diego-Taboada, A.: Pollen and spore shells – Nature's microcapsules. In *Microencapsulation in the Food Industry*; Gaonkar, A.G., Vasisht, N., Khare, A.R., Sobel, R., Eds.; Academic Press: Cambridge, Massachusetts, **2014**; pp 283-297.

[2] Mackenzie, G.; Boa, A.N.; Diego-Toboada, A.; Atkin, S.L.; Sathyapalan, T.: Sporopollenin, the least known yet toughest natural biopolymer. *Front Mater*. **2015**, *2*(66), 1–5.

[3] Ahlers, F.; Bubert, H.; Steuernagel, S.; Wiermann, R.: The nature of oxygen in sporopollenin from the pollen of *Typha angustifolia* L. *Z. Naturforsc. C*. **2000**, *55*(3–4), 129-136.

[4] Domínguez, E.; Mercado, J.A.; Quesada, M.A.; Heredia, A.: Isolation of intact pollen exine using anhydrous hydrogen fluoride. *Grana*. **1998**, *37*(2), 93-96.

[5] Brooks, J.; Shaw, G.: Chemical structure of the exine of pollen walls and a new function for carotenoids in nature. *Nature*. **1968**, *219*(5153), 532-533.

[6] Hayatsu, R.; Botto, R.E.; McBeth, R.L.; Scott, R.G; Winans, R.E.: Chemical alteration of a biological polymer" sporopollenin" during coalification: Origin, formation, and transformation of the coal maceral sporinite. *Energy Fuel*. **1988**, *2*(6), 843-847.

[7] De Leeuw, J.W.; Versteegh, G.J.M; Van Bergen, P.F.: Biomacromolecules of algae and plants and their fossil analogues. *Plant Ecol.* **2006**, *182*, 20-233.

[8] Watson, J.S.; Fraser, W.T.; Sephton, M.A.: Formation of a polyalkyl macromolecule from the hydrolysable component within sporopollenin during heating/pyrolysis experiments with *Lycopodium* spores. *J. Anal. Appl. Pyrolysis*. **2012**, *95*,138-144.

[9] Kim, S.S.; Douglas, C.J.: Sporopollenin monomer biosynthesis in *Arabidopsis*. *J. Plant Biol*. **2013**, *56*(1), 1-6.

[10] Wehling, K.; Niester, C.; Boon, J.J.; Willemse, M.T.M.; Wiermann, R.: *p*-Coumaric acid – A monomer in the sporopollenin skeleton. *Planta*. **1989**, *179*(3), 376-380.

[11] Dutta, S.; Hartkopf-Fröder, C.; Wilkes, H.; Greenwood, P.; Littke, R.; Mann, U.: The molecular composition of sporopollenin from fossil megaspores as revealed by micro-FTIR and pyrolysis-GC-MS. In *23rd International Meeting on Organic Geochemistry*; P88-MO: Torquay, UK, **2006**.

[12] Sáiz-Jiménez, C.: Production of alkylbenzenes and alkylnaphthalenes upon pyrolysis of unsaturated fatty acids. *Naturwissenschaften*. **1994**, *81*(10), 451-453.

[13] Conesa, J.A.; Font, R.; Marcilla, A.; Garcia, A.N.: Pyrolysis of polyethylene in a fluidized bed reactor. *Energy Fuel*. **1994**, *8*(6),1238-1246.

[14] Nierop, K.G.; Versteegh, G.J.; Filley, T.R.; De Leeuw, J.W.: Quantitative analysis of diverse sporomorph-derived sporopollenins. *Phytochemistry*. **2019**, *162*, 207-215.

[15] Li, F.S.; Phyo, P.; Jacobowitz, J.; Hong, M.; Weng, J.K.: The molecular structure of plant sporopollenin. *Nat. Plants*. **2019**, 5(1), 41-46.

[16] Bernard, S.; Benzerara, K.; Beyssac, O.; Balan, E.; Brown, G.E.Jr.: Evolution of the macromolecular structure of sporopollenin during thermal degradation. *Heliyon*. **2015**, *1*(2):1–28.

[17] Stolle, A.; Szuppa, T.; Leonhardt, S.E.; Ondruschka, B.: Ball milling in organic synthesis: Solutions and challenges. *Chem. Soc. Rev*. **2011**, *40*(5), 2317-2329.
[18] Brunow, G.; Lundquist, K.; Gellerstedt, G.: Lignin. In *Analytical Methods in Wood Chemistry, Pulping, and Papermaking*; Sjöström, E., Alén, R., Eds.; Springer Series in Wood Science; Springer: Berlin, Heidelberg, **1999**; pp 77-124.
[19] Kim, S.S.; Grienenberger, E.; Lallemand, B.; Colpitts, C.C.; Kim, S.Y.; de Azevedo Souza, C.; et al.: LAP6/POLYKETIDE SYNTHASE A and *LAP5/POLYKETIDE SYNTHASE B* encode hydroxyalkyl α-pyrone synthases required for pollen development and sporopollenin biosynthesis in *Arabidopsis thaliana*. *Plant Cell*. **2010**, *22*(12), 4045-4066.
[20] Grienenberger, E.; Kim, S.S.; Lallemand, B.; Geoffroy, P.; Heintz, D.; de Azevedo Souza, C.; et al.: Analysis of TETRAKETIDE α-PYRONE REDUCTASE function in Arabidopsis thaliana reveals a previously unknown, but conserved, biochemical pathway in sporopollenin monomer biosynthesis. *Plant Cell.* **2010**, *22*(12), 4067-4083.
[21] Lou, Y.; Zhu, J.; Yang, Z.: Molecular cell biology of pollen walls. In *Applied Plant Cell Biology*; Nick, P., Opatrny, Z, Eds.; Plant Cell Monographs; Springer: Berlin, Heidelberg, **2014**; *Vol.22*; pp179-205.
[22] Fraser, W.T.; Lomax, B.H.; Jardine, P.E.; Gosling, W.D.; Sephton, M.A.: Pollen and spores as a passive monitor of ultraviolet radiation. *Front. Ecol. Evol.* **2014**, *2*(12), 1–3.
[23] Chu, M.; Mierzwa, R.; Xu, L.; He, J.; Terracciano, M.; Patel, W.; et al.: Structure of Sch 419560, a novel α- pyrone antibiotic produced by Pseudomonas fluorescens. *J. Antibiot.* **2002**, *55*(2), 215-218.
[24] Barrier, S.; Diego-Taboada, A.; Thomasson, M.J.; Madden, L.; Pointon, J.C.; Wadhawan, J.D.; et al.: Viability of plant spore exine capsules for microencapsulation. J. *Mater. Chem.* **2011**, 21(4), 975-981.
[25] Fisher, G.L.; Bruinen, A.L.; Ogrinc, P.N.; Hammond, J.S.; Bryan, S.R.; Larson, P.E.; et al.: New method and mass spectrometer design for TOF-SIMS parallel imaging MS/MS. *Anal. Chem*. **2016**, *88*(12), 6433-6440.
[26] Fisher, G.L.; Hammond, J.S.; Bryan, S.R.; Larson, P.E.; Heeren, R.M.A.: The composition of poly(ethylene terephthalate) (PET) surface precipitates determined at high resolving power by tandem mass spectrometry imaging. *Microsc. Microanal*. **2017**, *23*(4), 843-848.
[27] Chini, C.E.; Fisher, G.L.; Johnson, B.; Tamkun, M.M.; Kraft, M.L.: Observation of endoplasmic reticulum tubules via TOF-SIMS tandem mass spectrometry imaging of transfected cells. *Biointerphases*. **2018**, *13*(3), 1–9.
[28] Fu, T.; Touboul, D.; Della-Negra, S.; Houël, E.; Amusant, N.; Duplais, C.; et al.: Tandem MS imaging and in situ characterization of bioactive wood metabolites in Amazonian tree species *Sextonia rubra*. *Anal. Chem*. **2018**, *90*(12), 7535-7543.
[29] Fu, T.; Houël, E.; Amusant, N.; et al.: Biosynthetic investigation of γ-lactones in *Sextonia rubra* wood using in situ TOF-SIMS MS/MS imaging to localize and characterize biosynthetic intermediates. *Nat. Sci. Rep*. **2019**, *9*(1), 1928-1938.
[30] Johnson, R.L.; Schmidt-Rohr, K.: Quantitative solid-state 13C NMR with signal enhancement by multiple cross polarization. *J. Magn. Reson*. **2014**, *239*, 44-49.

[31] Zetzsche, F.; Kälin, O.: Untersuchungen über die membrane der sporen und pollen V. 4. Zur autoxidation der sporopollenine. *Helv. Chim. Acta*. **1931**, *14*(1), 517-519.
[32] Gonzalez-Cruz, P.; Uddin, M.J.; Atwe, S.U.; Abidi, N.; Gill, H.S.: Chemical treatment method for obtaining clean and intact pollen shells of different species. *ACS Biomater. Sci. Eng*. **2018**, *4*(7), 2319-2329.
[33] Barrier, S.: Physical and chemical properties of sporopollenin exine particles. *Doctoral dissertation.* University of Hull: Hull, UK; **2008**.
[34] Gilli, G.; Bertolasi, V.; Ferretti, V.; Gilli, P.: Resonance-assisted hydrogen bonding. III. Formation of intermolecular hydrogen-bonded chains in crystals of β-diketone enols and its relevance to molecular association. *Acta Crystallogr. B*. **1993**, *49*(3), 564-576.
[35] Pan, M.H.; Huang, T.M.; Lin, J.K.: Biotransformation of curcumin through reduction and glucuronidation in mice. *Drug Metab. Dispos*. **1999**, *27*(4), 486-494.
[36] Thomas, A.; Charbonneau, J.L.; Fournaise, E.; Chaurand, P.: Sublimation of new matrix candidates for high spatial resolution imaging mass spectrometry of lipids: Enhanced information in both positive and negative polarities after 1,5-diaminonapthalene deposition. *Anal. Chem*. **2012**, *84*(4), 2048-2054.
[37] Hughey, C.A.; Hendrickson, C.L.; Rodgers, R.P.; Marshall, A.G.; Qian, K.: Kendrick mass defect spectrum: A compact visual analysis for ultrahigh-resolution broadband mass spectra. *Anal. Chem*. **2001**, *73*(19), 4676-4681.
[38] Kretzmann, J.A.; Evans, C.W.; Norret, M.; Iyer, K.S.: Supramolecular assemblies of dendrimers and dendritic polymers in nanomedicine. In *Comprehensive Supramolecular Chemistry II*; Atwood, J.L., Ed.; Academic Press: Cambridge, Massachusetts (Elsevier Inc.); **2017**, pp 237-256.
[39] Molin, L.; Seraglia, R.; Dani, F.R.; Moneti, G.; Traldi, P.: The double nature of 1,5-diaminonaphthalene as matrix-assisted laser desorption/ionization matrix: Some experimental evidence of the protonation and reduction mechanisms. *Rapid Commun. Mass Spectrom*. **2011**, *25*(20), 3091-3096.
[40] Neubert, H.; Halket, J.M.; Ocana, M.F.; Patel, R.K.: MALDI post-source decay and LIFT-TOF/TOF investigation of α-cyano- 4-hydroxycinnamic acid cluster interferences. *J. Am. Soc. Mass Spectrom*. **2004**, *15*(3), 336-343.
[41] Awad, H.; Stoudemayer, M.J.; Usher, L.; Amster, I.J.; Cohen, A.; Das, U.; et al.: The unexpected formation of $[M-H]^+$ species during MALDI and dopant-free APPI MS analysis of novel antineoplastic curcumin analogues. *J. Mass Spectrom*. **2014**, *49*(11), 1139-1147.
[42] Sugiyama, Y.; Kawakishi, S.; Osawa, T.: Involvement of the β-diketone moiety in the antioxidative mechanism of tetrahydrocurcumin. *Biochem. Pharmacol*. **1996**, *52*(4), 519-525.
[43] Uddin, M.J.; Abidi, N.; Warzywoda, J.; Gill, H.S.: Investigation of the fate of proteins and hydrophilicity/hydrophobicity of *Lycopodium clavatum* spores after organic solvent-base-acid treatment. *ACS Appl. Mater. Interfaces*. **2019**, *11*(23), 20628-20641.
[44] Raska, C.S.; Parker, C.E.; Huang, C.; Han, J.; Glish, G.L.; Pope, M.; et al.: Pseudo-MS^3 in a MALDI orthogonal quadrupole-time of flight mass spectrometer. *J. Am. Soc. Mass Spectrom*. **2002**, *13*(9), 1034-1041.

[45] Banoub, J.H.; Demian, W.L.; Piazzetta, P.; Sarkis, G.; Kanawati, B.; Lafont, D.; et al.: The in situ gas-phase formation of a C-glycoside ion obtained during electrospray ionization tandem mass spectrometry. A unique intramolecular mechanism involving an ion-molecule reaction. *Rapid Commun. Mass Spectrom.* **2015**, *29*(19), 1717-1732.
[46] Joly, N.; Aneed, A.E.; Martin, P.; Cecchelli, R.; Banoub, J.: Structural determination of the novel fragmentation routes of morphine opiate receptor antagonists using electrospray ionization quadrupole time of- flight tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **2005**, *19*(21), 3119-3130.
[47] Wesdemiotis, C.; Solak, N.; Polce, M.J.; Dabney, D.E.; Chaicharoen, K.; Katzenmeyer,
B.C.: Fragmentation pathways of polymer ions. *Mass Spectrom. Rev.* **2011**, *30*(4), 523-559.
[48] Francke, W.; Lübke, G.; Schröder, W.; Reckziegel, A.; Imperatriz-Fonseca, V.; Kleinert, A.; et al.: Identification of oxygen containing volatiles in cephalic secretions of workers of Brazilian stingless bees. *J. Braz. Chem.* **2000**, *11*(6), 562-571.
[49] Tengö, J.; Groth, I.; Bergström, G.; Schröder, W.; Krohn, S.; Francke, W.: Volatile secretions in three species of Dufourea (Hymenoptera: Halictidae) bees: Chemical composition and phylogeny. *Z. Naturforsch. C.* **1985**, *40*(9–10), 657-660.
[50] Knochenmuss, R.: MALDI and related methods: A solved problem or still a mystery? *Mass Spectrom. (Tokyo)*. **2013**, *2*(Special Issue), S0006- S0006.
[51] Niehaus, M.; Schnapp, A.; Koch, A.; Soltwisch, J.; Dreisewerd, K.: New insights into the wavelength dependence of MALDI mass spectrometry. *Anal. Chem.* **2017**, *89*(14), 7734-7741.
[52] Galesio, M.; Rial-Otero, R.; Capelo-Martínez, J.L.: Comparative study of matrices for their use in the rapid screening of anabolic steroids by matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **2009**, *23*(12), 1783- 1791.
[53] Clark, A.E.; Kaleta, E.J.; Arora, A.; Wolk, D.M.: Matrix-assisted laser desorption ionization–time of flight mass spectrometry: A fundamental shift in the routine practice of clinical microbiology. *Clin. Microbiol. Rev.* **2013**, *26*(3), 547-603.
[54] Batoy, S.; Akhmetova, E.; Miladinovic, S.; Smeal, J.; Wilkins, C.L.: Developments in MALDI mass spectrometry: The quest for the perfect matrix. *Appl. Spectrosc. Rev.* **2008**, *43*(6), 485-550.
[55] Yang, J.; Caprioli, R.M.: Matrix precoated targets for direct lipid analysis and imaging of tissue. *Anal. Chem.* **2013**, *85*(5), 2907-2912.
[56] Leopold, J.; Popkova, Y.; Engel, K.; Schiller, J.: Recent developments of useful MALDI matrices for the mass spectrometric characterization of lipids. *Biomolecules*. **2018**, *8*(173), 1–25.
[57] Zhang, H.; Olson, D.J.; Van, D.; Purves, R.W.; Smith, M.A.: Rapid identification of triacylglycerol-estolides in plant and fungal oils. *Ind. Crop. Prod.* **2012**, *37*(1), 186-194.
[58] Moulder, J.F.; Stickle, W.F.; Sobol, P.E.; Bomben, K.D.: *Handbook of X-Yay Photoelectron Spectroscopy*; Physical Electronics: Eden Prairie, Minnesota; **1992**.
[59] Joule, J.A.; Mills, K.: *Heterocyclic Chemistry*; John Wiley and Sons: Chichester, West Sussex, UK; **2010**.

[60] Wang, R.; Dobritsa, A.A.: Exine and aperture patterns on the pollen surface: Their formation and roles in plant reproduction. *Annu. Plant Rev. online*. **2018**, *1*(2):1-40.
[61] Frisch, M.J.; Trucks, G.W.; Schlegel, H.B.; Scuseria, G.; Robb, M.; Cheeseman, J.; et al.: Gaussian 09. Gaussian, Inc.: Wallingford, CT; **2010**.
[62] Gillies, E.R.; Frechet, J.M.: Dendrimers and dendritic polymers in drug delivery. *Drug Discov. Today*. **2005**, *10*(1), 35-43.
[63] Piffanelli, P.; Ross; J.H.E.; Murphy D.J.: Biogenesis and function of the lipidic structures of pollen grains. *Sex. Plant Reprod*. **1998**, *11*(2), 65-80.
[64] Ariizumi, T.; Toriyama, K.: Genetic regulation of sporopollenin synthesis and pollen exine development. *Annu. Rev. Plant Biol.* **2011**, *62*, 437-460.
[65] Brooks, J.; Shaw, G.: The post-tetrad ontogeny of the pollen wall and the chemical structure of the sporopollenin of *Lilium Henryi*. *Grana*. **1968**, *8*(2–3), 227-234.
[66] Wang, K.; Guo, Z.L.; Zhou, W.T.; Zhang, C.; Zhang, Z.Y.; Lou, Y.; et al.: The regulation of sporopollenin biosynthesis genes for rapid pollen wall formation. *Plant Physiol*. **2018**, *178*(1), 283-294.

Chapter 3

MALDI-TOF/TOF-MS/MS (negative ion mode) of French Oak Lignin: A Novel Series of Lignin and Tricin Derivatives attached to Carbohydrate and Shikimic acid Moieties.

This Chapter is reproduced with permission from Mikhael, A.; Jurcic, K.; Fridgen, T.D.; Delmas, M.; Banoub, J.: **Matrix-assisted laser desorption/ionization time-of-flight/time-of-flight tandem mass spectrometry (negative ion mode) of French Oak lignin: A novel series of lignin and tricin derivatives attached to carbohydrate and shikimic acid moieties**. *Rapid Commun. Mass Spectrom*. 2020; 34: e8841. https://doi.org/10.1002/rcm.8841

3.1. Abstract

We report the top-down lignomic analysis of the virgin released lignin (VRL) small oligomers, obtained from the French Oak wood, using a matrix-assisted laser desorption/ionization time-of-flight/time-of-flight (MALDI-TOF/TOF) instrument.

Twenty compounds were identified using MALDI-TOF-MS/MS of the virgin released lignin (VRL) extracted from the French Oak wood. Seven tricin derivatives and/or flavonoids, three syringylglycerol derivatives, two syringol derivatives, two flavonolignin derivatives, and six miscellaneous compounds: luteoferol, lariciresinol Isomer, 5-hydroxy guaiacyl derivative, syringyl -$C_{10}H_{10}O_2$ dimer, trihydroxy benzaldehyde derivative, and

aryl tetralin lignan derivative. Most of the identified compounds were in the form of carbohydrate and/or shikimic acid complexes.

This finding supports the presence of lignin-carbohydrate complexes in the isolated VRL. Also, these analyses showed that the French Oak lignin is abundant in syringol moieties present in the lignin syringyl-units or tricin derivatives. Moreover, the identification of some lignin-carbohydrate and /or flavonoid-shikimic acid complexes could provide new insight into the relation between the biosynthesis of lignin and tricin.

3.2. Introduction

The 3',5'-di-O-methyl ether of tricetin, commonly known as tricin, has been found in wheat straw lignin and other monocot samples. [1,2] It was also found that tricin cross-coupled with monolignols to form tricin (4-O-β) related dimers in the presence of peroxidase/hydrogen peroxide (biomimetic oxidation condition).[2] Ralph and coworkers have suggested that the role of tricin was only to initiate the formation of the lignin chain, as it could function as a potential nucleation site for the lignification of the monocots.[2] Also, it was proposed that the presence of tricin in lignin suggests that there is a correlation between the biosynthesis of lignin and tricin that could redefine the conventional lignin biosynthesis.[3] According to Lan *et al.*[2], it is recommended that the term flavonolignin must be used for lignin oligomers that contain tricin and /or other flavonoids in their structure.

It is well known that all the usual building blocks of the lignin oligomer can be produced by the general phenylpropanoid pathway.[4] In this pathway, phenylalanine acts

as a substrate for all plants, whereas tyrosine (Tyr) is the substrate for the grasses [5]. A major unique characteristic of this phenylpropanoid pathway is the presence of shikimate ester intermediates in which the *p*-coumaroyl shikimate ester becomes the preferred substrates for 3- hydroxylation by the enzyme C3H, to be converted to caffeoyl shikimate (Figure 3.1).[6,7] The caffeoyl shikimate esterase (CSE) can hydrolyze the product caffeoyl shikimate back to the corresponding caffeate.[6,7] Therefore, it was concluded that the occurrence of shikimate ester intermediates might have a role in the regulation of the phenylpropanoid pathway. [6] Additionally, it is essential to mention that shikimic acid derivatives are known to be merely intermediate in the lignin biosynthesis, and their presence was never described to exist in the final wood network and/or as an intermediate in the flavonoid's biosynthesis.[6]

For the last century, it was recognized that lignin was a heterogeneous phenylpropanoid macromolecule possessing a three-dimensionally branched architecture formed from cross-linked monomeric units (monolignols). [8] However, we have introduced a new model that originally stated there were no native lignin biomolecules existing as either a single huge biopolymer and/or a series of long polymers. [9] We proposed that native lignin biomolecules were composed of different lengths of short oligomers that are connected to cellulose and hemicellulose in a criss-cross pattern forming the network of wood.[9] We also suggested that accurate structural studies of lignins must be done on the series of constituting lignin oligomers released from the glycolignin (polysaccharide-lignin) complex, without any further purification or chemical

transformations. This type of released lignin oligomers is called virgin released lignin (VRL).[9]

Matrix-assisted laser desorption/ionization (MALDI) can be used for the rapid screening of the major components in complex lignin mixtures.[10] Also, it is accepted that the negative ion mode, with respect to the positive ion mode, would enhance the ionization of the analyte due to the formation of stable aromatic phenoxide ions.[10] Also, it is known that using the negative ion mode suppresses the formation of metal adducts that can add complexity to the MALDI- spectrum.[10] Additionally, in the present work, we have used the **"**Compagnie Industrielle de la Matière Végétale**"** (CIMV) extraction method for the extraction of the French Oak lignin from the vegetal biomass. This method uses only acidic solvents (HCOOH/CH_3COOH), which contains no salts, that could complicate the MALDI-MS spectrum.[11]

In this manuscript, we present the MALDI-TOF-MS (-ve ion mode) and high-energy collision CID-TOF/TOF-MS/MS analyses that were used to investigate the structure of the VRL lignins of the French Oak. Accordingly, we were capable of characterizing a series of diverse new compounds present in this studied complex mixture of the Oak lignin sample.

We described herein the first structural identification of novel tricin, and lignin derivatives attached to different shikimic acid and/or carbohydrate derivatives in the VRL lignin of the gymnosperm French Oak tree. Furthermore, we are also showing the presence of degraded flavonolignan or falvonosugars compounds, degraded or oxidized lignin derivatives lignans, and the flavonoid luteoferol.

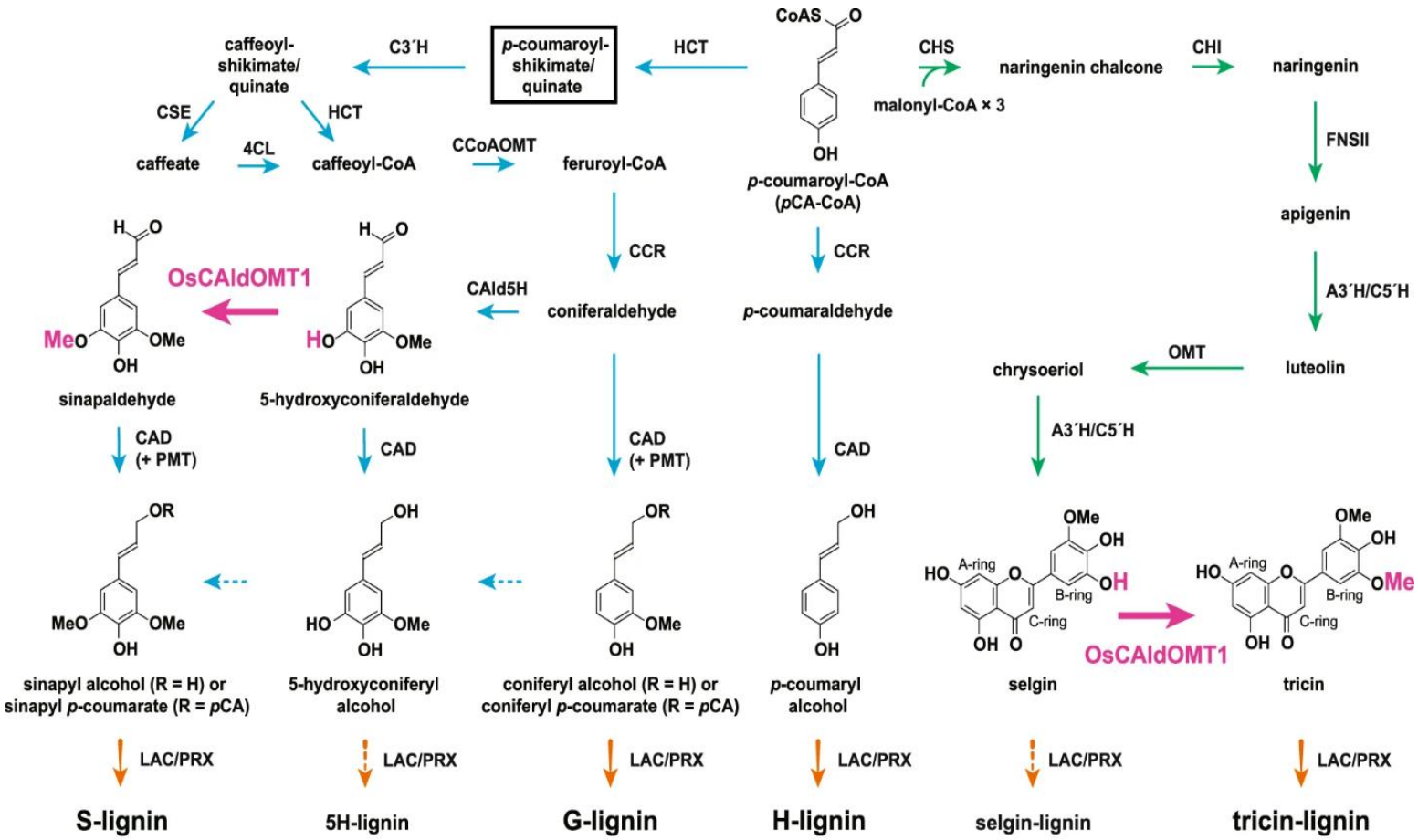


Figure 3.1. proposed biosynthesis of Lignin and tricin in rice adapted with permission from *Sci. Rep.* **2019**, *9*(1), 1-13. This figure shows that the shikimic and/or quinic acid derivatives are intermediates in the lignin biosynthesis. Also, it showed that these shikimic and/or quinic acid derivatives are not one of the final lignin monomers, and they are not intermediates in the tricin biosynthesis.

3.3. Experimental

3.3.1. Materials

The French Oak VRL sample was provided by Professor Michel Delmas, The "Compagnie Industrielle de la Matière Végétale" (CIMV). The Oak lignin powder was dissolved as 1 mg/mL in methanol (Fisher), chloroform (Sigma-Aldrich), and 1,4-dioxane (Sigma-Aldrich) at 1:1:1 ratio. The MALDI matrix, 2,5-dihydroxybenzoic acid, DHB (Sigma-Aldrich), was prepared as 20 mg/mL in acetonitrile (BDH Chemicals) and ethanol (Greenfield Global) at 1:1 ratio and acidified using trifluoroacetic acid, TFA (Fisher), at 0.1%. The sample was premixed with the matrix at a 1:1 ratio and 0.75 uL deposited onto the MALDI plate.

3.3.2. Lignin extraction

The lignin sample was extracted using the CIMV procedure, which selectively separates the cellulose, hemicellulose, and lignin at atmospheric pressure, and allows the destruction of the vegetable matter.[11] The catalyst solvent system used was a mixture of formic acid/acetic acid/water (30/50/20) which produced, after precipitation with water and filtration, the French Oak VRL lignin. In this method, formic acid hydrolyzes and destructure the vegetal biomass to cellulose, lignin, and hemicellulose. The presence of the acetic acid is to act as the solvent for the released fragments of lignin and hemicellulose portions. Lastly, the water presence helps the hydrolysis of hemicellulose and enhances the ionization of the acids during the cooking process.[11] One of the best advantages of the

CIMV solvolysis method is that is releases relatively intact lignin oligomers that can be used as a potential source for valuable aromatic chemicals.[11]

3.3.3. MALDI-TOF-MS/MS (Negative ion mode) and High energy-CID-TOF/TOF-MS/MS Analyses

Mass Spectrometric data were obtained using an AB Sciex 5800 MALDI TOF/TOF System (Framingham, MA, USA). Data acquisition and data processing were respectively done using a TOF TOF Series Explorer and Data Explorer (both from AB Sciex). The instrument is equipped with a 349 nm Nd:YLF OptiBeam On-Axis laser, and the laser pulse rate was 400 Hz. Reflectron negative mode was used, and MS acquisitions were acquired in the mass range from 100-1000. Reflectron and MSMS modes were externally calibrated at 50 ppm mass tolerance. Each MS mass spectrum was collected as a sum of 500 laser shots, while MSMS mass spectra were obtained as a sum of 900 shots. Raw files were viewed by using the Mmass program [12]. In addition, this latter program was used to find possible formulas for the selected ions. The search for possible formulas was done within +/- 0.2 error Da from the experimental masses. The resulting list of formulas was arranged according to their isotopic distribution fitting percentage to help us in assigning the most likely formulas for the studied ions. All assigned chemical structures are supported by tandem mass spectrometry, especially for the ones with low isotopic distribution fitting percentage.

3.3.4. Solid-state ^{13}C-NMR and 2D-Experiments

Solid-state 13C NMR experiments were performed on a Bruker AVANCE-400 Wide Bore NMR spectrometer (Bruker BioSpin, Rheinstetten, Germany) operating at a ^{13}C resonance frequency of 106 MHz and using a commercial solid probe H/X cross-polarization/magic angle spinning (CP/MAS) instrument (Bruker BioSpin). About 90 mg of the sample was placed in a zirconium dioxide rotor with an outer diameter of 4 mm, which was spun at 10 kHz at the magic angle.[11,13,14] The cross-polarization (CP) technique was applied with a ramped 1 H-pulse starting at 100% power and decreasing to 50% power. This procedure was used during the contact time of 2 ms, to circumvent the Hartmann-Hahn mismatches.[15] The experiments were performed at ambient temperature, and 16K scans were accumulated using a delay of 2 s. The ^{13}C chemical shifts were referenced to tetramethylsilane and calibrated with the glycine carbonyl signal, set at 176.5 ppm.[13,14, 16] For the 2D ^{1}H-^{13}C HSQC experiment, the French oak VRL acetylated lignin sample was used to overcomes the solubility issues as reported before for lignin samples.[17] 2D-HSQC measurements were performed according to the method published before by Yuan et al. [18]

3.4. Results and Discussion

3.4.1. ^{13}C-NMR and 2D-Experiments obtained for the VRL French Oak

In this section, we will focus on the supporting elements that indicate the presence of the syringol moiety as it is identified in the majority of the precursor ions studied in this manuscript. The solid-state ^{13}C-NMR of the French oak VRL sample showed *interalia* two major peaks. The first peak was identified at 57 ppm, indicating the presence of aromatic

methoxy groups.[19] This supports the existence of guaiacyl, syringyl lignin units and/or tricin. The second peak was identified at 150 ppm, supporting the presence of C3 and C5 of syringyl and/or tricin units, as demonstrated in Figure 3.2.[19] Other solid-state ^{13}C-NMR peaks are listed in Table 3.1.

As additional support for the presence of syringyl lignin units, the aromatic region of the 2D ^{1}H-^{13}C HSQC experiment of the acetylated VRL oak lignin was examined (Figure 3.3). Two important cross-peaks region at approximately δ_C/ δ_H 104/6.75 and δ_C/ δ_H 105.5/7.35 were assigned as $C_{2,6}$–$H_{2,6}$ in syringyl units (S) and $C_{2,6}$–$H_{2,6}$ in oxidized syringyl units (S').[20]

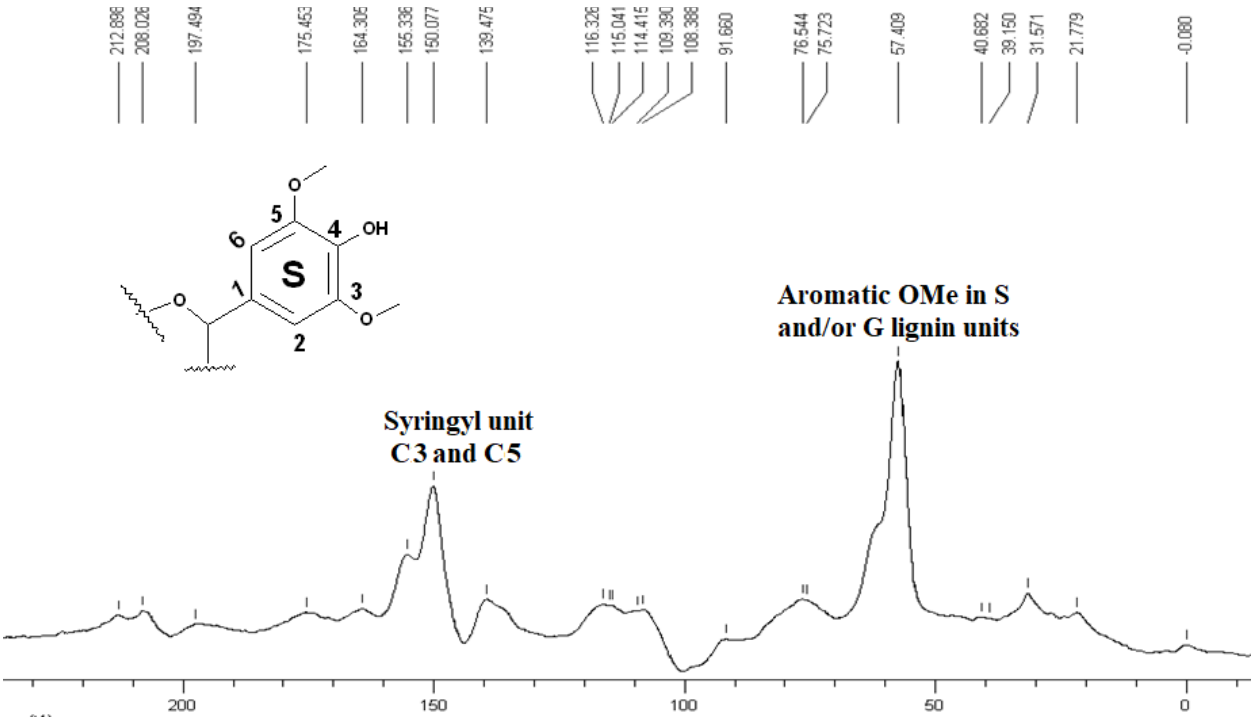


Figure 3.2. Solid state ^{13}C- NMR of the French oak VRL

Table 3.1. Summary of the solid state ^{13}C- NMR peaks of the OaK VRL

C-13 NMR shift regions and/or peaks	Possible Assignment	Reference
20-40 ppm	Aliphatic carbons	21
57 ppm	Aromatic OMe	19
75-90 ppm	Aliphatic Carbons-O-Cα or Cβ or Cγ	22
108-109 ppm	Aromatic C-2 of G lignin	19
114-116 ppm	Aromatic C-3 and C-5 of H lignin	19
139 ppm	Aromatic C-1 and C-4 of S-Lignin	23
150 ppm	Aromatic C-3 and C-5 of S unit	19
155 ppm	Aromatic C-O of H unit	19
164 ppm	Tricin C-2 and C-7 or Conjugated C=O	24,25
175-200 ppm	Different types of C=O	21

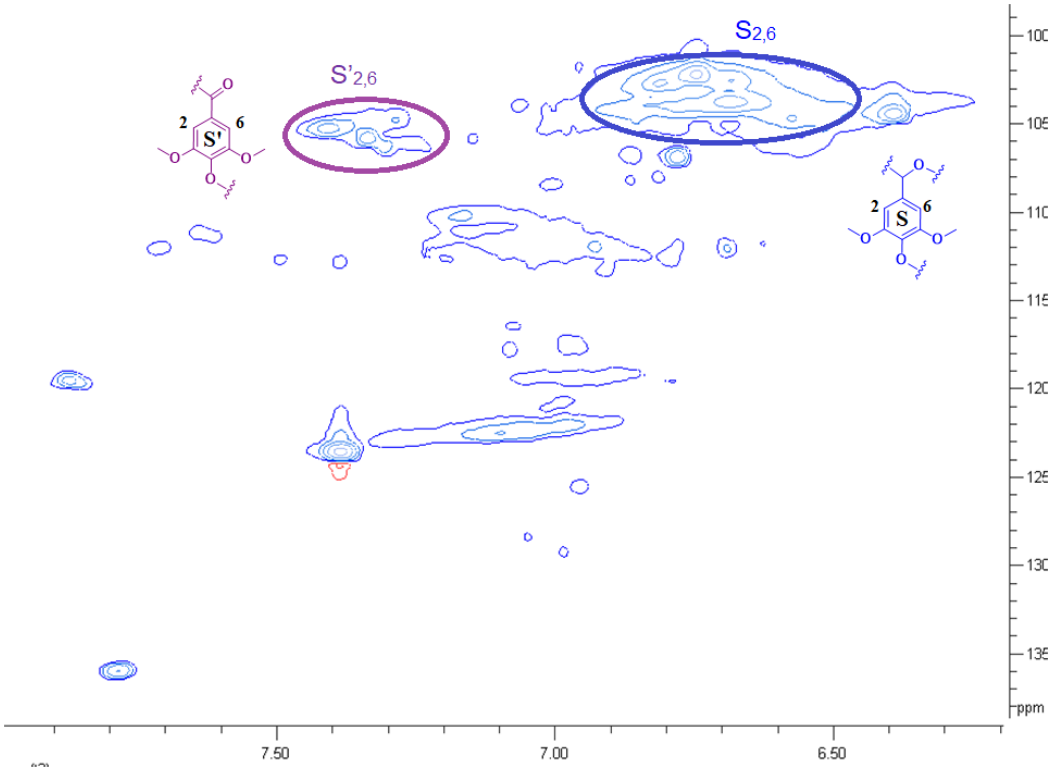


Figure 3.3. 2D ^{1}H-^{13}C HSQC aromatic region of the acetylated French oak VRL

3.4.2. MALDI-TOF-MS of the VRL Oak

In this work, we have used DHB as a matrix, as a recent study by Qi and Volmer [10] showed that DHB displayed a very diverse and distinctive distribution of lignin compounds. The MALDI-TOF-MS (- ion mode) spectrum of the French Oak lignin sample gave a series of ions *inter alia* at *m/z* 289.03, 307.04, 315.05, 329.04, 345.02, 359.97, 368.97, 382.96, 417.07, 417.07, 499.03, 535.97, 544.97, 552.95, 558.97, 568.93, 590.90, 653.05, 661.04 and 675.03 (Figure 3.4, and Table 3.2). Please note that this MALDI-TOF-MS was also descrambled by enlarging the different *m/z* regions to indicate the complexity of this spectrum (Figure A2.1).

The identity of the most abundant ions obtained from this complex MALDI-TOF mass spectrum will provide distinguishing information about the most important structural units of this complex mixture of native Oak lignin oligomers (VRLs). However, due to the intricacy and presence of various isobaric lignin oligomers derived from different lignol units and different linkages, this is not a straightforward task. To complicate matters further, the possibility of the presence of tricin (flavonoids) and lignin-carbohydrate and/or shikimic acid compounds adds to the complexity of the structural elucidation of this VRL Oak sample. For these reasons, accurate mass assignments are challenging because of multiple isobaric interferences (complex overlapping isotopic distribution) and centroid mass shifts.[26] However, a reasonable MS/MS fragmentation mechanism can be used to deduce the possible structures of the studied ions. This information will also provide more support to the suggested formulas, and it will help to decreases the number of possible candidates that could be attributed to each ion in the spectrum.[27]

In this work, the direct analysis of the complex lignin mixture was undertaken without any chromatographic pre-separation. The identification of the following defined twenty oligomers was performed according to the presence of one heteroatom (oxygen), a mass error within 0.2 Da, isotopic distributions of the precursor ions, the possible combinations of lignol, flavonoids and carbohydrates units, and reasonable MS/MS fragmentation patterns. Also, the level of unsaturation, also called the double bond equivalent (DBE), was calculated for the precursor ions to determine the number of unsaturation.

It is important to mention that the MALDI-MS analysis of this complex lignin mixture resembles the MALDI-MS analysis of a complicated mixture of pigments where the error between the calculated and observed *m/z* values could be up to 0.1 Da and in some cases from 0.01 to 0.25 Da as shown in previous literature.[28,29] Also, it should be understood that the structure elucidation of unknown complex mixtures remains a challenge despite the development of advanced mass spectrometry instruments.[30] According to Kind et al.,[30], seven rules were developed to select the most likely formula out of a thousand possible candidates in case of analyzing unknown complex mixtures. From the application of these rules, it was deduced that isotopic distribution fitting percentage is more important than the mass accuracy as it has more impact in constraining the number of possible formulas.[30] However, in the case of complex overlapping isotopic distributions, tandem mass spectrometry is essential to deduce the most likely structure.[26,27]

The LDI-spectrum of the DHB matrix only was compared to the MALDI-MS of the spotted sample to investigate the regions of strong matrix interferences (Figure A2.2). There was a strong overlap or interference from the matrix with one of the studied ions at *m/z* 315, which will be discussed later in section 3.7.1.

For the purpose of simplification of this manuscript, we have gathered all structurally similar types of lignin oligomer derivatives in separate sections, as shown here below.

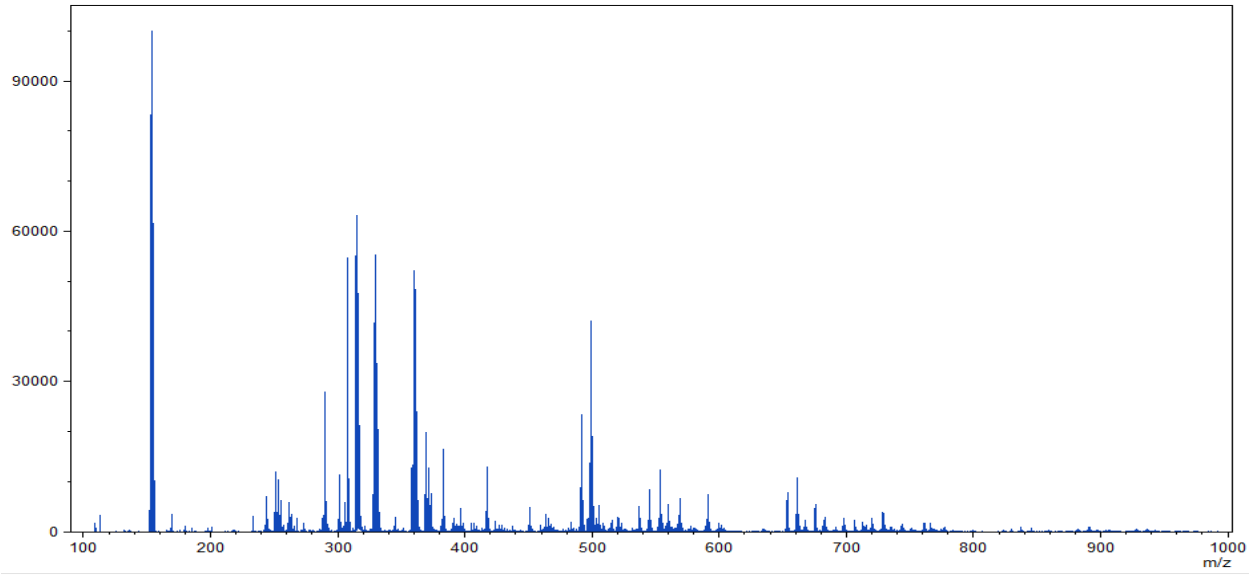


Figure 3.4. MALDI-TOF-MS of the lignin oligomers extracted from French Oak wood from *m/z* 100-1000

Table 3.2. Lignin oligomers identified in the negative ion mode MALDI-TOF-MS of the French Oak VRL lignin

Ion	Chemical Formula	Experimental *m/z*	Calc. *m/z*	Error (*m/z* units)	DBE	Pattern	MS/MS
1	$[C_{24}H_{20}O_{12} - H]^-$	499.03	499.09	-0.06	15.5	86.5	345, 315,153
2	$[C_{24}H_{24}O_{14}]^{\bullet-}$	535.97	536.11	-0.14	13.0	83.7	500, 493, 483, 383, 375, 367, 345, 153
3	$[C_{24}H_{26}O_{14} + Na - 2H]^-$	558.97	559.11	-0.14	12.5	79.8	537, 523, 509, 399, 345, 153
4	$[C_{27}H_{28}O_{15} - H]^-$	590.90	591.13	-0.23	14.5	97.9	559, 537, 531, 431, 345, 153
5	$[C_{28}H_{30}O_{18} - H]^-$	653.05	653.13	-0.08	14.5	99.2	617, 602, 493, 345, 315, 153
6	$[C_{29}H_{26}O_{18} - H]^-$	661.04	661.10	-0.06	16.5	98.5	503, 315
7	$[C_{31}H_{32}O_{17} - H]^-$	675.03	675.16	-0.13	16.5	95.6	521, 499, 491, 315
8	$[C_{18}H_{24}O_9 - H]^-$	382.96	383.13	-0.17	7.5	99.0	353, 351, 339, 337, 335, 323, 315, 243, 227, 215, 201, 179, 153
9	$[C_{23}H_{22}O_{16} - H]^-$	552.95	553.08	-0.13	13.5	94.5	537, 493, 383, 153
10	$[C_{25}H_{30}O_{15} - H]^-$	568.93	569.11	-0.22	11.5	97.1	539, 509. 407. 389, 383, 345, 315, 281, 153
11	$[C_{15}H_{16}O_7 - H]^-$	307.04	307.08	-0.04	8.5	97.8	261, 153
12	$[C_{14}H_{18}O_9 - H]^-$	329.04	329.09	-0.05	6.5	59.9	283, 153
13	$[C_{17}H_{16}O_6 - H]^-$	315.05	315.09	-0.04	10.5	52.8	289, 273, 179, 153, 135
14	$[C_{23}H_{24}O_{12} - H]^-$	491.05	491.12	-0.07	12.5	99.3	465, 437, 345, 329, 311, 315, 153
15	$[C_{15}H_{14}O_6 - H]^-$	289.03	289.07	-0.04	9.5	95.6	273, 245, 175, 153, 135
16	$[C_{14}H_{18}O_{10} - H]^-$	345.02	345.08	-0.06	6.5	89.2	315, 301, 153
17	$[C_{20}H_{24}O_6]^{\bullet-}$	359.97	360.16	-0.18	9.0	40.3	345. 327, 317, 301, 267, 225, 181, 153
18	$[C_{21}H_{22}O_6 - H]^-$	368.97	369.13	-0.16	11.5	56.6	317, 225, 153
19	$[C_{17}H_{22}O_{12} - H]^-$	417.07	417.10	-0.03	7.5	98.8	389, 315, 233, 153, 125
20	$[C_{31}H_{30}O_9 - H]^-$	544.97	545.19	-0.22	17	94.2	493, 383, 360, 345, 153

3.4.3. High Energy CID-TOF/TOF-MS/MS of the Oak VRL Tricin and/or related flavonoids

In this rationale, we have identified seven ions composed of tricin, or related compounds (flavonoids), attached to the following carbohydrates or shikimic acid derivatives: dehydroshikimic acid (*m/z* 499.03), acetylated penturonic acid (*m/z* 535.97), methylated hexuronic acid (*m/z* 558.97), disaccharide tetrose – ketohexose derivative (*m/z* 590.90), disaccharide hexose lactone – penturonic acid (*m/z* 653.05), ascorbic acid - glucascorbic acid (*m/z* 661.04), and quinic acid-deoxyshikimic acid (*m/z* 675.03).

3.4.3.1. Assignment of the precursor ion at *m/z* 499.03

The formula $[C_{24}H_{20}O_{12} - H]^-$ (DBE=15.5) was assigned tentatively to the novel deprotonated ion *1* at *m/z* 499.03 (Figure 3.4, Figure A2.1, and Table 3.2). This deprotonated molecule appeared to be composed of tricin attached to 3-dehydroshikimic acid.

It should be noted that this is the first identification in the literature in which tricin is attached to a shikimic acid derivative. Also, as mentioned previously in the introduction section, shikimic acid derivatives were described previously to exist as an intermediate in the lignin biosynthesis.[6] On the other hand, shikimic acid derivatives have never been described to be an intermediate in the tricin biosynthesis.[6] Therefore, the identification of tricin attached to deoxyshikimic acid may help biochemists to find an answer to how the lignin and tricin biosynthesis could be linked.

The product ion scan of the precursor ion *1* gave the product ion at *m/z* 345.10 by the neutral loss of the 3-dehydroshikimic acid derivative, which resembles the loss of a neutral sugar moiety in many MS/MS studies in the literature (Figure 3.5 and Scheme 3.1). [31]

It should be noted that in Scheme 3.1, the nomenclature developed for flavonoids by Mabry and Markham was used to describe rings A, B, and C in the flavonoid structure.[32] Lastly, the product ion scan of the precursor ion *1* experience ring C contraction to give the ion at *m/z* 315, as shown in Scheme 3.1. Lastly, the cleavage of the bond between ring B and ring C in the precursor ion *1* leads to the formation of the ion at *m/z* 153.

It is important to note that the precursor ion *1* at *m/z* 499 ion was not detected in the recorded LDI-TOF-MS of the matrix only (Figure A2.2). However, Schiller *et al.*, [33] proposed that this ion can be attributed to the DHB matrix cluster with the formula $[3DHB+K-2H]^-$. Nevertheless, the product ion scan of ion *1* showed the formation of two major ions at *m/z* 345 and *m/z* 315. If we assume that the precursor ion at *m/z* 499 was an adduct cluster of the matrix, we could very well consider the presence of the product ion at *m/z* 345 as a legitimate product ion that could originate from the [2DHB+K-2H] formula. This means that the ion at *m/z* 345 can be formed by the loss of one neutral DHB molecule (-154 Da) from the precursor ion at *m/z* 499. Fortunately, there is no logical way to explain the other formed product ion at *m/z* 315; this ion cannot be formed from the fragmentation of this matrix cluster $[3DHB+K-2H]^-$. This reasoning indicates that the ion at *m/z* 499 obviously belong to our proposed formula $[C_{24}H_{20}O_{12} - H]^-$.

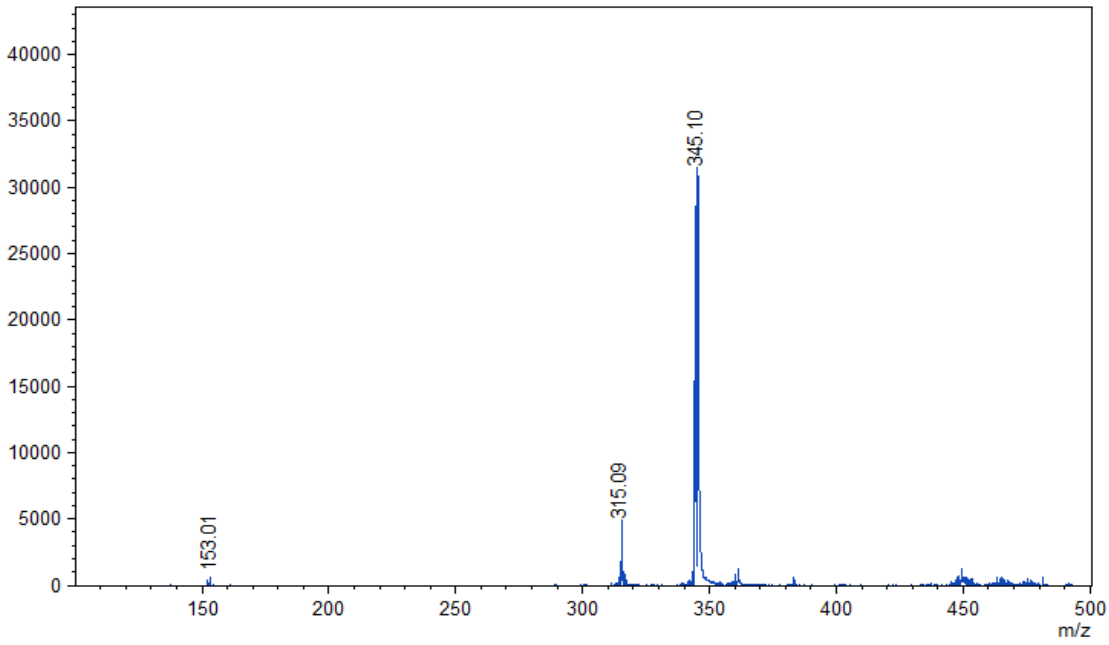
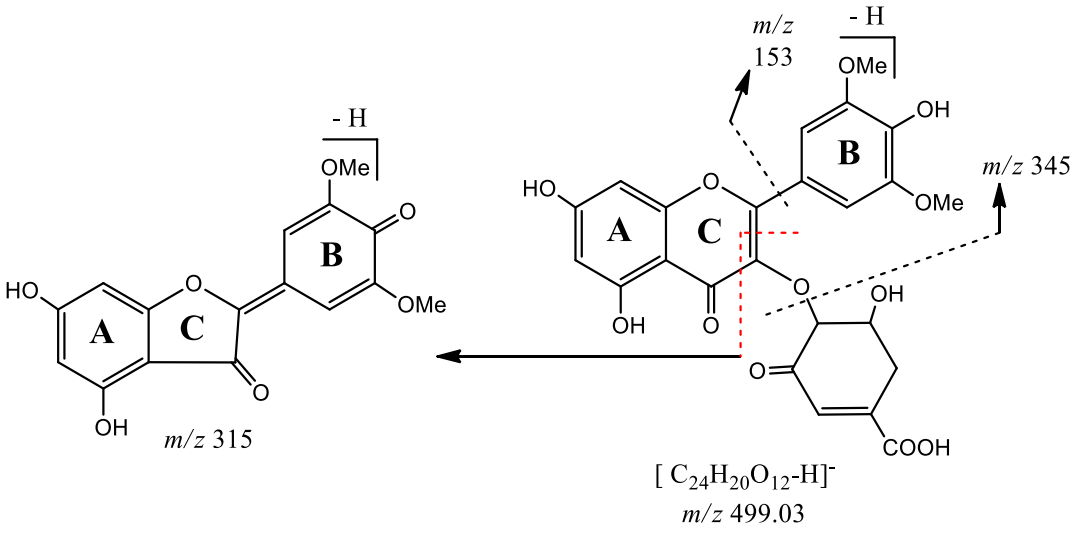


Figure 3.5. Product ion scan of the precursor ion *1* at *m/z* 499.03



Scheme 3.1. High energy CID-MS/MS fragmentation mechanism of the precursor anion *1* at *m/z* 499.03

3.4.3.2. Assignment of the precursor ion at *m/z* 535.97

The $[C_{24}H_{24}O_{14}]^{\bullet-}$ formula (DBE=13) was assigned to the radical ion *2* at *m/z* 535.97 (Figure 3.4, Figure A2.1, and Table 3.2). This radical ion *2* is composed of the tricin derviative connected to an acetylated penturonic acid. The product ion scan of *2* created the ion at *m/z* 500 by the simultaneous loss of two water molecules from the sugar moiety (Figure 3.6 and Scheme 3.2). This latter product ion at *m/z* 500 loses one phenolic oxygen to yield the ion at *m/z* 483. It should be noted that this kind of oxygen loss has been reported before for phenolic compounds.[34]

Furthermore, the radical ion *2* is subjected to the loss of an acetyl group radical (-43 Da) to yield the carboxylate ion at *m/z* 493, supporting the presence of the acetyl group on the acetylated sugar. This later product ion at *m/z* 493 experience a cleavage in the ring C of the tricin structure to yield the ion at *m/z* 367.

Moreover, the acetylated penturonic acid moiety in the precursor ion *2* experiences a ring cleavage to produce the ion at *m/z* 375. In addition, the simultaneous cleavage of the C1-O and the C1-C2 bond in the sugar ring lead to the formation of the radical ion at m/z 360. Furthermore, the loss of the sugar moiety from the precursor ion *2* gives the ion at *m/z* 345. Also, the ring C of the tricin moiety in the precursor ion *2* experiences an RDA reaction to produce the ion at *m/z* 383. Lastly, the cleavage of the bond connecting ring B and C in the precursor ion *2* lead to the formation of the ion syringol at *m/z* 153.

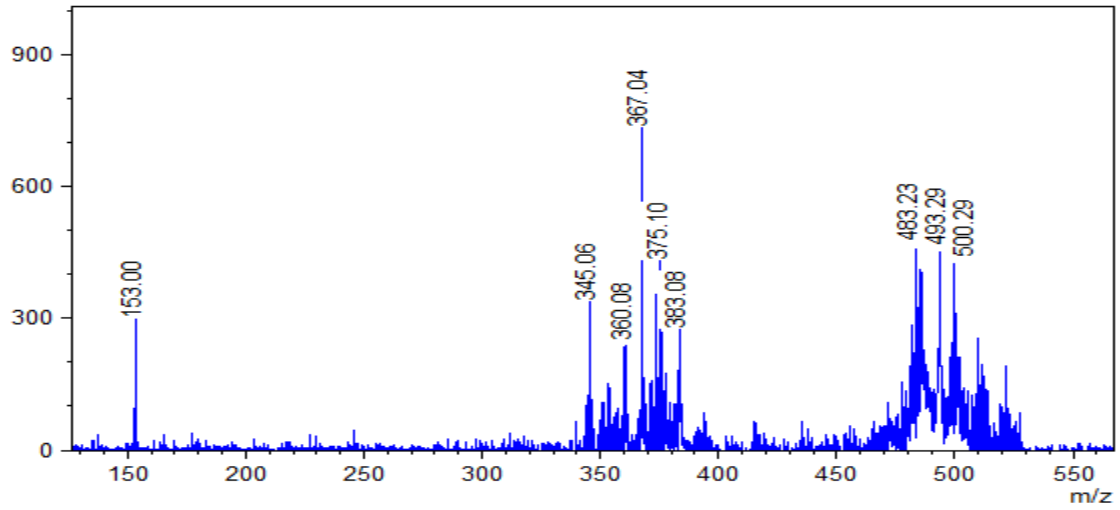
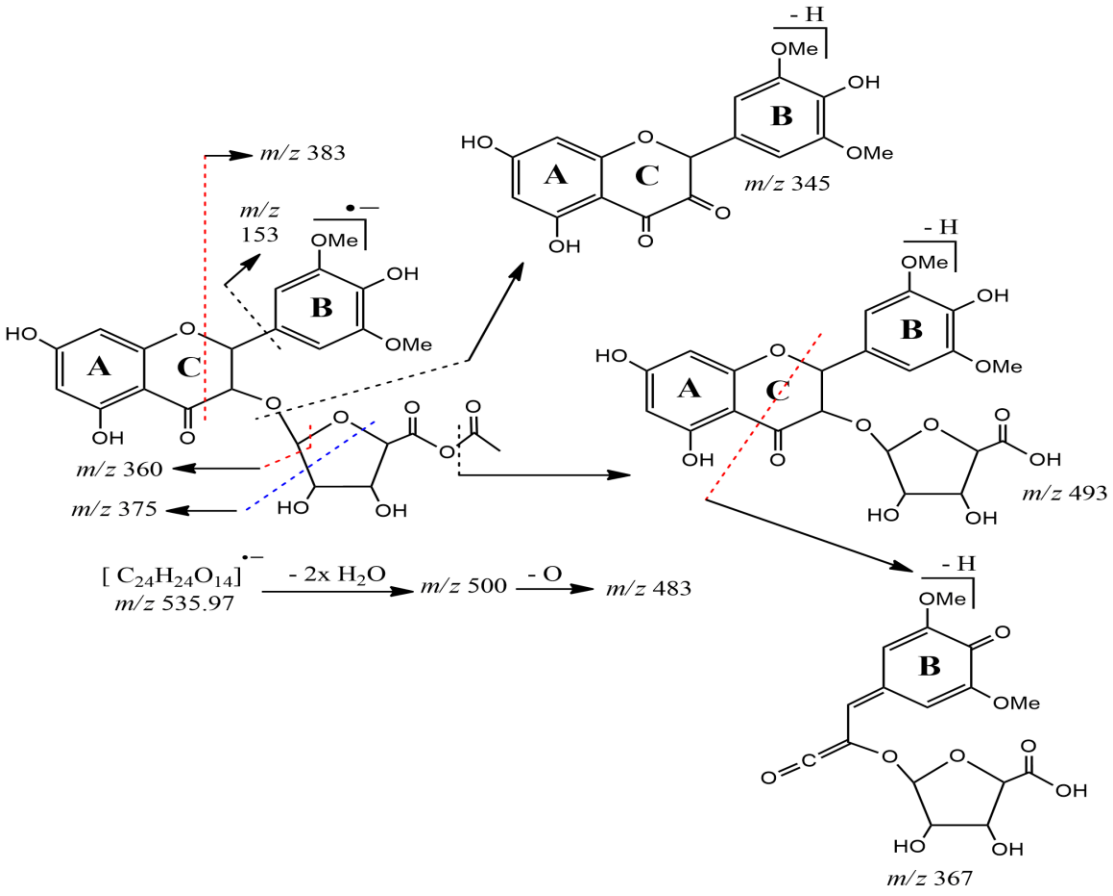


Figure 3.6. Product ion scan of the precursor ion at *m/z* 535.97



Scheme 3.2. High energy CID-MS/MS fragmentation mechanism of the precursor ion *2* at *m/z* 535.97

3.4.3.3. Assignment of the precursor ion at *m/z* 558.97

The formula $[C_{24}H_{26}O_{14} + Na - 2H]^-$ (DBE=12.5) was assigned to the precursor ion *3* at *m/z* 558.97 (Figure 3.4, Figure A2.1, and Table 3.2). This precursor ion *3* is composed of a tricin derivative connected to a methylated hexuronic acid. Unexpectedly, this ion *3* appeared to contain sodium, which origin was attributed to a possible minor contaminant in the chemicals/glassware used in the experimental procedure.

The product ion scan of this precursor ion showed the loss of 22 Da which accounts for the insource replacement of Na (-23 Da) in the structure of this ion with hydrogen (+1 Da) to create the ion at *m/z* 537 which formula was calculated as $[C_{24}H_{26}O_{14} - H]^-$ (Figure 3.7 and Scheme 3.3). To our knowledge, this is the first time that such an ion was described. Similarly, the product ion at *m/z* 537 undergoes phenolic ring contraction through the loss of carbon monoxide to give the ion at *m/z* 509. Additionally, the product ion at *m/z* 537 loses 44 Da ($-CO_2$) from the hexuronic acid moiety to yield the ion at *m/z* 493. This latter ion loses a carbon monoxide molecule through the phenolic ring contraction to create the ion at *m/z* 465. Furthermore, the ring C of the tricin derivative present in the ion at m/z 537 undergoes a ring cleavage to give the ion at *m/z* 399, as shown in Scheme 3.3. Moreover, the simultaneous loss of the sugar moiety and one hydrogen molecule from the ion at *m/z* 537 leads to the creation of the ion at *m/z* 345.

In addition, the precursor ion *3* loses 36 Da by the simultaneous elimination of two water molecules to form the product ion at *m/z* 523. Finally, the cleavage of the bond between ring C and ring B in the ion *3* leads to the formation of the syringol ion at *m/z* 153.

Needless to say, that the position of the methoxy group in the methylated hexuronic acid is tentative, as evidently it could be present in the other O-3 and O-4 positions.

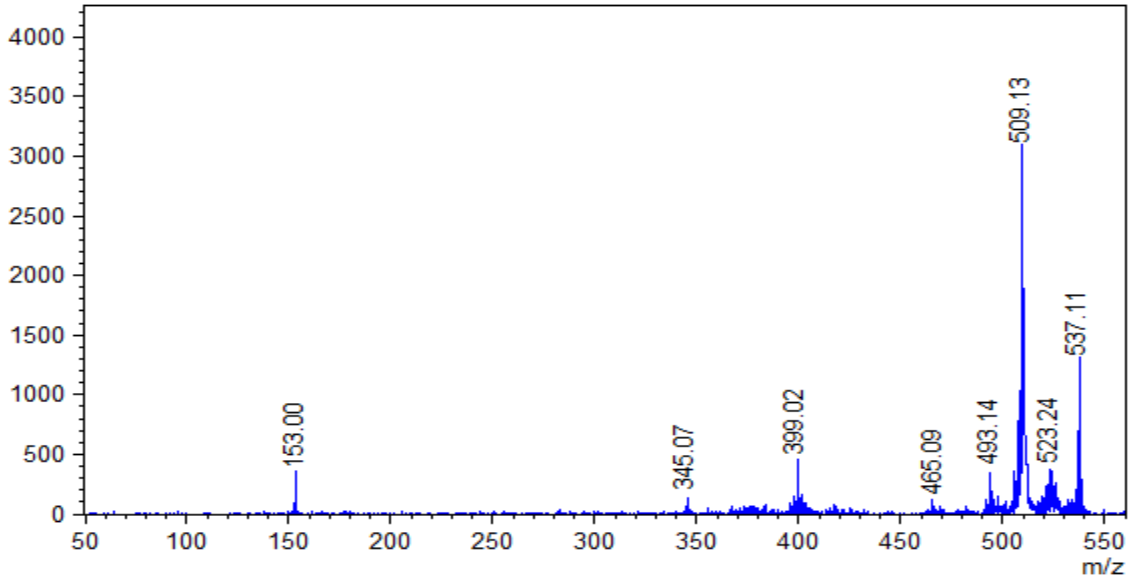
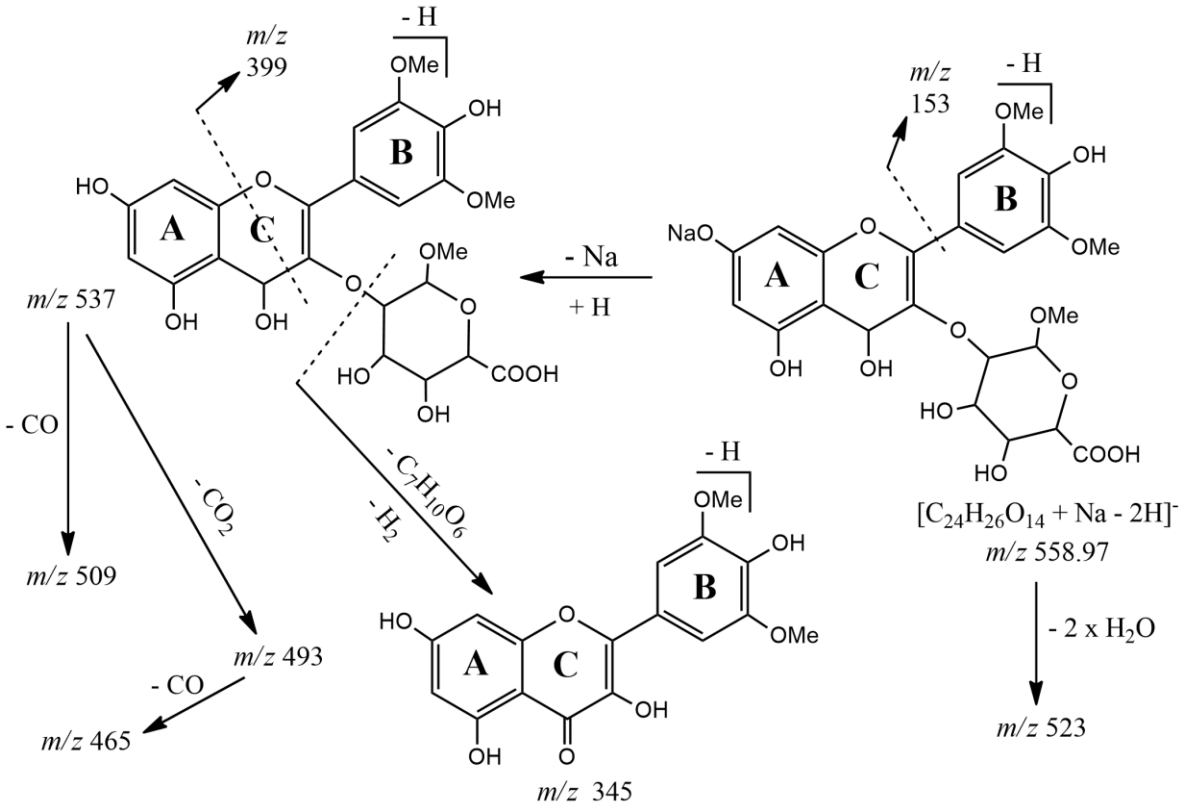


Figure 3.7. Product ion scan of the precursor ion *3* at *m/z* 558.97



Scheme 3.3. High energy CID-MS/MS fragmentation mechanism of the precursor ion *3* at *m/z* 558.97

3.4.4. Assignment of the precursor ion at *m/z* 590.90

The $[C_{27}H_{28}O_{15} - H]^-$ formula (DBE= 14.5) was assigned to the deprotonated molecule *2* at *m/z* 590.90 (Figure 3.4, Figure A2.1, and Table 3.2).This deprotonated molecule *2* is tentatively composed of a tricin derivative connected to disaccharide composed of a tetrose sugar linked to a ketohexose sugar derivative. The product ion scan of this ion *4* was initiated by the loss of 54 Da from the ketose derivative moiety to yield the ion at *m/z* 537 (Figure 3.8 and Scheme 3.4). Additionally, the precursor ion *4* experience a ring contraction in the ketohexose moiety (-60 Da) to yield the ion *m/z* 531. Furthermore, the precursor ion *4* experiences the loss of one methanol molecule from the ketohexose

derivative ring to produce the ion at *m/z* 559. The sequential loss of the two sugar moieties from the precursor ion *4* gives the ions at *m/z* 431 and 345, respectively. It should be noted that the loss of the disaccharide was followed by the loss of one hydrogen molecule to yield the ion at *m/z* 345, as shown in Scheme 3.4. Lastly, the cleavage of the bond connecting rings B and C of the tricin moiety in the precursor ion *4* leads to the formation of the ion of syringol at *m/z* 153.

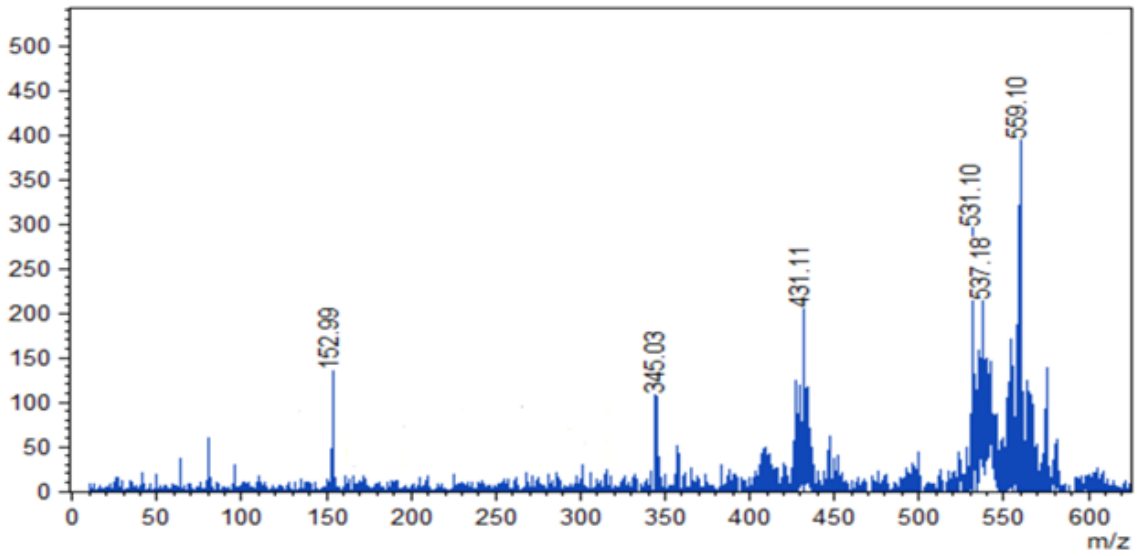
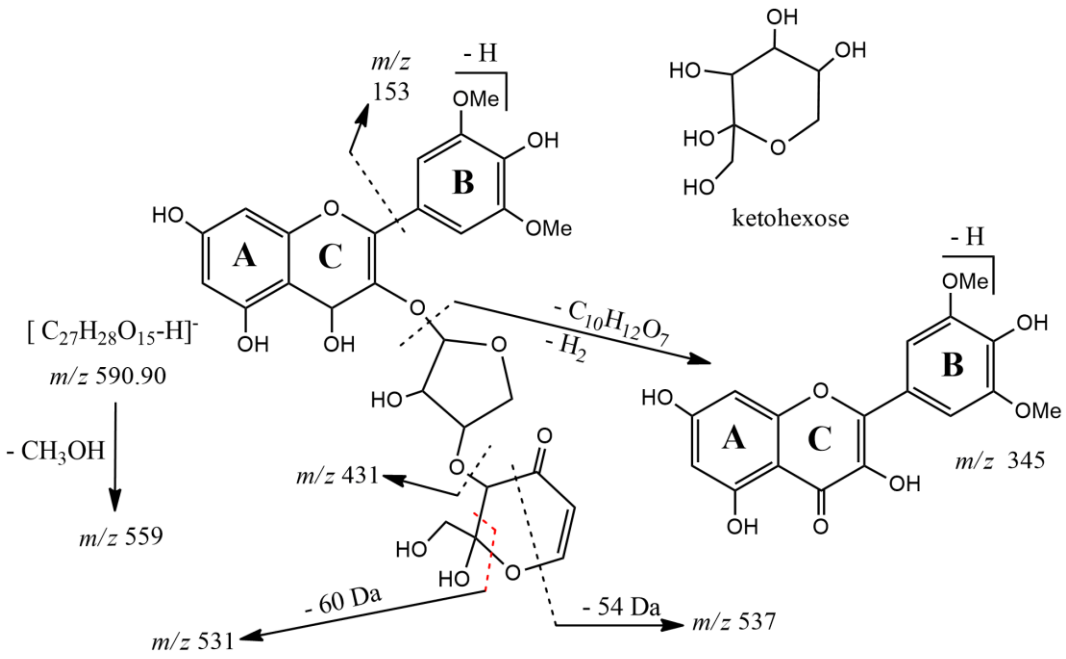


Figure 3.8. Product ion scan of the precursor ion *4* at *m/z* 590.90



Scheme 3.4. High energy CID-MS/MS fragmentation mechanism of the precursor ion *4* at *m/z* 590.90

3.4.3.5. Assignment of the precursor ion at *m/z* 653.05

The $[C_{28}H_{30}O_{18} - H]^-$ formula (DBE= 14.5) was assigned to the precursor ion *5* at *m/z* 653.05 (Figure 3.4, Figure A2.1, and Table 3.2). This precursor ion *5* was composed of a tricin derviative connected to two different sugar moieties: a hexose lactone and a penturonic acid. The product ion scan of this ion *5* showed the simultaneous loss of two water molecules to yield the ion at *m/z* 617 (Figure 3.9 and Scheme 3.5). This latter product ion at *m/z* 617 loses one methyl radical to yield the ion at *m/z* 602. The precursor ion *5* fragments by loss of the sugar lactone to from the ion at *m/z* 493. This latter product ion at *m/z* 493 can further fragment by the loss of the penturonic acid sugar moiety to produce the ion at *m/z* 345. Also, the product ion at *m/z* 493 can undergo a ring C cleavage to yield

the ion at *m/z* 315. Lastly, the cleavage of the bond connecting ring B and C of the tricin moiety in the precursor ion *5* lead to the formation of the ion of syringol at *m/z* 153.

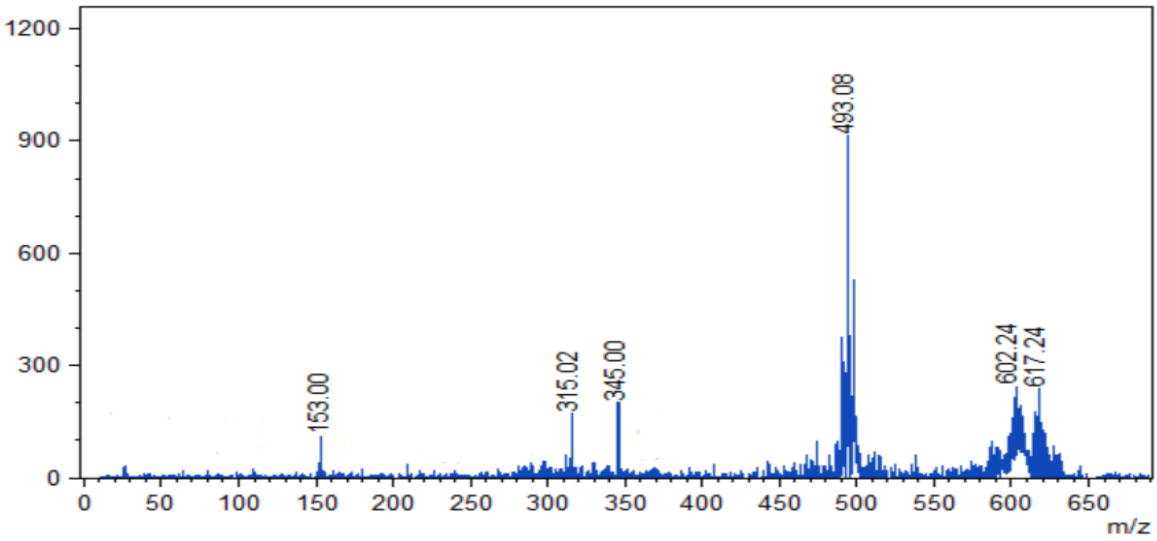
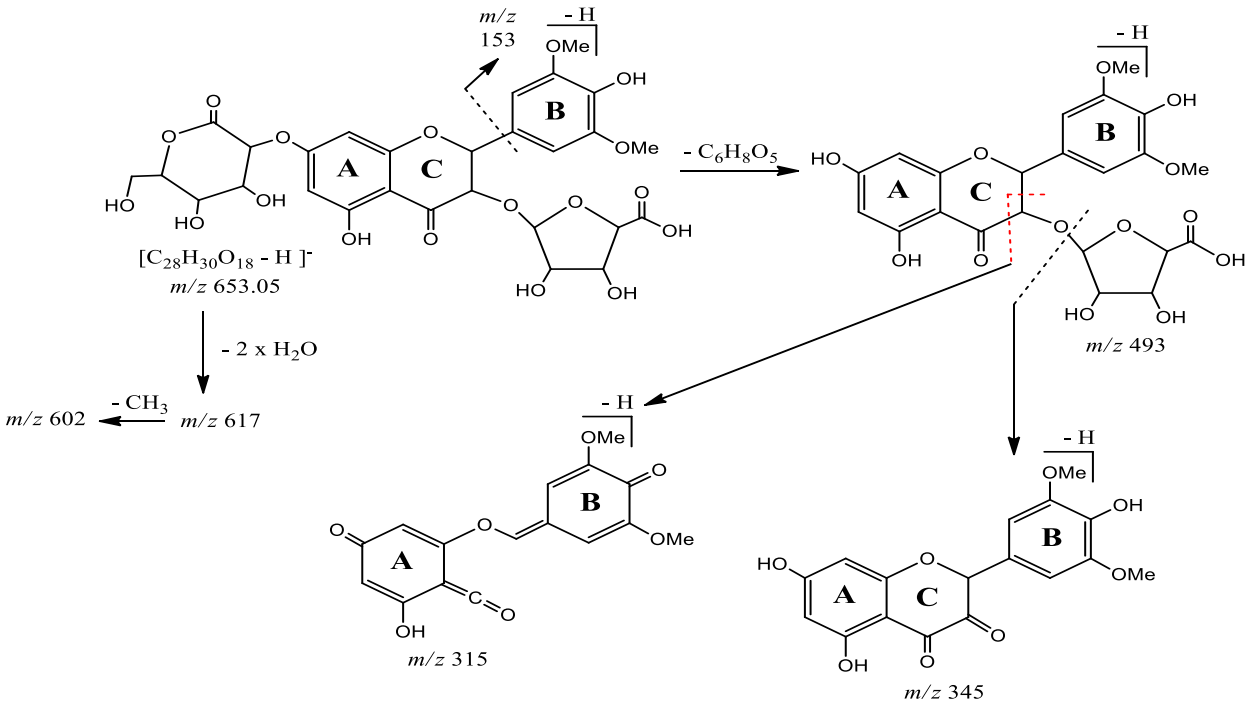


Figure 3.9. Product ion scan of the precursor ion *5* at *m/z* 653.05



Scheme 3.5. High energy CID-MS/MS fragmentation mechanism of the precursor ion *5* at *m/z* 653.05

3.4.3.6. Assignment of the precursor ion at *m/z* 661.04

The $[C_{29}H_{26}O_{18} - H]^-$ formula (DBE=17.5) was assigned to the precursor ion *6* at *m/z* 661.04 (Figure 3.4, Figure A2.1, and Table 3.2). This precursor ion *6* is composed of selgin (Demethylated tricin) connected to two different sugar derivatives: glucascorbic acid and ascorbic acid. The precursor ion 6 fragments by loss of the ascorbic acid moiety to form the product ion at *m/z* 503 (Figure 3.10 and Scheme 3.6). The loss of both sugar moieties from the precursor ion *6* produces the ion at *m/z* 315.

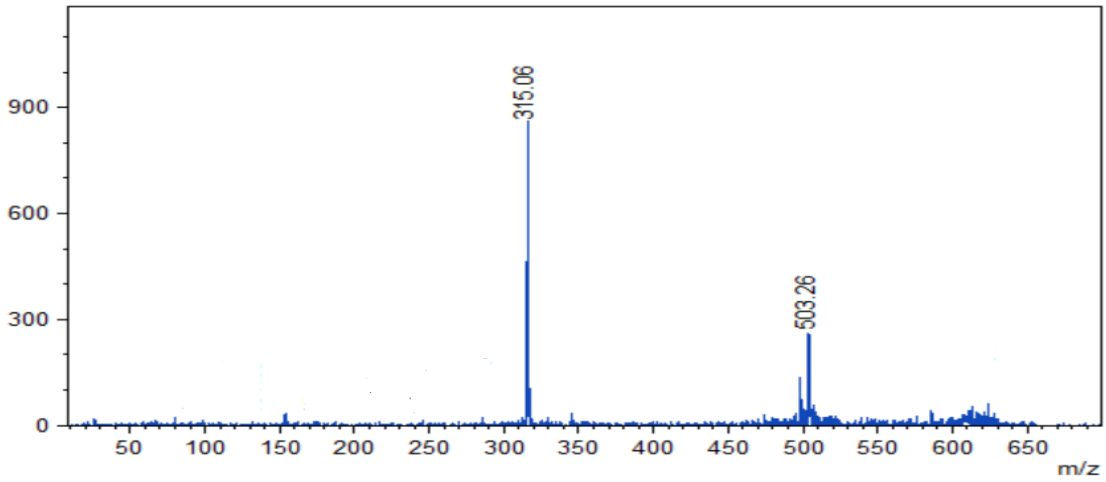
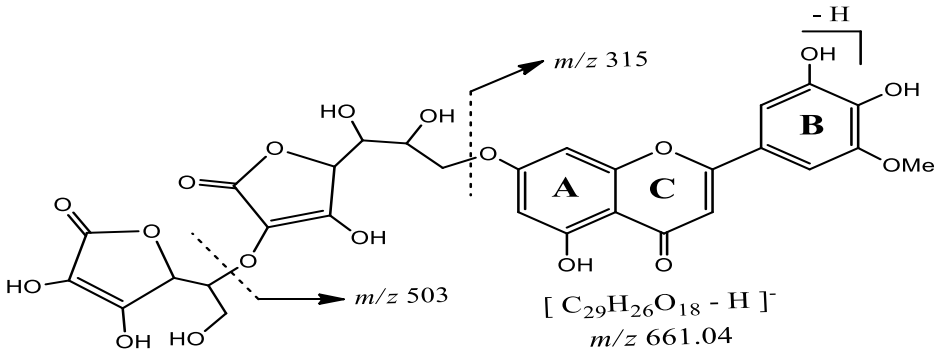


Figure 3.10. Product ion scan of the precursor ion *6* at *m/z* 661.04



Scheme 3.6. High energy CID-MS/MS fragmentation mechanism of the precursor ion *6* at *m/z* 661.04

3.4.3.7. Assignment of the precursor ion at *m/z* 675.03

The formula $[C_{31}H_{32}O_{17} - H]^-$ (DBE=16.5) was assigned to the ion *7* at *m/z* 675.03 (Figure 3.4, Figure A2.1, and Table 3.2). This complex precursor ion *7* contains a tricin derivative with an extra hydroxyl group in ring A of the flavonoid structure. This tricin skeleton is connected to both quinic acid and dehydroshikimic acid rings. It should be noted that quinic acid has been described before to exist in chlorogenic acid (caffeoyl quinate), which is considered as an intermediate in the lignin biosynthesis, but it was never described to be related and/or attached to flavonoids.

Again, as we mentioned before, for the ion *1* at *m/z* 499 (tricin + deoxyshikimic acid), the presence of quinic and/or shikimic acid derivatives attached to tricin, could provide a new idea about the relation between lignin and tricin biosynthesis.

The precursor ion *7* loses the dehydroshikimic acid moiety to form the product ion at *m/z* 521 (Figure 3.11 and Scheme 3.7). Also, the loss of the quinic acid moiety from the precursor ion *7* leads to the formation of the ion at *m/z* 499.

Moreover, the ring contraction of the central ring C of the flavonoid skeleton presented in the precursor ion *7* leads to the creation of the product ion at *m/z* 491. This latter ion at *m/z* 491 loses the quinic acid moiety to form the aromatic ion at *m/z* 315.

Lastly, although this ion was not detected in the LDI of the matrix only spectrum. We found that it is worth to mention that this ion was proposed by Schiller et al. [33] to be composed of the matrix cluster $[4DHB+ Na + K-3H]^-$. If we consider the presence of this matrix cluster, it can form only two product ions at *m/z* 521 $[3DHB+ Na + K-3H]^-$ and *m/z* 499 $[3DHB+ K-2H]^-$, while the other product ions at *m/z* 315 and 491 cannot be formed

from this matrix cluster formula. This supports that this ion at m/z 675 originates from the sample.

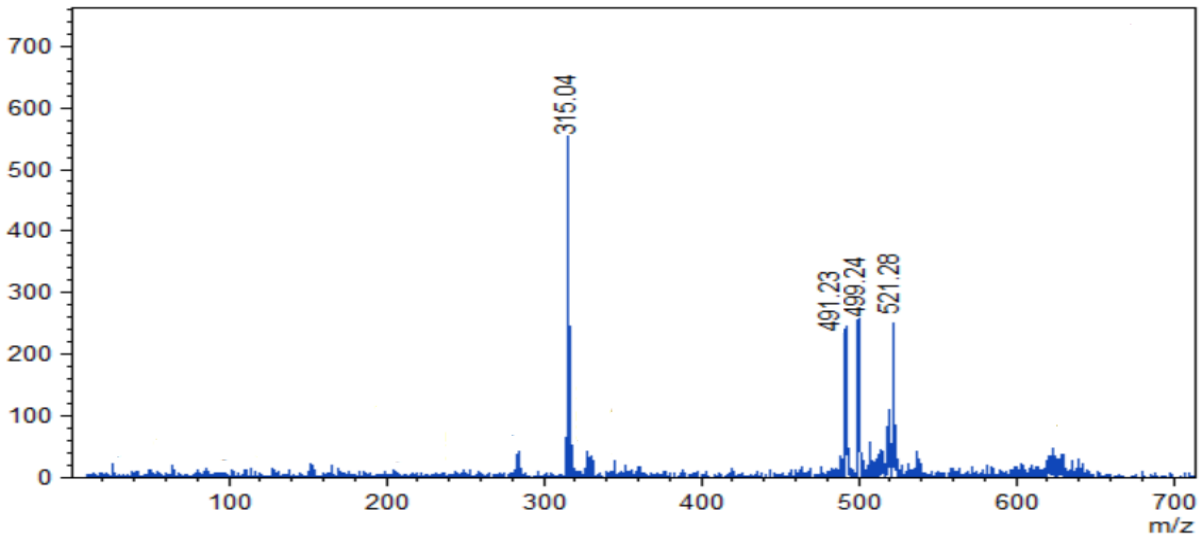
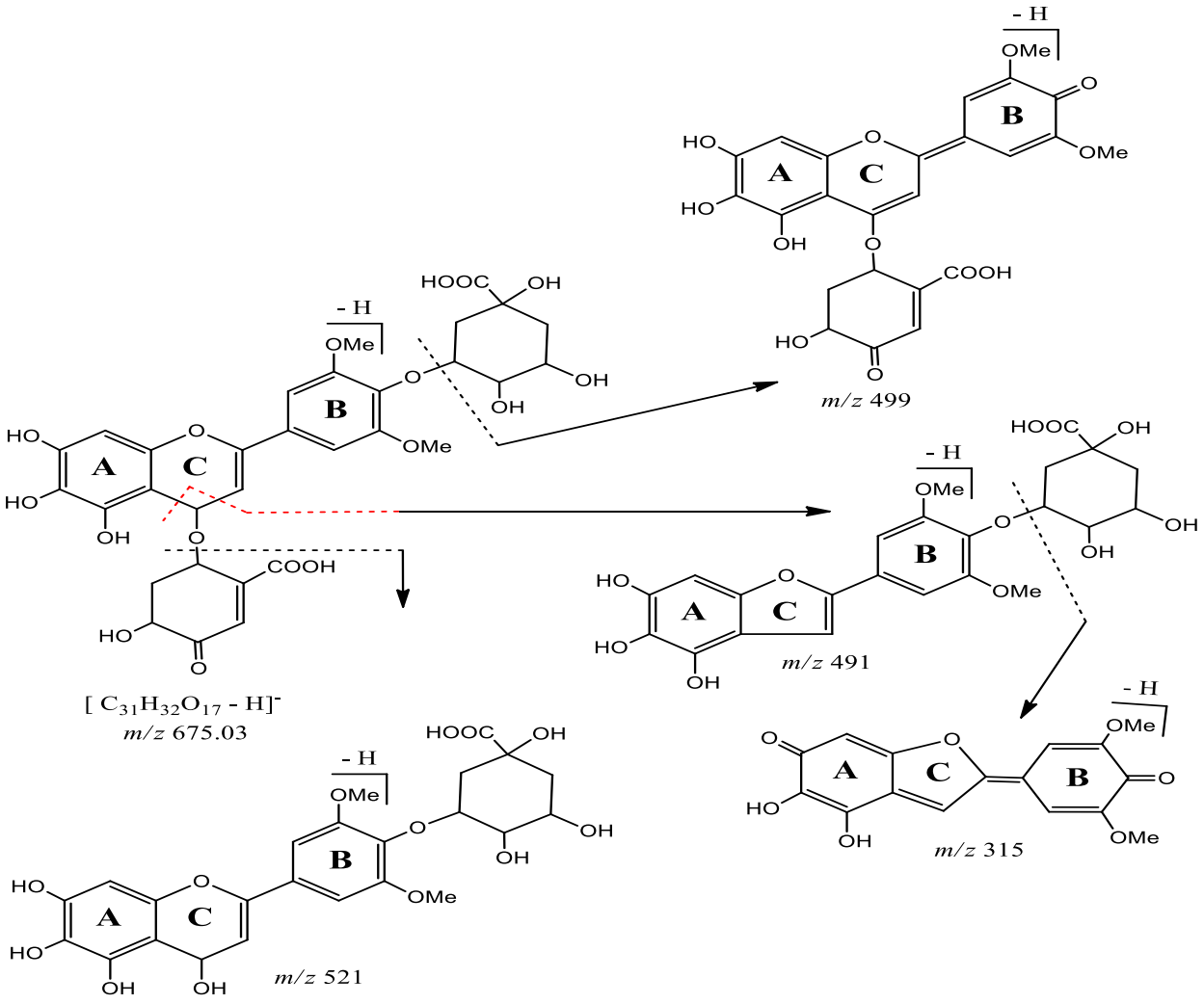


Figure 3.11. Product ion scan of the precursor ion *7* at *m/z* 675.03



Scheme 3.7. High energy CID-MS/MS fragmentation mechanism of the precursor ion *7* at *m/z* 675.03

3.4.4. High Energy CID-TOF/TOF-MS/MS of the Oak VRL Syringylglycerol derivatives

Three ions were composed of syringylglycerol derivative attached to the following carbohydrates or shikimic acid moieties: deoxyshikimic acid (*m/z* 382.96), oxidized ascorbic acid derivative - dehydrated hexose (*m/z* 552.93), and deoxyshikimic acid - daucic acid (*m/z* 568.93).

3.4.4.1. Assignment of the precursor ion at *m/z* 382.96

The formula $[C_{18}H_{22}O_9 - H]^-$ (DBE = 8.5) was assigned to the precursor ion *8* at *m/z* 382.96 (Figure 3.4, Figure A2.1, and Table 3.2). It is composed of the syringylglycerol unit connected to deoxyshikimic acid. The product ion scan of the ion *8* experienced the sequential loss of two formaldehyde molecules ($-CH_2O$) from the two methoxy groups presented in the syringyl unit to yield the ions at *m/z* 353 and 323, respectively (Figure 3.12 and Scheme 3.8).

The product ion scan of the precursor ion *8* was initiated by the loss of carbon dioxide (-44Da) to give the ion at *m/z* 339, which supports the presence of the carboxylic acid group in the deoxyshikimic acid moiety. This latter product ion at *m/z* 339 was subjected to the consecutive losses of two hydrogen molecules from the deoxyshikimic acid moiety to yield the ions at *m/z* 337 and 335, respectively. These recent consecutive losses of CO_2 and $2H_2$ promote the conversion of the deoxyshikimic acid to an aromatic phenol ring.

Moreover, the precursor ion *8* loses a methanol molecule to yield the product ion at *m/z* 351. This latter product ion at *m/z* 351 simultaneously loses two water molecules to produce the ion at *m/z* 315. Moreover, this last product ion at *m/z* 315 experiences a ring cleavage in the deoxyshikimic acid moiety to yield the ion at *m/z* 201.

Furthermore, the loss of the deoxyshikimc acid moiety from the precursor ion *8* leads to the formation of the ion at *m/z* 243. This latter product ion at *m/z* 243 undergoes phenolic ring contraction accomplished by the loss of a carbon monoxide molecule to

produce the ion at *m/z* 215. This last product ion at *m/z* 215 losses two water molecules to give the ion at *m/z* 179.

Finally, the product ion scan of the precursor ion *8* showed the sequential bond cleavages of the C1-C7 bond of the S-unit and the Cα–O bond of the S-unit to produce the ions at *m/z* 153 and 227, respectively.

It should be noted that the product ion scan of precursor ion *8* looks very complicated, this is due to the presence of several isobars that are selected in the same time frame during the MS/MS analysis. However, we were capable of finding and suggesting one structure that can create all the major product ions, which may represent the major precursor of this isobaric mixture.

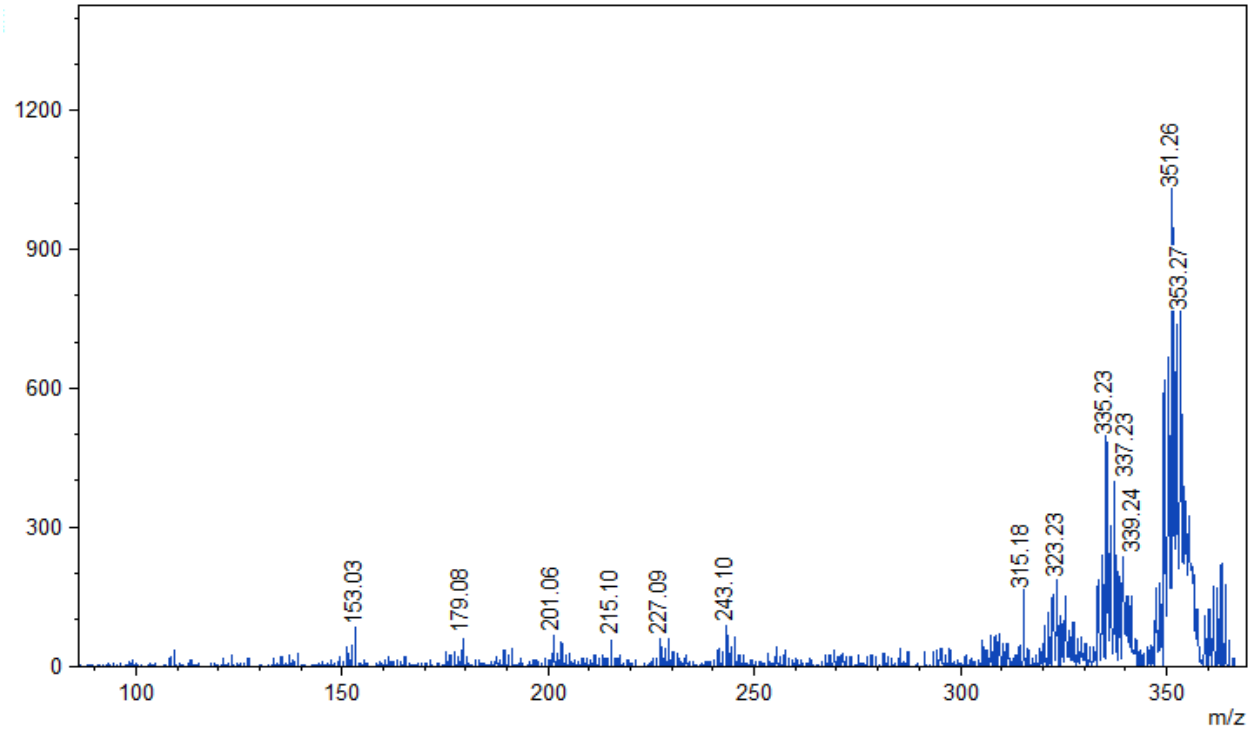
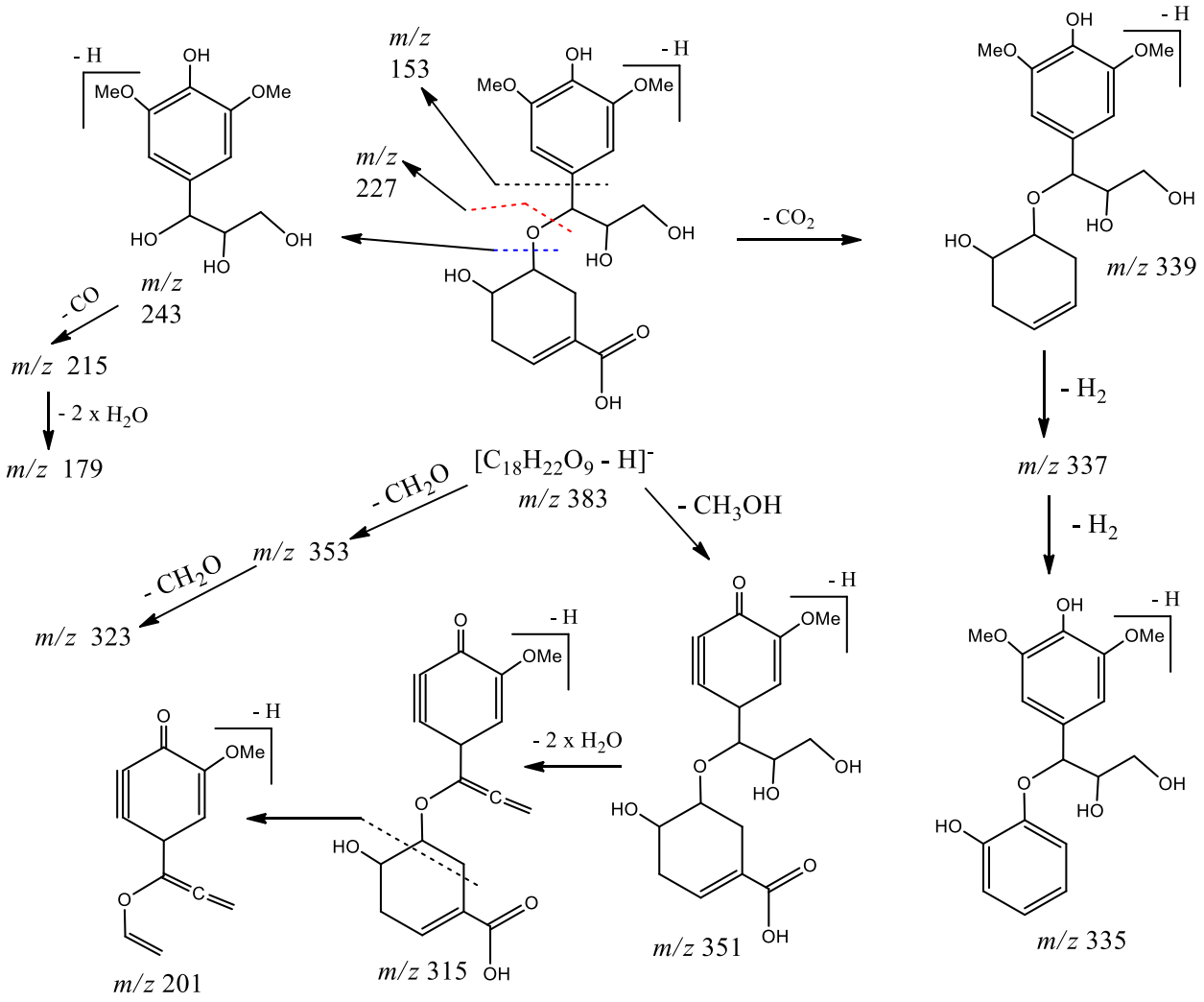


Figure 3.12. Product ion scan of the precursor ion at *m/z* 382.96



Scheme 3.8. High energy CID-MS/MS fragmentation mechanism of the precursor ion *8* at *m/z* 382.96

3.4.4.2. Assignment of the precursor ion at *m/z* 552.95

The $[C_{23}H_{22}O_{16}- H]^-$ formula (DBE=13.5) was assigned to the precursor ion *9* at *m/z* 552.95 (Figure 3.4, Figure A2.1, and Table 3.2). This ion is composed of a syringylglycerol derivative, which is glycosylated by a disaccharide consisting of dehydrated hexose and oxidized ascorbic acid derivatives. The product ion scan of this ion

9 created the ion at *m/z* 537 by the loss of the phenolic oxygen (Figure 3.13 and Scheme 3.9). Also, the product ion scan of the precursor ion *9* was subjected to the loss of 44 Da (-CO_2) from the oxidized ascorbic acid derivative to yield the ion at *m/z* 509.

Additionally, the product ion scan of this ion *9* revealed the loss of one methanol molecule (-32 Da) from the syringyl unit to yield the ion at *m/z* 521. Furthermore, the precursor ion *9* can lose 60 Da (- $C_2H_4O_2$) either from the dehydrated hexose ring or from the syringyl unit, as shown in scheme 3.9 to yield the ion at *m/z* 493.

Furthermore, the cleavage of the C5-O bond of the oxidized ascorbic acid moiety in the precursor ion *9* leads to the formation of the ion at *m/z* 383, while the cleavage of the C6-O bond of the dehydrated hexose in ion *9* leads to the formation of the ion at *m/z* 367. Finally, the C1-C7 bond cleavage of the S-unit in ion *9* leads to the creation of the syringol ion at *m/z* 153.

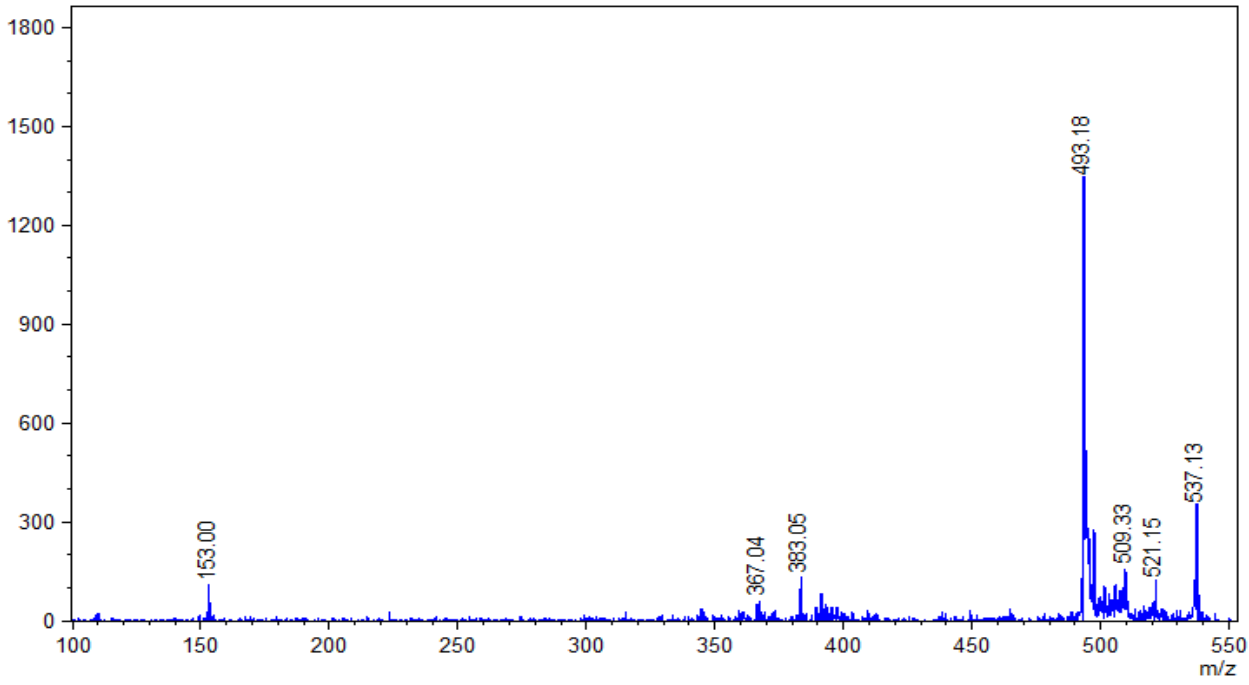
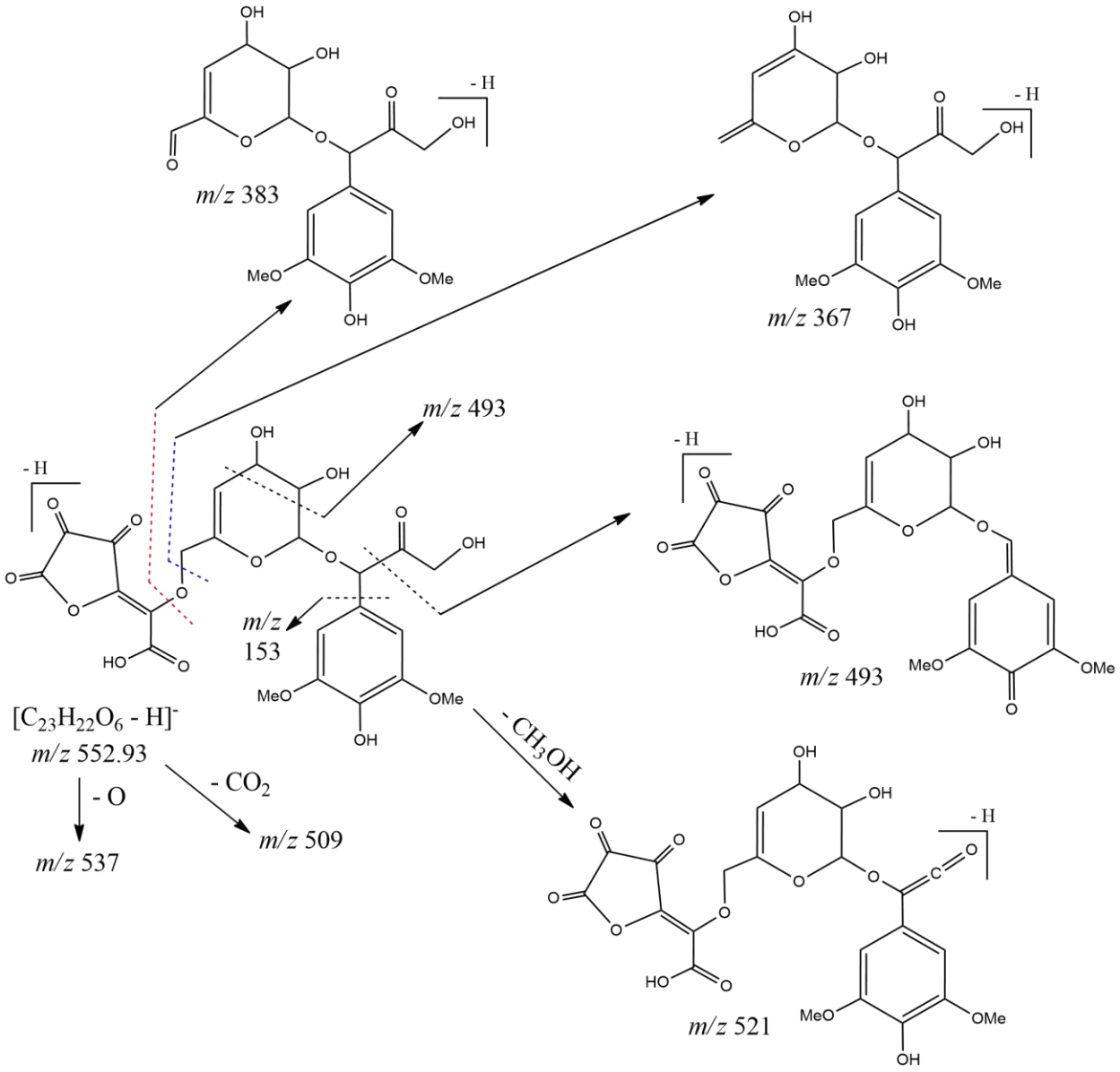


Figure 3.13. Product ion scan of the precursor ion *9* at *m/z* 552.95



Scheme 3.9. High energy CID-MS/MS fragmentation mechanism of the precursor ion *9* at *m/z* 552.93

3.4.4.3. Assignment of the precursor ion at *m/z* 568.93

The formula $[C_{25}H_{30}O_{15} - H]^-$ (DBE=11.5) was assigned to the precursor ion *10* at *m/z* 568.93 (Figure 3.4, Figure A2.1, and Table 3.2). This precursor ion *10* is composed of lignin S-unit derivative connected to deoxyshikimic acid, which is attached to a very rare sugar-acid derivative called daucic acid. The product ion scan of this precursor ion *10* was initiated by the loss of CO_2 to give the product ion at *m/z* 525, supporting the presence of the carboxylic acid groups of daucic acid (Figure 3.14 and Scheme 3.10). In addition, the combined loss of one water molecule and one methanol molecule from the precursor ion *10* leads to the formation of the ion at *m/z* 519. Furthermore, the product ion scan of this ion *10* indicated the sequential loss of two formaldehyde molecules ($-CH_2O$) from the syringyl unit methoxy groups that respectively led to the formation of the ions at *m/z* 539 and 509. This latter product ion at *m/z* 539 can be subjected to an RDA decomposition in the deoxyshikimic acid moiety to form the ion at *m/z* 281.

Also, the cleavage of the daucic acid ring in the precursor ion *10* at *m/z* 568.93 leads to the formation of the product ion at *m/z* 407, as shown in scheme 3.10. This latter product ion at *m/z* 407 loses one water molecule to yield the ion at *m/z* 389. This last product ion at *m/z* 389 showed the loss of 44 Da ($-C_2H_4O$) from the deoxyshikimic acid ring to create the ion at *m/z* 345. Moreover, the loss of sugar daucic acid ring from the precursor ion *10* leads to the formation of the ion at *m/z* 383. Finally, the C1-C7 bond cleavage in the S-unit leads to the creation of the syringol ion at *m/z* 153.

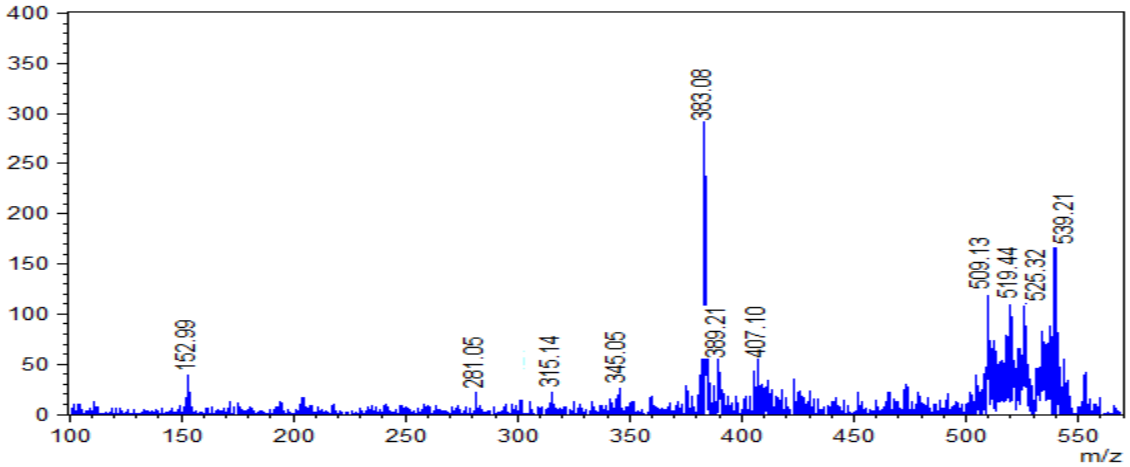
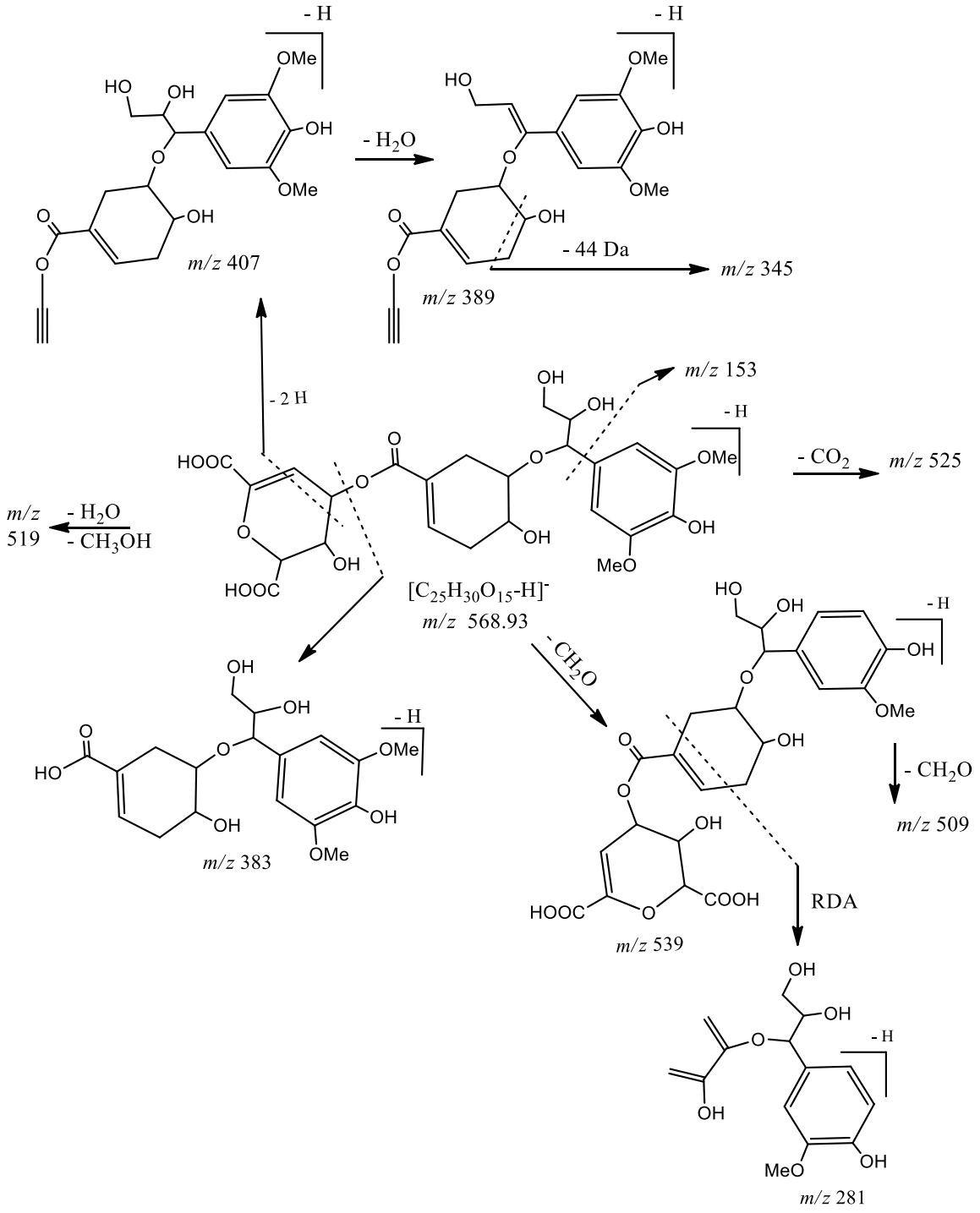


Figure 3.14. Product ion scan of the precursor ion *10* at *m/z* 568.93



Scheme 3.10. High energy CID-MS/MS fragmentation mechanism of the precursor ion *10* at *m/z* 568.93

3.4.5. High Energy CID-TOF/TOF-MS/MS of the Oak VRL Syringol derivatives

In this MALDI-TOF-MS of the VRL rationale, we characterized two dimeric precursor ions which were composed of a syringol unit attached to either a dehydroshikimic acid (*m/z* 307.04) and/or a hexuronic acid (*m/z* 329.04) residues.

3.4.5.1. Assignment of the precursor ion at *m/z* 307.04

The formula $[C_{15}H_{16}O_7 - H]^-$ (DBE= 8.5) was assigned to the precursor ion *11* at *m/z* 307.04 (Figure 3.4, Figure A2.1, and Table 3.2). This dimeric ion *11* is composed of a syringol moiety connected to 3-dehydroshikimic acid.

It should be noted that syringol, which is a derivative of the syringyl lignin, could originally exist as it is, or it could result from the degradation of syringyl units during the extraction process. Once more, we would like to reiterate that the shikimic acid derivatives are known to be intermediate in the lignin biosynthesis, but they were never described to be incorporated into the wood network.[6] The presence of shikimic acid attached to syringol may indicate that the shikimic acid derivatives are not just intermediates in the biosynthesis of lignin, but they can be a part of the final wood network that may present to perform other biosynthetic tasks when needed.

The product ion scan of this precursor ion *11* was initiated with the loss of 46 Da in the form of formic acid to produce the product ion at *m/z* 261 (Figure 3.15 and Scheme 3.11). This loss supports the presence of a carboxylic acid group, which in turn confirms the presence of the shikimic acid derivative. The precursor ion *11* was also found to

experience a neutral loss of the 3-dehydroshikimic acid moiety to yield the major syringol ion at *m/z* 153.

Although the precursor ion *11* at *m/z* 307.04 was not detected in the LDI-TOF-MS of the matrix only, there is a remote possibility that this precursor ion could exhibit matrix contamination during the MALDI-TOF-MS analysis of the sample. This isobaric contamination could well originate from the $[2DHB-H]^-$ formula.[35,36] It should be noted that this cluster cannot be subjected to a small neutral loss of 46 Da, which was observed in the product ion scan of this precursor ion at m/z 307.04. [37] On the other hand, this $[2DHB-H]^-$ cluster can form only the product ion at m/z 153 through the gas-phase neutral evaporation of one DHB molecule. [37] However, when we investigated the isotopic fitting percentage of the product ion at *m/z* 153, we found that this ion could be assigned as either $[C_7H_6O_4-H]^-$ (matrix) and/or $[C_8H_{10}O_3- H]^-$ (Sample); we found that our proposed C8 formula has an excellent isotopic fitting percentage 97.5% with respect to the DHB ion (30.0%).

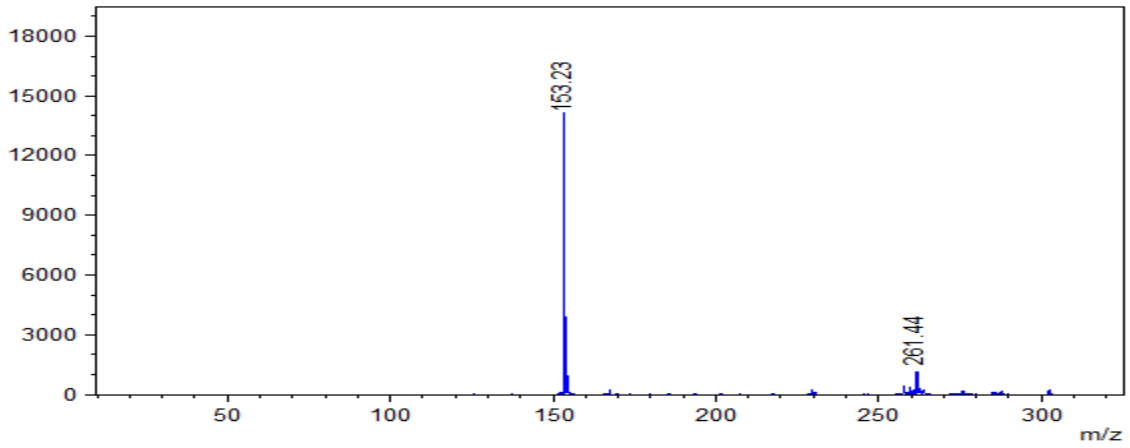
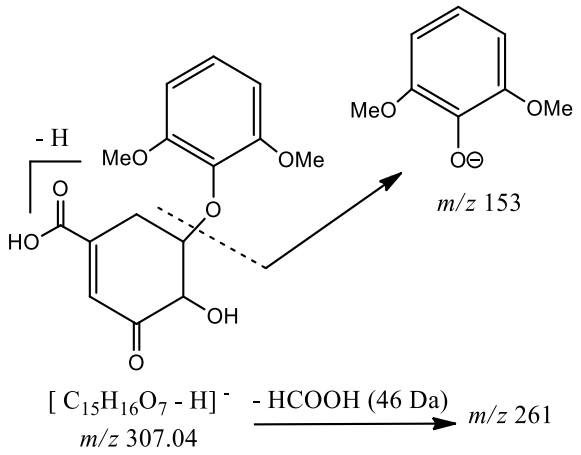


Figure 3.15. Product ion scan of the precursor ion *11* at *m/z* 307.04



Scheme 3.11. High energy CID-MS/MS fragmentation mechanism of the precursor ion *11* at *m/z* 307.04

3.4.5.2. Assignment of the precursor ion at *m/z* 329.04

The formula $[C_{14}H_{18}O_9 - H]^-$ (DBE= 6.5) was assigned to the precursor ion *12* at *m/z* 329.04 (Figure 3.4, Figure A2.1, and Table 3.2). It is composed of syringol residue glycosylated by a hexuronic acid moiety. The product ion scan of this precursor ion *12* was initiated by the loss of 46 Da in the form of formic acid to produce the product ion at *m/z* 283 (Figure 3.16 and Scheme 3.12). This loss supports the presence of a carboxylic acid group, which supports the presence of the hexuronic acid moiety. Also, the precursor ion *12* was subjected to a neutral loss of the hexuronic acid moiety to yield the syringol ion at *m/z* 153.

It should be noted that in general, O-glycosides/glucuronides phenolic compounds MS/MS is characterized by the neutral loss of the sugar ring. [15]

Once more, we would like to point out, that the ion at *m/z* 329, was not detected in the LDI-TOF-MS of the matrix alone. However, its presence can be assigned to the isobaric

matrix cluster [2DHB+Na-2H]$^-$; this kind of assumption could well complicate the MS/MS analysis of this precursor ion. [35,36] However, If this were the case, it would be reasonable to expect that this [2DHB+Na-2H]$^-$ cluster, through the neutral evaporation of one sodiated DHB molecule to form the product ion at *m/z* 153. Again, this cluster cannot undergo a neutral loss of 46 Da, as demonstrated in the product ion scan of this precursor ion at *m/z* 329. [37] So, in theory, the ion at *m/z* 153 can be assigned as [$C_7H_6O_4$-H]$^-$ (matrix) and/or [$C_8H_{10}O_3$- H]$^-$ (Sample); we found that the C8 formula had an excellent isotopic fitting percentage 91.0% with respect to the DHB ion (29.7%).

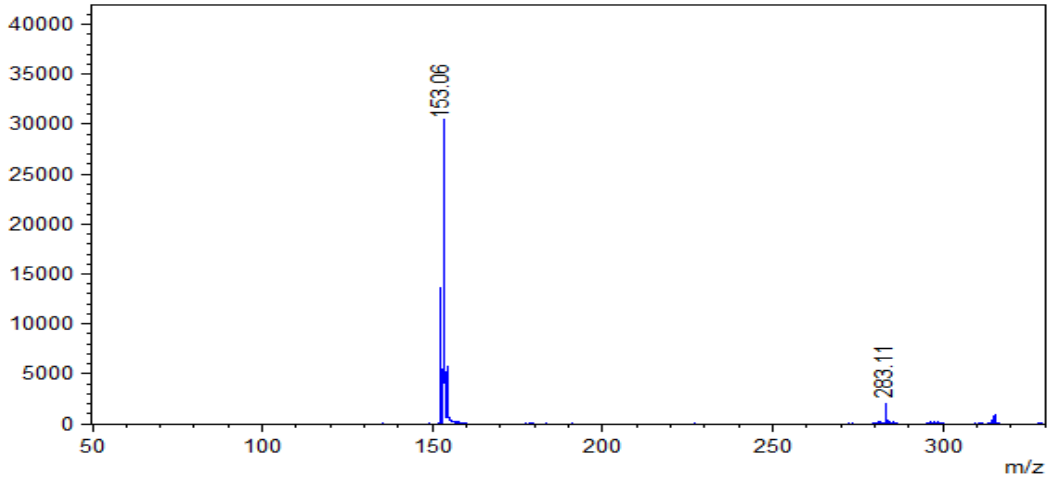
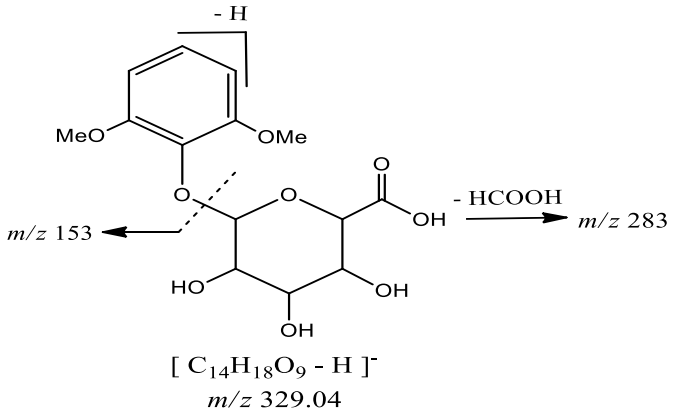


Figure 3.16. Product ion scan of the precursor ion *12* at *m/z* 329.04



Scheme 3.12. High energy CID-MS/MS fragmentation mechanism of the precursor ion *12* at *m/z* 329.04

3.4.6. High Energy CID-TOF/TOF-MS/MS of the Oak VRL Flavonolignins and /or Flavonosugars derivatives

We have also identified two ions that could be formed from degraded flavonolignans or flavonosugar derivatives: the first ion at *m/z* 315.05 with no sugar moieties in its proposed structure and the second one at *m/z* 491.05 which contained a tetrose sugar.

3.4.6.1. Assignment of the precursor ion at *m/z* 315.05

The formula $[C_{17}H_{16}O_6 - H]^-$ (DBE= 10.5) was assigned to the precursor ion *13* at *m/z* 315.05 (Figure 3.4, Figure A2.1, and Table 3.2). This ion *13* was hypothesized to originate from a flavonolignin structure which can be formed through the addition of the two hydroxyl groups on ring A to the propanoid double bond of a lignin unit. Similarly, it can originate from a falvonosugar structure which can be formed through a dehydration reaction between the two hydroxyl groups of ring A and the contiguous C_1OH-C_2OH position of a sugar ring. (Scheme A2.1).

The product ion scan of the precursor ion *13* gives the product ion at *m/z* 289 through the loss of 26 Da (- C_2H_2), as shown in Figure 3.17 and Scheme 3.13. The product ion at *m/z* 289 loses one phenolic oxygen to give the ion at *m/z* 273. In addition, the precursor ion *13* experiences the cleavage of the bond between Cβ with respect to ring B and Cα with respect to ring A to give the two ions at *m/z* 135 and 179 as demonstrated in Scheme 3.13. Lastly, the product ion at *m/z* 179 loses an acetylene molecule 26 Da (C_2H_2) to yield the ion at *m/z* 153.

As we mentioned previously in section 3.3**,** there is an obvious overlap from the matrix with this precursor ion *13* at *m/z* 315. For this reason, the MS/MS of the matrix ion at *m/z* 315 formed in the LDI of Matrix spectrum was compared to the product ion scan of the same ion originated from the MALDI-MS spectrum of the sample (Figure A2.3, Scheme A2.2). This MS/MS comparison is essential to differentiate between the sample ions and the matrix ions, in case they have the same nominal mass.[37] The product ion scan of the proposed isobaric $[2DHB\text{-}2H_2O + 2\ Na - 3H]^-$ matrix cluster gave two major ions at *m/z* 178 and 135, (Scheme A2.2) whereas the MS/MS of our $[C_{17}H_{16}O_6 - H]^-$ assigned precursor ion at m/z 315, gave a series of product ions at *m/z* 289, 273, 179, 153 and 135 (Scheme 3.13). This evidence precludes the absence of isobaric contamination.

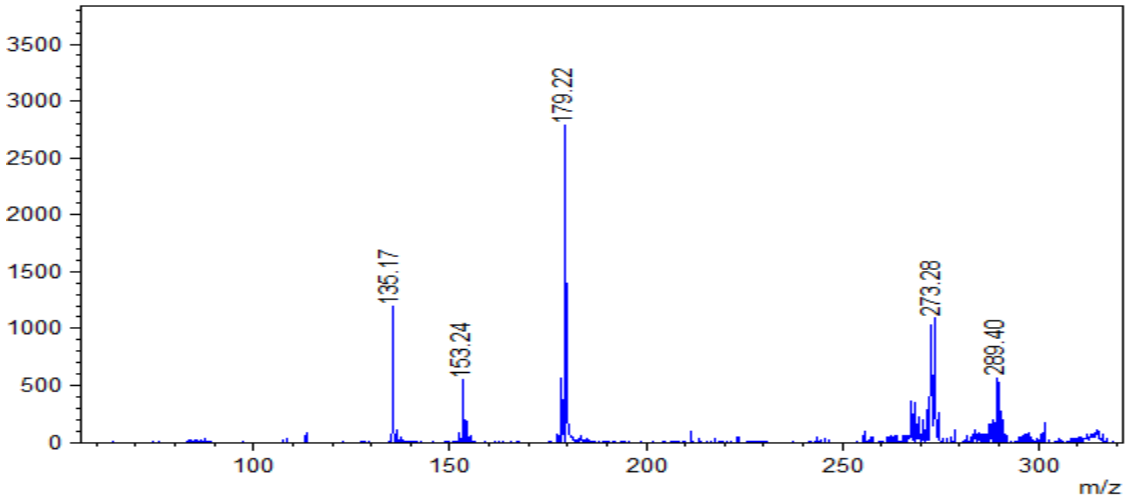
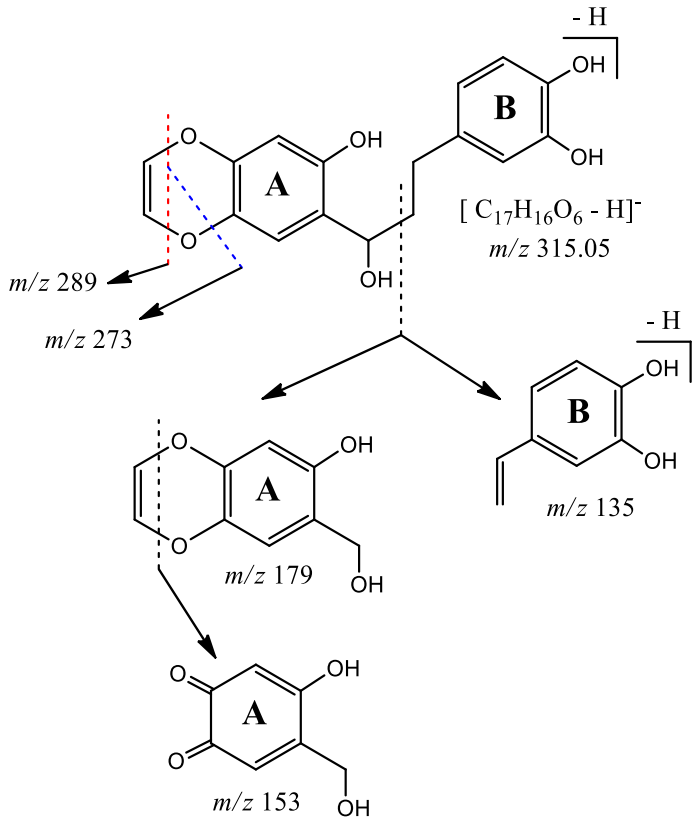


Figure 3.17. Product ion scan of the precursor anion *13* at *m/z* 315.05



Scheme 3.13. High energy CID-MS/MS fragmentation mechanism of the precursor ion 13 at *m/z* 315.05

3.4.6.2. Assignment of the precursor ion at *m/z* 491.05

The precursor ion *14* at *m/z* 491.05 was assigned a formula $[C_{23}H_{24}O_{12} - H]^-$ (DBE=12.5) (Figure 3.4, Figure A2.1, and Table 3.2). This precursor ion *14* is a tricin derivative with an extra hydroxyl group on ring A, which is connected to tetrose sugar. Also, two out of the three hydroxy groups of ring A are attached to -CH=CH- forming a six-membered ring indicating the presence of flavonolignan or flavonosugar in the original structure.

The MS/MS fragmentation of the precursor ion *14* was initiated by the loss of a molecule of acetylene C_2H_2 (-26 Da) to give the product ion at *m/z* 465 (Figure 3.18 and Scheme 3.14). This latter ion at *m/z* 465 experienced the loss of one molecule of carbon monoxide to yield the ion at *m/z* 437. Also, the loss of the tetrose sugar from the product ion at *m/z* 465 leads to the ion at *m/z* 345. Moreover, the cleavage of the bond connecting ring B and ring C in the product ion at m/z 465 lead to the formation of the ions at *m/z* 315 and *m/z* 153. In addition, the precursor ion *14* experience a ring C cleavage to yield the ion at *m/z* 329.

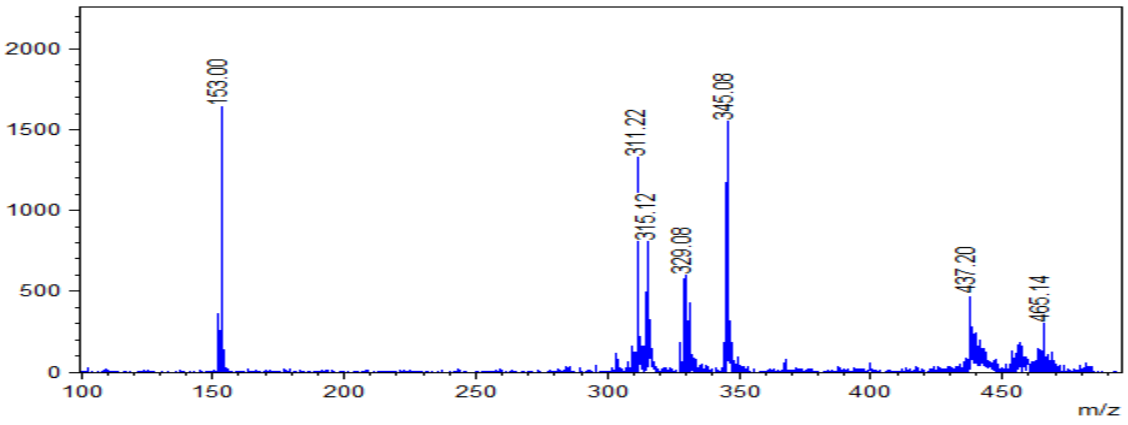
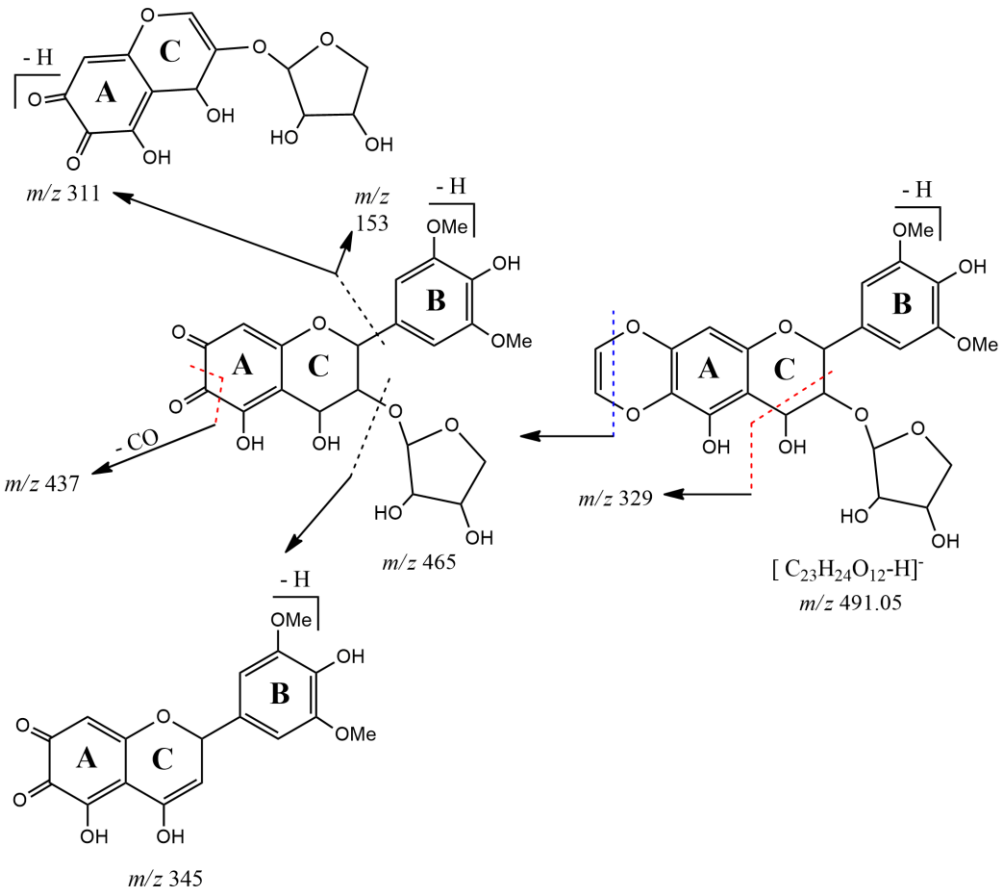


Figure 3.18. Product ion scan of the precursor ion *14* at *m/z* 491.05



Scheme 3.14. High energy CID-MS/MS fragmentation mechanism of the precursor ion *14* at *m/z* 491.05.

3.4.7. High Energy CID-TOF/TOF-MS/MS of the Oak VRL miscellaneous compounds

The last six ions identified were as follows: Luteoferol (*m/z* 289.03), lariciresinol isomer (*m/z* 359.97), 3,4-dihydroxy-5-methoxy benzyl alcohol (5-hydroxy guaiacyl derivative) attached to hexuronic acid (*m/z* 345.02), syringyl lignin unit attached to $C_{10}H_{10}O_2$ lignin derivative (*m/z* 368.97), trihydroxy benzaldehyde attached to the disaccharide dipentose (*m/z* 417.07), and aryl tetralin lignan structure composed of syringyl unit, guaiacyl unit, and $C_{10}H_{10}O_2$ lignin derivative (*m/z* 545.97).

3.4.7.1. Assignment of the precursor ion at *m/z* 289.03

The formula $[C_{15}H_{14}O_6 - H]^-$ (DBE= 9.5) was assigned to the precursor ion *15* at *m/z* 289.03 (Figure 3.2, Figure A2.1, and Table 3.2). We suggest that this ion is the known flavonoid, luteoferol.

It should be mentioned that Luteoferol has never been found to be incorporated or attached to any lignin structure. According to previous literature, this ion can also be assigned to the isobaric matrix cluster $[2DHB\text{-}H_2O]^-$ that can be formed through the esterification of two DHB molecules.[38] These two possible formulas $[C_{15}H_{14}O_6 - H]^-$ and $[2DHB\text{-}H_2O]^-$ have isotopic distribution percentage 95.6% and 94.9%, respectively. These latter percentages slightly favored the assigned formula of luteoferol. However, the origin of this ion can be judged through the comparison between the intensity of the ion in the MALDI spectrum of the sample with its intensity in the LDI spectrum of matrix only (Figure A2.4). [39]

The product ion scan of the precursor ion *5* afforded the ion at *m/z* 273, which was formed by loss of 16 Da in the form of a phenolic oxygen atom, as shown in Figure 3.19 and Scheme 3.15. In addition, the loss of C_2H_4O (- 44 Da) from the precursor ion*15* creates the ion at *m/z* 245. This latter ion loses another 44 Da by the cleavage of the ring C bonds (cleavage of the bonds denoted by 2 and 4, Scheme 3.15), as indicated by the blue line in scheme 3.15. Also, the precursor ion *15* decomposes into the two product ions at *m/z* 153 ($^{1,3}A^-$) and 135 ($^{1,3}B^-$) through the RDA fragmentation indicated by the red line in Scheme 3.15. Finally, the primary product ion at *m/z* 245 undergoes cleavage in ring C by the

elimination of $C_3H_2O_2$ (- 70 Da), as shown in Scheme 3.15 to yield the secondary product anion at *m/z* 175.

Another point to support that this ion does not belong to the matrix is that if we consider that the product ion at m/z 245 is $[2DHB\text{-}H_2O\text{-}CO_2\text{-}H]^-$, it cannot lose $C_3H_2O_2$ (- 70 Da) to yield the secondary product ion at m/z 175, as demonstrated briefly in Figure 3.20.

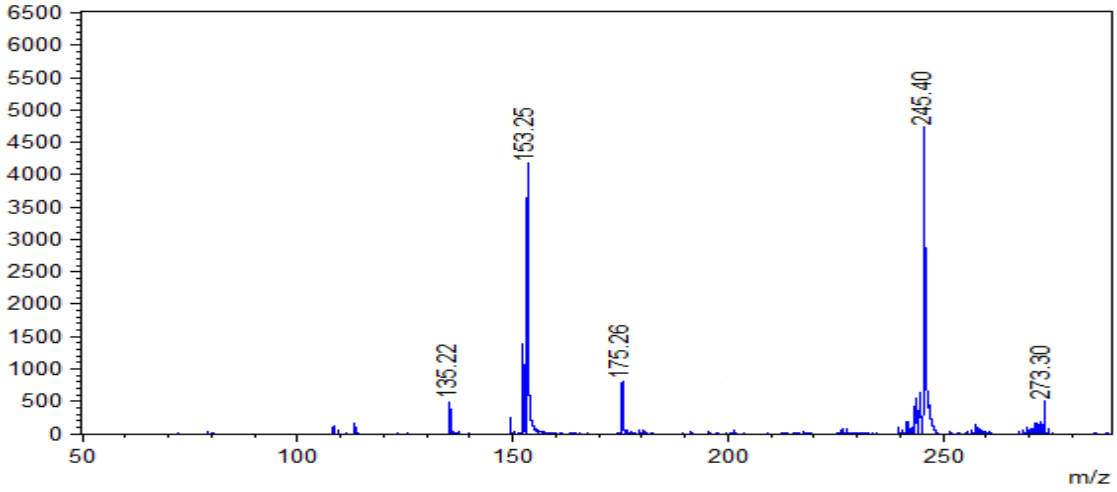
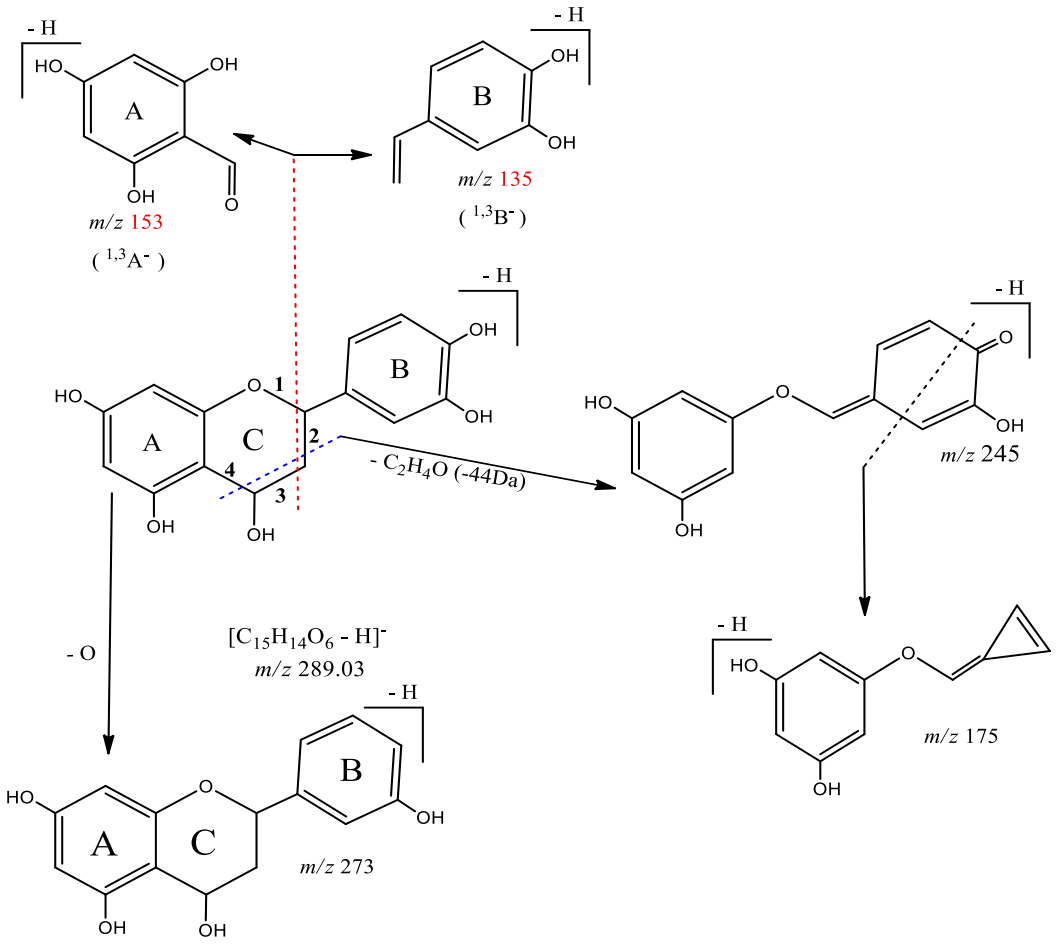


Figure 3.19. Product ion scan of the precursor anion 15 at *m/z* 289.03



Scheme 3.15. High energy CID-MS/MS fragmentation mechanism of the precursor anion *15* at *m/z* 289.03

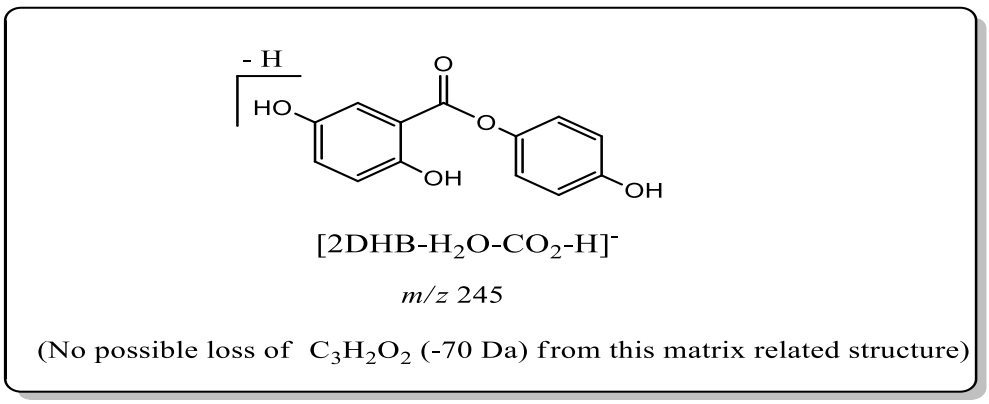


Figure 3.20. The possible matrix product ion structure at *m/z* 245 showed that it cannot form the secondary product ion at m/z 175 through the loss of $C_3H_2O_2$ moiety.

3.4.7.2 Assignment of the precursor ion at *m/z* 345.02

The formula $[C_{14}H_{18}O_{10} - H]^-$ (DBE = 6.5) was assigned to the precursor ion *16* at *m/z* 345.02 (Figure 3.4, Figure A2.1, and Table 3.2). This ion seems to be composed of 3,4-dihydroxy-5-methoxy benzyl alcohol attached to a hexuronic acid. We presume that this benzyl alcohol derivative can initially originate from an F- Lignin unit (5-hydroxy guaiacyl) through the cleavage of its Cα-Cβ bond.

The product ion scan of precursor ion *16* was initiated by the loss of neutral formaldehyde molecule (- 30 Da) from the aromatic methoxy group to produce the ion at *m/z* 315 (Figure 3.21 and Scheme 3.16). Also, the precursor ion *16* loses 44 Da (- CO_2), supporting the presence of the hexuronic acid moiety. In addition, the loss of the hexuronic acid moiety from the precursor ion *16* yields the ion at *m/z* 153.

It should be noted that this ion was not detected in the LDI-spectrum of the matrix only. However, this ion can be attributed to the $[2M+K-2H]^-$matrix cluster [34,35] Nevertheless, as discussed previously, the product ion scan of this precursor ion *17* is characterized by the loss of formaldehyde (- 30 Da) to form the product ion at *m/z* 315. Obviously, this loss cannot occur from the suggested matrix cluster formula. Also, it should be noted that this kind of matrix clusters are not supposed to show small neutral losses like 30 Da and 44 Da as seen in the product ion scan of this ion (Figure 3.22); instead, it is expected to undergo neutral evaporation of the matrix cluster components.[37] Also, If we consider that the product ion at *m/z* 153 can be assigned as $[C_7H_6O_4-H]^-$ (matrix) and/or $[C_8H_{10}O_3- H]^-$ (Sample), the C8 formula has a higher isotopic fitting percentage 97.7% with respect to the DHB ion (91.8%).

Lastly, in this example, we would like to explain how we decided on choosing the best structure for this precursor ion, this is demonstrated in Figure A2.5 and Scheme A2.3. In Scheme A2.3, we compared the fragmentation pathway of the top four possible formulas for this precursor ion at m/z 345, which ended with the choice of the assigned chemical formula $[C_{14}H_{18}O_{10} - H]^-$. It should be noted that even the chosen formula can be presented as two isomers A and B. The MS/MS of the precursor ion designated as isomer A (Scheme 3.16) can produce the stable benzylic ion at *m/z* 153. Whereas the MS/MS of precursor ion designated as isomer B is expected to show a neutral loss of dehydrated glucuronic acid moiety (-176 Da), which will not form the product ion at *m/z* 153, as presented in Scheme A2.3. For these reasons, we preferred to choose Isomer A to represent the structure of this ion 17 at *m/z* 345.

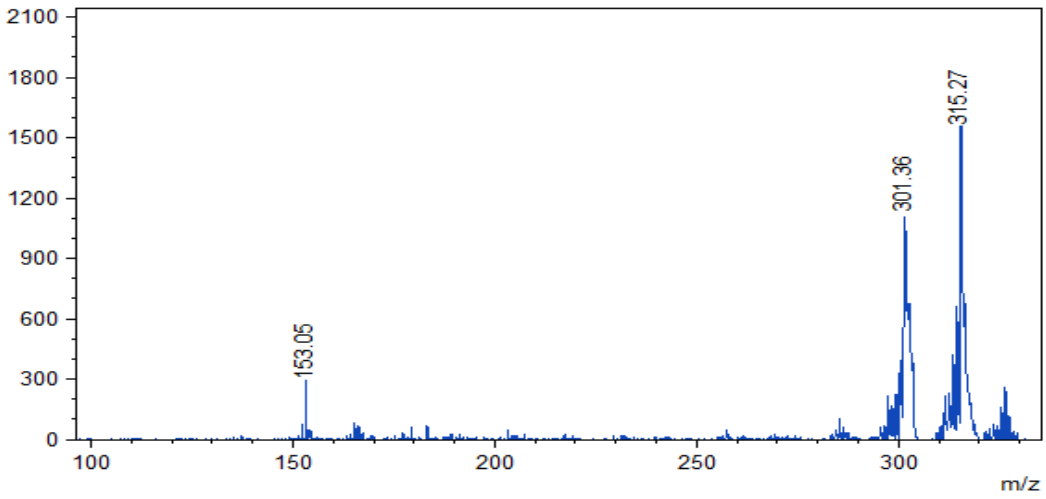
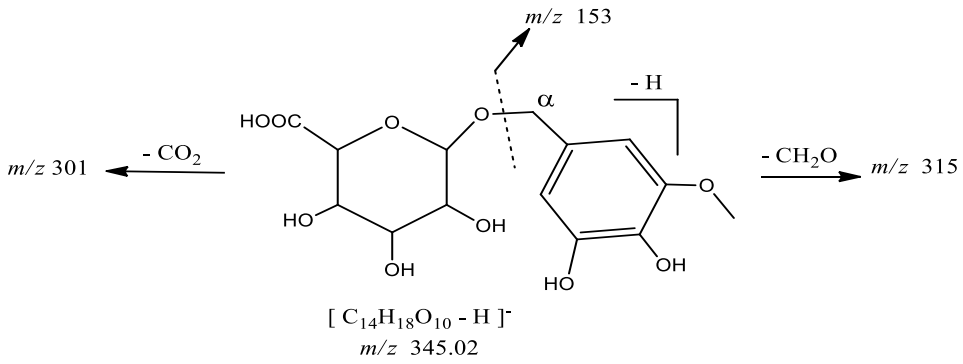


Figure 3.21. Product ion scan of the precursor ion *16* at *m/z* 345.02



Scheme 3.16. High energy CID-MS/MS fragmentation mechanism of the precursor ion *16* at *m/z* 345.02

3.4.7.3. Assignment of the precursor ion at *m/z* 359.97

The formula $[C_{20}H_{24}O_6]^{\bullet -}$ (DBE= 9) was assigned to the radical ion *17* at *m/z* 359.97. This ion was presumed to be a different isomer of lariciresinol lignan, which is composed of syringyl and coumaryl lignin unit instead of two gauiacyl lignins (Figure 3.4, Figure A2.1, and Table 3.2).

The MS/MS fragmentation of this ion is very straight forward and shows strong support for the proposed structure. The product ion scan of *16* was initiated by the loss of the methyl radical (-15 Da) from one of the two aromatic methoxy groups to yield the ion at *m/z* 345 (Figure 3.22 and Scheme 3.17). It should be noted that the loss of methyl radical is characteristic of lignins with methoxylated units.[40] This latter ion at *m/z* 345 is subjected to consecutive losses of a water molecule, carbon monoxide molecule, and formaldehyde molecules to yield the ions at *m/z* 327, *m/z* 317 and *m/z* 315, respectively. Furthermore, the precursor ion 17 can lose the aromatic phenol moiety from the H unit to give the ion at *m/z* 267. Likewise, the precursor ion 17 can be subjected to C1-C7 cleavage of the S-unit to form the syringol ion at *m/z* 153. Moreover, the central five-membered aliphatic rings of

the precursor ion 17 at *m/z* 359.97 experience several types of cleavage indicated by red and blue lines, as shown in Scheme 3.17 to produce the ions at *m/z* 225 and *m/z* 181. It should be noted that the product ion at m/z 225 is a characteristic ion for the sinapyl lignin unit.[40]

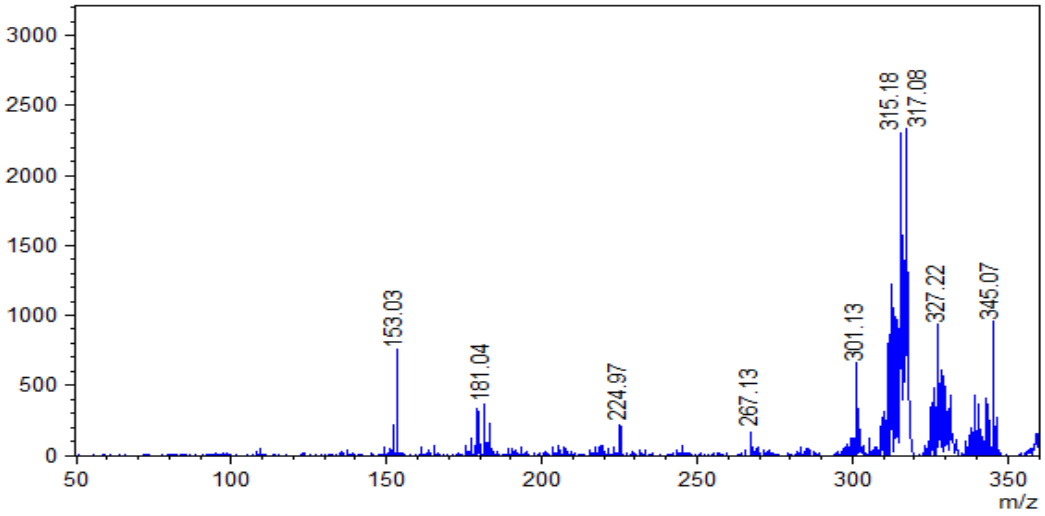
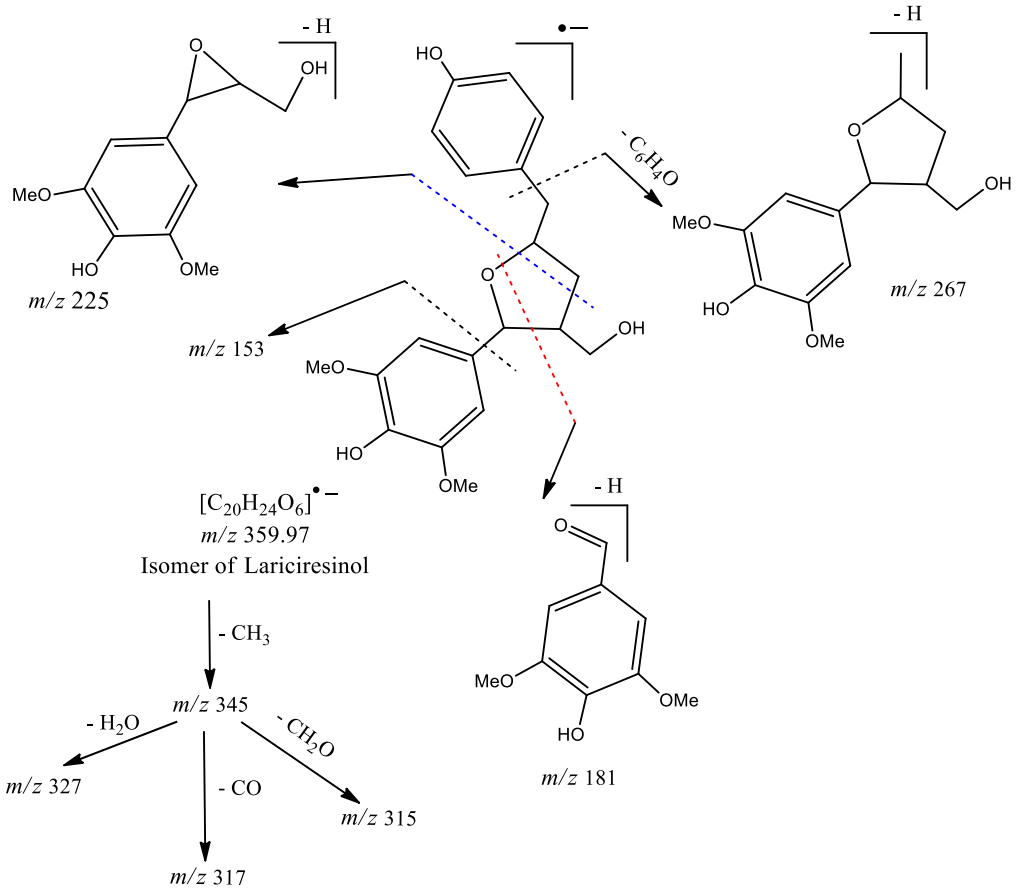


Figure 3.22. Product ion scan of the precursor radical anion *17* at *m/z* 359.97



Scheme 3.17. High energy CID-MS/MS fragmentation mechanism of the precursor radical anion *16* at *m/z* 359.97

3.4.7.4. Assignment of the precursor ion at *m/z* 368.97

The formula $[C_{21}H_{22}O_6 - H]^-$ (DBE=11.5) was assigned to the precursor ion *18* at *m/z* 368.97 (Figure 3.4, Figure A2.1, and Table 3.2). This ion appears to be composed of syringyl lignol unit S, linked through a β-5 bond to a novel $C_{10}H_{10}O_2$ lignin derivative. It should be noted that this $C_{10}H_{10}O_2$ structure has been mentioned previously as an unidentified product formed during the lignin pyrolysis by Oudia et al.[41] and as 4-propynylguaiacol by Ralph and Hatfield.[42]

The product ion scan of precursor ion *18* is initiated by the loss of 52 da (- C_4H_4) to create the ion at *m/z* 317 (Figure 3.23 and scheme 3.18). Also, the precursor ion *18* is subjected to two simultaneous cleavages of the β-5 linkage and the C4-O bond presented in the $C_{10}H_{10}O_2$ unit to form the ion at *m/z* 225. Once more, it should be noted that the ion at *m/z* 225 is known to be characteristic of the presence of syringyl lignin unit, as reported before in literature.[40] Lastly, the cleavage of the C1-Cα bond of the syringyl unit in ion *18* gives the aromatic ion syringol at *m/z* 153.

Lastly, we would like to give a hint on how we found the best structure for this precursor ion 18 at *m/z* 368.97 (Figure A2.6 and Scheme A2.4). In this latter scheme SI-4, we showed the most likely two isobaric structures that can be assigned to this ion at *m/z* 368.97. We compared the expected fragmentation pathway of the two different isobaric formulas. The first isobar A (our proposed formula) can form the product ion at *m/z* 317 through one easy neutral loss of C_4H_4, while isobar B containing laevoglucosan moiety can form the product ion at *m/z* 317 through three complex combined losses (Scheme A2.4). This showed why we preferred to choose isobar A, simply, because it forms the product ion at *m/z* 317 through one neutral loss instead of a combination of three losses (Scheme A2.4). Needless to say, that these two isobars are possible to exist; however, we choose the most likely precursor ion that fragments smoothly.

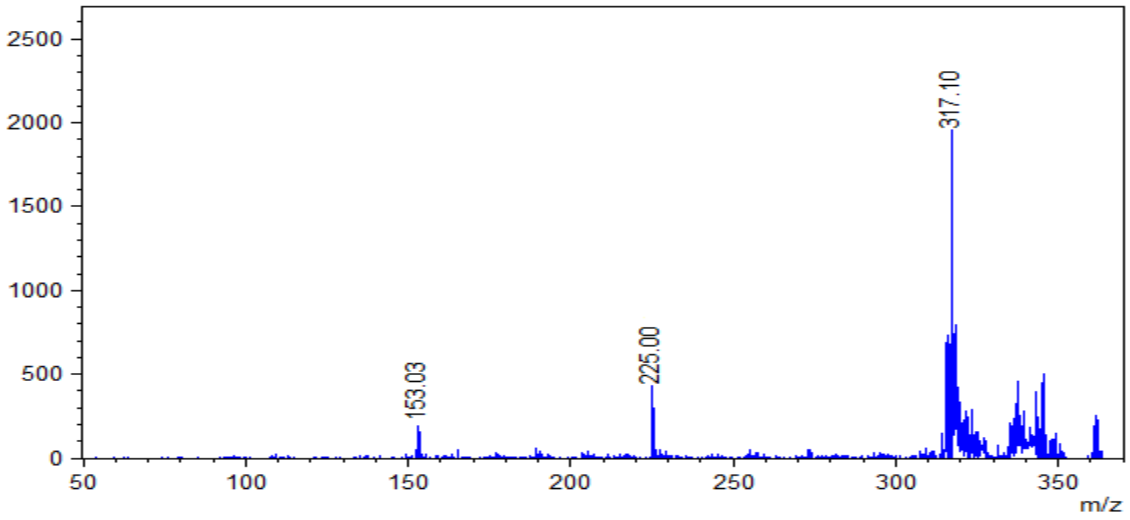
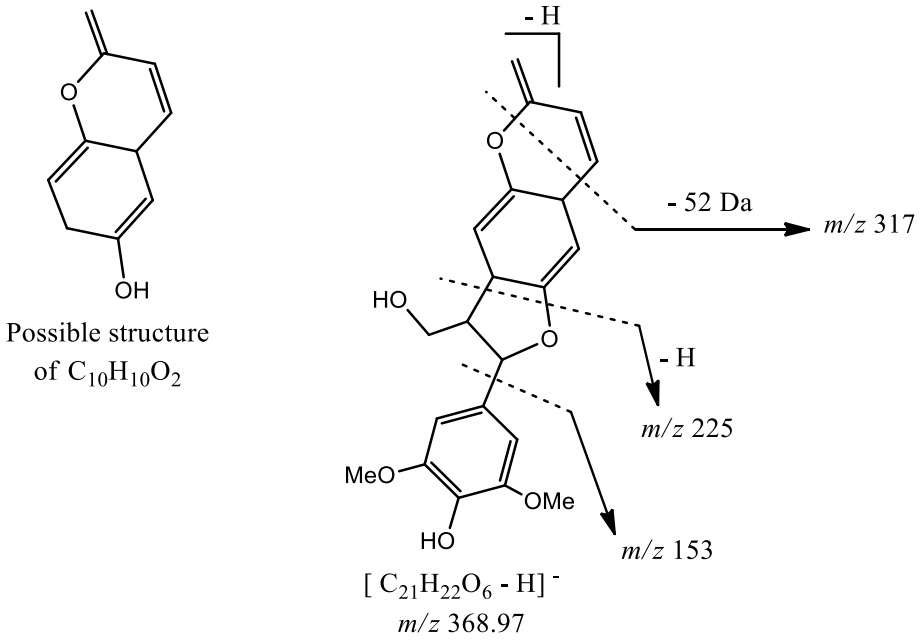


Figure 3.23. Product ion scan of the precursor ion *18* at *m/z* 368.97



Scheme 3.18. High energy CID-MS/MS fragmentation mechanism of the precursor ion *18* at *m/z* 368.97

3.4.7.5. Assignment of the precursor ion at *m/z* 417.07

The formula $[C_{17}H_{22}O_{12} - H]^-$ (DBE = 7.5) was assigned to the precursor ion *19* at *m/z* 417.07 (Figure 3.4, Figure A2.1, and Table 3.2). This precursor ion *19* was composed of two pentose sugars (disaccharide) attached to trihydroxy benzaldehyde (gallaldehyde). It should be noted that this gallaldehyde may originally exist as L gallyl lignol unit that was hypothetically proposed before by Vanholme *et al.* [43] The loss of the two pentoses sugars from the ion *19* leads to the formation of the ion at *m/z* 153 (Figure 3.24 and Scheme 3.19). Likewise, the product ion scan of this precursor ion *19* showed the loss of carbon monoxide (-28Da) from any one of the two free hydroxyl groups in the phenolic unit to yield the ion at *m/z* 389. This latter product ion experiences the loss of the two pentoses to produce the ion at *m/z* 125. Moreover, this last product ion at *m/z* 389 experiences a ring cleavage in the first pentose sugar to generate the ion $^{2,4}X_1$ at *m/z* 315.

Furthermore, the precursor ion *19* experiences two ring cleavages to yield the ions $^{0,2}X_1$ and $^{1,4}A_2$ at *m/z* 327 and 233, respectively. It should be noted that for the various sugar ring cleavages, we have used the nomenclature developed by Domon and Costello.[44] We used here the Letter A to label the fragments containing the non-reducing sugar end, while the letter X is used to describe the ions containing the aromatic ring (aglycone). Also, the superscripts denote the position of bond cleavage.

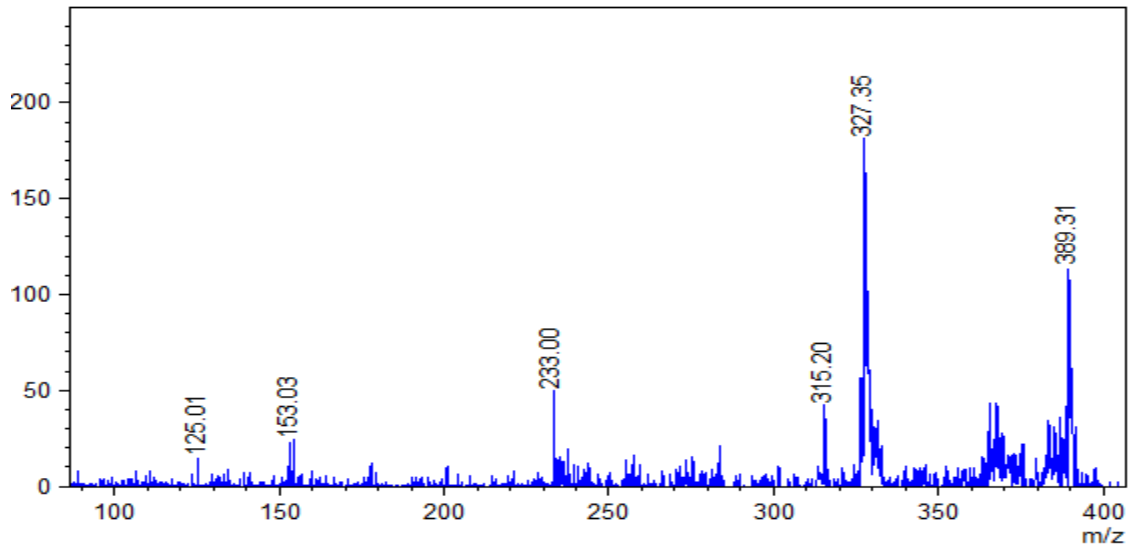
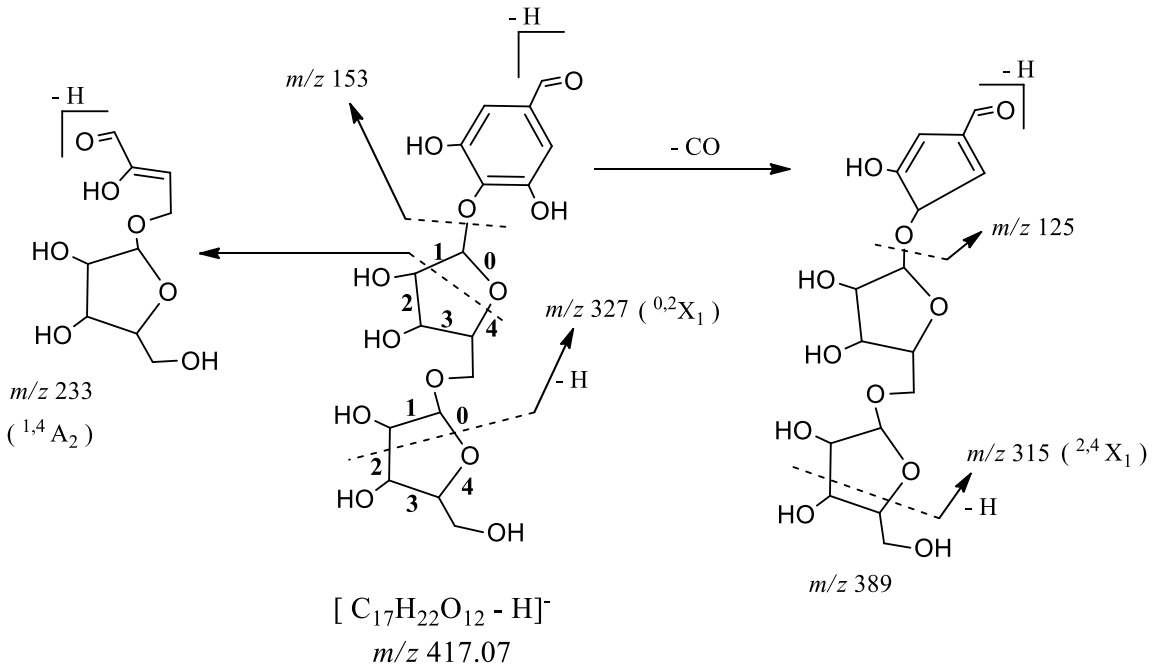


Figure 3.24. Product ion scan of the precursor ion *19* at *m/z* 417.07



Scheme 3.19. High energy CID-MS/MS fragmentation mechanism of the precursor ion 19 at m/z 417.07

3.4.7.6. Assignment of the precursor ion at *m/z* 544.97

The formula $[C_{31}H_{30}O_9 - H]^-$ (DBE=17) was assigned to the precursor ion *20* at *m/z* 544.97 (Figure 3.4, Figure A2.1, and Table 3.2). We have tentatively assigned the structure of this latter precursor ion to a tetralin lignan dimer composed of S and G lignin units attached to the lignin derivative $C_{10}H_{10}O_2$, which was described before for the precursor ion *18* at *m/z* 368.97. The product ion scan of the precursor ion *20* showed the loss of 52 Da as observed previously in the MS/MS of the precursor ion *18* (*m/z* 368.97) to yield the ion at *m/z* 493 (Figure 3.25 and Scheme 3.20). The complete loss of the $C_{10}H_{10}O_2$ moiety from the precursor ion *20* leads to the formation of the ion at *m/z* 383. The precursor ion at *m/z* 545.97 experience the cleavage of the C1-Cα bond in the G-unit and the β-β bond connecting the S and G lignin units. This latter two bonds cleavages can be accompanied by either the cleavage of the Cβ-Cγ in the G unit to yield the ion at *m/z* 360 or the cleavage of Oγ - Cγ bond between the S and G lignin units to produce the ion at *m/z* 345 (Scheme 3.20). Lastly, the cleavage of the C1- C7 bond of the S unit in the precursor ion *20* leads to the formation of the syringol ion at *m/z* 153.

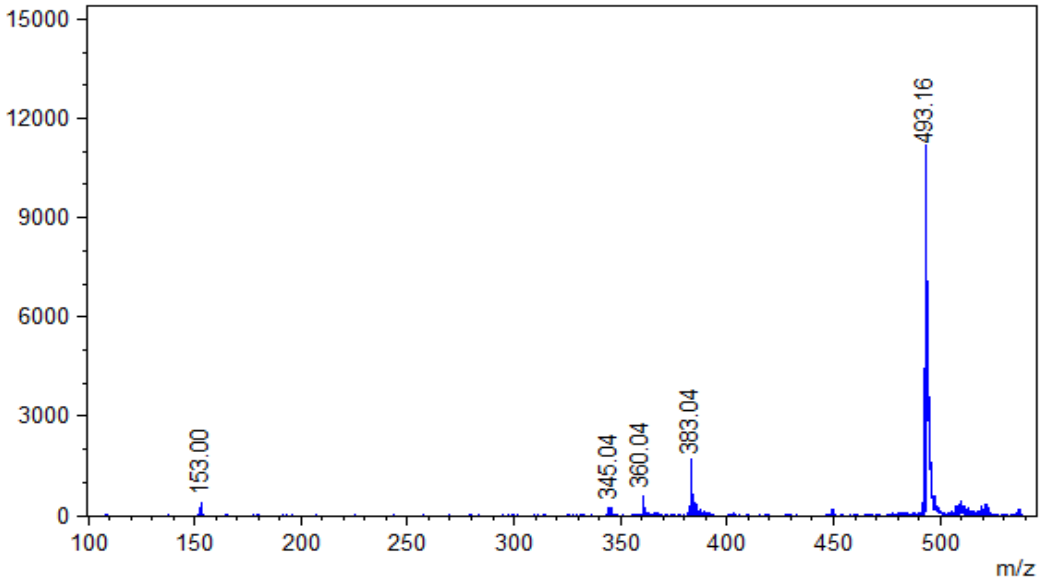
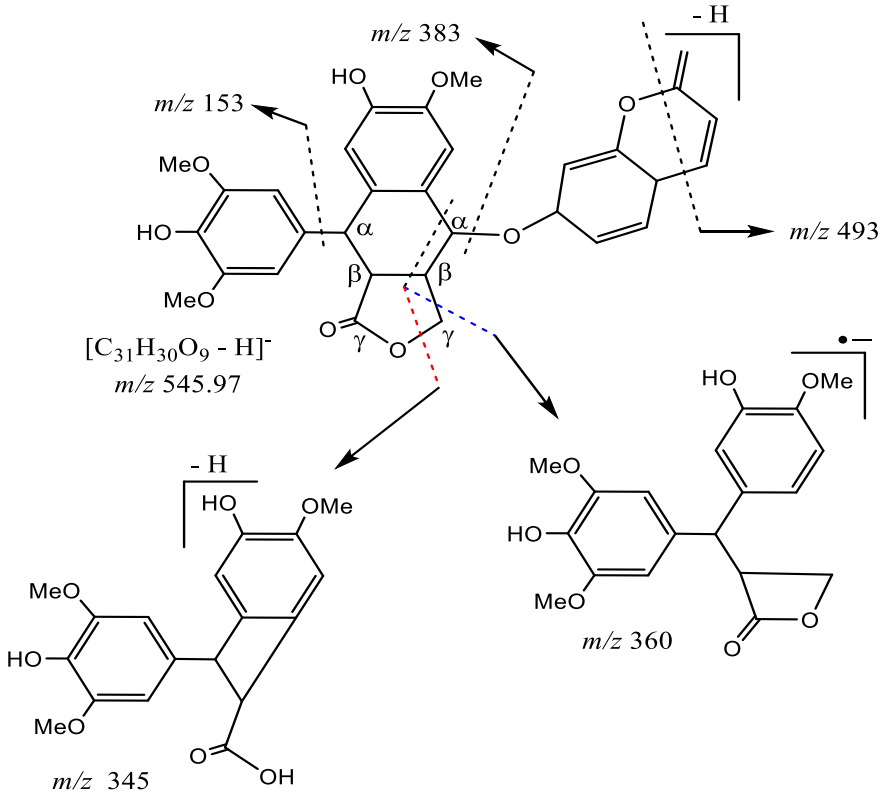


Figure 3.25. Product ion scan of the precursor ion *20* at *m/z* 544.97



Scheme 3.20. High energy CID-MS/MS fragmentation mechanism of the precursor ion *20* at *m/z* 544.97

3.5. Conclusion

In this manuscript, we showed the diversity of the compounds present in the studied complex lignin sample. In the French Oak wood tree (angiosperm), we described the first structural identification of novel tricin, and lignin derivatives attached to different shikimic acid and/or carbohydrate derivatives, degraded flavonolignan or falvonosugars compounds, degraded or oxidized lignin units, lignans, and the flavonoid luteoferol. Overall, the compounds identified in this study showed that the French Oak VRLs are rich in syringol moieties presented in syringyl lignin units and tricin derivatives.

We have described in this manuscript the complexity of the isolated Oak VRL lignin mixture due to the presence of a large variety of the possible lignin monomers and their attachment linkages. In addition, the diverse presence of tricin derivatives and lignin carbohydrate complexes adds more to the complexity of this isolated VRL mixture. Furthermore, it is very easy to comprehend that this series of lignin derivatives were never found by other researchers using an outdated method of lignin extractions and purification, which increase the possibility of degradation and/or modification during the lignin chemical modifications. Nevertheless, though we consider the solvolysis CIMV method to be a very mild extraction procedure, it would also be beneficial to consider the possibility of that even this lignin extraction process, could add some more disparity in the modifications in the complex VRL mixture composition. These modifications can be summarized as sugar acetylation, dehydration, oxidation, and the cleavage of lignin units to benzyl alcohol and/or benzaldehyde derivatives.

For these reasons and based on this structural characterization complexity, the identification of the components of the complex VRL mixture, is unquestionably not an easy task. Thus, the structural identification of these complex lignin derivatives cannot depend only on the mass accuracy and the complex overlapping isotopic distribution. Considering that we have found the purification of the VRLs and impossible task, we had no recourse left, except using tandem mass spectrometry, which allowed us to tentatively support the unequivocally assigned formulas for the reported chemical structures.

Lastly, the identification of some new compounds such as luteoferol, tricin, or lignin- shikimic acid complexes and degraded flavonolignans or flavosugars could provide further information to expert biologists on the relation between flavonoids and lignin biosynthesis.

It is primordial to realize the impact of the detection of the flavonoids and/or lignin - shikimic acid complexes, especially since it is well known that shikimate esters are intermediates in the lignin biosynthesis and they have never been described before to be involved in the wood network. Also, shikimic acid was never described previously to be attached to tricin or flavonoids derivatives and /or to be incorporated in any intermediates of the flavonoid's biosynthesis.

In conclusion, it seems appropriate to mention that the structural elucidation of lignins cannot depend anymore on calculating the ratio of the constituent lignols H: G:S by pyrolysis GC-EI-MS and by performing 2D ^{1}H-NMR experiments trying to recognize known expected linkages.

A wealth of scientific information is hidden inside the VRLs, which have not been modified by further chemical purifications or derivatizations.

References

[1] Del Río, J.C.; Rencoret, J.; Prinsen, P.; Martínez, A.T.; Ralph, J.; Gutiérrez, A.: Structural characterization of wheat straw lignin as revealed by analytical pyrolysis, 2D-NMR, and reductive cleavage methods. *J. Agr. Food Chem*. **2012**, *60*(23), 5922-5935.

[2] Lan, W.; Lu, F.; Régner, M.; Zhu, Y.; Rencoret, J.; Ralph, S.A.; et al.: Tricin, a flavonoid monomer in monocot lignification. *Plant Physiol.* **2015**, 167(4), 1284-1295.

[3] Li, M.; Pu, Y.; Yoo, C.G.; Ragauskas, A.J.: The occurrence of tricin and its derivatives in plants. *Green Chem*. **2016**, *18*(6): 1439-1454.

[4] Vanholme, R.; De Meester, B.; Ralph, J.; Boerjan, W.: Lignin biosynthesis and its integration into metabolism. *Curr. Opin. Biotechnol.* **2019**, *56*, 230-239.

[5] Maeda, H.A.: Lignin biosynthesis: tyrosine shortcut in grasses. *Nat plants*. **2016**, *2*(6), 1-2.

[6] Vanholme, R.; Cesarino, I.; Rataj, K.; Xiao, Y.; Sundin, L.; Goeminne, G.; et al.: Caffeoyl shikimate esterase (CSE) is an enzyme in the lignin biosynthetic pathway in Arabidopsis. *Science*. **2013**, 341(6150), 1103-1106.

[7] Lam, P.Y.; Tobimatsu, Y.; Matsumoto, N.; Suzuki, S.; Lan, W.; Takeda, Y.; et al.: OsCAldOMT1 is a bifunctional O-methyltransferase involved in the biosynthesis of tricin-lignins in rice cell walls. *Sci. Rep.* **2019**, *9*(1), 1-13.

[8] Doherty, W.O.; Mousavioun, P.; Fellows, C.M.: Value-adding to cellulosic ethanol: Lignin polymers. *Ind. Crops Prod*. **2011**, *33*(2), 259-276.

[9] Banoub, J.; Delmas, G.; Joly, N.; Mackenzie, G.; Cachet, N.; Delmas, M.: A critique on the structural analysis of lignins and application of novel tandem mass spectrometric strategies to determine lignin sequencing. *J. Mass Spectrom*. **2015**, *50*, 5-48.

[10] Qi, Y.; Volmer, D.A.: Chemical diversity of lignin degradation products revealed by matrix- optimized MALDI mass spectrometry. *Anal. Bioanal. Chem.* **2019**, *411*(23), 6031-6037.

[11] Lam, H.Q.; Le Bigot, Y.; Delmas, M.; Avignon, G.: A new procedure for the destructuring of vegetable matter at atmospheric pressure by a catalyst/solvent system of formic acid/acetic acid. Applied to the pulping of triticale straw. *Ind. Crops Prod*. **2001**, *14*, 139- 144.

[12] Strohalm, M.; Hassman, M.; Košata, B.; Kodíček, M.: mMass data miner: an open source alternative for mass spectrometric data analysis. *Rapid Commun. Mass Spectrom.* **2008**, *22*(6), 905-908.

[13] Schaefer, J.; Stejskal, E.O.: Carbon-13 nuclear magnetic resonance of polymers spinning at the magic angle. *J. Am. Chem. Soc.* **1976**, *98*(4), 1031-1032.

[14] Peersen, O.B; Wu, X.L.; Kustanovich, I.; Smith, S.O.: Variable-amplitude cross-polarization MAS NMR. *J. Magn. Res*. **1993**, *104*, 334-339.

[15] Atalla, R.H.; Vanderhart, D.L.: Native cellulose: a composite of two distinct crystalline forms. *Science*. **1984**, *223*(4633), 283-285.
[16] Morcombe, C.R.; Zilm, K.W.: Chemical shift referencing in MAS solid state NMR. *J. Magn. Res.* **2003**, *162*(2), 479-486.
[17] Sette, M.; Lange, H; Crestini, C.: Quantitative HSQC analyses of lignin: a practical comparison. *Comput. Struct. Biotechnol. J.* **2013**, *6*(7): 1-7.
[18] Yuan, T.Q.; Sun, S.N.; Xu, F.; Sun, R.C.: Characterization of lignin structures and lignin–carbohydrate complex (LCC) linkages by quantitative 13C and 2D HSQC NMR spectroscopy. *J. Agric. Food Chem.* **2011**, *59*(19), 10604-10614.
[19] Fox, S.C.; McDonald, A.G.: Chemical and thermal characterization of three industrial lignins and their corresponding lignin esters. *BioResources.***2010**, *5*(2), 990-1009.
[20] Wen, J.L.; Sun, S.L.; Xue, B.L.; Sun, R.C.: Recent advances in characterization of lignin polymer by solution-state nuclear magnetic resonance (NMR) methodology. *Materials.* **2013**, *6*(1), 359-391.
[21] Bianchi, T.S.: *Biogeochemistry of Estuaries*; Oxford University Press: New York, NY; **2007**.
[22] Baldock, J.A.; Nelson, P.N.: Soil Organic Matter, Section B. In *Handbook of Soil Science*; Sumner, M.E, Ed.; CRC Press: Boca Raton, FL, **2000**; pp 25–84.
[23] Ramezani, N.; Sain, M.: Thermal and Physiochemical Characterization of Lignin Extracted from Wheat Straw by Organosolv Process. *J. Polym. Environ*. **2018**, *26*(7), 3109-3116.
[24] Wang, Y.; Shen, J.Z.; Chan, Y.W.; Ho, W.S.: Identification and Growth Inhibitory Activity of the Chemical Constituents from Imperata Cylindrica Aerial Part Ethyl Acetate Extract. *Molecules.* **2018**, *23*(7), 1-13.
[25] Jiang, B.; Zhang, Y.; Guo, T.; Zhao, H.; Jin, Y.: Structural characterization of lignin and lignin-carbohydrate complex (LCC) from ginkgo shells (Ginkgo biloba L.) by comprehensive NMR spectroscopy. *Polymers.* **2018**, *10*(7), 1-17.
[26] Sleno, L.; Volmer, D.A.; Marshall, A.G.: Assigning product ions from complex MS/MS spectra: the importance of mass uncertainty and resolving power. *J. Am. Soc. Mass Spectrom.* **2005**, *16*(2), 183-198.
[27] Demarque, D.P.; Crotti, A.E.; Vessecchi, R.; Lopes, J.L.; Lopes, N.P. Fragmentation reactions using electrospray ionization mass spectrometry: an important tool for the structural elucidation and characterization of synthetic and natural products. *Nat. Prod. Rep*. **2016**, *33*(3), 432-455.
[28] Schreiver, I.; Eschner, L.M.; Luch, A.: Matrix-assisted laser desorption/ionization tandem mass spectrometry for identification of organic tattoo pigments in inks and tissue samples. *Analyst*. **2018**, *143*(16), 3941-3950.
[29] Niederer, M.; Hauri, U.; Kroll, L.; Hohl, C.: Identification of organic pigments in tattoo inks and permanent make-ups using MALDI-TOF mass spectrometry. *F1000Research.* **2017**, *6*: 1-14.
[30] Kind, T.; Fiehn, O.: Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC bioinform*. **2007**, *8*(1), 1-20.

[31] Ma, Y.; Kind, T.; Yang, D.; Leon, C.; Fiehn, O.: MS^2Analyzer: a software for small molecule substructure annotations from accurate tandem mass spectra. *Anal. Chem.***2014**, *86*(21), 10724-10731.
[32] Mabry, T.J.; Markham, K.R.: Mass Spectrometry of Flavonoids. In *The Flavonoids*; Harborne, J.B., Mabry, T.J., Mabry, H, Eds.; Springer: Boston, MA, **1975**, pp 78-126.
[33] Schiller, J.; Süß, R.; Fuchs, B.; Müller, M.; Petković, M.; Zschörnig, O.; et al.: The suitability of different DHB isomers as matrices for the MALDI-TOF MS analysis of phospholipids: which isomer for what purpose?. *Eur. Biophys. J.* **2007**, *36*(4-5), 517-527.
[34] Bajpai, L.; Varshney, M.; Seubert, C.N.; Stevens, S.M.; Johnson, J.V.; Yost, R.A.; et al.: Mass spectral fragmentation of the intravenous anesthetic propofol and structurally related phenols. *J. Am. Soc. Mass Spectrom*. **2005**, *16*(6), 814-824.
[35] Hao, C.; Ma, X.; Fang, S.; Liu, Z.; Liu, S.; Song, F.; et al. Positive-and negative-ion matrix-assisted laser desorption/ionization mass spectrometry of saccharides. *Rapid Commun. Mass Spectrom.* **1998**, *12*(7), 345-348.
[36] Petković, M.; Schiller, J.; Müller, M.; Benard, S.; Reichl, S.; Arnold, K.; Arnhold, J.: Detection of individual phospholipids in lipid mixtures by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry: phosphatidylcholine prevents the detection of further species. *Anal. Biochem.* **2001**, *289*(2), 202-216.
[37] Neubert, H.; Halket, J.M.; Ocana, M.F.; Patel, R.K.: MALDI post-source decay and LIFT-TOF/TOF investigation of α-cyano-4-hydroxycinnamic acid cluster interferences. *J. Am. Soc. Mass Spectrom.* **2004**, *15*(3), 336-343.
[38] Wallace, W.E.; Arnould, M.A.; Knochenmuss, R.: 2,5-Dihydroxybenzoic acid: laser desorption/ionisation as a function of elevated temperature. *Int. J. Mass Spectrom.* **2005**, *242*(1), 13-22.
[39] Keller, B.O.; Li, L.: Discerning matrix-cluster peaks in matrix-assisted laser desorption/ionization time-of-flight mass spectra of dilute peptide mixtures. *J. Am. Soc. Mass Spectrom.* **2000**, *11*(1), 88-93.
[40] Morreel, K.; Dima, O.; Kim, H.; Lu, F.; Niculaes, C.; Vanholme, R.; et al.: Mass spectrometry-based sequencing of lignin oligomers. *Plant physiol.* **2010**, *153*(4), 1464-1478.
[41] Oudia, A.; Mészáros, E.; Jakab, E.; Simões, R.; Queiroz, J.; Ragauskas, A.; et al.: Analytical pyrolysis study of biodelignification of cloned Eucalyptus globulus (EG) clone and Pinus pinaster Aiton kraft pulp and residual lignins. *J. Anal. Appl. Pyrol.* **2009**, *85*(1-2), 19-29.
[42] Ralph, J.; Hatfield, R.D.: Pyrolysis-GC-MS characterization of forage materials. *J. Agr. Food Chem.***1991**, *39*(8), 1426-1437.
[43] Vanholme, R.; Morreel, K.; Darrah, C.; Oyarce, P.; Grabber, J.H.; Ralph, J.; et al.: Metabolic engineering of novel lignin in biomass crops. *New Phytol.* **2012**, 196(4), 978-1000.
[44] Domon, B.; Costello, C.E.: A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. *Glycoconj. J.* **1988**, 5(4), 397-409.

Chapter 4

Top-down lignomics analysis of the French pine lignin by atmospheric pressure photoionization quadrupole time-of-flight tandem mass spectrometry: Identification of a novel series of lignin-carbohydrate complexes

This Chapter is reproduced with permission from Mikhael, A.; Fridgen, T.D.; Delmas, M.; Banoub, J.: **Top-down lignomics analysis of the French pine lignin by atmospheric pressure photoionization quadrupole time-of-flight tandem mass spectrometry: Identification of a novel series of lignin–carbohydrate complexes.** *Rapid Commun. Mass Spectrom*. 2020; 34: e8910. https://doi.org/10.1002/rcm.8910.

4.1. Abstract

We report the top-down lignomics analysis of the virgin released lignin (VRL) extracted from the French pine wood by using atmospheric pressure photoionization quadrupole time-of-flight APPI-QqTOF-MS and low-energy collision-induced dissociation tandem mass spectrometry CID-MS/MS.

Fourteen novel lignin-carbohydrate complexes (LCC) were identified using the APPI-QqTOF-MS/MS of the very complex mixture of virgin released lignins (VRLs), directly extracted from the French pine wood without any kind of purification. The low-

energy CID-MS/MS analyses allowed us to establish the fragmentation patterns of the precursor ions and to identify the complex structures of the identified protonated LCCs molecules. This novel identified series of LCCs were composed of one or two carbohydrate rings to which one, two, or three lignin units were covalently attached. In addition to the fourteen LCCs, acetyl eugenol was identified in the French pine VRL sample. The identification of acetyl eugenol indicates the possible lignin degradation and modification (acetylation) during the mild extraction method developed by the Compagnie Industrielle de la Matière Végétale (CIMV). These series of LCC provide evidence that lignins are covalently linked to carbohydrates in the native wood network and act as a cross-linkers between cellulose and hemicellulose components of wood.

4.2. Introduction

All vascular plants and woods cell walls (CW) are basically composed of cellulose, hemicelluloses, and lignins, which are formed in large quantities using solar energy.[1] Lignin is the second most abundant natural biopolymer and constitutes the most abundant natural resource of aromatic moieties. **[2-5]** Lignins are generally incorporated with hemicelluloses in the spaces between cellulose microfibrils to harden the wood and to provide hydrophobicity for water movement.[2-5] Lignin oligomers are composed mainly of three monomers: *p*-coumaryl (H), coniferyl (G), and sinapyl (S) alcohols linked covalently with various types of lignin linkages. [3-7] In addition, recently, some uncommon lignin monomers were identified, such as caffeyl alcohol (C), 5-hydroxyguaiacyl unit (F).[8]

In addition, gallyl lignin or 5-hydroxy caffeyl alcohol (L) was hypothetically proposed to exist as a possible lignin monomer.[8]

In the last century, lignin oligomers were speculated to exist incorporated in one single-molecule structure that may never ceased to grow.[9] As a result, the chemical structure of lignin was described as a highly cross-linked and three-dimensional biopolymer.[10] For this reason, lignin researchers felt obliged to depolymerize the raw extracted lignin before analysis.[11] However, because of the harsh conditions used in the extraction of lignin from the CW, and further attempts of depolymerization, obtaining lignin in its unaltered form is indeed a myth. Therefore, there is still much debate on whether any lignin extract adequately represents the native lignin structure. The structural analysis of lignin, therefore, has remained a challenge that presented a real burden on analytical scientists.

We have introduced a new paradigm that stated that there are no huge lignin macromolecules; instead, we proposed that the extracted lignins were composed of linear oligomers that are connected to cellulose and hemicellulose in a grid pattern forming the network of wood.[11] We also have asserted that for lignomics structural elucidation, the analytical work should be completed on the series of the lignin oligomers released from the wood polysaccharide-lignin network, without any further depolymerization, purification or chemical modifications; this type of extracted lignin oligomers are called virgin released lignin (VRL).[11]

It has been proposed that lignin-carbohydrate complexes (LCC) are formed by lignin oligomers that are attached to arabino-glucuronoxylan (Hemicellulose) through five

different types of linkages.[12] The most important linkages found were of the benzylic ethers (BE) bond types that are formed between the lignin benzylic carbon (α-position) and the hydroxyl group of the sugar rings. Likewise, gamma ester bonds (GE) could be formed between the hydroxyl group at the lignin γ position and the carboxyl group of sugar acids. Other different types of carbohydrate-lignin linkages (LCC) are shown in Figure A3.1.[12-14] All investigative research to deduce the linkages between lignin, cellulose, and hemicellulose, has indicated that ferulic acid residues can be linked to arabinoxylan via ester linkages between the carboxylic acid group of ferulic acid and the C-5 primary alcohol of the arabinose moiety.[15] This latter ferulic acid residue attached to the sugar ring can be reconnected to another lignin oligomer via an ether bond (oxidative coupling) to form the ferulate-polysaccharide–lignin complexes.[15,16-19]

Nevertheless, the presence of covalently linked lignin-carbohydrate complexes has always continued to be one of the most controversial topics in lignocellulose chemistry.[12] This is mainly due to the fact that researchers have had enormous difficulty in isolating and characterizing the LCCs. This is not surprising as these LCCs are most probably lost (hydrolyzed and decomposed) during the depolymerization, degradation, and purification attempts before analysis. For these reasons, the majority of information on the LCC structures was usually obtained from model synthetic compound experiments.[12] Recently, some methods to "quantify" the different lignin structures and LCC linkages by using a combination of quantitative ^{13}C- and 2D-HSQC NMR techniques were suggested to be effective. However, the accurate assignment of the associated LCCs carbohydrates and linkages have not yet been established.[12]

To our knowledge, there is only one example of the analysis of LCC using mass spectrometry in literature. In this example, chloride doping, ESI-QTOF-MS (- ion mode), and low-energy CID tandem mass spectrometry were used to analyze some undegraded LCCs.[20] These complexes ranged in mass from *m/z* 326−714 Da and exhibited both carbohydrate and lignin characteristics. [20]

In this manuscript, we have used APPI-QqTOF-MS (+ ion mode) and Low-Energy CID-MS/MS analyses to investigate the structure of the VRLs extracted from French pine. We have used the CIMV extraction method to isolate the virgin French pine lignin.[21] This method uses acidic solvents ($HCOOH/CH_3COOH$), which decrease the possibility of metal adduct formation. Interestingly, from the analysis of this VRL mixture, we have identified fifteen compounds: acetylated eugenol and fourteen novel LCCs.

4.3. Experimental

4.3.1. Sample preparation

The French pine lignin sample was provided by Professor Michel Delmas. The lignin sample was extracted using the CIMV procedure, which selectively separates the cellulose, hemicellulose, and lignin at atmospheric pressure, and allows the destructuring of the vegetable matter.[21] The solvent system used was a mixture of formic acid/acetic acid/water (30/50/20) which produced, after precipitation with water and filtration, the French pine VRL lignin. Approximately, 0.1 mg of pine lignin powder was dissolved in 1 ml of dioxane-methanol (2:1) for MS analysis.

4.3.2. Solid-state ^{13}C-NMR

Solid-state ^{13}C NMR experiments were performed on a Bruker AVANCE-400 Wide Bore NMR spectrometer (Bruker BioSpin, Rheinstetten, Germany) operating at a ^{13}C resonance frequency of 106 MHz and using a commercial solid probe H/X cross-polarization/magic angle spinning (CP/MAS) instrument (Bruker BioSpin). About 90 mg of the sample was placed in a zirconium dioxide rotor with an outer diameter of 4 mm, which was spun at 10 kHz at the magic angle.[21-23] The cross-polarization (CP) technique was applied with a ramped 1 H-pulse starting at 100% power and decreasing to 50% power. This procedure was used during the contact time of 2 ms, to circumvent the Hartmann-Hahn mismatches.[24] The experiments were performed at ambient temperature, and 16K scans were accumulated using a delay of 2 s. The ^{13}C chemical shifts were referenced to tetramethylsilane and calibrated with the glycine carbonyl signal, set at 176.5 ppm. [22,23, 25]

4.3.3. APPI-QqTOF-MS and Low-Energy CID-MS/MS

Mass spectrometry was performed using an Applied Biosystems (Foster City, CA, USA) API-QSTAR-XL MS/MS quadrupole orthogonal time-of-flight (QqTOF)-MS/MS hybrid instrument. APPI was performed with a PhotoSpray ion source (Applied Biosystems) operated at 1300V at a temperature of 400 ºC, with all acquisitions performed in the positive ion mode. Samples were infused into the mass spectrometer with an integrated Harvard syringe pump at a rate of 0.1 mL/min. The auxiliary nebulizer gas pressure setting was fixed at 25 psi, and the nebulizer gas pressure at 74 psi. The curtain gas pressure was set at 30 psi. The declustering potential (DP) was set at + 100 eV. The focus potential (FP) was adjusted to +100 V. Toluene was selected as the dopant for its

ability to undergo trouble-free photoionization at 8.83 eV. The eluent was composed of dioxane/methanol (2:1). No modifier was used to enhance ion production.

The mass calibration of the ToF analyzer in the positive ion mode was performed with the PhotoSpray ion source, using 1,2,3,5-tetra-O-acetyl-β −D-ribofurnanose and checking for the exact masses of the $[M+H]^+$ ion $[C_{13}H_{18}O_9 + H]^+$ at *m/z* 319.1029 and the $[M+H–AcOH]^+$ ion $[C_{11}H_{14}O_7 + H]^+$ at *m/z* 259.0817. Calibration for higher masses was performed with octa-O-acetyl-ß-D-lactopyranose and checking for the $[M+H]^+$ ion $[C_{28}H_{38}O_{19} + H]^+$ at *m/z* 679.2085. Product ion spectra were obtained on the same instrument as described above. Nitrogen was used as the collision gas for MS/MS analyses with collision energies varying between 10 and 35 eV. Collision energy (CE) and CID gas conditions were adjusted in each acquisition such that the precursor ion remained abundant in the product ion spectra. We have used in general a 1 m/z unit resolution for the MS/MS selection of the precursor ion for simplification of the analysis.

4.4. Result and Discussion

4.4.1. Solid-state ^{13}C-NMR obtained for the French pine VRL

The solid-state ^{13}C-NMR of the French pine VRL sample (Figure A3.2) showed *inter alia* two major peaks. The first peak was detected at 57 ppm, indicating the presence of aromatic methoxy groups in the guaiacyl lignin units.[26] The second peak was detected at 150 ppm, supporting the existence of aromatic C-3 of guaiacyl units.[26] Other solid-state ^{13}C-NMR peaks are listed in Table A3.1. It should be noted that most of the assignments in the aromatic regions were attributed to G lignin units because pine (softwood) lignins

are characterized by the abundance of guaiacyl lignin units.[27] Furthermore, previous TOF-SIMS studies of milled wood pine lignin supported this conclusion. [27] Lastly, it should be noted that the major aliphatic peak at 21 ppm may be attributed to the methyl group of acetic acid that was used in the CIMV extraction. It should be noted that during the CIMV extraction process, acetic and/or formic acid may react with the free hydroxyl groups in carbohydrate rings and/or phenolic units (formylation and acetylation).This indicates that even mild extraction processes can alter the structure of the obtained lignin oligomers.

4.4.2. APPI-QqTOF-MS of the French pine VRLs

The VRLs mixture obtained by the CIMV extraction of French pine wood was subject to APPI-MS (+ ion mode) using a QqTOF-MS instrument. The VRLs mixture was introduced by direct infusion in the ionization source without any prior separation, as this deemed to be an impossible task.

A quick perusal of the APPI-QqTOF-MS (+ ion mode) French pine VRLs indicates a very complex mixture of intertwining ions, which obviously would complicate the measuring of isotopic patterns (Figures A3.3 and Figure A3.4). The complexity of the APPI-MS most probably arises from the variation in the presence of different possible lignin monomers, isomeric lignin oligomers, and the presence of novel lignin-carbohydrate complexes (LCC). Moreover, as mentioned before, in section 3.2, possible acetylation

and/or formylation during the extraction procedure may complicate matters further. Just to explain the ambiguity of Figure A3.3 adequately, it should be noted for the recorded MS from *m/z* 200 to *m/z* 600, we should expect 400 ion signals. Unfortunately, this is not the case, as we definitely see more than 400 ion signals (ca > 800 signals). This is due to the strong likelihood of presence of isobaric interferences at each *m/z* (centroid mass shift and overlapping isotopic distribution). Also, it is important to note that the magnification of any single recorded ion mass indicates that it is definitely composed of more than one ion (Figure A3.5). For these reasons, we have enlarged the APPI-QqTOF-MS into various *m/z* ranges, in an attempt to descramble the ions distributions and to pinpoint the ions chosen for further low-energy CID-MS/MS analyses. (Figure A3.4). For comparison sake, our APPI-MS look alike the one described by Qi and Volmer using the high- resolution LDI-FTICR-MS of an alkali lignin sample which was recorded from *m/z* 100-799 and was indeed of 3370 monoisotopic formulas.[28]

In this study, we have fished-out and identified some of the major ions (high intensities) from this complicated APPI-MS that contains overlapped isobars at every *m/z* unit. Also, it is essential to mention that the APPI-QqTOF-MS recorded at different *m/z* value ranges gave different MS blueprints for the same concentration of the infused lignin analyte (see Figure A3.3).

In this manuscript, and for the sake of brevity, we are only discussing specifically the series of protonated molecules at *m/z* 207.10, 279.14, 293.15, 323.18, 409.31, 433.16, 461.17, 489.17, 517.18, 533.18, 611.24, 637.25, 651.25, 663.46, 683.26, 695.26 and

711.27, obtained after APPI-QqTOF-MS of the French pine VRLs (Figures A3.3, Figure A3.4 and Table 4.1).

The major ions at *m/z* 409.31 and 663.46 (Figure A3.4) are omitted from the study. The exact mass of the ion at *m/z* 409.31 indicates that this ion has a lower DBE, which suggests that this ion may be related to fatty acids, as shown before by Jarrell et al. in the analysis of organosolv switchgrass lignin. [29] The other ion at *m/z* 663.46 is suggested to be the oxidized form of the contaminant orgafos-168, as shown before in literature.[30]

The identification of the lignin oligomers was performed according to the presence of one heteroatom (Oxygen), mass error (within 0.1 *m/z* unit), the possible combinations of the different lignol units, and reasonable MS/MS fragmentation patterns. Furthermore, the double bond equivalent (DBE), was calculated for the precursor ions to determine the number of unsaturation. Lastly, it should be noted that the values of DBE and C/O ratio for the identified formulas were consistent with the values reported before in literature.[29, 31] Typically, lignin monomers possess the following values: DBE $\geq$ 4 and C/O ratio $\geq$ 2.5, while lignin dimers possess these values: DBE $\geq$ 8 and C/O $\geq$ 2.5. [29, 31] For a lignin carbohydrate dimer ion, typical values are DBE $\leq$ 6 and C/O $\leq$ 2, which indicates that replacing a lignin monomer with a carbohydrate ring leads to a lower DBE and lower C/O ratio. [29, 31]

Table 4.1. Identification of lignin Oligomers in the Mass Spectrum of French pine lignin VRLs extracted from French pine Wood using positive ion mode APPI-QqTOF-MS and Low-Energy CID-MS/MS

Ion	Chemical Formula	Experimental *m/z*	Theoretical *m/z*	Error (*m/z*)	DBE	MS/MS
1	$[C_{12}H_{14}O_3 + H]^+$	207.10	207.10	0.00	5.5	192, 179, 165, 137, 123, 109, 97, 91, 69
2	$[C_{15}H_{18}O_5 + H]^+$	279.14	279.12	+ 0.02	6.5	264, 249, 219, 191, 177, 173, 163, 149, 145, 135, 123, 117, 105, 95, 57
3	$[C_{16}H_{20}O_5 + H]^+$	293.15	293.14	+ 0.01	6.5	261, 233, 229, 135, 57
4	$[C_{16}H_{18}O_7 + H]^+$	323.18	323.11	+ 0.07	7.5	291, 263, 249, 245, 207, 189, 165, 128, 119, 91, 57
5	$[C_{19}H_{28}O_{11} + H]^+$	433.16	433.17	- 0.01	5.5	401, 386, 375, 359, 341, 309, 295, 283, 277, 249, 167, 163, 149, 137, 131
6	$[C_{20}H_{28}O_{12} + H]^+$	461.17	461.17	0.00	6.5	446, 443. 428, 417, 401, 383, 368, 337, 309, 295, 277, 151, 137
7	$[C_{25}H_{28}O_{10} + H]^+$	489.17	489.18	-0.01	11.5	471, 446, 443, 428, 411, 397, 383, 365, 351, 337, 323, 309, 294, 173, 163, 161, 151, 137
8	$[C_{26}H_{28}O_{11} + H]^+$	517.18	517.17	+0.01	12.5	499, 485, 471, 457, 443, 428, 411, 393, 379, 363, 337, 321, 307, 297, 289, 277, 203, 191, 175, 161, 151, 137
9	$[C_{26}H_{28}O_{12} + H]^+$	533.18	533.17	+0.01	12.5	515, 487, 472, 455, 441, 427, 413, 363, 321, 307, 297, 277, 203, 175, 161, 151, 137
10	$[C_{31}H_{40}O_{14} + H]^+$	637.25	637.25	0.00	11.5	619, 513, 457, 333, 323, 307, 151, 137
11	$[C_{31}H_{38}O_{15} + H]^+$	651.25	651.23	+0.02	12.5	633, 621, 605, 514, 424, 321, 309, 295, 283, 175, 151,137
12	$[C_{33}H_{42}O_{16} + H]^+$	695.26	695.25	+ 0.01	12.5	677, 649, 621, 587, 361, 337, 321, 307, 295, 283, 175, 161, 151, 137
13	$[C_{33}H_{38}O_{11} + H]^+$	611.24	611.25	-0.01	14.5	593, 562, 487, 356, 343, 321, 307, 297, 283, 277, 161, 147, 137
14	$[C_{35}H_{38}O_{14} + H]^+$	683.26	683.23	+ 0.03	16.5	637, 321, 309, 297, 283, 265, 151, 137
15	$[C_{36}H_{38}O_{15} + H]^+$	711.27	711.23	+ 0.04	17.5	665, 459, 349, 333, 321, 309, 297, 283, 265, 161, 151, 137

4.4.3. Low-Energy CID-MS/MS Analyses

Forthrightly, the APPI-QqTOF-MS of the French pine VRLs (Figures A3.3 and Figure A3.4) may look more like a horror image for the analyst! The complexity of this intricate spectrum indicates a very complicated mixture of released VRL oligomers, obtained from one of the supposedly mildest solvolysis extraction methods. For these reasons, assignments of chemical formulas based on mass accuracy solely remain challenging because of the complex overlapping isotopic distribution and centroid mass shifts.[32] However, a reasonable low-energy CID-MS/MS fragmentation mechanism for the proposed chemical structure of each precursor ion would provide more support to the assigned formulas.[33] This will decrease the number of possible candidates that could be attributed to each ion in the conventional APPI-MS.

In the present work, it became apparent that in some cases, the precursor ions were indeed a mixture of different isobars. To add more complexity to this situation, each possible isobar that could be assigned to one precursor ion can also be represented by different isomeric structures (isomers). All these factors lead to the formation of "messy" product ion scans, which lead to the formation of many diagnostic product ions; this will be discussed later in this section. Accordingly, we will describe several low-energy CID-MS/MS measurements of some major protonated molecules, in a quest to deduce their molecular structures and identities. This study allowed us to identify acetylated eugenol and fourteen novel lignin-carbohydrate complexes. The summary of the lignin derivatives nomenclature used to describe the lignin moieties in this series of LCC is shown in Figure A3.6.

In general, the MS/MS spectrum of each identified LCC was composed of three types of product ions. The first type of product ions resulted from the cleavage of the carbohydrate ring and/or resulting small losses. The second type resulted from the cleavage of the propanoid aliphatic part of the lignin units, and finally the third type which were produced from the aromatic product ions, these were evidently diagnostics for the presence of the aromatic lignin units. It essential to mention that our described MS/MS cleavages resemble the ones described for the oxidized dicaffeoyl quinic acid, feruolyted sugars, and some recently reported LCCs. [20,34,35]

Please note that most of the identified LCCs MS/MS showed the formation of the important aromatic product ions at *m/z* 137 ($C_8H_9O_2^+$) and *m/z* 151 ($C_9H_{11}O_2^+$), which are typical for the presence of guaiacyl lignol portions and/or its related isomers. [27] It is important to note that these aromatic product ions were previously identified as the characteristic ions in the TOF-SIMS of milled pine wood lignin.[27]

It will become evident that all MS/MS analyses performed on this series of protonated molecules displayed almost the same gas-phase fragmentation mechanisms and supported the presence of sugar moiety attached to lignin monomers (Lignin – Carbohydrate complexes LCC).

The conventional Kendrick mass defect (KMD) (using CH_2 units) of the APPI-MS spectra at different *m/z* ranges were plotted (Figure A3.7). The KMD plots showed that these series of ions are closely related to each other as they appeared as one bundle of closely related ions. This indicates that the recorded MS is not composed of a random series of ions, but these ions are structurally related to each other. Therefore, it was not surprising

that most of the selected ions in this work are related to different classes of LCCs. In general, the KMD plots can be represented by different horizontal lines. The ions that differ in the number of CH_2 and possess the same mass defect are always aligned on the same horizontal line.[36]

In addition, KMD plots are not restricted to the use of CH_2 units only, but they can be modified by using other units that are suitable for the studied class of compounds. [37] For example, the [OCH_2] units were used to align a series of lignin monomers that differ in the number of methoxy groups such as coumaric acid, ferulic acid, and sinapic acid. [37] For this reason, the KMD plot using [OCH_2] units of the APPI-MS spectrum in the range of *m/z* 400-800 were plotted, as shown in Figure A3.8a. The first benefit of this plot is that it can be used as a mass defect filtering tool. [38] For example, the major ions at *m/z* 409 and *m/z* 663 that were omitted from this study have different mass defect range with respect to the major bundle of ions (Figure A3.8a). The second benefit from the KMD plot is that it allows the prediction of the ions formulas and/or structures that are aligned on the same horizontal line (same class).[37] For example, the KMD plot using [OCH_2] units align some ions that differ by 30 Da (OCH_2) on the same horizontal ion such as the ions at *m/z* 487 and *m/z* 517, and the ions at *m/z* 581 and *m/z* 611 (Figure A3.8b). Based on CID-MS/MS analysis, the chemical formulas $[C_{26}H_{28}O_{11} + H]^+$ and $[C_{33}H_{38}O_{11} + H]^+$ were assigned to the precursor ions at *m/z* 517 and *m/z* 611, respectively (Table 4.1). As a result, the chemical formulas of the ions at *m/z* 487 and *m/z* 581 can be predicted easily by removing one [OCH_2] unit from the formulas of the ions at *m/z* 517 and *m/z* 611, respectively. Accordingly, the formulas $[C_{25}H_{26}O_{10} + H]^+$ and $[C_{32}H_{36}O_{10} + H]^+$ will be assigned to the

ions at *m/z* 487 and *m/z* 581, respectively. Moreover, it should be noted that all lignin units (Figure A3.6) identified in this series of LCCs can be selected as the base unit for plotting various KMD plots, which can be used to predict more formulas and classes of this complex lignin mixture. This will be represented latter in a separate work focusing on KMDs plots and more CID MS/MS analyses.

4.4.3.1. Product ion scan of the simple lignol monomer *1*

The formula $[C_{12}H_{14}O_3 + H]^+$ (DBE=5.5) was assigned to the protonated molecule *1* at *m/z* 207.10 (Figure A3.9 and Scheme A3.1). This protonated molecule *1* was identified as acetylated eugenol *1*.

The product ion scan of precursor ion *1* showed the loss of methyl radical to yield the protonated radical product ion at *m/z* 191. The loss of ethylene molecule (-28 Da) from ion *1* leads to the formation of the benzylic cation at *m/z* 179. This latter product ion at *m/z* 179 loses the acetyl group, likely in the form of ketene CH_2=C=O (- 42 Da), to produce the secondary product ion at *m/z* 137. This latter product ion at *m/z* 137 loses one molecule of carbon monoxide to generate the product ion at *m/z* 109. Also, the product ion at *m/z* 137 losses one methanol molecule to yield the product ion at *m/z* 105.

Moreover, the product ion scan of the precursor ion *1* showed the loss of the three-carbon aliphatic chain to yield the aromatic product ion at *m/z* 165. This latter ion eliminates the acetyl group in the form of ketene CH_2=C=O (-42 Da) to form the secondary product ion at *m/z* 123. This latter ion can lose either a methanol molecule to produce the secondary ion at *m/z* 91 or undergo an RDA reaction to produce the ion at *m/z* 97. This last

ion at *m/z* 97 loses one molecule of carbon monoxide to yield the product ion at *m/z* 69, which contains the aromatic cyclopropenyl ring. Lastly, the product ion at *m/z* 69 losses one methylene carbene to yield the product ion at *m/z* 55.

4.4.3.2. The product ion scan of LCC composed of a glycose residue attached to one lignin unit

The precursor ions *2*, *3* and *4* at *m/z* 279.14, 293.15 and 323.18, respectively, are tentatively assigned to be composed of a glycose residue attached to one lignin unit. For the sake of brevity, we are discussing here the MS/MS fragmentation pathway of the precursor ion *2* at *m/z* 279.14, while the other precursor ions *3* and *4* are described in detail in section A3.1. The formula $[C_{15}H_{18}O_5 + H]+$ (DBE=6.5) was assigned to the protonated molecule 2 at m/z 279.14 (Figure A3.10 and Scheme A3.2a and A3.2b). In order to explain the very complex top-down analysis of the French Pine VRL, we were faced in this work, with a unique challenge, as each selected precursor ion, could indeed be composed of a mixture of isomeric structures (identical isobars). This is why the product ion scan of the precursor ion *2* gives a very complex spectrum arising from both identical molecular weight isomers. Hence, the first isomer component of this precursor ion was attributed to isomer A, which was attributed to being composed of a dehydrated deoxy-pentose moiety covalently attached to coniferyl alcohol, whereas, the second isomer B was attributed to a dideoxy-pentose moiety covalently attached to 3,4-methylenedioxy cinnamyl alcohol residue.

Consequently, the product ion scan of precursor ion *2* (isomer A) displayed the loss of 15 Da in the form of methyl radical to produce the radical product ion at *m/z* 264

(Scheme A3.2a). Similarly, this precursor ion *2* can lose a molecule of formaldehyde (CH_2O) from the aromatic ring, to produce the secondary ion at *m/z* 249. This latter ion at *m/z* 249 underwent further cleavage of the C_1-C_7 bond of the lignin unit to generate the secondary product ion at *m/z* 95 ($C_6H_7O^+$), which supports the presence of the lignin phenolic unit. This latter ion losses one carbon monoxide molecule to form the ion at *m/z* 67.

In addition, this precursor ion *2* (Isomer A) can suffer a cleavage in the lignin unit Cγ-O bond to form the product ion at *m/z* 163; this fragmentation supports the presence of guaiacyl unit as reported before in literature.[39] Moreover, Isomer A can also fragment by cleavage of the lignin unit Cβ- Cγ bond to form the product ion at *m/z* 149. This last ion can eliminate 32 Da one methanol molecule to form the secondary ion at *m/z* 117.

Moreover, Isomer A, once more, can be subjected to the cleavage of the C_1-Cα bond in the coniferyl alcohol unit to create the product ion at *m/z* 123. Lastly, the sugar ring in the precursor ion undergoes a ring cleavage to yield the product ion at *m/z* 57 ($^{0,3}A$), as demonstrated in scheme A3.2a. It should be noted this notation of product ions obtained from the sugar ring cleavage was defined according to Domon and Costello nomenclature.[40] Also, it should be notes that m/z 57 could be assigned as $C_4H_9^+$ that can originate from another possible isomer and/or isobar containing a butyl chain.

To complicate the matter further, in the same time frame, the precursor ion *2* (Isomer B) can fragment by a sugar ring cleavage to produce the product ion at *m/z* 219 ($^{2,4}X$) as shown in Scheme A3.2b. This latter ion loses 46 Da (- CH_2O_2) from the

methylenedioxy moiety to produce the secondary product ion at *m/z* 173. This latter fragmentation was reported before for methylenedioxy derivatives. [41] Moreover, the $^{2,4}X$ ion at *m/z* 219 can also eliminate an ethylene molecule to yield the secondary ion at *m/z* 191.

Also, this precursor ion 2 (isomer B) is subjected to cleavage of the C_5-O bond of the sugar moiety to form the ion at *m/z* 177. This latter product ion loses a molecule of methanol to form the ion at *m/z* 145. Furthermore, the cleavage of the Cα-Cβ bond of the lignin unit that is present in the precursor ion *2* (Isomer B) leads to the production of the benzylic ion at *m/z* 135. Lastly, this latter product ion loses 30 Da (CH_2O) to yield the ion at *m/z* 105. Once more, we would like to point out that the finding of two isomeric precursor ions that have identical formulas really show the complexity of the VRL French pine lignin mixture.

To add more complexity to this situation, we would like to discuss in this example the possibility of the presence of other isobars (another formula) that could be assigned to the precursor ion 2. The formula $[C_{12}H_{22}O_7 + H]^+$ could be assigned to the precursor ion *2* at *m/z* 279.14, which is composed of a hexose attached to hexanoic acid through ester bond formation. As shown in Scheme A3.3 , this ion can form the product ions at *m/z* 249, *m/z* 219, *m/z* 163, *m/z* 145, *m/z* 135, *m/z* 117, *m/z* 105, *m/z* 57. This indicates that some of the product ions obtained in that MS/MS could be a mixture of at least two isobaric product ions formed from two different isobars (two different formulas). However, this formula cannot lose methyl radical as this kind of fragmentation is known for the aromatic methoxy group. Additionally, it cannot form aromatic product ions at *m/z* 177, *m/z* 149, *m/z* 123,

and most importantly, the protonated phenol at *m/z* 95. These facts do not omit the presence of this second isobar $[C_{12}H_{22}O_7 + H]^+$, but it shows that the assigned isobar $[C_{15}H_{18}O_5 + H]^+$ (Isomer A and B) are the most suitable assignment as it gives all major and characteristic product ions smoothly. This exemplifies the birth of what we have resolved to qualify as "Messy MS/MS".

4.4.3.3. The product ion scan of LCC composed of two glycoses attached to one lignin unit

The precursor ions *5* and *6* at *m/z* 433.16 and 461.17, respectively, are tentatively assigned to be composed of two glycoses attached to one lignin unit. For the sake of brevity, we are discussing here the MS/MS fragmentation pathway of the precursor ion *6* at *m/z* 461.17, while the other precursor ions *5* is described in detail in section A3.2. The formula $[C_{20}H_{28}O_{12} + H]^+$ (DBE=6.5) was assigned to the protonated molecule *6* at *m/z* 461.17 and appeared to be a trimer composed of penturonic acid, pentose and dihydroconiferyl lignin unit (Isomer A). Also, another possible isomer B was proposed to be composed of deoxyhexose, oxidized pentose, and 5-hydroxy dihydrocaffeyl alcohol. (Figure 4.1 and Scheme 4.1).

The product ion scan of the precursor ion at *m/z* 461.17 was initiated by the loss of methyl radical from the aromatic guaiacyl unit to create the radical product ion at *m/z* 446. Similarly, the product ion scan of this precursor ion eliminated a carbon dioxide molecule from the penturonic acid ring to produce the ion at *m/z* 417. Likewise, the product ion scan can also be initiated by the loss of a water molecule to create the product ion at *m/z* 443. This latter ion at *m/z* 443 can lose a methyl radical from the guaiacyl unit to produce the

radical product ion at *m/z* 428. Furthermore, this latter product ion at *m/z* 428 losses 60 Da through the cleavage of the pentose sugar ring (-$C_2H_4O_2$) to yield the product ion at *m/z* 368, as demonstrated in Scheme 4.1.

Moreover, the precursor ion at *m/z* 461.17 (Isomer A) can be subjected to a ring cleavage in the pentose sugar ring to yield the product ion at *m/z* 401. This latter ion loses one water molecule to produce the product ion at *m/z* 383. In addition, the gas-phase fragmentations of this precursor ion *6* can also happen by three different cleavages occurring in the guaiacyl lignin moiety. Indeed, the precursor ion can be subjected to a Cβ-Cγ bond cleavage to form two product ions at *m/z* 309 and *m/z* 151. Additionally, the second type of cleavage occurs in the Cα-Cβ bond to create the product ions at *m/z* 137 and 323. Furthermore, the third type of cleavage occurring in the C_1-Cα bond of the guaiacyl unit leads to the formation of the product ion at *m/z* 337. Finally, the different isomer B was proposed to exist to explain the formation of the product ion at *m/z* 295, as shown in Scheme 4.1.

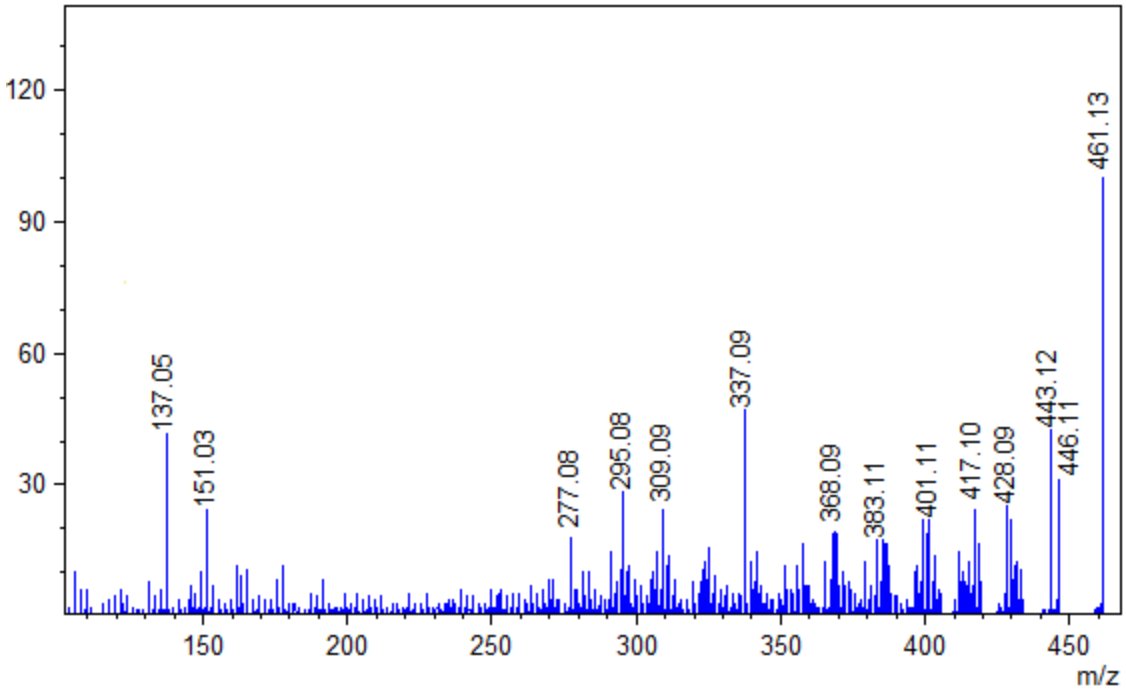
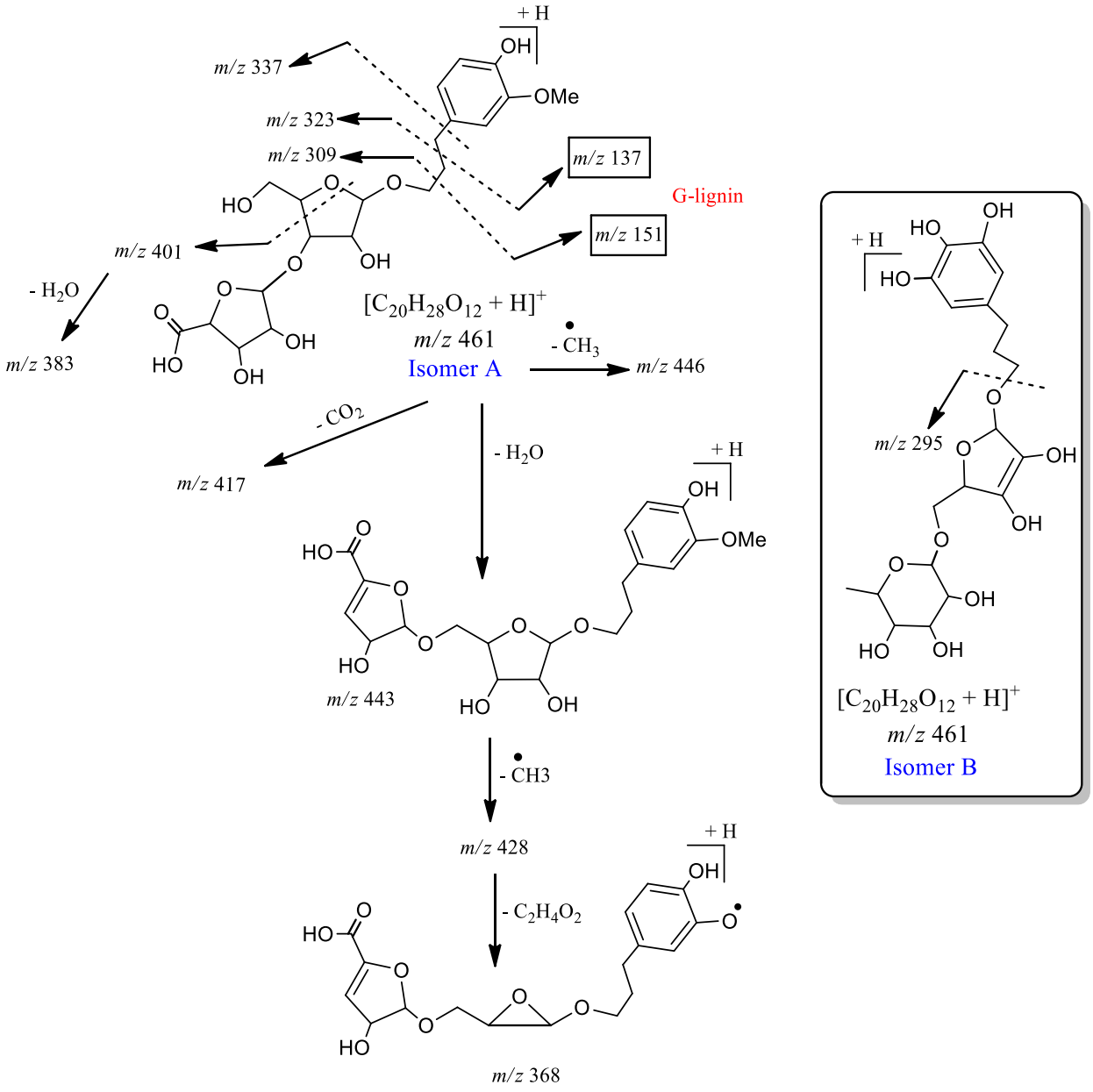


Figure 4.1. Product ion scan of the protonated precursor ion *6* at *m/z* 461.17.



Scheme 4.1. Low energy CID-MS/MS fragmentation mechanism of the protonated precursor ion *7* at *m/z* 461.17

4.4.3.4. The product ion scan of LCC composed of a glycose attached to two lignin units

The precursor ions *7, 8* and *9* at *m/z* 489.17, 517.18 and 533.18, respectively, are tentatively assigned to be composed of a glycose residue attached to two lignin units. For the sake of brevity, we are discussing here the MS/MS fragmentation pathway of the precursor ion *7* at *m/z* 489.17, while other precursor ions *8* and *9* are described in detail in section A3.3. The formula $[C_{25}H_{28}O_{10} + H]^+$ (DBE=11.5) was assigned to the protonated molecule *7* at *m/z* 489.17 (Figure 4.2 and Scheme 4.2a and 4.2b). Once more, this protonated molecule can be represented by three different isomers A, B, and C. Please note that the isomers A and B are positional isomers composed of penturonic acid connected to two lignin units: 3,4-methylenedioxy cinnamyl alcohol and coniferyl alcohol derivative. Whereas, the isomer C is composed of dehydroshikimic acid attached to 5-hydroxy dihydrocaffeyl alcohol (L gallyl lignin derivative) and dihydrocaffeyl alcohol

The product ion scan of the precursor ion *at m/z* 489.17 (Isomer A) was initiated by the loss of a water molecule to generate the product ion at *m/z* 471 (Scheme 4.2a). Similarly, the product ion scan of this precursor ion could also be initiated by the loss of formic acid from the penturonic acid ring to produce the product ion at *m/z* 443. This latter Ion eliminates a methyl radical to generate the radical ion at *m/z* 428. Also, the product ion at *m/z* 443 loses 46 Da (- CH_2O_2) from the methylenedioxy ring to generate the ion at *m/z* 397.[41] Additionally, the product ion at *m/z* 443 loses one methanol molecule to create the product ion at *m/z* 411. This last ion loses one carbonyl group to yield the secondary ion at *m/z* 383. This latter ion *at m/z* 383 loses a water molecule to give the ion at *m/z* 365.

It should be mentioned that the MS/MS of the precursor ion *at m/z* 489.17 (Isomer A) can also occur by two different bond cleavages happening in the dihydroconiferyl alcohol unit. Firstly, the Cα-Cβ bond cleavage generates both product ions at *m/z* 351 and *m/z* 137. The second mechanism occurs by the Cβ-Cγ cleavage to create the product ion at *m/z* 337 and *m/z* 151. Regarding the 3,4-methylenedioxy cinnamyl alcohol unit, the cleavage of the Cγ-O bond in this unit leads to the formation of the aromatic product ion at *m/z* 161.

Concerning isomer B, we propose that the product ion scan of this precursor ion at *m/z* 489.17 (Isomer B) is initiated by cleavage of the C_2-O bond in the sugar ring to create the product ion at *m/z* 309 as shown in Scheme 4.2a. This last ion loses one methyl radical from the G lignin unit to give the radical ion at *m/z* 294. In addition, MS/MS of the precursor ion at *m/z* 489.17 (isomer B) can also occur by the Cγ-O bond cleavage of the lignin unit containing the methylenedioxy moiety to form the aromatic product ion at *m/z* 163.

Lastly, the product ion scan of precursor ion *8* (third isomer C) showed the cleavage of the Cγ-O bond in the gallyl unit to form the product ion at *m/z* 323. This latter ion at *m/z* 323 could form the product ion at *m/z* 173. This last product ion was assigned to the protonated molecule of dehydroshikimic acid, as shown in Scheme 4.2b.

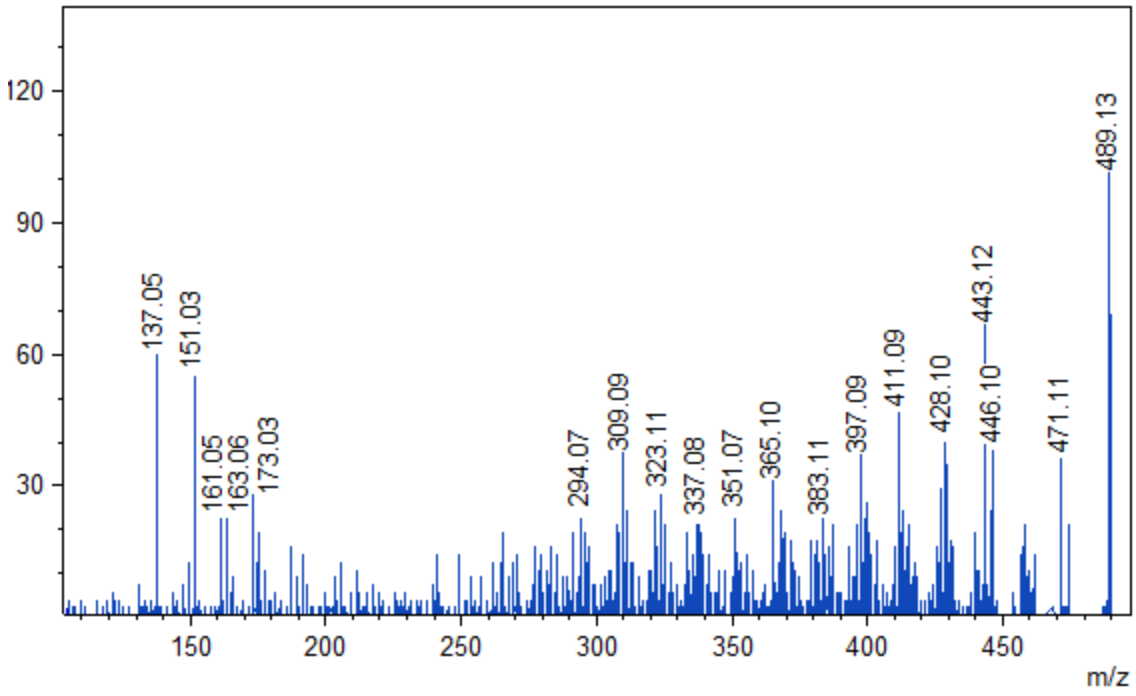
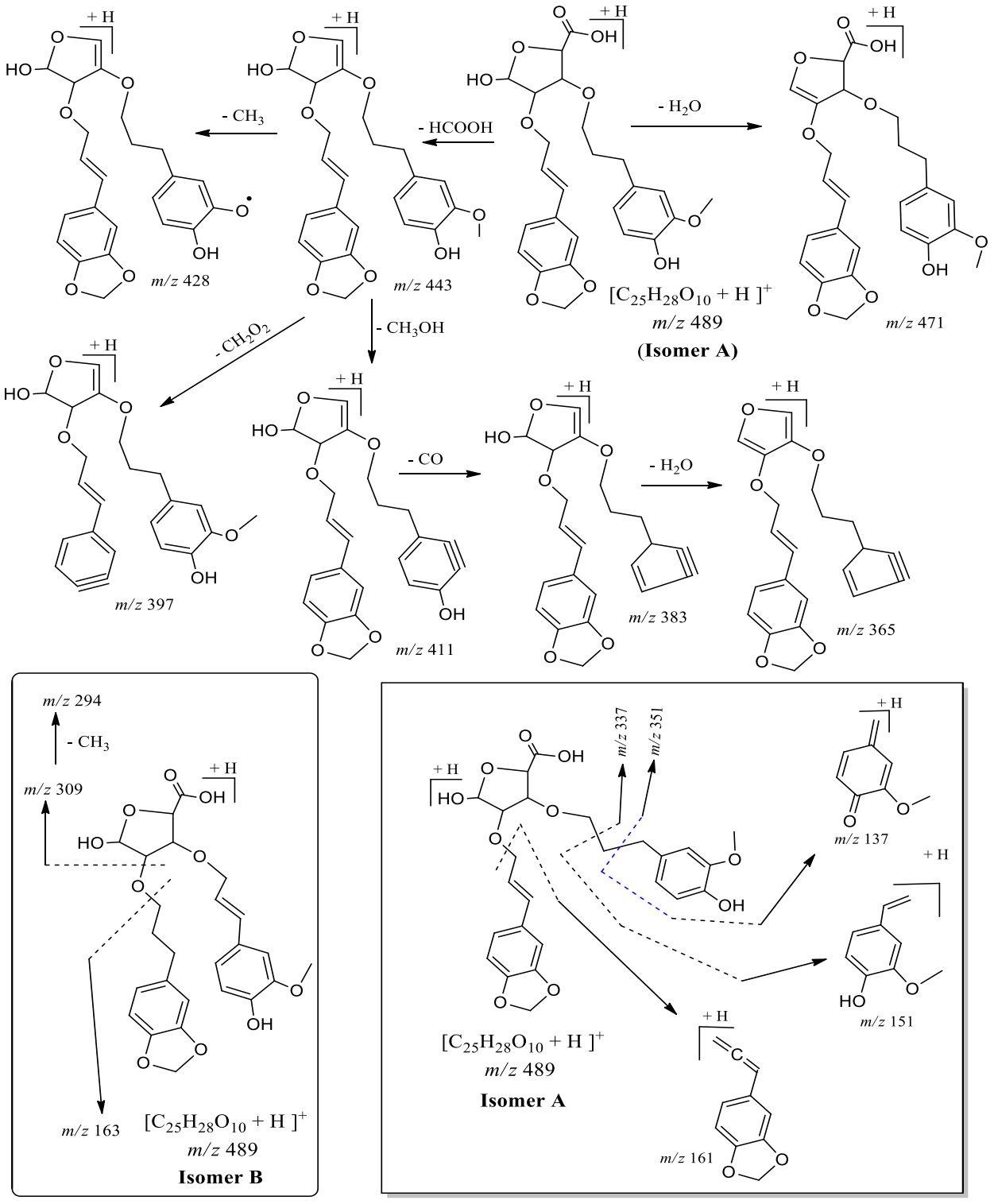
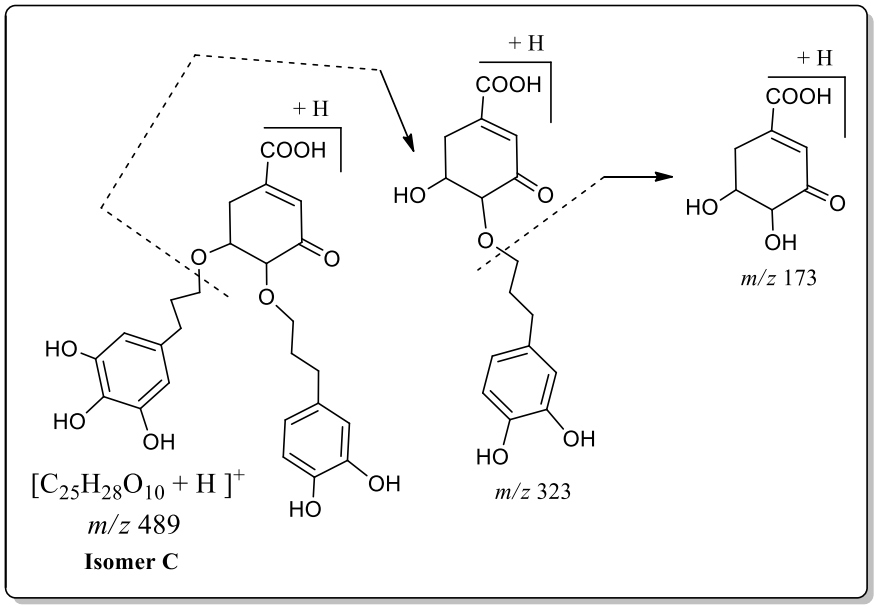


Figure 4.2. Product ion scan of the protonated precursor ion *7* at *m/z* 489.17.



Scheme 4.2a. Low energy CID-MS/MS fragmentation mechanism of the protonated precursor ion *7* at *m/z* 489.17.



Scheme 4.2b. Low energy CID-MS/MS fragmentation mechanism of the protonated precursor ion at *7 m/z* 489.17

4.4.3.5. The product ion scan of LCC composed of two glycoses attached to two lignin units

The precursor ions *10*, 11 and *12* at *m/z* 637.25, 651.25 and 695.26, respectively, are tentatively assigned to be composed of two glycoses attached to two lignin units. For the sake of brevity, we are discussing here the MS/MS fragmentation pathway of the precursor ion *10* at *m/z* 637.25, while other precursor ions *11* and *12* are described in detail in section A3.4. The $[C_{31}H_{40}O_{14}+ H]^+$ formula (DBE=11.5) was allocated to the protonated molecule *10* at *m/z* 637.25, and it was tentatively proposed to exist as a tetramer composed of hexose, deoxyhexose, dihydroferulic acid, and coumaryl alcohol. (Figure 4.3 and Scheme 4.3).

The product ion scan of the precursor ion *10* at *m/z* 637.25 is initiated by the loss of a water molecule to produce the product ion at *m/z* 619. Also, the precursor ion *10*

undergoes the cleavage of the C_1-Cα bond of the dihydroferulic acid generates the product ion at *m/z* 513. In addition, this precursor ion *10* experiences the loss of the hexose moiety (- $C_6H_{12}O_6$) to give the product ion at *m/z* 457. This latter ion at *m/z* 457 experience the cleavage of the C1-Cα bond in the dihydroferulic acid unit (loss of methoxy phenol) to produce the ion at *m/z* 333. Also, the product ion at *m/z* 457 experience the cleavage of the C1-O bond of the deoxyhexose ring to yield the product ion at *m/z* 307. In addition, another possible positional isomer of the product ion at *m/z* 457 that can originate from another isomeric precursor ion was drawn to demonstrate the formation of the product ion at *m/z* 323, as shown in Scheme 4.3.

Lastly, this precursor ion *10* can be subjected to two different simultaneous cleavages of the Cβ- Cγ and the Cα-Cβ bonds present in the dihydroferulic acid unit, to create the aromatic product ions at *m/z* 151 and 137, respectively.

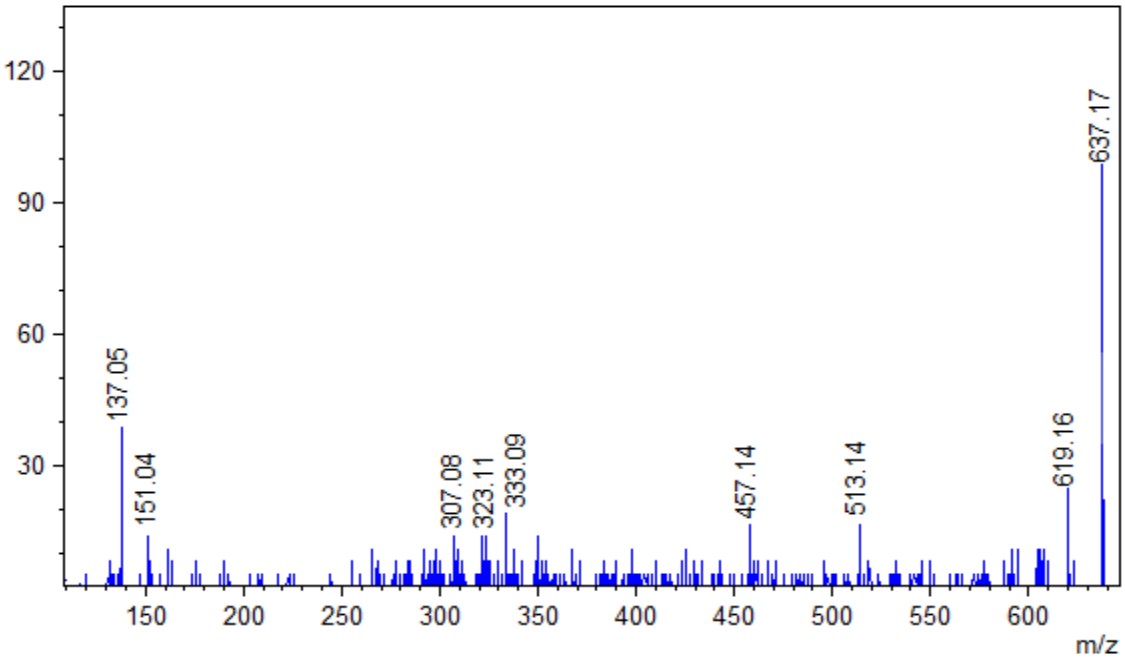
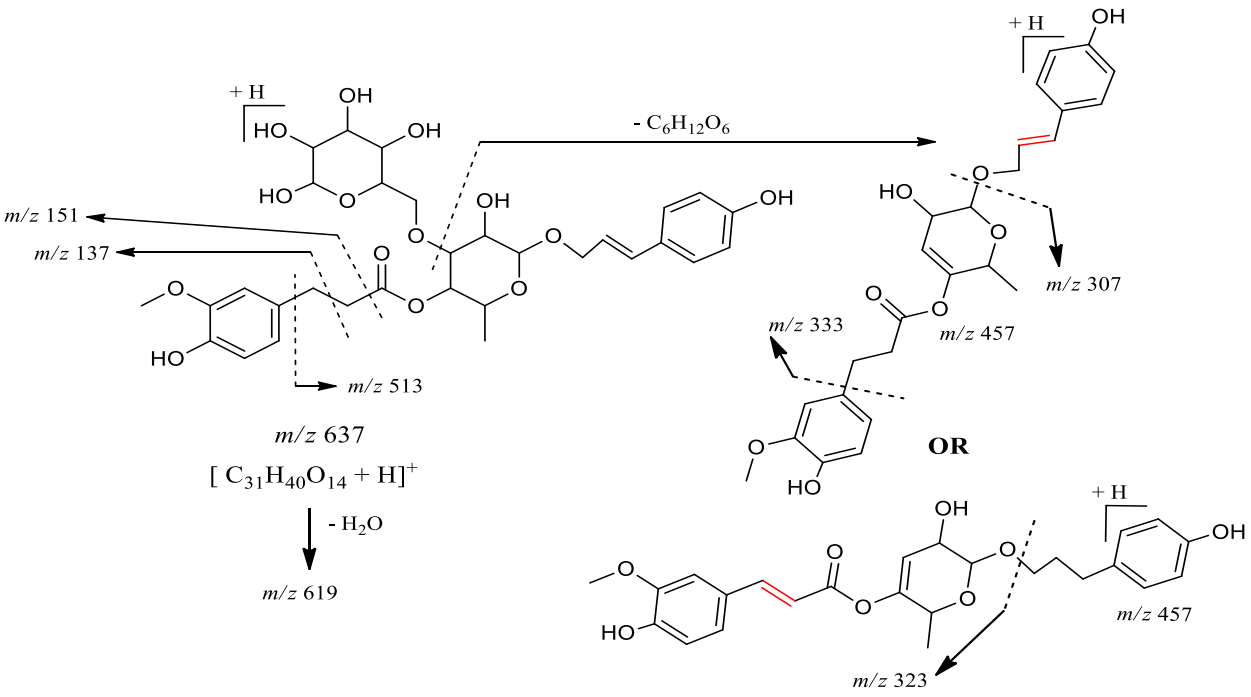


Figure 4.3. Product ion scan of the protonated precursor ion *10* at *m/z* 637.25.



Scheme 4.3. Low energy CID-MS/MS fragmentation mechanism of the protonated precursor ion *10* at *m/z* 637.25. Red double bonds indicate the point of difference between the two positional isomers.

4.4.3.6. LCC composed of one glycose residue attached to three lignol units.

The precursor ions *13*, *14* and *15* at *m/z* 611.24, 683.26 and 711.27, respectively, are tentatively assigned to be composed of a glycose residue attached to three lignin units. For the sake of brevity, we are discussing here the MS/MS fragmentation pathway of the precursor ion *14* at *m/z* 683.26, while other precursor ions *13* and *15* are described in detail in section A3.5. The formula $[C_{35}H_{38}O_{14}+ H]^+$ (DBE=16.5) was assigned to the protonated molecule *14* at *m/z* 683.26 (Figure 4.4 and Scheme 4.4). The protonated molecule *14* at *m/z* 683.26 is proposed to be composed of a hexose moiety attached to three lignin derivatives: dihydrocoumaric acid, dihydroferulic acid, and 3,4-(methylenedioxy) dihydrocinnamic acid.

The product ion scan of *m/z* 683.26 was initiated by loss 46 Da (-CH_2O_2) from the methylenedioxy moiety, to produce the product ions at *m/z* 637 (Scheme 4.4). This latter ion was subjected to two consecutive cleavages of the Cβ- Cγ bond and the Cα-Cβ bond in the dihydro- ferulic acid unit to yield the aromatic product ions at *m/z* 151 and *m/z* 137, respectively. Furthermore, this product ion scan is subjected to two simultaneous cleavages of the C2-O and C3-O bonds in the sugar ring to yield the product ion at *m/z* 321.

Likewise, this precursor ion (Isomer B) is subjected to a sugar ring cleavage and creates the product ion at *m/z* 297 ($^{1,3}X$). This latter ion can lose a methylene carbene from the methylenedioxy moiety to yield the product ion at *m/z* 283. Lastly, this later ion at *m/z* 283 loses one water molecule to produce the product ion at *m/z* 265.

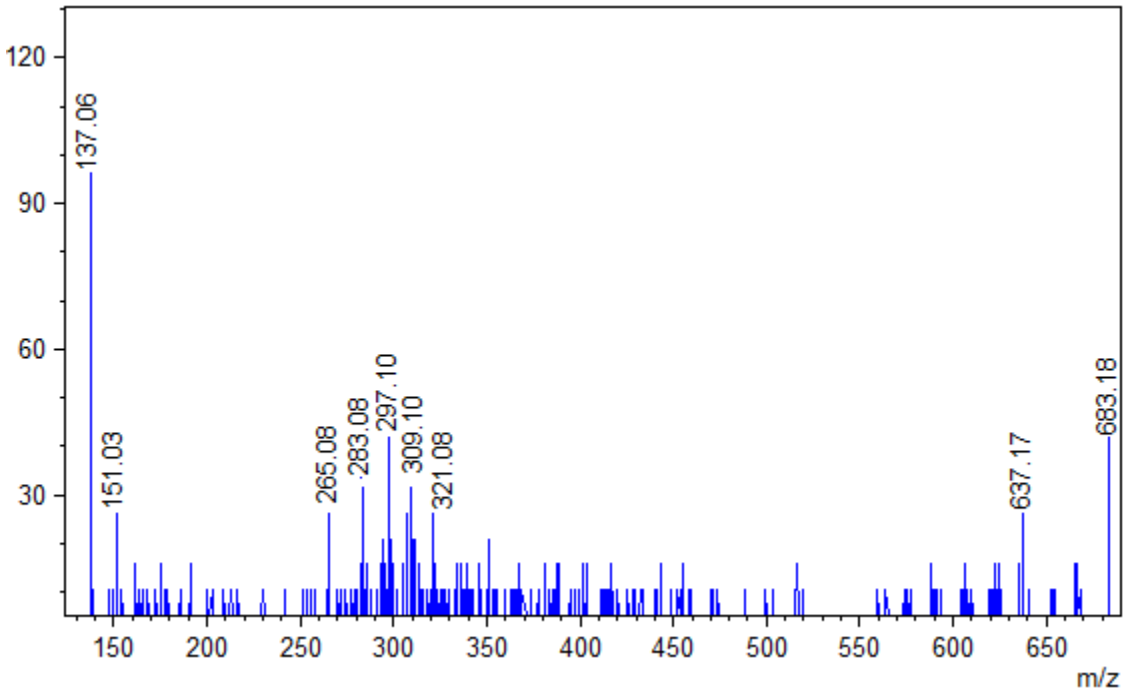
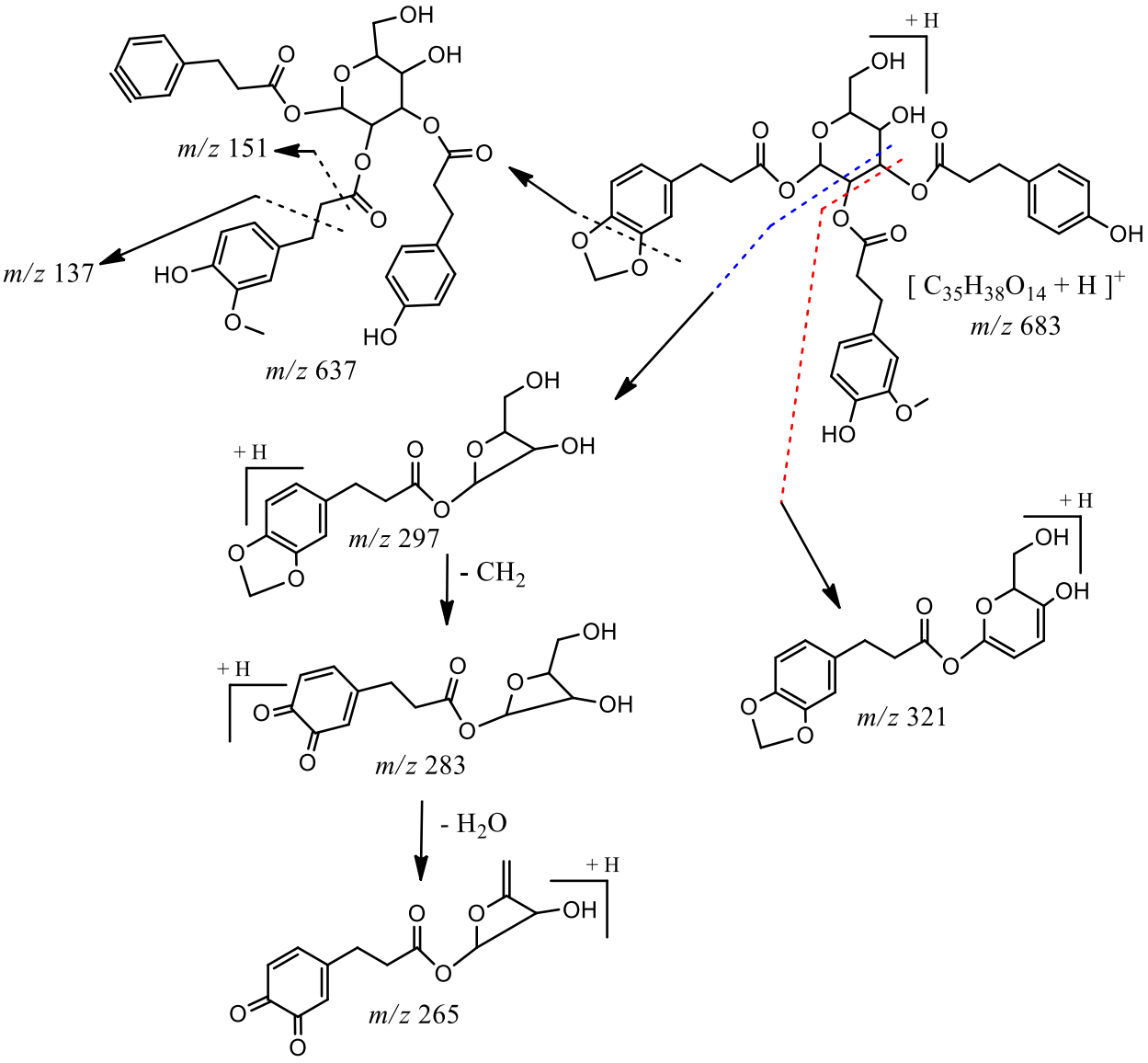


Figure 4.4. Product ion scan of the protonated precursor ion *14* at *m/z* 683.26.



Scheme 4.4. Low energy CID-MS/MS fragmentation mechanism of the protonated precursor ion *14* at *m/z* 683.26

4.5. Conclusion

In this very complex structural study by APPI-QqTOF-MS and MS/MS of the French pine VRl, we were capable of fishing-out fifteen ions form a very complex mixture of ions, to obtain a blueprint of the pine VRL structures. The structure elucidation of these

selected series of ions leads us to identify a series of novel isomeric and isobaric lignin-carbohydrate complexes. These complexes are containing different pentose and hexose moieties attached to one, two, and three lignin monomers. In addition to these series of novel LCCs, a lignin degradation product eugenol was detected in its acetylated form. This indicates that even mild CIMV extraction method can degrade lignin and alter its structure through acetylation and/or formylation.

The work presented in this rationale should be considered as a primary screening effort to reveal the major structural features of the pine VRLs sample using APPI-QqTOF-MS and low energy CID MS/MS. However, a more in-depth study is definitely needed using higher-resolution mass spectrometer instruments (Orbitrap and/or FTICR) to discriminate closely the overlapping isobars. In addition, to help differentiate the presence of isomeric mixture resulting from one possible chemical formula, a more elaborate study using ion mobility is needed to distinguish the different isomers from isobars at every *m/z* in this complicated pine VRLs mixture. Also, we are working on the synthesis of the herein identified different LCCs is order to confirm their identification.

The identified series of the novel LCCs indicates that the unique G, H, and C lignin monomers and their derivatives are attached directly to carbohydrate rings. This also means specifically indicate that neither the ferulic acid and /or coumaric acid, are the sole residues, which can crosslink the lignin oligomers and carbohydrate chains in the wood network. Overall, this identified series of LCCs support the presence of different lignin monomers as cross-linkers in the wood network.

The major question that we should ask ourselves is the following: how come such a series of novel LCC derivatives were never reported before? The answer is quite obvious, structural investigation of lignin oligomers/polymers (?) should never be completed on modified, depolymerized and purified: lignin. Indeed, the usual way of extracting lignin by the Kraft-, the Sulphite- and the soda lignin processes use very harsh conditions. This obviously results in obtaining modified lignin mixtures, which structures without doubts have been severely altered. This is, of course, followed by futile attempts to purify the obtained modified lignins before subjecting it to more degradation procedures in an attempt to "depolymerize" the lignin, which is assumed to be a huge biopolymer! Needless to say, that all attempts of depolymerisation, purification and chemical degradation procedures cause more alteration to the structure of lignins prior their structural investigations.

Indeed, the isolation of lignin in its unaltered form is a highly unlikely process.[11] For comparison sake, the old extraction methods, coupled to their depolymerisation, degradations and purification procedures, could be viewed as using a wrecking ball to break a very brittle oligomer assembly. The reactivity of the released fragments may lead to more complex reactive species that can further rearrange and condense to the more artefactual altered polymeric structure. [42] We have already proposed that extracted VRLs mixtures were composed of short linear oligomers that were connected to cellulose and hemicellulose in a grid pattern forming the network of wood.[11]

In conclusion, we would like to indicate that the short VRL oligomers can be compared to the packing peanuts shipping anti-static foam present in a parcel, trying to protect the good inside it (Cellulose rods). For these reasons, we would like to reiterate that for

lignomics structural elucidation, the analytical work should be completed on series of the lignin oligomers released from the wood polysaccharide-lignin network, without any further purification or chemical modifications, this type of extracted lignin oligomers we have designated as virgin released lignin (VRL).[11]

References

[1] Bidlack, J.; Malone, M.; Benson, R.: Molecular structure and component integration of secondary cell walls in plants. *Proc. Oklahoma Acad. Sci.* **1992**, *72*, 51-56.

[2] Barros, J.; Serk, H.; Granlund, I.; Pesquet, E.: The cell biology of lignification in higher plants. *Ann. Bot*. **2015**, *115*(7), 1053-1074.

[3] Dolk, M.; Pla, F.; Yan, J.F.; McCarthy, J.L.: Lignin. 22. Macromolecular characteristics of alkali lignin from western hemlock wood. *Macromolecules*. **1986**, *19*(5), 1464-1470.

[4] Campbell, M.M.; Sederoff, R.R.; Variation in lignin content and composition. Mechanisms of control and implications for genetic improvement of plants. *Plant Physiol*. **1996**, *110* (1), 3-13.

[5] MacKay, J.; Dimmel, D.R.; Boon, J.J.: Pyrolysis mass spectral characterization of wood from CAD-deficient pine. *J. Wood Chem. Technol*. **2001**, *21*, 19-29.

[6] Vanholme, R.; Demedts, B.; Morreel, K.; Ralph, J.; Boerjan, W.: Lignin biosynthesis and structure. *Plant physiol.* **2010**, *153*(3), 895-905.

[7] Stevanovic, T.: Renewable Materials for Today's Environment. In *Lignocellulosic Fibers and Wood Handbook*; Belgacem, M.N., Pizzi, A., Eds.; Scrivener Publishing: Salem, Massachusetts, USA, **2016**; pp 49-110.

[8] Vanholme, R.; Morreel, K.; Darrah, C.; Oyarce, P.; Grabber, J.H.; Ralph, J.; et al.: Metabolic engineering of novel lignin in biomass crops. *New Phytol*. **2012**, 196(4), 978-1000.

[9] Glasser, W.G.: Lignin. In *Pulp and Paper: Chemistry and Chemical Technology*; Casey, J.P., Ed.; John Wiley and Sons, Inc: New York, 1980, *Vol. 1*, 3rd ed.; pp 39–111.

[10] Davin, L.B.; Lewis, N.G.: Lignin primary structures and dirigent sites. *Curr. Opin. Biotechnol*. **2005**, *16*, 407-415.

[11] Banoub, J.; Delmas, G.; Joly, N.; Mackenzie, G.; Cachet, N.; Delmas, M.: A critique on the structural analysis of lignins and application of novel tandem mass spectrometric strategies to determine lignin sequencing. *J. Mass Spectrom*. **2015**, *50*, 5-48.

[12] Giummarella, N.; Pu, Y.; Ragauskas, A.J.; Lawoko, M.; A critical review on the analysis of lignin carbohydrate bonds. *Green Chem.* **2019**, *21*(7), 1573-1595.

[13] Katahira, R.; Elder, T.J.; Beckham, G.T.: Chapter 1: A Brief Introduction to Lignin Structure. In *Lignin Valorization: Emerging Approaches*; Beckham, G.T., Ed. Royal Society of Chemistry, 2018; pp 1-20

[14] Watanabe, T.; Koshijima, T.: Evidence for an ester linkage between lignin and glucuronic acid in lignin–carbohydrate complexes by DDQ-oxidation. *Agric. Biol. Chem*. **1988**, *52*(11), 2953-2955.

[15] Ralph, J.; Helm, R.F.: Lignin/Hydroxycinnamic Acid/Polysaccharide Complexes: Synthetic Models for Regiochemical Characterization. In *Forage Cell Wall Structure and Digestibility*; Jung, H.G., Buxton, D.R., Hatfield, R.D., Ralph, J., Eds.; ASA-CSSA-SSSA, 677 S: Segoe Rd., Madison, WI 53711, USA, **2015**; pp 201-246.
[16] Brett, C.T.; Wende, G.; Smith, A.C.; Waldron, K.W.: Biosynthesis of cell-wall ferulate and diferulates. *J. Sci. Food Agric.* **1999**, *79*(3), 421-424.
[17] Bunzel, M.; Ralph, J.; Kim, H.; Lu, F.; Ralph, S.A.; Marita, J.M.; et al.: Sinapate dehydrodimers and sinapate-ferulate heterodimers in cereal dietary fiber. *J. Agric. Food Chem.* **2003**, *51*(5), 1427-1434.
[18] Kondo, T.; Mizuno, K.; Kato, T.: Cell wall-bound p-coumaric and ferulic acids in Italian ryegrass. *Can. J. Plant Sci*.**1990**, *70*(2), 495-499.
[19] Marcia, M.D.O.: Feruloylation in grasses: current and future perspectives. *Mol. Plant.* **2009**, *2*(5): 861-872.
[20] Boes, K.S.; Narron, R.H.; Park, S.; Vinueza, N.R.: Mass spectrometry exposes undocumented lignin-carbohydrate complexes in biorefinery pretreatment stream. *ACS Sustain. Chem. Eng*. **2018**, *6*(8), 10654-10659.
[21] Lam, H.Q.; Le Bigot, Y.; Delmas, M.; Avignon, G.: A new procedure for the destructuring of vegetable matter at atmospheric pressure by a catalyst/solvent system of formic acid/acetic acid. Applied to the pulping of triticale straw. *Ind. Crops Prod*. **2001**, *14*, 139- 144.
[22] Schaefer, J.; Stejskal, E.O.: Carbon-13 nuclear magnetic resonance of polymers spinning at the magic angle. *J. Am. Chem. Soc.* **1976**, *98*(4), 1031-1032.
[23] Peersen, O.B.; Wu, X.L.; Kustanovich, I.; Smith, S.O.: Variable-amplitude cross-polarization MAS NMR. *J. Magn. Res*. **1993**, *104*, 334-339.
[24] Atalla, R.H.; Vanderhart, D.L.: Native cellulose: a composite of two distinct crystalline forms. *Science*.**1984**, *223*(4633), 283-285.
[25] Morcombe, C.R.; Zilm, K.W.: Chemical shift referencing in MAS solid state NMR. *J. Magn. Res.* **2003**, *162*(2), 479-486.
[26] Fox, S.C.; McDonald, A.G.: Chemical and thermal characterization of three industrial lignins and their corresponding lignin esters. *BioResources.***2010**, *5*(2), 990-1009.
[27] Saito, K.; Kato, T.; Tsuji, Y.; Fukushima, K.: Identifying the characteristic secondary ions of lignin polymer using ToF− SIMS. *Biomacromolecules*. **2005**, *6*(2), 678-683.
[28] Qi, Y.; Volmer, D.A.: Chemical diversity of lignin degradation products revealed by matrix- optimized MALDI mass spectrometry. *Anal. Bioanal. Chem.* **2019**, 411(23), 6031-6037.
[29] Jarrell, T.M.; Marcum, C.L.; Sheng, H.; Owen, B.C.; O'Lenick, C.J.; Maraun, H.; et al.: Characterization of organosolv switchgrass lignin by using high performance liquid chromatography/high resolution tandem mass spectrometry using hydroxide-doped negative-ion mode electrospray ionization. *Green Chem.* **2014**, *16*(5), 2713-2727.
[30] Butovich, I.A.: On the lipid composition of human meibum and tears: comparative analysis of nonpolar lipids. *Investig. Ophthalmol. Vis. Sci.* **2008**, *49*(9), 3779-3789.
[31] Prothmann, J.; Spégel, P.; Sandahl, M.; Turner, C.: Identification of lignin oligomers in Kraft lignin using ultra-high-performance liquid chromatography/high-resolution

multiple-stage tandem mass spectrometry (UHPLC/HRMS^n). *Anal. Bioanal. Chem.* **2018**, *410*(29), 7803-7814.
[32] Sleno, L.; Volmer, D.A.; Marshall, A.G.; Assigning product ions from complex MS/MS spectra: the importance of mass uncertainty and resolving power. *J. Am. Soc. Mass Spectrom.* **2005**, *16*(2), 183-198.
[33] Demarque, D.P.; Crotti, A.E.; Vessecchi, R.; Lopes, J.L.; Lopes, N.P.: Fragmentation reactions using electrospray ionization mass spectrometry: an important tool for the structural elucidation and characterization of synthetic and natural products. *Nat. Prod. Rep*. **2016**, *33*(3): 432-455.
[34] Malunga, L.N.; Beta, T.: Isolation and identification of feruloylated arabinoxylan mono-and oligosaccharides from undigested and digested maize and wheat. *Heliyon*. **2016**, *2*(5), 1-18.
[35] Santos, M.D.; Lopes, N.P.; Iamamoto, Y.; HPLC-ESI-MS/MS analysis of oxidized di-caffeoylquinic acids generated by metalloporphyrin-catalyzed reactions. *Quím.* **2008**, *31*(4), 767-770.
[36] Hughey, C.A.; Hendrickson, C.L.; Rodgers, R.P.; Marshall, A.G.; Qian, K.: Kendrick mass defect spectrum: a compact visual analysis for ultrahigh-resolution broadband mass spectra. *Anal. Chem*. **2001**, *73*(19), 4676-4681.
[37] Qi, Y.; Hempelmann, R.; Volmer, D.A.: Two-dimensional mass defect matrix plots for mapping genealogical links in mixtures of lignin depolymerisation products. *Anal. Bioanal. Chem.* **2016**, *408*(18), 4835-4843.
[38] Sleno, L.: The use of mass defect in modern mass spectrometry. *J. Mass. spectrom*. **2012**, *47*(2), 226-236.
[39] Graven, P.; De Koster, C.G.; Boon, J.J.; Bouman, F.: Structure and macromolecular composition of the seed coat of the Musaceae. *Ann. Bot*. **1996**, *77*(2), 105-122.
[40] Domon, B.; Costello, C.E.: A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. *Glycoconj. J.* **1988**, *5*(4), 397-409.
[41] Krotulski, A.J.; Mohr, A.L.; Papsun, D.M.; Logan, B.K.: Dibutylone (bk-DMBDB): intoxications, quantitative confirmations and metabolism in authentic biological specimens. *J. Anal. Toxicol.* **2018**, *42*(7): 437-445.
[42] Gosselink, R.J.A.: Lignin as a renewable aromatic resource for the chemical industry. PhD Thesis: Wageningen University, Wageningen, NL, **2011.**

Chapter 5

Top-down lignomics analysis of the French Oak lignin by atmospheric pressure photoionization and electrospray ionization quadrupole time-of-flight tandem mass spectrometry: Identification of a novel series of lignans

This chapter is reproduced with permission from Mikhael, A.; Fridgen, T.D.; Delmas, M.; Banoub, J.: **Top-down lignomics analysis of the French Oak lignin by atmospheric pressure photoionization and electrospray ionization quadrupole time-of-flight tandem mass spectrometry: Identification of a novel series of lignans**. *J Mass Spectrom*. 2020, e4676 https://doi.org/10.1002/jms.4676.

5.1. Abstract

We report herein the top-down lignomic analysis of virgin released lignin (VRL) extracted from the French oak wood using APPI- QqTOF-MS (+ ion mode). Eight major protonated lignin oligomers were identified using the APPI-QqTOF-MS/MS of this complex VRL mixture without any kind of purification. This series of protonated oligomer ions were identified as neolignan cedrusin *(1)*, five different aryltetralin lignans dimers *(2-6)*, one lignan-dehydroshikimic acid complex *(7)*, and a lignan trimer *(8)*. Similarly, ESI-QqTOF-MS (+ ion mode) allowed us to identify three extra aryltetralin lignan derivatives *(9-11)*. The Kendrick mass defect analysis was used for the simplification of this complex APPI-QqTOF-MS into a compositional map, which displayed clustering points of

associated ions possessing analogous elemental composition. This series of novel protonated molecules were selected and subjected to low-energy CID-MS/MS analyses. The obtained gas-phase fragmentation patterns helped to tentatively assign their most likely structures. Also, it was found that the use of different APPI and ESI ambient ionization techniques enhances the ionization of different types of lignin oligomers.

5.2. Introduction

The lignin constituent of the lignocellulose biomass is considered the most abundant renewable source composed of aromatic units in nature.[1] Lignin is the product of the oxidative enzymatic polymerization of three monomeric aromatic compounds (monolignols): *p*-coumaryl (H), coniferyl (G), sinapyl (S) alcohols. [2-4] Also, some lignin units are not commonly incorporated in lignin oligomers such as caffeyl (C), 5-hydroxyguaiacyl (F), and the hypothetically proposed gallyl unit (L).[5]

There is a particular class of lignin dimers called lignans, which is widely distributed in the plant kingdom. The lignans are composed of two phenylpropanoids (C-9 units) and also involved biochemically to phenylalanine metabolism.[6] These two C-9 phenylpropanoid units are attached through the C_8-C_8' bond.[7] According to the IUPAC nomenclature, in the absence of the C_8-C_8' bond between the two monomers, the resulting dimers are classified as neolignans. [7] In general, lignans are known to possess important pharmacological properties such as anti-platelet, antiviral, anti-tumor, anti-depressant, and pesticidal activities. [8-12]

An important physical property of extracted lignin oligomers is that they exist as a racemic mixture with no optical activity.[13] In contrast, lignans are known to have a specific optical activity supposedly controlled by a special type of proteins called dirigent proteins.[14,15] However, Some lignan dimers and/or structures such as pinoresinol was identified in its racemic form (optically inactive) as a degradation product from lignin samples.[16]

Previously, we have studied the analysis of the same virgin released lignin (VRL) Oak mixture by MALDI-TOF-MS (- ion mode) and high-energy CID-TOF/TOF-MS/MS.[17] In this work, we have identified tricin derivatives, syringylglycerol derivatives, syringol derivatives, flavonolignin derivatives, and other miscellaneous compounds such as luteoferol, lariciresinol lignan isomer, and an aryl tetralin lignan derivative.[17] It is interesting to note that the series of derivatives identified by MALDI-TOF-MS was completely different with respect to the series presented herein. This confirms the notion that analyses of the same VRL mixture with different ionization modes allow the characterization of different types of lignin compounds and establish the diversity of compounds within the VRL mixture to different ambient ionization methods.[18,19] The chemical diversity of lignin was also discussed by Volmer and Qi when analyzing alkali lignin by MALD-TOF-MS with different MALDI matrices.[20] Likewise, chemical lignin diversity was found when different ionization modes such as MALDI, APPI, ESI, and APCI were used to investigate wheat straw VRL mixture and the degradation products of alkali lignin.[18,19, 21]

This manuscript describes the APPI- and ESI-QqTOF-MS (+ ion mode) and low-energy CID-MS/MS analyses used to investigate the structure of the VRL extracted from

French Oak. In the APPI-MS study, we have selected eight major oligomers, which were identified as follows: the neolignane cedrusin (*1*), five different aryltetralin lignans (*2*-*6*), one lignan-dehydroshikimic acid complex derivative (*7*), and a lignan trimer (*8*). Lastly, the use of ESI-QqTOF-MS and low-energy CID-MS/MS permitted to identify three additional aryltetralin lignan derivatives (*9*-*11*) to those identified by APPI. All the proposed structures were supported by their CID-MS/MS fragmentation pathways.

5.3. Experimental

5.3.1. Sample preparation

The French Oak lignin VRL sample used for this study was provided by Professor Michel Delmas. The lignin sample was extracted using the CIMV procedure, which selectively separates the cellulose, hemicellulose, and lignin at atmospheric pressure, and allows the destructuring of the vegetable matter.[22] The solvent system was a mixture of formic acid/acetic acid/water (30/50/20), which produced, after precipitation with water and filtration, the French Oak VRL lignin. Approximately 0.1 mg of Oak lignin powder was dissolved in 1 ml of dioxane-methanol (2:1) for MS analysis.

5.3.2. APPI-QqTOF-MS, ESI-QqTOF-MS, and Low-Energy CID-MS/MS

Mass spectrometry was performed using an Applied Biosystems (Foster City, CA, USA) API-QSTAR-XL MS/MS quadrupole orthogonal time-of-flight (QqToF)-MS/MS hybrid instrument. APPI was performed with a PhotoSpray ion source (Applied

Biosystems) operated at 1300V at a temperature of 400 ºC, with all acquisitions completed in the positive ion mode. Samples were infused into the mass spectrometer with an integrated Harvard syringe pump at a rate of 0.1 mL/min. The auxiliary nebulizer gas pressure setting was fixed at 25 psi, and the nebulizer gas pressure at 74 psi. The curtain gas pressure was set at 30 psi. The declustering potential (DP) was established at + 100 eV. The focus potential (FP) was adjusted to +100 V. Toluene was selected as the dopant for its ability to undergo trouble-free photoionization at 8.83 eV. The eluent was composed of dioxane/methanol (2:1). No modifier was used to enhance ion production. ESI-QqTOF-MS (+ ion mode) analysis was performed with the same MS instrument equipped with a TurboIonspray source operated at 4.5 kV at a temperature of 200 °C.

The mass calibration of the TOF analyzer in the positive ion mode was performed with the PhotoSpray ion source, using 1,2,3,5-tetra-O-acetyl-β −D-ribofurnanose and checking for the exact masses of the $[M+H]^+$ ion $[C_{13}H_{18}O_9 + H]^+$ at *m/z* 319.1029 and the $[M+H–AcOH]^+$ ion $[C_{11}H_{14}O_7 + H]^+$ at *m/z* 259.0817. Calibration for higher masses was performed with octa-O-acetyl-ß-D-lactopyranose and checking for the $[M+H]^+$ ion $[C_{28}H_{38}O_{19} + H]^+$ at *m/z* 679.2085. Product ion spectra were obtained on the same instrument, as described above. Nitrogen was used as the collision gas for MS/MS analyses with collision energies varying between 10 and 35 eV. Collision energy (CE) and CID gas conditions were adjusted in each acquisition such that the precursor ion remained abundant in the product ion spectra. In general, we have used a 1 *m/z* unit resolution for the MS/MS selection of the precursor ion for the simplification of the analysis.

5.4. Results and Discussion

5.4.1. APPI-QqTOF-MS of the French Oak VRLs

APPI-QqTOF-MS (+ ion mode) of the Oak VRLs mixture was measured with several mass ranges (*m/z* 350-700 and 450-800) to investigate all oligomers (Figure 5.1). The sample was directly infused in the ionization source without any chromatographic separation, as we realized that any attempts to separate such heterogenous mixture were futile. A quick perusal of the APPI-MS spectrum indicates a very complex mixture of intertwining ions, which of course, complicate the measuring of isotopic patterns (Figure 5.1). The complexity of this APPI-MS results was credited to the probable presence of a multitude of different possible isobaric and/or isomeric lignin compounds such as lignin/lignan oligomers, lignin-carbohydrate complexes (LCC), and possible flavonoids derivatives (e.g., tricin).[23] For this reason, the APPI-MS was recorded at different mass ranges and was enlarged and zoomed in an attempt to descramble the ions distributions and to pinpoint the ions chosen for further low-energy CID-MS/MS analyses.

Just to explain the ambiguity of the measured APPI-MS adequately, it is essential to note that when reexamining an APPI-MS recorded with a mass range from *m/z* 350 to *m/z* 700, ideally, we ought to expect at least 350 ion signals. Unfortunately, this is not the case, as the MS detector sees more than 700 ion signals. Needless to say that this number of ions signals increases as the concentration of the analyte increases. For comparison, Qi and Volmer reported that when using high- resolution LDI-FTICR-MS of an alkali lignin sample recorded from *m/z* 100-799, they obtained 3370 monoisotopic formulas.[18] This large multitude of measured ion signals indicates the presence of isobaric interferences at

each *m/z*. Interestingly, to explain this observation, when any single ion mass was magnified, it was realized that it was composed of more than one ion (Figure A4.1).

For this reason, we understood that it would be impossible to make chemical assignments based solely on accurate mass and isotopic distributions using our QqTOF-MS instrument. In other words, centroid mass shifts and intricate overlapping isotopic patterns complicate the task of the lignin oligomers identification.

Consequently, we decided to use a general rule to identify this complex series of lignin oligomers. We proposed that this task should be realized by considering the presence of one type of heteroatoms (Oxygen) and that the identification of chemical formulas will reside within 0.1 *m/z* unit. Also, all possible combinations of different lignol units were used to propose chemical structures confirmed by reasonable MS/MS fragmentation patterns. Lastly, all proposed chemical formulas suggested for this series of protonated molecules were consistent with DBE and C/O ratio values for lignin monomers (DBE $\geq$ 4 and C/O ratio $\geq$ 2.5), lignin dimers (DBE $\geq$ 8 and C/O $\geq$ 2.5) and so on.[24,25]

In this manuscript, and for the sake of brevity, we have chosen to discuss specifically the major protonated molecules at *m/z* 347.13 (*1*), 355.16 *(2)*, 369.10 *(3)*, 399.10 *(4)*, 401.11 *(5)*, 417.11*(6)*, 537.14 *(7)*, 643.16 *(8)* {(Figure 5.1a and 5.1b), Figure SI-2 and Table 5.1}. It is essential to mention that the major ion at *m/z* 663.38 was assigned as the oxidized form of the contaminant Orgafos-168 $[C_{42}H_{63}O_4P + H]^+$ (Figure 5.1b).[26] As suggested by Volmer *et al.*, [21] the presence of phosphorus compounds in the analyzed lignin samples may have resulted due to the contamination from agriculture effluent, leaks from chemical plants or waste, and dumps from manufacturers.

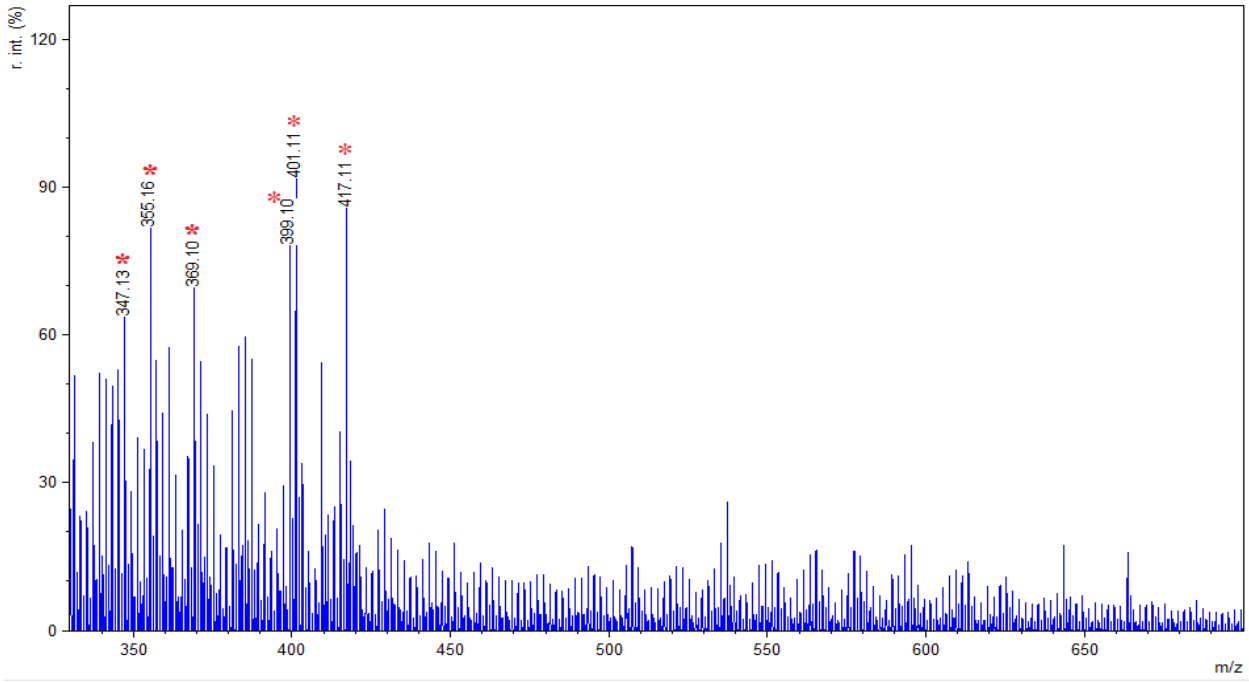


Figure 5.1a. APPI-QqTOF-MS of the French Oak lignin from *m/z* 340-700

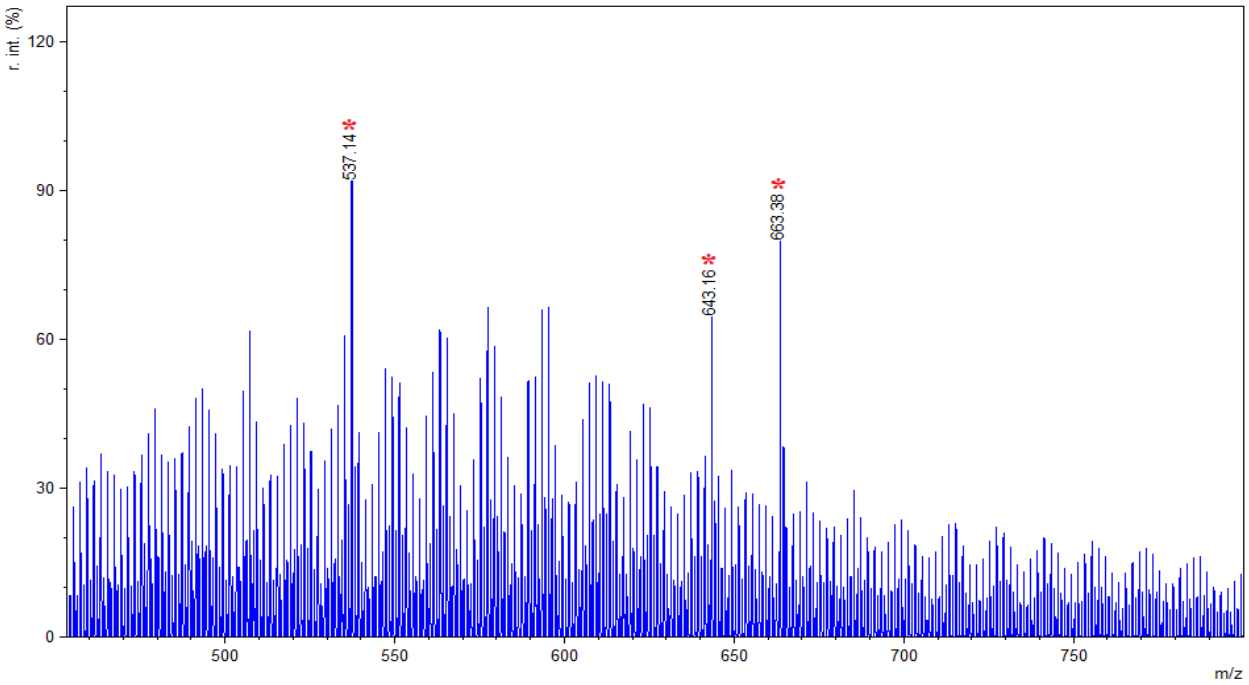


Figure 5.1b. APPI-QqTOF-MS of the French Oak lignin from *m/z* 450-800

Table 5.1. APPI-QqTOF-MS and Low-Energy CID-MS/MS analyses of the extracted VRLs from the French Oak Wood

Ion	Chemical Formula	Experimental *m/z*	Theoretical *m/z*	Error (*m/z* units)	DBE	MS/MS
1	$[C_{19}H_{22}O_6 + H]^+$	347.13	347.15	-0.02	8.5	332, 319, 317, 315, 301, 289, 287, 273, 259, 219, 199, 197, 161
2	$[C_{21}H_{22}O_5 + H]^+$	355.16	355.16	0.00	10.5	340, 323, 295, 243, 235, 211, 179, 151
3	$[C_{20}H_{16}O + H]^+$	369.10	369.10	0.00	12.5	354, 337, 309, 293, 277, 249, 231, 167, 137, 121
4	$[C_{21}H_{18}O_8 + H]^+$	399.10	399.10	0.00	12.5	384, 371, 356, 341, 339, 325, 321, 307, 279, 261, 247, 245, 233, 217, 213, 185, 181, 167, 157
5	$[C_{21}H_{20}O_8 + H]^+$	401.11	401.13	- 0.02	11.5	386, 383, 369, 355, 351, 323, 313, 309, 299, 291, 279, 271, 263, 247, 231, 219, 217, 187, 181, 167, 159, 107
6	$[C_{21}H_{20}O_9 + H]^+$	417.11	417.12	- 0.01	11.5	399, 233, 217, 205, 193, 181, 167
7	$[C_{28}H_{24}O_{11} + H]^+$	537.14	537.14	0.00	16.5	383, 351, 323, 299, 291, 263, 215, 207, 193, 181, 167
8	$[C_{33}H_{38}O_{13} + H]^+$	643.16	643.24	-0.08	14.5	417, 263, 225, 191, 181, 167

5.4.2. Kendrick Mass Defect Plots

The Kendrick mass defect analysis permits the simplification of this complex APPI-QqTOF-MS spectrum into a compositional map, which displays associated ions clustering points possessing analogous elemental composition.[27] The resulting display maintains a rough spacing of 1 Da between odd and even masses, which corresponds to either even vs odd number of nitrogen atoms and $^{12}C_{\boldsymbol{c}}$ vs. $^{12}C_{\boldsymbol{c}-1}{}^{13}C_1$ elemental compositions. Similarly, 2-Da separations, corresponding to a double bond or ring; and 14 Da separations, corresponding to one CH_2 group. [27]

Similarly, the conventional Kendrick mass defect (KMD) plot, obtained from the APPI-MS recorded in the *m/z* range 340-700, was constructed (using CH_2 units) (Figure A4.3). The KMD plot showed that a series of intimately related ions looked like one bundle

of closely related ions. This finding suggests that the recorded MS is not composed of a random series of diverse ions, but that these ions are structurally related to each other. In general, the KMD plots are characterized by different horizontal lines. The ions that vary in the number of CH_2 groups and have the same mass defect are always aligned on the same horizontal line.[27]

Also, the plot of the resulting KMD analyses are not limited to the use of CH_2 units only, but they can be adapted using other functional units appropriate for the studied class of compounds. [28] For example, the [OCH_2] units were used to align a series of lignin monomers that differ in the number of methoxy groups such as coumaric acid, ferulic acid, and sinapic acid. [28] For this reason, the KMD plot using [OCH_2] units of the APPI-MS spectrum in the range of *m/z* 340-700 was constructed (Figure A4.4). Moreover, it should be noted that the KMD plot can be used as a mass defect filtering tool. [29] For example, the major ion at *m/z* 663 excluded from this study has a different mass defect value compared to the main bundle of ions (Figure A4.4).

Furthermore, the KMD displayed maps allowed us to rationalize and predict correlated precursor ions, which possessed analogous empirical formulas. For example, in the magnified part of the KMD plot using [OCH2] units, it was evident that the precursor ions $[C_{28}H_{24}O_{11} + H]^+$ and $[C_{33}H_{38}O_{13} + H]^+$ at *m/z* 537 (*7*) and *m/z* 643 *(8)* (Table 5.1) were related to the ions at *m/z* 507 and *m/z* 613, respectively (Figure 5.2). This was deduced due to the alignment of the ions at *m/z* 507 and *m/z* 613 on the same horizontal line with the ions at *m/z* 537 (*7*) and *m/z* 643 *(8)*, respectively (Figure 5.2). Accordingly, the formulas of the ions at *m/z* 507 and *m/z* 613 can be obtained by removing one [OCH_2] unit from the

assigned chemical formulas of the ions at *m/z* 537 *(7)* and *m/z* 643 *(8)*. For this reason, the formulas $[C_{27}H_{22}O_{10} + H]^+$ and $[C_{32}H_{36}O_{12} + H]^+$ were assigned to the ions at *m/z* 507 and *m/z* 613, respectively.

Likewise, the same reasoning can be applied to the KMD plots using CH_2 as a base unit. For example, the KMD plot using $[OCH_2]$ units align all the ions that differ by 14 Da (CH_2) on the same horizontal line; this was the case for the ions at *m/z* 399 *(4)* and *m/z* 413, and the ions at *m/z* 417 *(6)* and *m/z* 431 (Figure 5.3).

Based on CID-MS/MS analysis, which will be presented below in Sections 3.3.4 and 3.3.6, the chemical formulas $[C_{21}H_{18}O_8 + H]^+$ and $[C_{21}H_{20}O_9 + H]^+$ were assigned respectively to the precursor ions at *m/z* 399 *(4)* and *m/z* 417*(6)* (Table 5.1). Consequently, the formulas $[C_{22}H_{20}O_8 + H]^+$ and $[C_{22}H_{22}O_9 + H]^+$ were assigned to the ions at *m/z* 413 and *m/z* 431. This was predicted easily by adding one $[CH_2]$ unit to the related ions formulas at *m/z* 399*(4)* and *m/z* 417*(6)*.

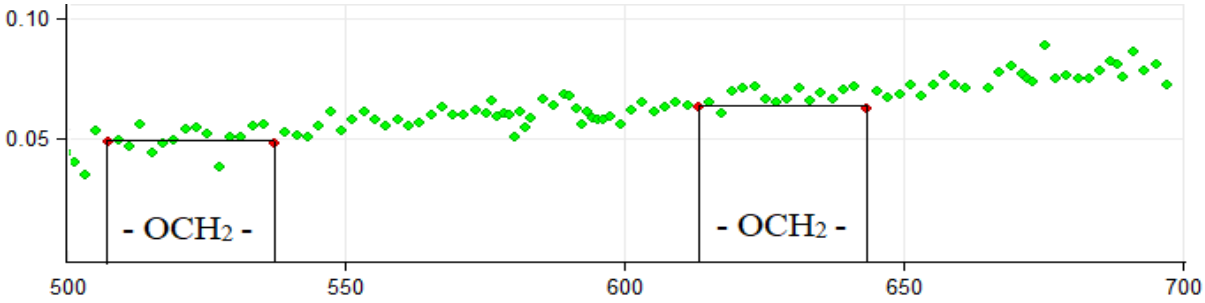


Figure 5.2. Zooming of the Figure A4.4-KMD plot using OCH_2 units to show the alignment of the ions at *m/z* 507 and *m/z* 537, and the ions at *m/z* 613 and *m/z* 643.

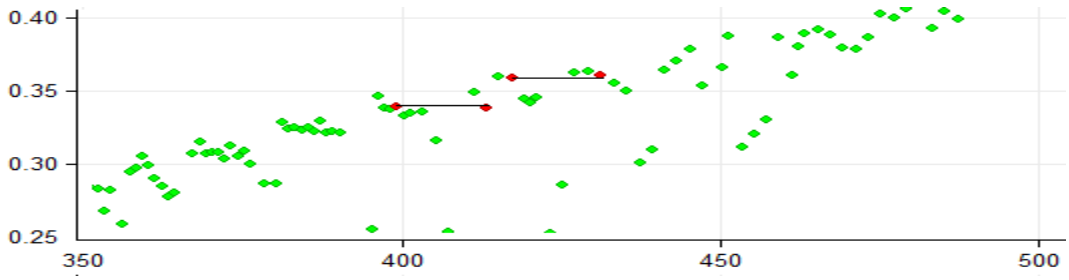


Figure 5.3. Zooming of the Figure A4.3- KMD plot using CH_2 units to show the alignment of the ions at *m/z* 399 and *m/z* 413, and the ions at *m/z* 417 and *m/z* 431.

5.4.3. Low-Energy CID-MS/MS Analyses

As explained in the previous sections, the APPI-MS of the Oak VRL oligomers indicates a very complicated heterogeneous mixture of released VRL oligomers.

Due to the intricate presence of overlapping ion signals and additional overlapping isotopic distribution and centroid mass shifts, the accurate mass assignments remained an impossible challenging task to perform with our QqTOF-MS/MS instrument. [30] Nonetheless, measuring CID-MS/MS fragmentations will permit to postulate proposed chemical structure for each precursor ions and would also provide more support to the assigned formulas.[31] In addition, it will also decrease the number of possible candidates that could be attributed to each ion identified in the APPI-MS.

Accordingly, we will describe some low-energy CID-MS/MS measurements of eight characteristic major protonated molecules formed by the APPI ionization.

It should be mentioned that in most of the MS/MS studied; we noted the characteristic presence of the aromatic product ions at *m/z* 181 and 167 assigned respectively, as $C_9H_{11}O_3^+$ and $C_{10}H_{13}O_3^+$. These last two ions are characteristic of the presence of the syringyl monomer.[32] Furthermore, the identification of these syringyl ions

was consistent with our recent analysis by MALDI-TOF-MS/MS of the same oak VRL mixture. This indicates that the French oak sample is rich with syringyl lignin units.[17]

5.4.3.1. Product ion scan of the protonated molecule *1* at *m/z* 347.13

The $[C_{19}H_{22}O_6 + H]^+$ (DBE= 8.5) formula was assigned to the precursor ion *1* at *m/z* 347.13 (Figure 5.4 and Scheme 5.1). This precursor ion is likely to be the neolignane cedrusin (1a).[33] The product ion scan of the precursor ion *1* was initiated by the loss of methyl radical (-15 Da) to yield the product ion at *m/z* 332. Also, the precursor ion *1* was subjected to a phenolic ring contraction through the loss of one molecule of carbon monoxide to yield the ion at *m/z* 319. This latter ion loses one formaldehyde molecule to yield the product ion at *m/z* 289. As well, the product ion at *m/z* 319 can also lose either ethanol (-46 Da) or propanol (-60 Da) to yield the secondary product ions respectively at *m/z* 273 and *m/z* 259.

In addition, the precursor ion *1* can also fragment by loss of either formaldehyde (-30 Da) or methanol (-32 Da) molecules to create, respectively, the product ions at *m/z* 317 and *m/z* 315. Furthermore, the product ion scan of the precursor ion *1* eliminated an ethanol (-46 Da) molecule to yield the stable benzylic cation at *m/z* 301. Likewise, the precursor ion *1* eliminates one propanol molecule (-60 Da) to generate the product ion at *m/z* 287.

Lastly, due to the complexity of the product ion scan of this precursor ion *1* at *m/z* 347.13, some product ions cannot be explained uniquely from the structure of the neolignane cedrusin (1a). For this reason, another possible precursor isobaric ion (1b) with the chemical formula $[C_{22}H_{18}O_4 + H]^+$ is proposed. Isobar (1b) is an aryltetralin lignan

derivative containing the indene aromatic moiety. (Scheme A4.1). This isobar (1b) was used to explain the formation of the product ions at *m/z* 219, *m/z* 199, *m/z* 197, and *m/z* 161, as shown in Scheme A4.1. In an earlier work published, we have introduced a designation to this kind of very complicated crude MS/MS spectrum as "messy " MS/MS.[34] This nomenclature was used to justify the fact that the window of selection (1 Da) of the precursor ion the quadrupole of the QqTOF-MS/MS instrument (1 *m/z* unit) existed as a mixture of several isomers and/or isobars, which indeed complicate the MS/MS spectrum (Figure 5.4).

Nevertheless, this type of ambiguity was also demonstrated by Qi *et al.* while performing MS/MS analysis with the high-resolution FT-ICR-MS^n instruments and using a 0.2 *m/z* unit window for selecting the precursor ion.[21] They found that for this small window value, 20 isobaric compounds existed! [21] For this reason, a future work using a combination of ion mobility and high-resolution mass spectrometer instruments such as FTICR or Orbitrap is needed to distinguish between the closely overlapping isobars and/or isomers.

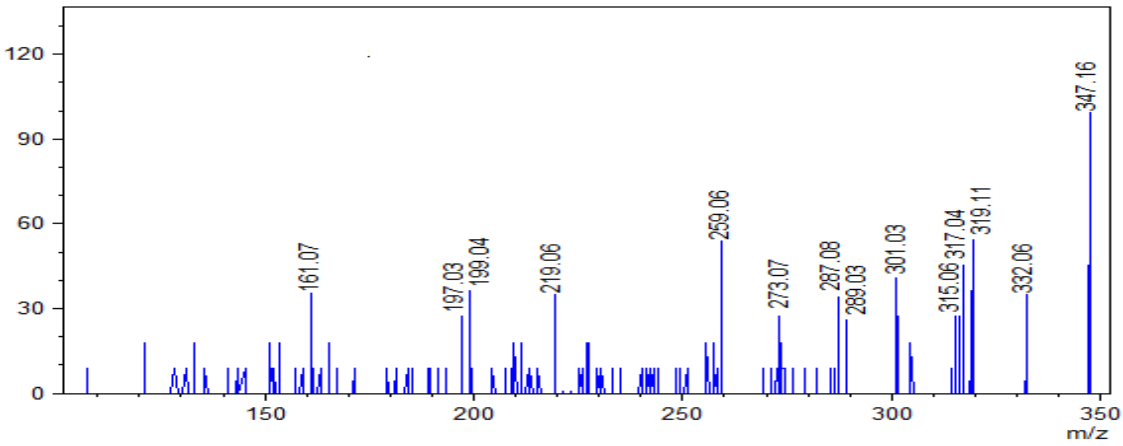
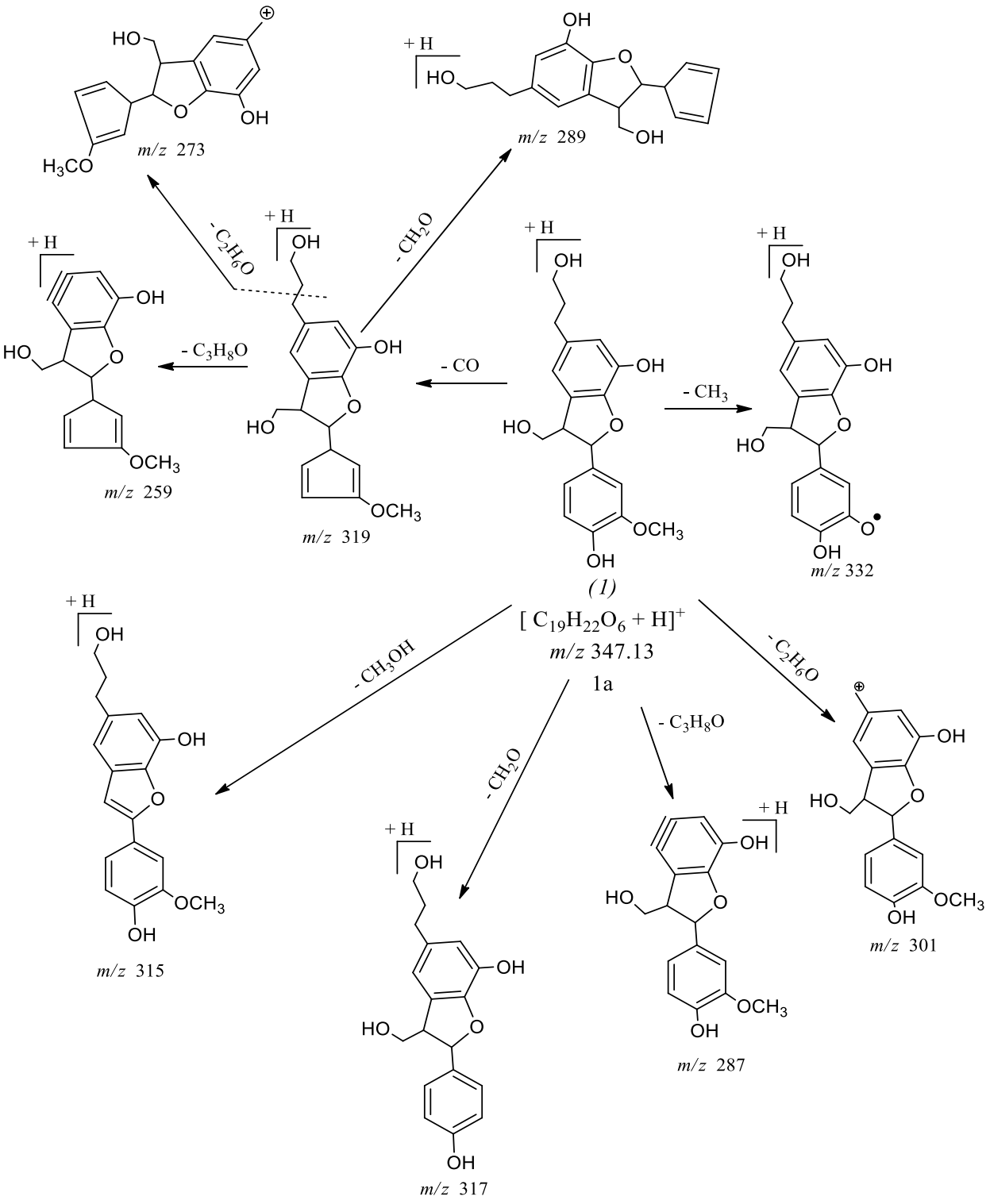


Figure 5.4. Product ion scan of the protonated precursor ion *1* at *m/z* 347.13



Scheme 5.1. Low energy CID-MS/MS fragmentation mechanism of the protonated precursor ion *1* at *m/z* 347.13

5.4.3.2. Product ion scan of the protonated molecule *2* at *m/z* 355.16

The $[C_{21}H_{22}O_5 + H]^+$ (DBE= 10.5) formula was assigned to the precursor ion *2* at *m/z* 355.16 (Figure 5.5 and Scheme 5.2). This precursor ion is tentatively assigned as an aryltetralin lignan derivative composed of methylated coumaric acid and methylated coniferyl alcohol (2a). The product ion scan of the precursor ion *2* was initiated by the loss of methyl radical (-15 Da) to yield the product ion at *m/z* 340 (Scheme 5.2). Also, the precursor ion *2* eliminates one methanol molecule to yield the product ion at *m/z* 323. Besides, the precursor ion 2 also experiences the loss of a methyl formate molecule ($-C_2H_4O_2$) to create the product ion at *m/z* 295. Moreover, a ring contraction in the central aliphatic ring of the lignan structure leads to the formation of the product ion at *m/z* 235, as demonstrated in Scheme 5.2. Lastly, the formation of the characteristic product ions at *m/z* 151 (dimethoxy benzyl cation) and *m/z* 179 (methylated coumaric acid protonated molecule) support the proposed lignan structure, as shown in Scheme 5.2. Lastly, another possible lignan isomer (2b) was proposed to explain the formation of the product ion at *m/z* 243.

Another possible isobar (2c) with the chemical formula $[C_{17}H_{22}O_8 + H]^+$ was proposed to explain the formation of the product ion at *m/z* 211. The isobar (2c) is a lignin carbohydrate complex composed of sinapyl alcohol and levoglucosan sugar. The loss of the sugar moiety leads to the formation of the protonated molecule of sinapyl alcohol at *m/z* 211 (Scheme A4.2). It should be noted that isobar (2c) can experience the loss of 60 Da ($-C_2H_4O_2$) from the sugar ring to explain the formation of the product ion at *m/z* 295.

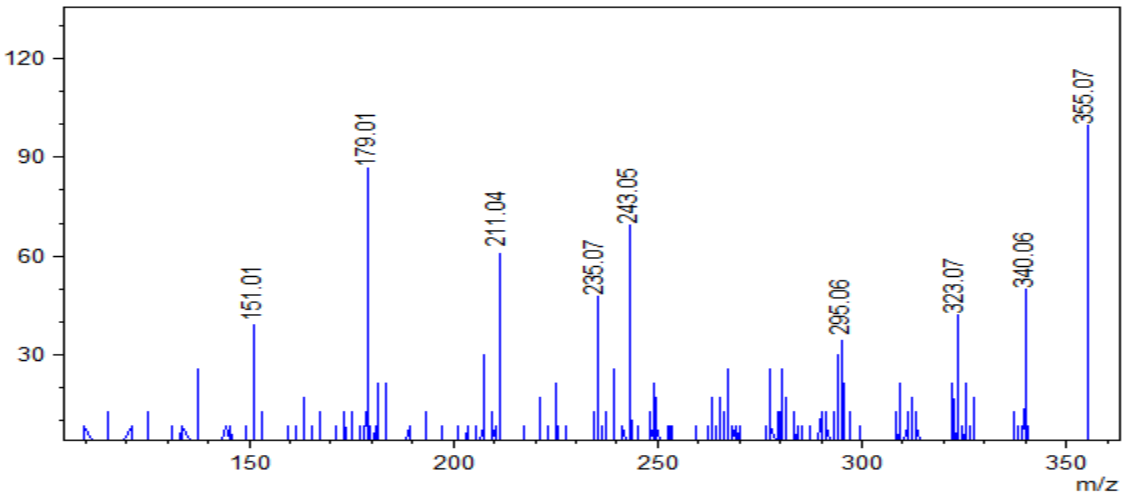
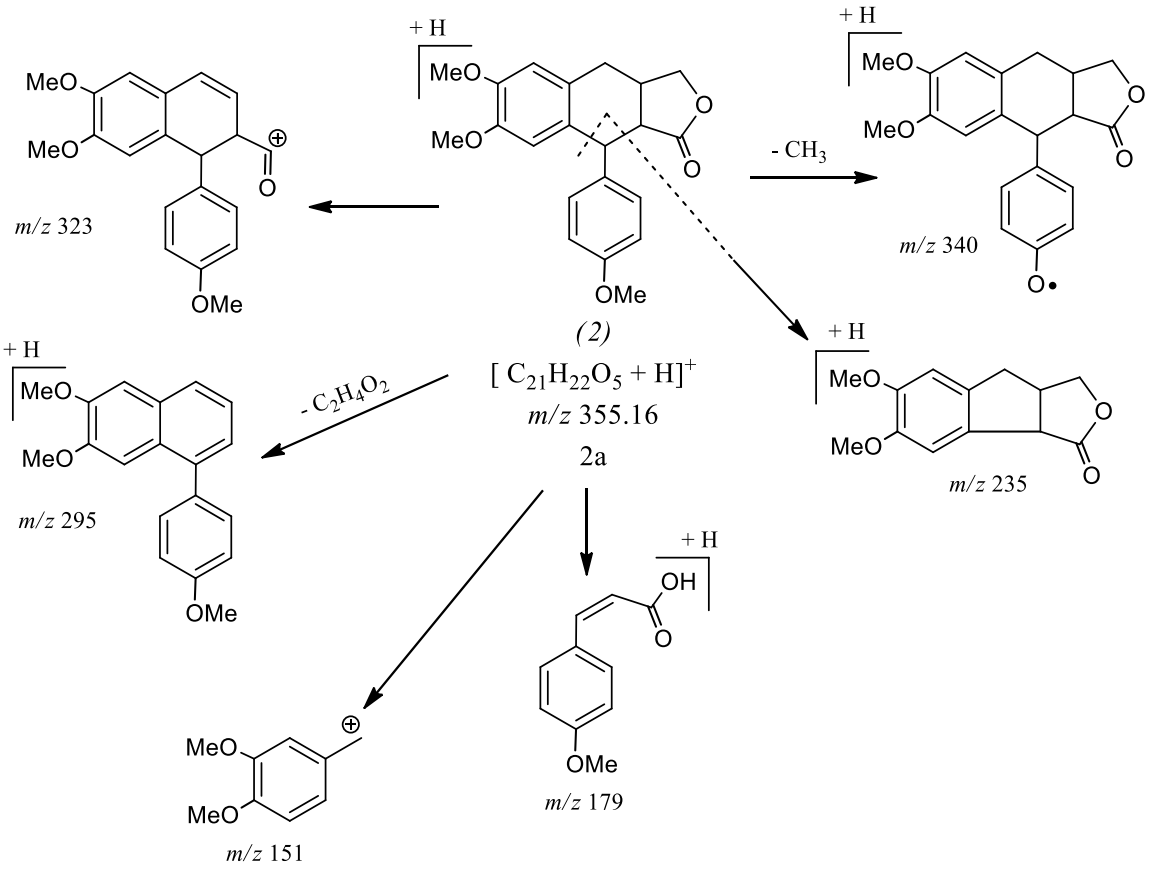
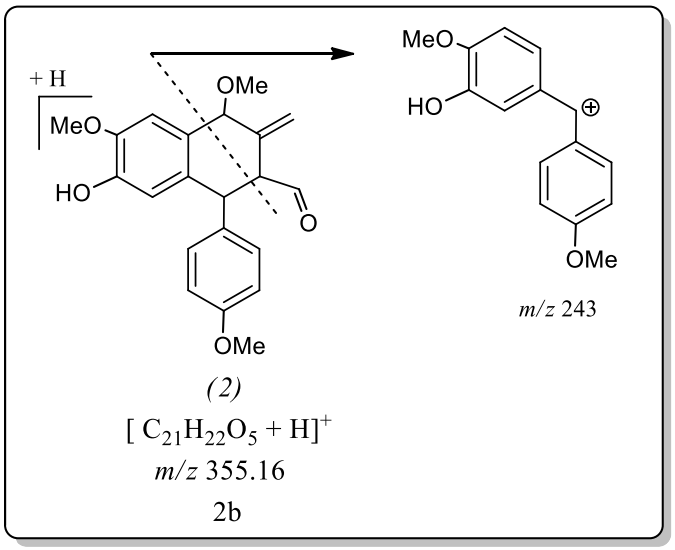


Figure 5.5. Product ion scan of the protonated precursor ion *2* at *m/z* 355.16





Scheme 5.2. Low energy CID-MS/MS fragmentation mechanism of the protonated precursor ion *2* at *m/z* 355.16

5.4.3.3. Product ion scan of the protonated molecule *3* at *m/z* 369.10

The $[C_{20}H_{16}O_7 + H]^+$ (DBE= 12.5) formula was assigned to the precursor ion *3* at *m/z* 369.10 (Figure 5.6 and Scheme 5.3). This precursor ion is tentatively assigned to an aryltetralin lignan derivative (3a). The product ion scan of the precursor ion *3* was initiated by the loss of methyl radical (-15 Da) to yield the product ion at *m/z* 354. Likewise, the precursor ion *3* eliminates one methanol molecule to yield the product ion at *m/z* 337. This latter ion loses a carbon monoxide molecule to yield the ion at *m/z* 309. Moreover, cleavage in the central aliphatic ring of the lignan structure leads to the formation of the product ions at *m/z* 121 and 249, as shown in Scheme 5.3. The product ion at *m/z* 249 loses a water molecule to yield the product ion at *m/z* 231. Lastly, the presence of the guaiacyl moiety in the lignan structure was supported by the formation of the characteristic product ion at *m/z* 137, as demonstrated in Scheme 5.3.[32]

Additionally, another two possible isomers (3b) and (3c) were proposed to explain the formation of the product ions at *m/z* 293 and *m/z* 277. Lastly, the formation of the product ion at *m/z* 167 supports the presence of a syringyl moiety, which leads us to propose another lignan isobar 3d with the chemical formula $[C_{21}H_{20}O_6 + H]^+$ to demonstrate its formation.

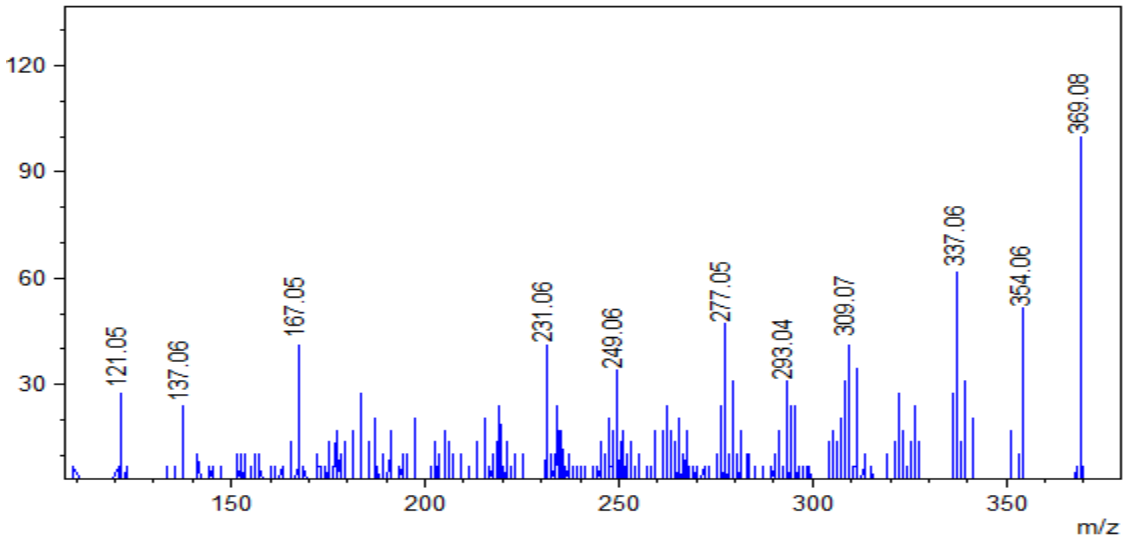
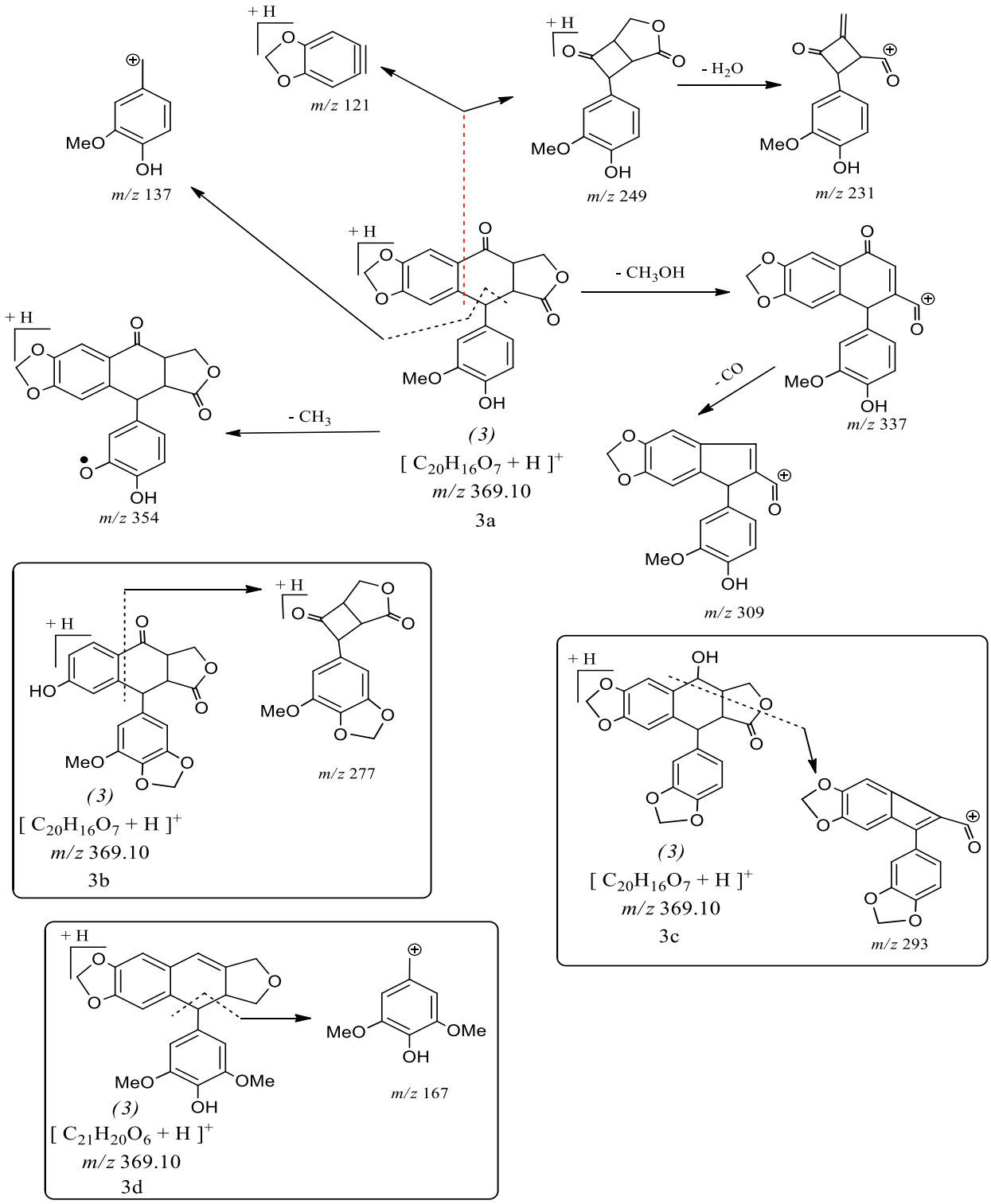


Figure 5.6. Product ion scan of the protonated precursor ion *3* at *m/z* 369.10



Scheme 5.3. Low energy CID-MS/MS fragmentation mechanism of the protonated precursor ion *3* at *m/z* 369.10

5.4.3.4. Product ion scan of the protonated molecule *4* at *m/z* 399.10

The $[C_{21}H_{18}O_8 + H]^+$ (DBE= 12.5) formula was assigned to the precursor ion *4* at *m/z* 399.10. This precursor ion is tentatively assigned to the oxidized form of the lignan 4'-demethylepipodophyllotoxin (4a) (Figure 5.7 and Scheme 5.4).[35] The product ion scan of the precursor ion *4* was initiated by the loss of methyl radical (-15 Da) to yield the product ion at *m/z* 384 (Scheme 5.4). This latter product ion can lose a carbon monoxide molecule to yield the product ion at *m/z* 356. Likewise, the precursor ion *4* is subjected to the loss of a carbon monoxide molecule to yield the product ion at *m/z* 371. This latter product ion loses dioxirane CH_2O_2 (-46 Da) from the methylenedioxy moiety to yield the product ion at *m/z* 325. [36] This latter product ion at *m/z* 325 loses one water molecule to create the product ion at *m/z* 307.

Furthermore, the precursor ion *4* can also be instigated by the neutral loss of the aromatic moiety $C_7H_4O_2$ (-120 Da) to form the product ion at *m/z* 279. Moreover, the precursor ion 4 can also eliminate a dimethoxy phenol molecule (-154 Da) to form the product ion at *m/z* 245. This latter ion eliminates a carbon monoxide molecule to yield the secondary product ion at *m/z* 217.

Additionally, the precursor ion *4* experiences a neutral loss of acetolactone molecule $[C_2H_2O_2]$ (-58 Da) to yield the product ion at *m/z* 341. This latter product ion loses a hydrogen molecule to produce the conjugated product ion at *m/z* 339. Lastly, the formation of the characteristic product ions at *m/z* 167 and 181 supported the presence of the syringyl lignin unit in the lignan structure, as presented in Scheme 5.4.

Another possible structure for lignan isomers (4b) and (4c) of the precursor ion *4* at *m/z* 399 was also proposed to explain the formation of the product ions at *m/z* 233, *m/z* 247, *m/z* 249, and *m/z* 261 (Scheme A4.3).

Lastly, isomer (4d) of the precursor ion *4* was proposed to be composed of an anthraquinone derivative attached to shikimic acid. This latter isomer (4d) was proposed to explain the formation of the product ions at *m/z* 321, *m/z* 213, *m/z* 185, and *m/z* 157 (Scheme A4.3). It should be noted that the product ions at *m/z* 213 and 185 were reported before to be formed from anthraquinone derivatives.[37]

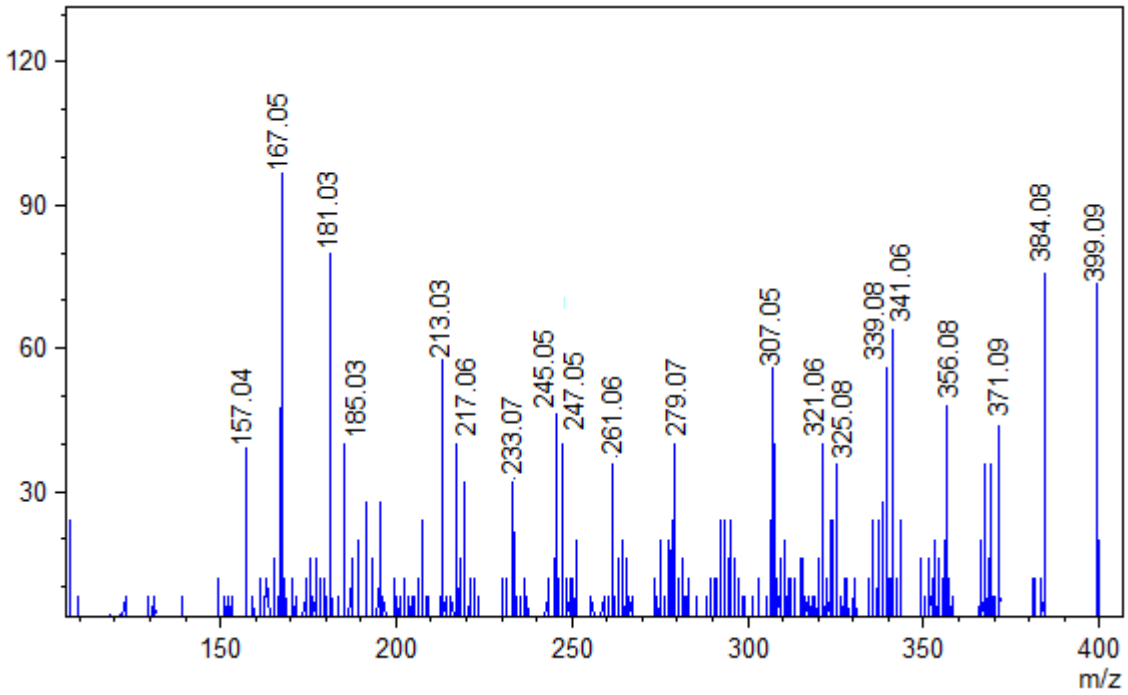
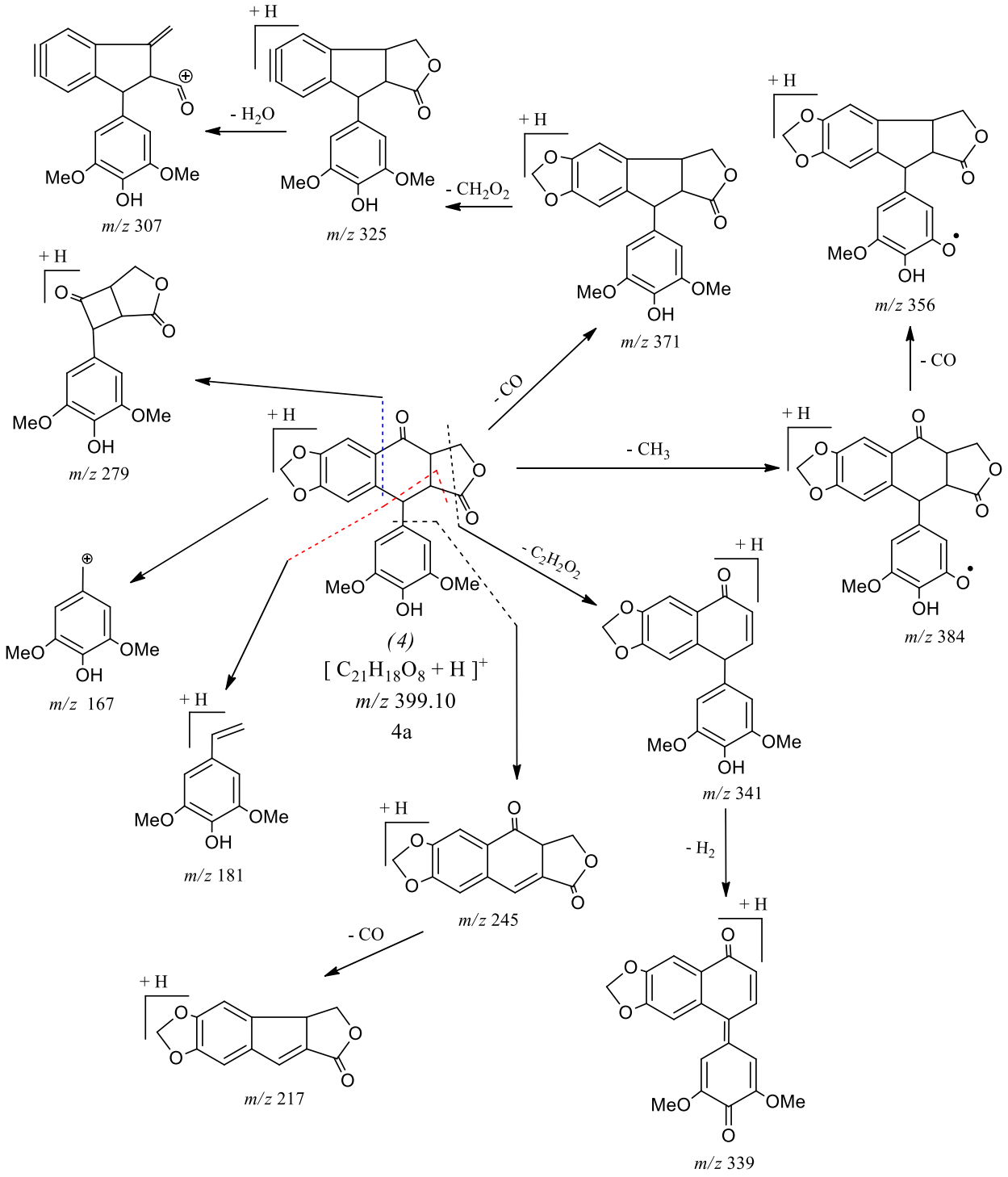


Figure 5.7. Product ion scan of the protonated precursor ion *4* at *m/z* 399.10



Scheme 5.4. Low energy CID-MS/MS fragmentation mechanism of the protonated precursor ion *4* at *m/z* 399.10

5.4.3.5. Product ion scan of the protonated molecule *5* at *m/z* 401.11

The $[C_{21}H_{20}O_8 + H]^+$ (DBE= 11.5) formula was assigned to the precursor ion *5* at *m/z* 401.11. This precursor ion is tentatively assigned to an isomer of the lignan 4'-demethylepipodophyllotoxin (5a) (Figure 5.8 and Scheme 5.5).[35]

The product ion scan of the precursor ion *5* was initiated by the loss of methyl radical to yield the product ion at *m/z* 386. Likewise, precursor ion *5* was subjected to a water molecule elimination to yield the product ion at *m/z* 383. In its turn, this latter product ion experiences the neutral loss of two different aromatic moieties, $C_9H_{10}O_3$ or $C_7H_4O_2$, to yield the product ions at *m/z* 217 and *m/z* 263. As mentioned before (section 3.3), the formation of the product ions at *m/z* 167 and 181 support the presence of the syringyl unit, as shown in Scheme 5.5. The latter ion at *m/z* 167 can lose two formaldehyde molecules to yield the product ion at *m/z* 107.

To complicate matters further, this precursor ion 5 can be subjected to another series of fragmentations shown, as well, in Scheme 5.5. The precursor ion *5* eliminates one dioxirane molecule ($-CH_2O_2$) from the methylenedioxy moiety of the lignan structure to yield the product ion at *m/z* 355.[36] This last ion at *m/z* 355 showed the loss of formic acid ($-CH_2O_2$) to produce the product ion at *m/z* 309. This latter ion at *m/z* 309 can lose one water molecule to yield the product ion at *m/z* 291. Furthermore, the precursor ion *5* can also lose a methanol molecule to create the product ion at *m/z* 369. This latter ion yields the product ion at *m/z* 351 by losing one water molecule. This last ion at *m/z* 351 loses one carbon monoxide molecule to produce the product ion at *m/z* 323. Moreover, the loss of

dimethoxy phenol moiety from the precursor ion 5 leads to the formation of the product ion at *m/z* 247 (Scheme 5.5).

Lastly, another isome (5b) could be described that can account for other characteristic product ions that cannot be formed from isomer (5a). The cleavage of the central aliphatic ring of the lignan structure in isomer (5b) leads to the formation of the product ion at *m/z* 219 (Scheme A4.4). This later ion at *m/z* 219 showed the consecutive losses of one methanol and one carbon monoxide molecules to yield the product ions at *m/z* 187 and 159. Lastly, another cleavage across the two aliphatic rings of the lignan structure leads to the formation of the product ion at *m/z* 313, as demonstrated in Scheme A4.4. Moreover, the loss of the aromatic moiety $C_7H_6O_2$ leads to the formation of the product ion at *m/z* 279. Lastly, isomer (5b) was used to demonstrate the formation of the product ions at *m/z* 299 and *m/z* 271, as shown in Scheme A4.4.

Additionally, another possible isomer (5c) was proposed to explain the formation of the product ion at *m/z* 231 (Scheme A4.4). It should be noted that this isomer (5c) contains an aromatic hydroxymethyl group that may be formed during the extraction process due to the presence of formic acid that can cause hydroxymethylation of G lignin units as demonstrated before in literature.[38]

It should be realized that once more, we are rationalizing the quality of this last MS/MS analysis as “Messy” by proposing three different gas-phase fragmentations for three isomers composing the precursor ion *5* at *m/z* 401.11.

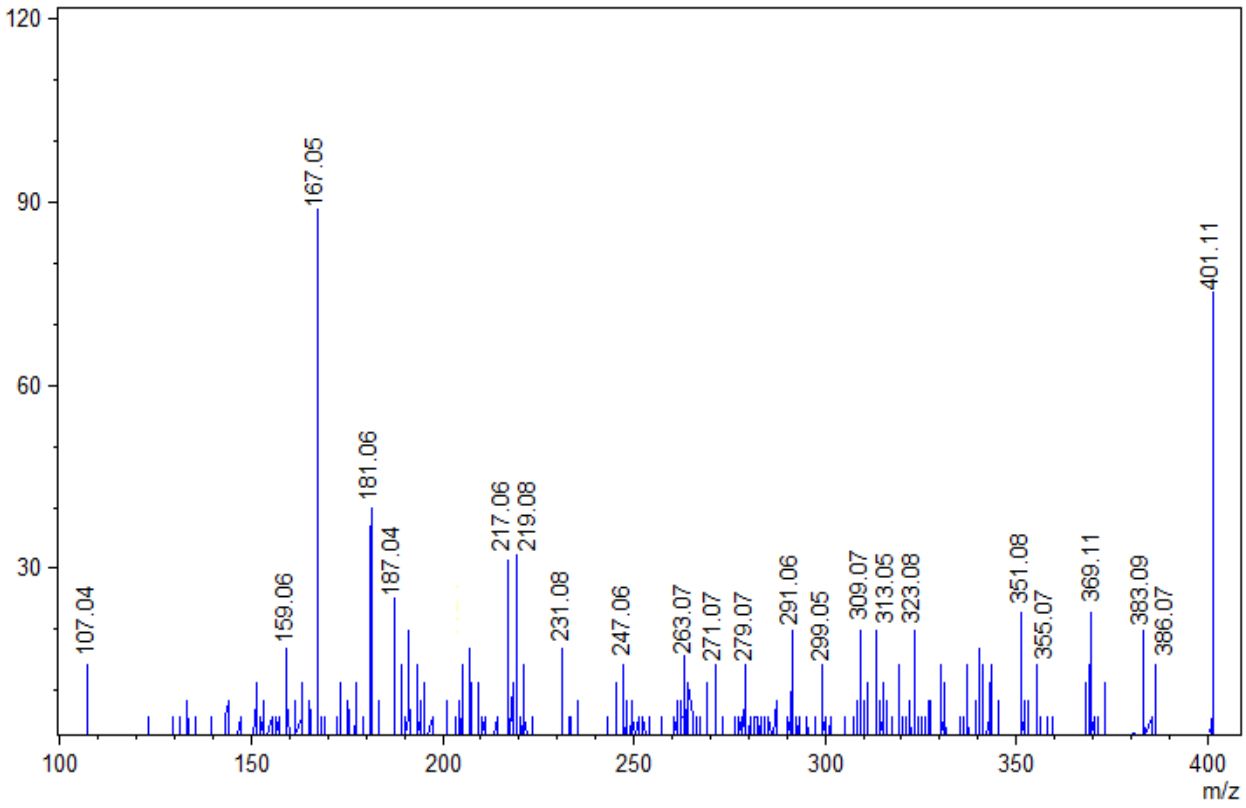
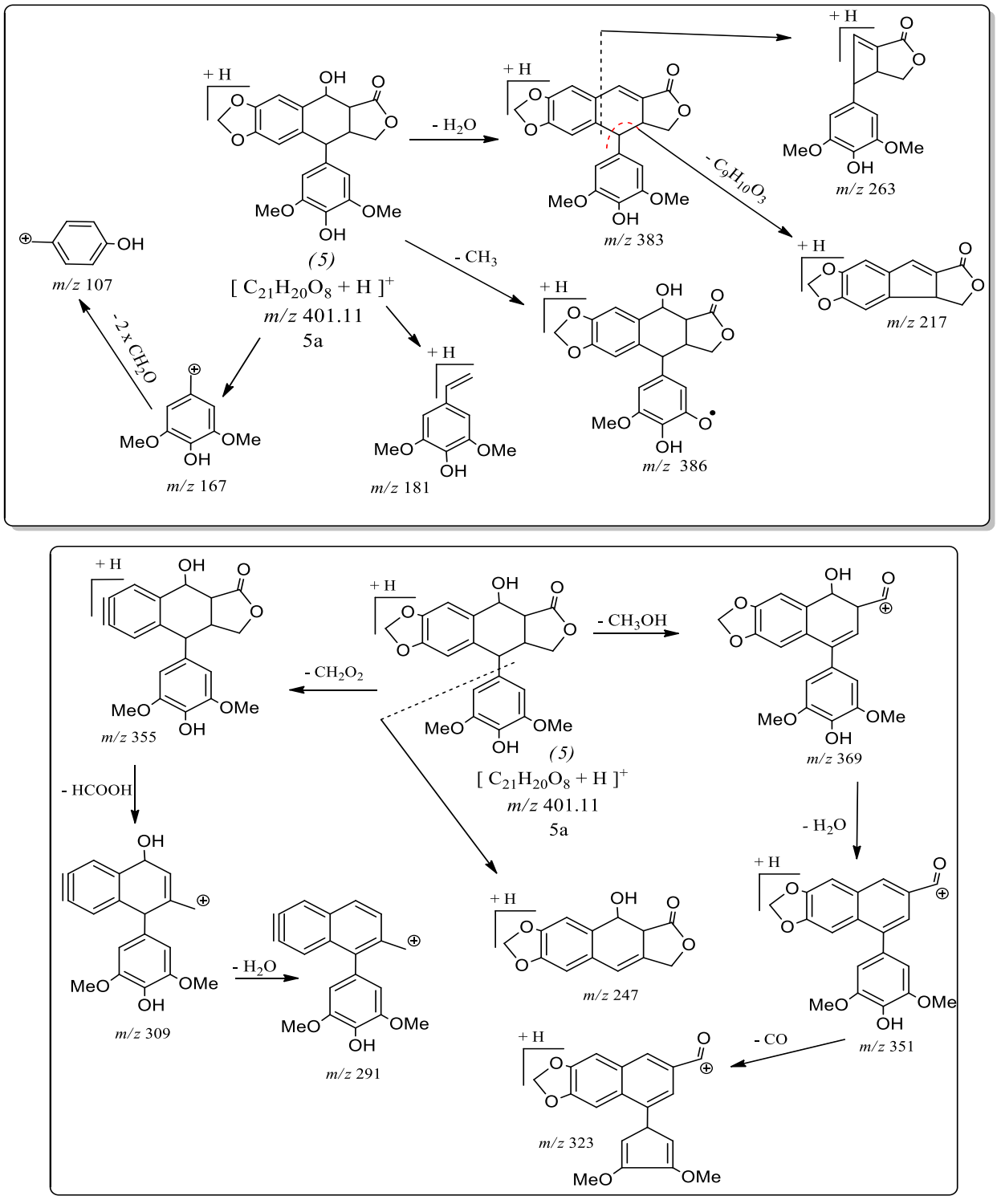


Figure 5.8. Product ion scan of the protonated precursor ion *5* at *m/z* 401.11



Scheme 5.5. Low energy CID-MS/MS fragmentation mechanism of the protonated precursor ion *5* at *m/z* 401.11

5.4.3.6. Product ion scan of the protonated molecule *6* at *m/z* 417.11

The $[C_{21}H_{20}O_9 + H]^+$ (DBE= 11.5) formula was assigned to the precursor ion *6* at *m/z* 417.11. This precursor ion is tentatively assigned to an aryltetralin lignan derivative (6a) (Figure 5.9 and Scheme 5.6).

The product ion scan of the precursor ion *6* was initiated by the loss of a water molecule to yield the product ion at *m/z* 399. This latter ion loses the aromatic moiety $C_9H_{10}O_3$ to yield the product ion at *m/z* 233. This latter product ion at *m/z* 233 loses a carbon monoxide molecule to yield the product ion at *m/z* 205. Furthermore, the precursor ion (6a) experiences a cleavage in its aliphatic rings that lead to the formation of the product ion at *m/z* 193, as demonstrated in Scheme 5.6.

Moreover, the formation of the product ions at *m/z* 167 and *m/z* 181 support the presence of the methylated syringyl lignin unit, as shown in scheme 5.6. Lastly, a possible isomer (6b) was proposed to demonstrate the formation of the product ion at *m/z* 217.

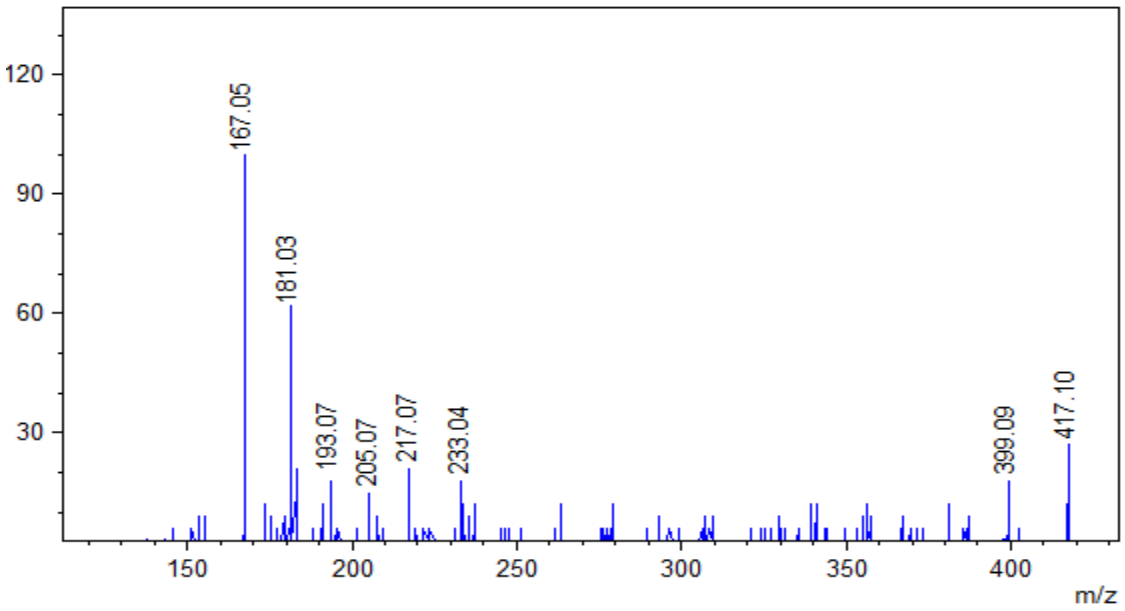
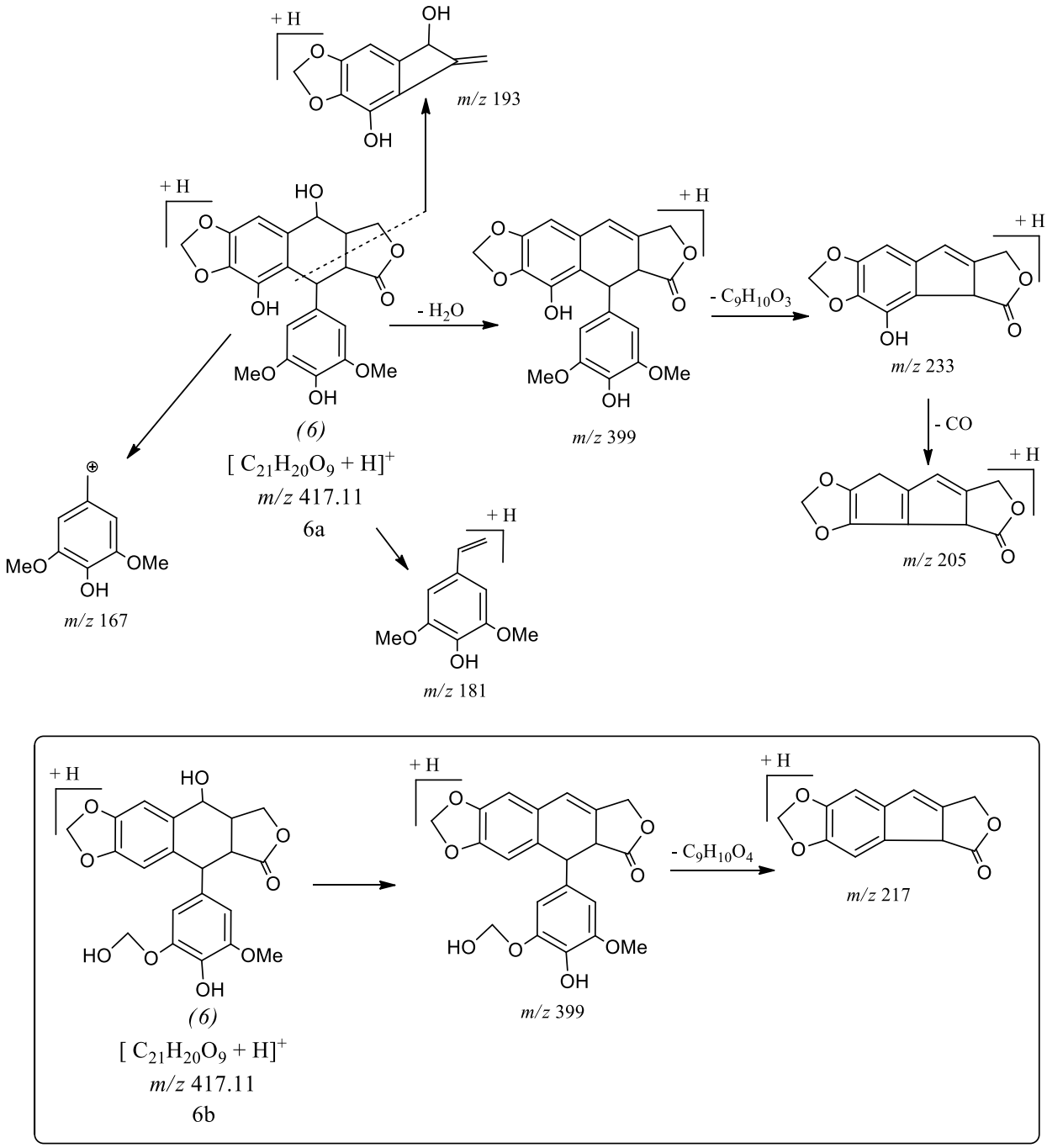


Figure 5.9. Product ion scan of the protonated precursor ion *6* at *m/z* 417.11



Scheme 5.6. Low energy CID-MS/MS fragmentation mechanism of the protonated precursor ion *6* at *m/z* 417.11

5.4.3.7. Product ion scan of the protonated molecule *7* at *m/z* 537.14

The $[C_{28}H_{24}O_{11} + H]^+$ (DBE= 16.5) formula was assigned to the precursor ion *7* at *m/z* 537.14. This precursor ion is tentatively assigned to the dehydrated form of an isomer of the lignan 4'-demethylepipodophyllotoxin attached to dehydroshikimic acid moiety (7a) (Figure 5.10 and Scheme 5.7). Also, it is important to mention that in our previous MALDI –TOF/TOF-MS/MS study of the Oak VRL mixture, we have identified a similar series of shikimic acid derivatives. [17]

The dehydroshikimic acid moiety is lost at the beginning of the product ion scan of the precursor ion *7* at *m/z* 537.14 to yield the product ion at *m/z* 383 (Scheme 5.7). This latter product ion at *m/z* 383 experiences the loss of a methanol molecule to create the product ion at *m/z* 351. Also, the product ion at *m/z* 383 undergoes a cleavage across its aliphatic rings to create the product ion at *m/z* 299. Moreover, the same product ion at *m/z* 383 losses the aromatic moiety $C_7H_4O_2$ to yield the product ion at *m/z* 263. Additionally, the formation of a series of characteristic product ions at *m/z* 167, *m/z* 181, *m/z* 193, and *m/z* 207 supported the presence of the syringyl unit in the lignan structure as presented in Scheme 5.7. Lastly, another two possible isomers (7b) and (7c) were proposed to demonstrate the formation of the product ions at *m/z* 215, *m/z* 291, and *m/z* 323, as shown in Scheme A4.5.

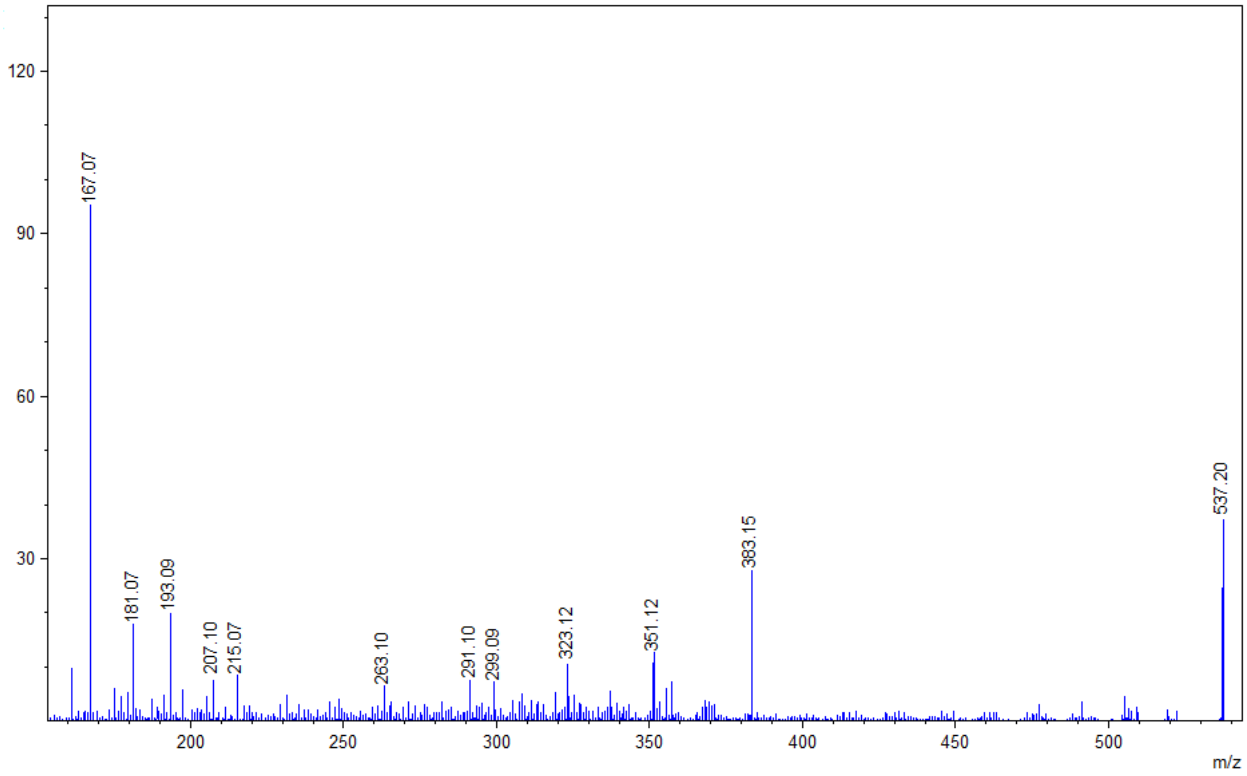
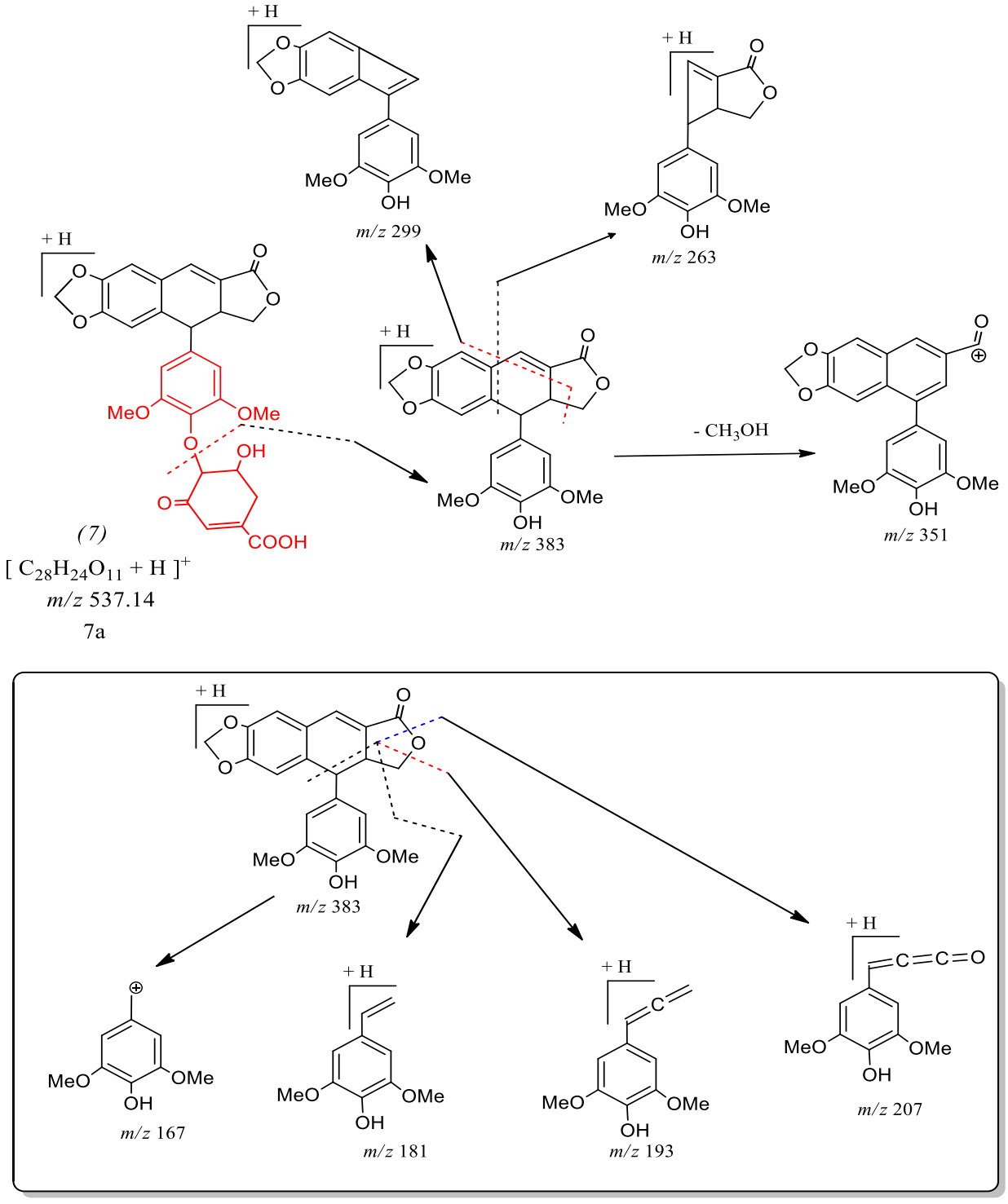


Figure 5.10. Product ion scan of the protonated precursor ion *7* at *m/z* 537.14



Scheme 5.7. Low energy CID-MS/MS fragmentation mechanism of the protonated precursor ion *7* at *m/z* 537.14. The red moiety in the precursor ion *7* was identified before in the analysis of the same sample using MALDI-TOF-MS/MS. [17]

5.4.3.8. Product ion scan of the protonated molecule *8* at *m/z* 643.16

The $[C_{33}H_{38}O_{13} + H]^+$ (DBE= 14.5) formula was assigned to the precursor ion *8* at *m/z* 643.16. This precursor ion is tentatively assigned to the novel lignan trimer composed of syringlyglycerol unit attached to the dibenzyl butyrolactone lignan $[C_{22}H_{24}O_8]$ reported previously by Li et al. (Figure 5.11 and Scheme 5.8).[39] It should be noted that similar syringyl glycerol derivatives were detected recently in the same Oak VRL sample, which was identified by using MALDI-TOF-MS/MS.[17]

The product ion scan of the precursor ion *8* at *m/z* 643.16 was initiated by the loss of the syringylglycerol unit to yield the lignan dimer at *m/z* 417. This latter ion at *m/z* 417 was subjected to cleavage of its aliphatic butyrolactone ring to yield the product ions at *m/z* 225 (protonated sinapinic acid) and *m/z* 191. Furthermore, the product ion scan of ion *8* fragments by loosing the $[C_{19}H_{24}O_8]$ moiety (-380 Da) to yield the product ion at *m/z* 263. Lastly, the formation of a series of characteristic product ions at *m/z* 167, 181, supported the presence of the syringyl lignin unit in the proposed structure, as presented in Scheme 5.8.

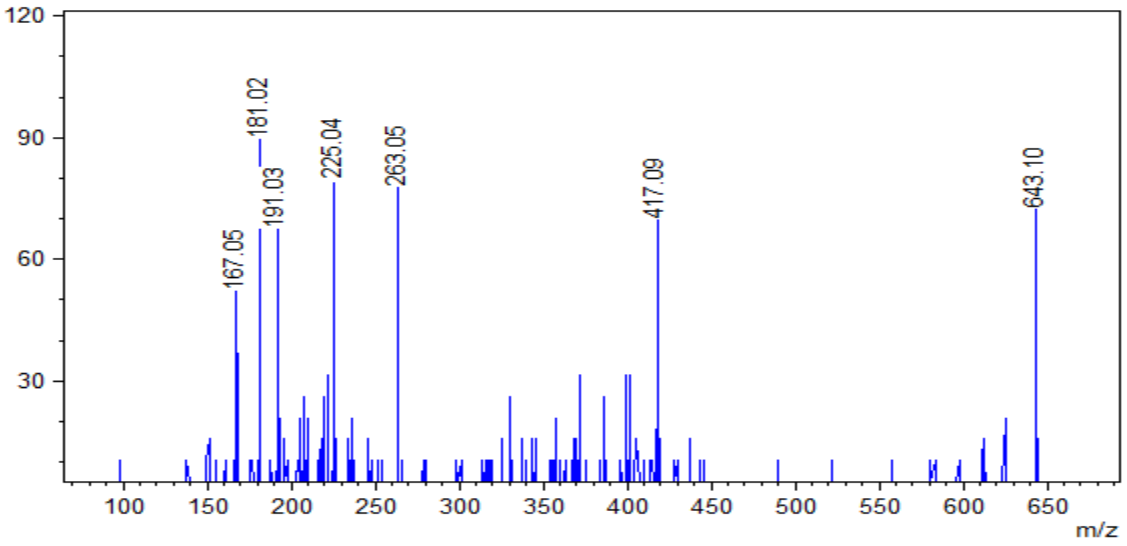
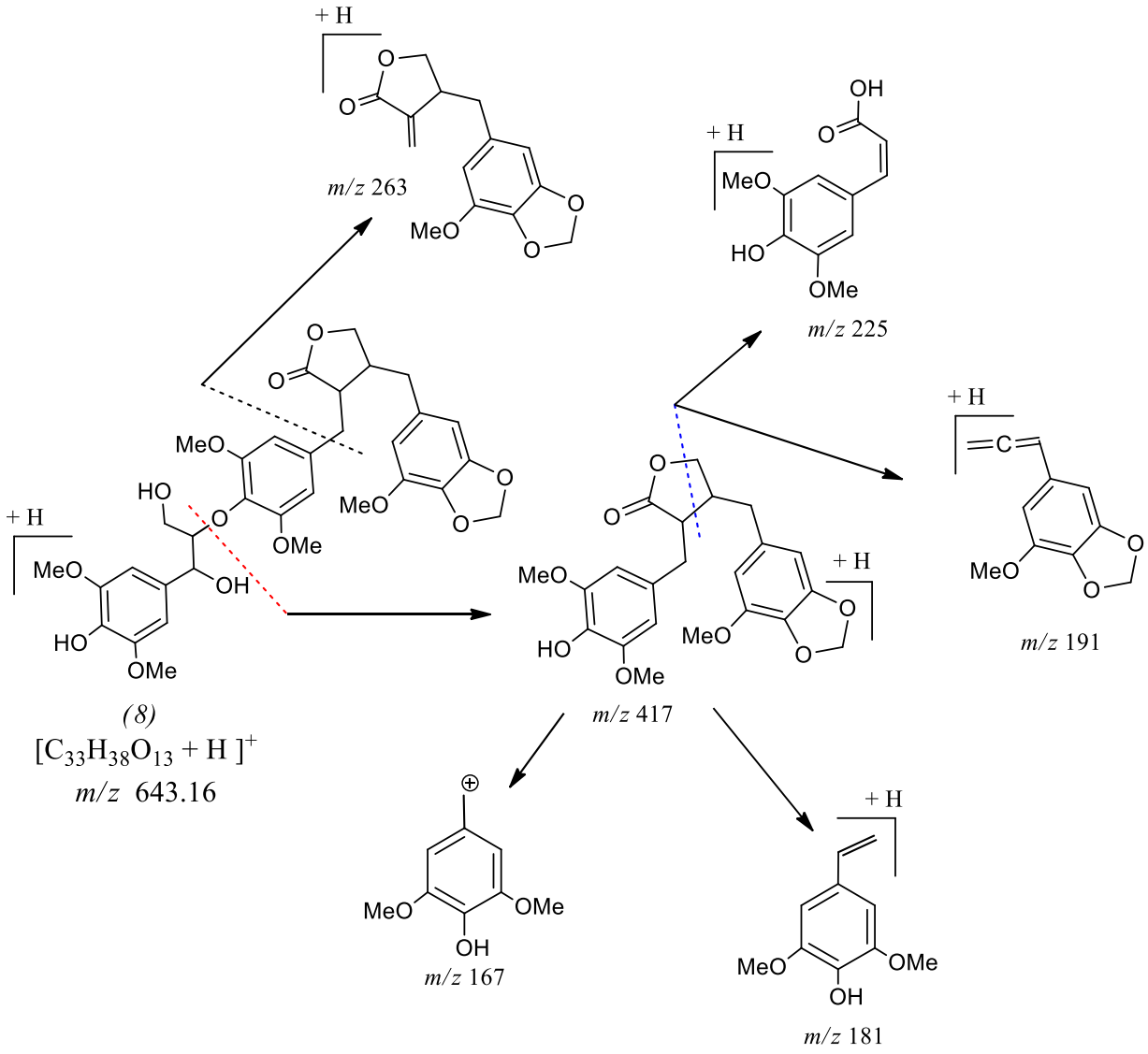


Figure 5.11. Product ion scan of the protonated precursor ion *8* at *m/z* 643.16



Scheme 5.8. Low energy CID-MS/MS fragmentation mechanism of the protonated precursor ion *8* at *m/z* 643.16

5.4.4. ESI-QqTOF-MS and Low energy CID MS/MS

Recently, a major effort was directed to comprehend whether the ion intensity in an ESI-MS spectrum actually reflects the measured lignin solution concentrations or not. This was needed in order to avoid problems in the analytical quantification of lignin.[40]

It was established that the ions intensities of any measured ESI spectrum do not necessarily reflect the real concentration of the analyte. This is simply due to the large dependence of the ESI response on other factors of the analyte molecular properties.[41] Recently, Asare and Lynn reported that while ESI-MS measuring of equimolar synthesized pure standard lignin, they noted that a slight change in the structure of the standard analytes resulted in recorded MS with different orders of magnitudes. Therefore, the high ESI response of some lignin oligomers in complex mixtures does not really mean that they have the highest concentration in the studied mixture, but their chemical properties allowed them to be ionized efficiently.

In this rationale, we have used another atmospheric pressure ionization ESI-MS (+ve ion mode) in a quest to pinpoint the presence of new VRL oligomers and in order to obtain additional structural information.

The ESI-QqTOF-MS of the Oak VRL mixture indicated the presence of two protonated molecules at *m/z* 399.12 *(4)* and *m/z* 537.17 *(7)*, which were already identified in the APPI-MS (Figure 5.12). In addition, it showed the presence of three extra novel VRL protonated molecules at *m/z* 441.12 *(9)*, *m/z* 471.13 *(10)*, and *m/z* 499.13 *(11)* (Figure 5.12, Figure A4.5 and Table 5.2). In this section, we will discuss the product ion scan of these last three novel oligomers.

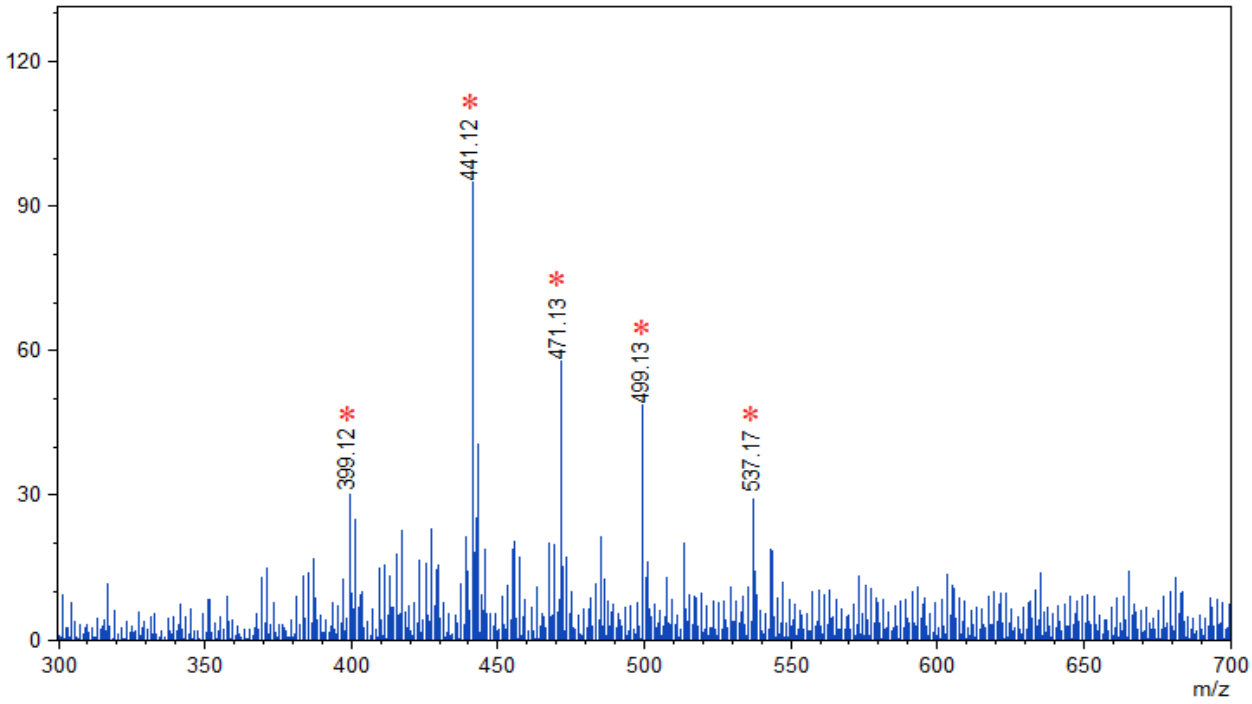


Figure 5.12. ESI-QqTOF-MS of the French Oak lignin from *m/z* 300-700

Table 5.2. ESI-QqTOF-MS and Low-Energy CID-MS/MS analyses of the extracted VRLs from French Oak Wood

Ion	Chemical Formula	Experimental *m/z*	Theoretical *m/z*	Error (*m/z* units)	DBE	MS/MS
9	$[C_{24}H_{24}O_8 + H]^+$	441.12	441.15	-0.03	12.5	399, 367, 357, 342, 325, 307, 296, 279, 181, 167
10	$[C_{24}H_{22}O_{10} + H]^+$	471.13	471.13	0.00	13.5	439, 425, 353, 337, 319, 277, 245, 233, 217, 193, 181, 167, 147, 137
11	$[C_{25}H_{22}O_{11} + H]^+$	499.13	499.12	+ 0.01	14.5	453, 319, 299, 279, 217, 193, 181, 167

5.4.4.1. The product ion scan of the precursor ion *9* at *m/z* 441.12

The chemical formula $[C_{24}H_{24}O_8 + H]^+$ was tentatively assigned to the precursor ion *9* at *m/z* 441.12, whose structure was proposed as diacetylated aryl tetralin lignan

derivative, which originated from the combination of sinapic acid and *p*-coumaryl alcohol (Figure 5.13 and Scheme 5.9). It should be noted that this diacetylated derivative, most probably, is due to the result of *in situ* acetylation, which occurs during the CIMV extraction procedure using acetic acid and formic acid mixture.

Definitely, the presence of two acetyl groups was supported by the product ion scan of the precursor ion *9* at *m/z* 441.12, which was initiated by the consecutive losses of two acetyl groups (2 x 42 Da) to yield the product ions at *m/z* 399 and *m/z* 357, respectively. The product ion at *m/z* 357 eliminates a methanol molecule to yield the product ion at *m/z* 325. In its turn, this latter product ion, loses a methyl radical to yield the radical ion at *m/z* 342. This latter product radical ion at *m/z* 342 can lose a molecule of formic acid to create the product radical ion at *m/z* 296.

Additionally, the precursor ion at *m/z* 441.12 experiences two different cleavages in the aliphatic rings of the lignan structure to yield the product ions at *m/z* 307 and *m/z* 279, as demonstrated in Scheme 5.9. Lastly, as for the APPI-MS/MS section, the characteristic products ion of the syringyl moiety (*m/z* 167 and *m/z* 181) originate from the product ion at *m/z* 399, as shown in Scheme 5.9.

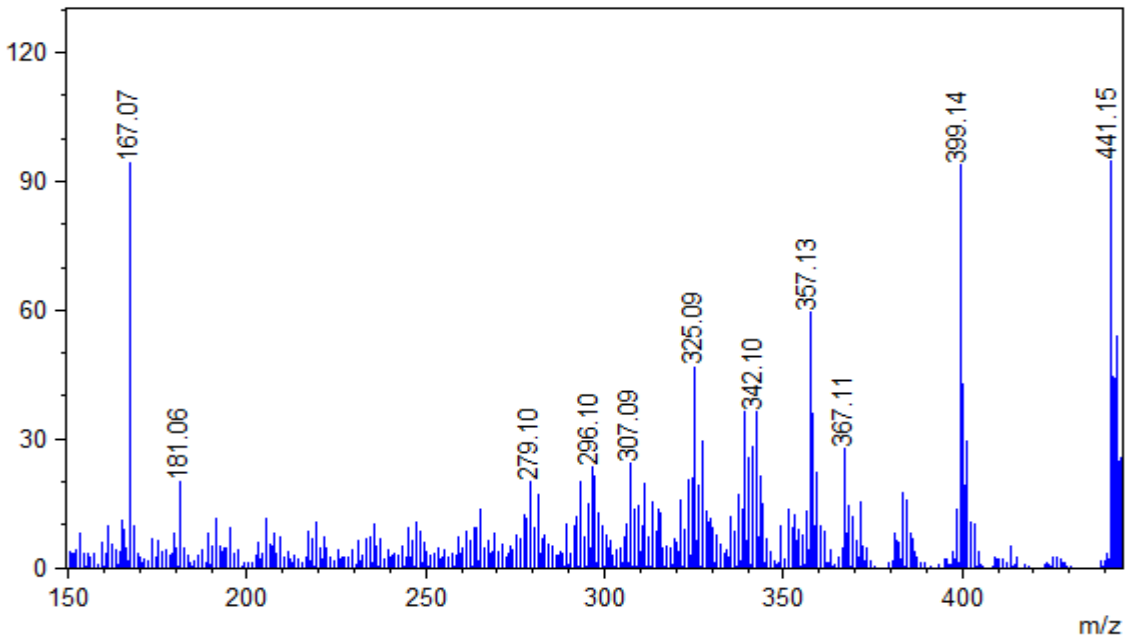
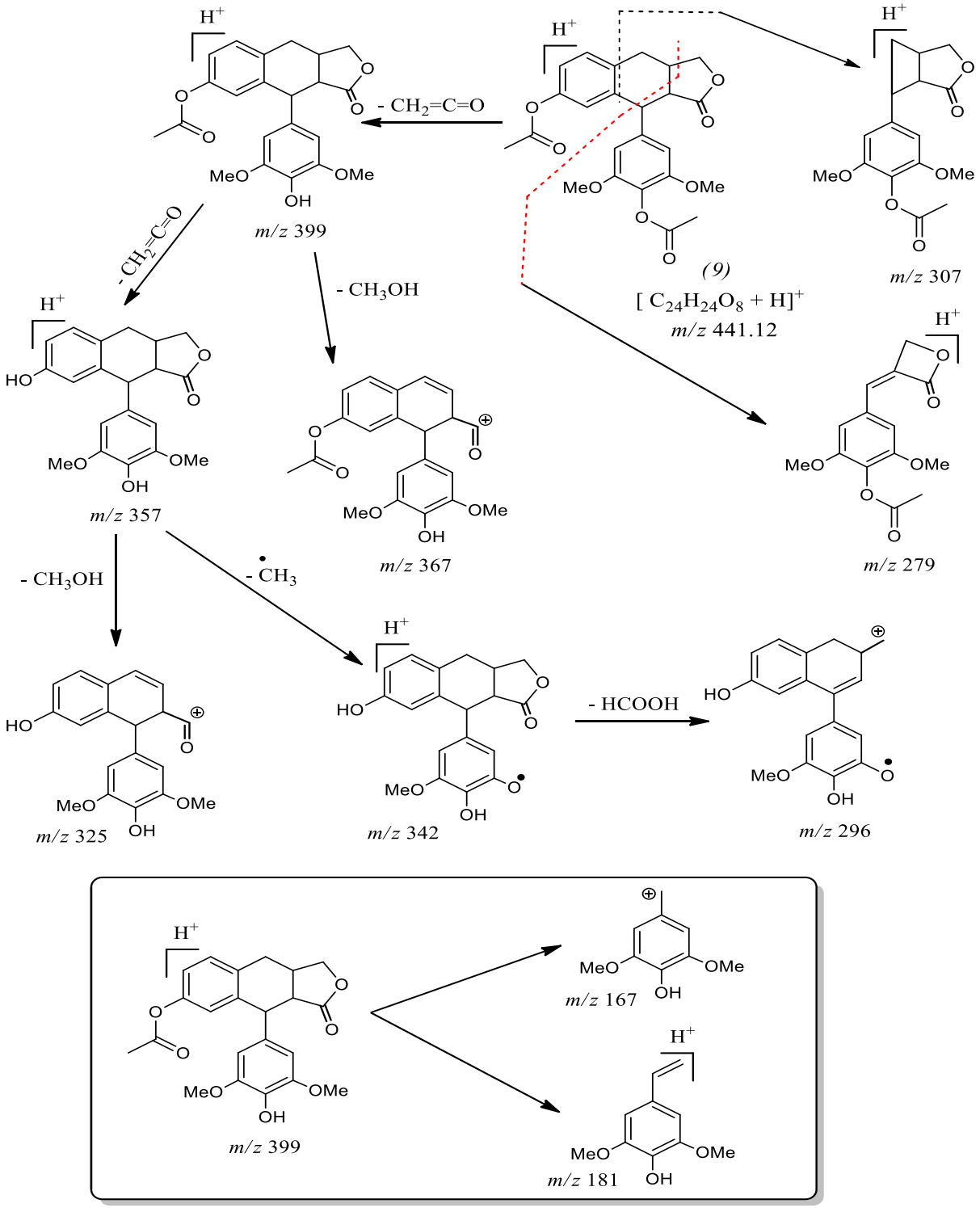


Figure 5.13. Product ion scan of the protonated precursor ion *9* at *m/z* 441.12



Scheme 5.9. Low energy CID-MS/MS fragmentation mechanism of the protonated precursor ion *9* at *m/z* 441.12

5.4.4.2. The product ion scan of the precursor ion *10* at *m/z* 471.13

The chemical formula $[C_{24}H_{22}O_{10} + H]^+$ was tentatively assigned to the precursor ion *10* at *m/z* 471.13, which is an oxidized form of the lignan 4'-demethylepipodophyllotoxin derivative containing a lactic acid moiety and denoted as isomer (10a) (Figure 5.14 and Scheme 5.10).

As already mentioned before and due to the complexity of the product ion spectrum (Messy MS/MS), we were forced to consider the presence of two other isomers (10b) and (10c), possessing the same chemical formula.

Therefore, the isomer (10b) was defined as being a lignin carbohydrate complex, composed of an oxidized pentose sugar derivative, which was attached to both dihydrocoumaryl alcohol and 3,4-methylenedioxycinnamic acid. As well, isomer 10c was composed of an aryl tetralin lignan derivative containing a pyruvic acid moiety.

Subsequently, the product ion scan of the precursor ion (10a) at *m/z* 471 was initiated by the loss of a methanol molecule to yield the product ion at *m/z* 439. Also, this precursor (10a) was subjected to the loss of a formic molecule to create the product ion at *m/z* 425. This latter ion can lose 72 Da from the lactic acid moiety ($-C_3H_4O_2$) to yield the product ion at *m/z* 353. Lastly, isomer (10a) showed two different neutral losses of $C_{12}H_{14}O_5$ (- 238 Da) and $C_{11}H_{14}O_5$ (-226 Da) to yield the product ions at *m/z* 233 and *m/z* 245, respectively (Scheme 5.10). This latter product ion at *m/z* 245 loses one carbon monoxide molecule to yield the product ion at *m/z* 217.

Additionally, the product ion scan of the precursor ion (10b) at *m/z* 471 could be initiated by cleavage of the Cγ-O of the dihydrocoumaryl unit to yield the product ion at

m/z 337 (Scheme 5.10). Also, another cleavage in this precursor ion B happened in the C3-O bond in pentose derivative affords the product ion at *m/z* 319. Similarly, the cleavage of the Cβ-Cγ bond of the precursor ion 3,4-methylenedioxycinnamic acid leads to the production of the product ion at *m/z* 147.

Lastly, the product ion scan of the precursor ion *10* formed from isomer 10c was proposed to explain the formation of the characteristic lignin product ions at *m/z* 137, *m/z* 167, and *m/z* 181 (Scheme A4.6). It should be noted that the ion at *m/z* 137 can originate from the ion at *m/z* 167 through the loss of a formaldehyde molecule. Also, it can originate from any other possible isomer containing the G lignin units as the ion at *m/z* 137 known specific to the presence of the G -Lignin unit. [32]

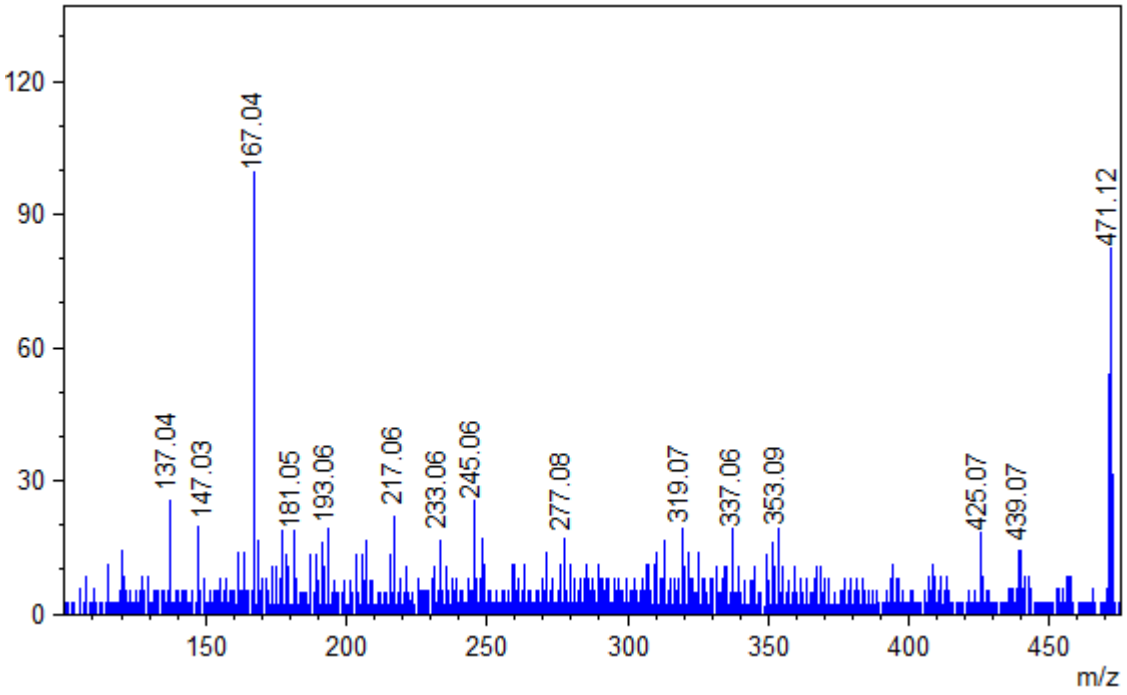
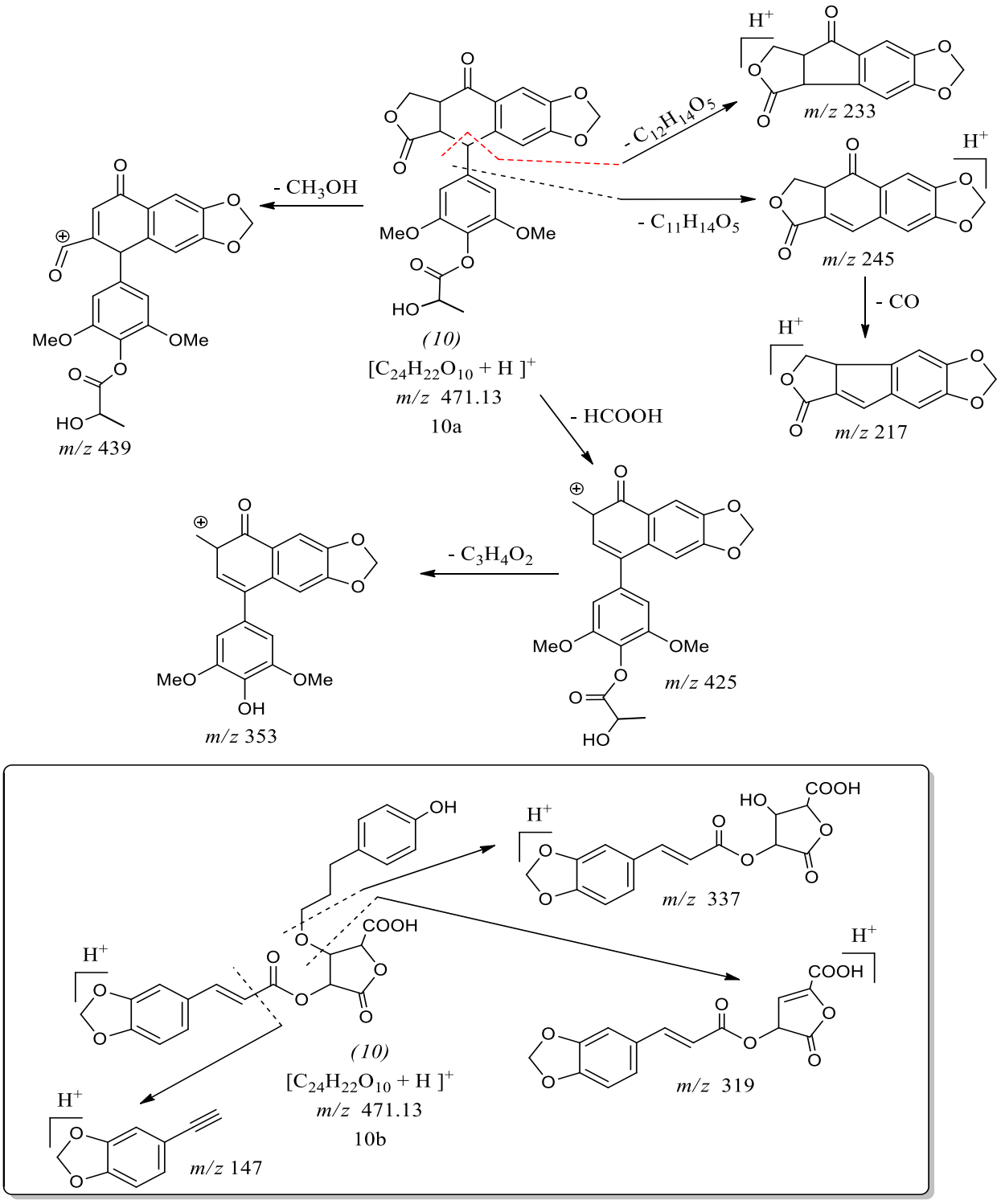


Figure 5.14. Product ion scan of the protonated precursor ion *10* at *m/z* 471.13



Scheme 5.10. Low energy CID-MS/MS fragmentation mechanism of the protonated precursor ion *10* at *m/z* 471.13

5.4.4.3. The product ion scan of the precursor ion *11* at *m/z* 499.13

The chemical formula $[C_{25}H_{22}O_{11} + H]^+$ was tentatively assigned to the precursor ion *11* at *m/z* 499.13, which was characterized as an aryl tetralin lignan derivative containing a butanedioic acid moiety and denoted as isomer (11a) (Figure 5.15 and Scheme 5.11). Similarly, due to the complexity of the product ion spectrum, we were obliged to consider the presence of another isomer 11b, which is composed of dehydroshikimic acid attached to caffeic acid and dihydrocaffeic acid aromatic units. The loss of caffeic acid moiety from this isomer 11b can lead to the formation of the product ion at *m/z* 319, as shown in Scheme 5.11.

The product ion scan of isomer (11a) was initiated by losing one molecule of formic acid to yield the product ion at *m/z* 453. This latter ion loses the $C_8H_{10}O_3$ moiety to yield the product ion at *m/z* 299. Likewise, this precursor ion (11a) can lose 46 Da ($-CH_2O_2$) from the methylenedioxy moiety to form another possible isomer of the product ion at *m/z* 453. This latter ion experiences the combined neutral loss of butanedioic acid $C_4H_4O_4$ (-116 Da) and acetolactone $C_2H_2O_2$ (-58 Da) moieties from the aliphatic rings of the lignan structure to produce a stable aromatic product ion at *m/z* 279. Moreover, the combined loss of the butanedioic moiety $C_4H_4O_4$ (-116 Da) and aromatic moiety $C_9H_{10}O_3$ (-166 Da) from the precursor ion (11a) lead to the formation of the product ion at *m/z* 217. Lastly, the characteristic product ions of the syringyl lignin unit at *m/z* 167, 181, and 193 can be formed from isomer 1, as shown in Scheme 5.11.

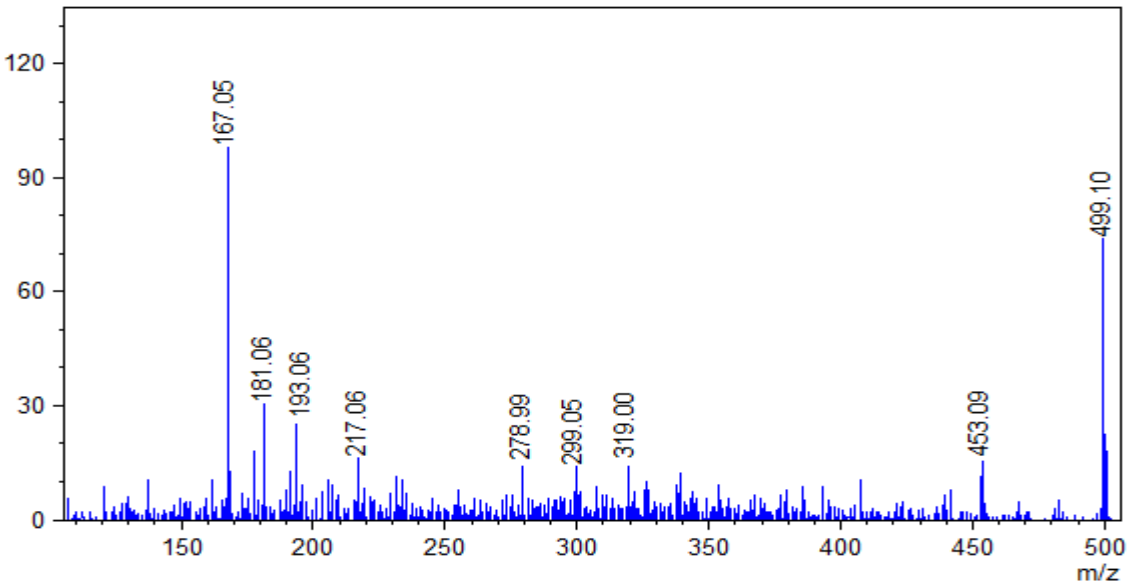
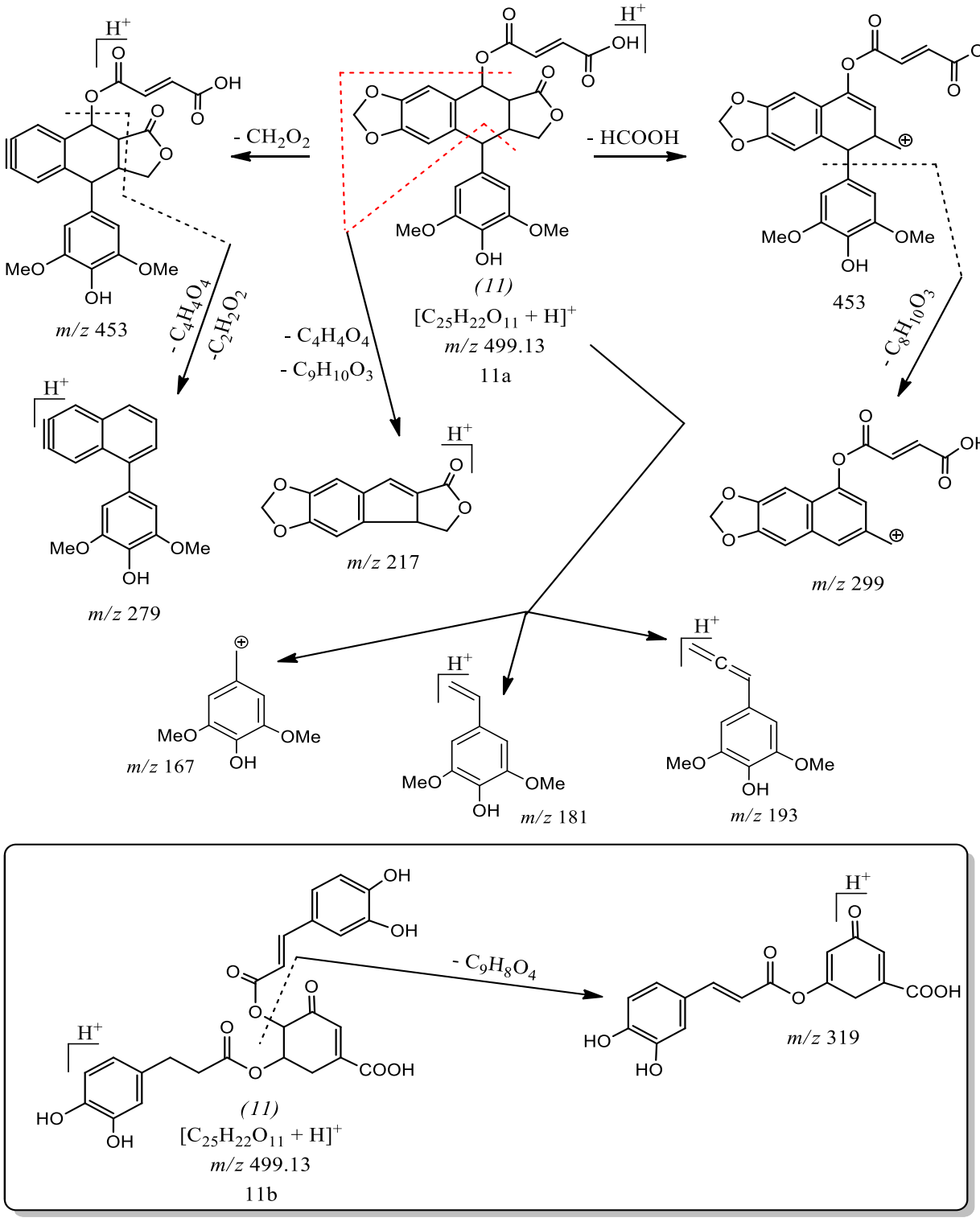


Figure 5.15. Product ion scan of the protonated precursor ion *11* at *m/z* 499.13



Scheme 5.11. Low energy CID-MS/MS fragmentation mechanism of the protonated precursor ion *11* at *m/z* 499.13

To sum it up, this ESI-MS/MS study showed that it was possible to identify different diagnostic Oak VRL oligomer ions beside other oligomers identified by APPI-MS/MS. This allows us to conclude that the MS analysis with varying methods of ionization can enhance the identification of different classes of VRL oligomers.

5.5. Conclusion

In this structural study of the Oak VRL oligomer by APPI and ESI-QqTOF-MS and low energy CID-MS/MS, we have investigated the structure of eleven major small oligomers present in the French Oak VRLs sample. All the ions identified were structurally related to a series of novel lignans. It is expected that the identification of novel lignan structures in this Oak VRL sample could provide new insight into the relationship between the complex biosynthesis of lignin and lignan. Also, it should be understood that this presented study indicated that the Oak VRL mixture composition is a very complex one. It contains various classes of compounds such as flavonoids, neolignans, lignans, lignin, lignin-shikimic acid derivatives, and lignin-carbohydrate complexes.[17]

The MS/MS analyses of VRL mixtures showed the intricate complexity in defining the obtained fragmentation patterns, which could not be solved unless we resorted to the formulation that each precursor ion was composed of different isobar and/or isomers. Evidently and for this reason, the herein presented work represents only the identification of the eleven VRL oligomers which were selected for CID-MS/MS. This should be considered as a starting point to uncover all structural features and the chemical diversity of the Oak VRLs mixture. For this reason, future work using a combination of ion mobility

and high-resolution mass spectrometer instruments such as FTICR or Orbitrap is needed to distinguish in order to separate and identify the closely overlapping isobars and/or isomers.

Similarly, these results showed that the tedious analysis of this kind of VRL mixtures contains a lot of valuable information that could be lost during lignin purification and degradation before doing the structural analysis.

For these reasons, we would like to reiterate that for lignomics structural elucidation, the analytical work should be completed on series of lignin oligomers released from the wood polysaccharide-lignin network, without any further purification or chemical modifications. This type of extracted lignin oligomers we have designated as virgin released lignin (VRL).[42]

In addition, it should be understood that analysis of lignin oligomer mixtures, whether industrially produced, degraded, and purified, cannot be accomplished by measuring G: H: S ratio of the lignol monomers. Especially by using antiquated methods such as pyrolysis GC-MS and by performing 1D- and 2D-NMR analyses, which do not take in account the incredible chemical diversity of lignins.

References

[1] Holladay, J.E.; White, J.F.; Bozell, J.J.; Johnson, D.: *Top value-added chemicals from biomass-Volume II—results of screening for potential candidates from biorefinery lignin (No. PNNL-16983)*. Pacific Northwest National Lab: Richland, WA, United States; **2007**.
[2] Vanholme, R.; Demedts, B.; Morreel, K.; Ralph, J.; Boerjan, W.: Lignin biosynthesis and structure. *Plant Physiol*. **2010**, *153*(3), 895-905.
[3] Doherty, W.O.; Mousavioun, P.; Fellows, CM.: Value-adding to cellulosic ethanol: lignin polymers. *Ind. Crop. Prod*. **2011**, *33*(2), 259-276.
[4] Strassberger, Z.; Tanase, S.; Rothenberg, G.: The pros and cons of lignin valorisation in an integrated biorefinery. *RSC Adv*. **2014**, *4*(48), 25310-25318.
[5] Vanholme, R.; Morreel, K.; Darrah, C.; Oyarce, P.; Grabber, J.H..; Ralph, J.; et al.: Metabolic engineering of novel lignin in biomass crops. *New Phytol.* **2012**, *196*(4), 978-1000.
[6] Lewis, N.G.; Davin, L.B.; Sarkanen, S.: Lignin and lignan biosynthesis: distinctions and reconciliations. *ACS Symp. Ser*. **1998**, *697*, 1-27.
[7] Moss, G.P.: Nomenclature of lignans and neolignans (IUPAC Recommendations 2000). *Pure Appl. Chem*. **2000**, *72*, 1493-1523.
[8] Weng, J.R.; Ko, H.H.; Yeh, T.L.; Lin, H.C.; Lin, C.N.: Two new arylnaphthalide lignans and anti-platelet, constituents from Justicia procumbens. *Arch. Pharm. Pharm. Med. Chem*. **2004**, *337*, 207-212.
[9] Hsu, H.Y.: *Treating Cancer with Chinese Herbs*. Oriental Healing Arts Institute: Los Angeles, **1982**; pp 238-250.
[10] Chen C.C.; Hsin, W.C.; Ko, F.N.; Huang, Y.L.; Ou, J.C.; Teng, C.M.: Antiplatelet arylnaphthalide lignans from Justicia procumbens. *J. Nat. Prod*. **1996**, *59*, 1149-1150.
[11] Chen C.C.; Hsin, W.C.; Huang, Y.L.: Six new diarylbutane lignans from Justicia procumbens. *J. Nat. Prod*. **1998**, *61*, 227-229.
[12] Fukamiya, N.; Lee, K.H.; Antitumor agents, 81. Justicidin-A and Diphyllin, two cytotoxic principles from Justicia procumbens. *J. Nat. Prod*. **1986**, *49*, 348-350.
[13] Ralph, J.; Peng, J.; Lu, F.; Hatfield, R.D.; Helm, R.F.: Are lignins optically active? *J. Agric. Food Chem*. **1999**, *47*(8), 2991-2996.
[14] Davin, L.B.; Lewis, N.G.: Lignin primary structures and dirigent sites. *Curr. Opin. Biotechnol*. **2005**, *16*(4), 407-415.
[15] Hatfield, R.; Vermerris, W.: Lignin formation in plants. The dilemma of linkage specificity. *Plant Physiol*. **2001**, *126*(4), 1351-1357.
[16] Freudenberg, K.; Chen, C.L.; Harkin, J.M.; Nimz, H.; Renner, H.: Observation on lignin. *Chem. Comm. (London)*. **1965**, *11*, 224-225.
[17] Mikhael, A.; Jurcic, K.; Fridgen, T.D.; Delmas, M.; Banoub J.: Matrix-assisted laser desorption/ionization time-of-flight/time-of-flight tandem mass spectrometry (negative ion mode) of French oak lignin: a novel series of lignin and tricin derivatives attached to carbohydrate and shikimic acid moieties. *Rapid Commun. Mass Spectrom*. **2020**, *34*(18):1-24.
[18] Banoub, J.H.; Delmas, M.: Structural elucidation of the wheat straw lignin polymer by atmospheric pressure chemical ionization tandem mass spectrometry and matrix-

assisted laser desorption/ionization time-of-flight mass spectrometry. *J. Mass Spectrom.* **2003**, *38*(8), 900-903.
[19] Banoub, J.H.; Benjelloun- Mlayah, B.; Ziarelli, F.; Joly, N.; Delmas, M.: Elucidation of the complex molecular structure of wheat straw lignin polymer by atmospheric pressure photoionization quadrupole time-of-flight tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **2007**, *21*(17), 2867-2888.
[20] Qi, Y.; Volmer, D.A.: Chemical diversity of lignin degradation products revealed by matrix-optimized MALDI mass spectrometry. *Anal. Bioanal. Chem.* **2019**, *411*(23), 6031-6037.
[21] Qi, Y.; Fu, P.; Li, S.; Ma, C.; Liu, C.; Volmer, D.A.: Assessment of molecular diversity of lignin products by various ionization techniques and high-resolution mass spectrometry. *Sci. Total Environ.* **2020**, *713*, 1-8.
[22] Lam, H.Q.; Le Bigot, Y.; Delmas, M.; Avignon, G.: A new procedure for the destructuring of vegetable matter at atmospheric pressure by a catalyst/solvent system of formic acid/acetic acid. Applied to the pulping of triticale straw. *Ind. Crop Prod.* **2001**, *14*, 139-144.
[23] Lan, W.; Yue, F.; Rencoret, J.; Del Río, J.C.; Boerjan, W.; Lu, F.; et al.: Elucidating tricin-lignin structures: assigning correlations in HSQC spectra of monocot lignins. *Polymers.* **2018**, *10*(8), 1-6.
[24] Jarrell, T.M.; Marcum, C.L.; Sheng, H.; Owen, B.C.; O'Lenick, C.J.; Maraun, H.; et al.: Characterization of organosolv switchgrass lignin by using high performance liquid chromatography/high resolution tandem mass spectrometry using hydroxide-doped negative-ion mode electrospray ionization. *Green Chem.* **2014**, *16*(5), 2713-2727.
[25] Prothmann, J.; Spégel, P.; Sandahl, M.; Turner, C.: Identification of lignin oligomers in Kraft lignin using ultra-high-performance liquid chromatography/high-resolution multiple-stage tandem mass spectrometry (UHPLC/HRMS^n). *Anal. Bioanal. Chem.* **2018**, *410*(29), 7803-7814.
[26] Butovich, I.A.: On the lipid composition of human meibum and tears: comparative analysis of nonpolar lipids. *Invest. Ophthalmol. Vis. Sci.* **2008**, *49*(9), 3779-3789.
[27] Hughey, C.A.; Hendrickson, C.L.; Rodgers, R.P.; Marshall, A.G.; Qian, K.:
Kendrick mass defect spectrum: a compact visual analysis for ultrahigh-resolution broadband mass spectra. *Anal. Chem.* **2001**, *73*(19), 4676-4681.
[28] Qi, Y.; Hempelmann, R.; Volmer, D.A.: Two-dimensional mass defect matrix plots for mapping genealogical links in mixtures of lignin depolymerisation products. *Anal. Bioanal. Chem.* **2016**, *408*(18), 4835-4843.
[29] Sleno, L.: The use of mass defect in modern mass spectrometry. *J. Mass Spectrom.* **2012**, *47*(2), 226-236.
[30] Sleno, L.; Volmer, D.A.; Marshall, A.G.: Assigning product ions from complex MS/MS spectra: the importance of mass uncertainty and resolving power. *J. Am. Soc. Mass Spectrom.* **2005**, *16*(2), 183-198.
[31] Demarque, D.P.; Crotti, A.E.; Vessecchi, R.; Lopes, J.L.; Lopes, N.P.:
Fragmentation reactions using electrospray ionization mass spectrometry: an important tool for the structural elucidation and characterization of synthetic and natural products. *Nat. Prod. Rep.* **2016**, *33*(3), 432-455.

[32] Saito, K.; Kato, T.; Tsuji, Y.; Fukushima, K.: Identifying the characteristic secondary ions of lignin polymer using ToF−SIMS. *Biomacromolecules*. **2005**, *6*(2):678-683.
[33] Chang, M.J.; Hung, T.M.; Min, B.S.; Kim, J.C.; Woo, M.H.; Choi, J.S.; et al.: Lignans from the fruits of *Forsythia suspensa* (Thunb.) Vahl protect high-density lipoprotein during oxidative stress. *Biosci. Biotechnol. Biochem*. **2008**, *72*(10), 2750-2755.
[34] Mikhael. A.; Fridgen, T.D.; Delmas, M.; Banoub, J.: Top-down lignomics analysis of the French pine lignin by atmospheric pressure photoionization quadrupole time-of-flight tandem mass spectrometry: identification of a novel series of lignin-carbohydrate complexes. *Rapid Commun. Mass Spectrom*. **2020**, *34*(22),1-13.
[35] Li, Q.; Yan, G.; Ge, T.: A fragmentation study of two compounds related to 4′-demethylepipodophyllotoxin in negative ion electrospray ionization by MS^n ion-trap time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom*. **2008**, *22*(3), 373-378.
[36] Krotulski, A.J.; Mohr, A.L.; Papsun, D.M.; Logan, B.K.: Dibutylone (bk-DMBDB): intoxications, quantitative confirmations and metabolism in authentic biological specimens. *J. Anal. Toxicol*. **2018**, *42*(7), 437-445.
[37] Dai, H.; Chen, Z.; Shang, B.; Chen, Q.: Identification and quantification of four anthraquinones in rhubarb and its preparations by gas chromatography–mass spectrometry. *J. Chromatogr. Sci*. **2018**, *56*(3), 195-201.
[38] Huang, J.; Fu, S.; Gan, L.: *Lignin Chemistry and Applications*. Elsevier: Amsterdam, Netherland, **2019**; pp 51-78.
[39] Li, N.; Wu, J.L.; Sakai, J.I.; Ando, M.: Dibenzylbutyrolactone and dibenzylbutanediol lignans from *Peperomia duclouxii*. *J. Nat. Prod*. **2003**, *66*(11), 1421-1426.
[40] Asare, S.O.; Lynn, B.C.: A comparative study of the electrospray ionization response of β-O-4′ lignin model compounds. *J. Mass Spectrom*. **2019**, *54*(6), 540-548.
[41] Du, L.; White, R.L.: Improved partition equilibrium model for predicting analyte response in electrospray ionization mass spectrometry. *J. Mass Spectrom*. **2009**, *44*(2), 222-229.
[42] Banoub, J.; Delmas, G.H. Jr.; Joly, N.; Mackenzie, G.; Cachet, N.; Benjelloun-Mlayah, B.; et al.: A critique on the structural analysis of lignins and application of novel tandem mass spectrometric strategies to determine lignin sequencing. *J. Mass Spectrom*. **2015**, *50*(1), 5-48.

Chapter 6. Summary and Future work

6.1. Summary

In this work, the molecular structure of the *lycopodium clavatum* sporopollenin was demystified and unraveled using secondary ion mass spectrometry (SIMS) and matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS) coupled with tandem mass spectrometry (MS/MS). These mass spectrometric techniques allowed the identification of the two major building blocks of the sporopollenin network. The first building unit is the crystalline liquid-like poly(hydroxyacid) network, while the second unit is the cyclic rigid backbone composed of polyhydroxylated tetraketide- like monomers.

The hard ionization technique SIMS-TOF-MS and high energy CID-MS/MS allowed the identification of the sporopollenin C-14 and C-16 fatty acid monomers and the glycerol core unit. Additionally, it helped in the identification of the beta diketone moiety in the identified fatty acids chains. After unraveling the small monomers of sporopollenin by the SIMS study, the soft ionization technique MALDI was used to identify larger fragments of sporopollenin compared to those identified by the SIMS study.

The use of different MALDI matrices helped greatly in unraveling the two major building blocks of sporopollenin. The use of 1,5-diaminonaphthalene (DAN) and α-cyano-4-hydroxycinnamic acid (CHCA) allowed the identification of the polyhydroxy acid network, while the use of the 2-(4-hydroxyphenylazo)benzoic acid (HABA) allowed the identification of the macrocyclic rigid backbone composed of the polyhydroxylated tetraketide-like monomers. Moreover, the identified sporopollenin units were supported

by several solid-state NMR studies. The most important one was the ^{1}H NMR consisted of two overlapped NMR spectra (sharp and broad spectra), supporting that the sporopollenin shell comprises two different building blocks.

It should be mentioned that the cyclic rigid backbone of sporopollenin contains the conjugated heterocyclic alpha pyrone rings, which was supported by the solid-state ^{13}C-NMR. However, high-resolution X-ray photoelectron spectroscopy (HR-XPS) indicates the absence of aromaticity and/or conjugation in the sporopollenin structure. The NMR and XPS results appear to be contradicting each other's; however, it indicates that the alpha pyrone rings are pseudo and/or weak aromatic compounds that are not capable of showing any pi-pi* shake-up aromatic peaks in XPS. Also, it should be mentioned that alpha pyrone rings can undergo chemical reactions that are not typical for aromatic compounds, such as Diels-Alder reaction, electrophilic additions, and ring openings. [1]

The second topic in this thesis is the structural elucidation of virgin released lignins (VRLs). These VRLs were extracted by one of the finest extraction methods using a simple mixture of acetic acid, formic acid, and water. This method was developed by the CIMV French company. [2] The VRLs structure is considered to be close enough to the native lignin structure. The analysis of these VRLs using different ionization techniques (Chapters 3-5) demonstrated the chemical diversity and the complexity of the analyzed complex lignin mixtures.

In Chapter 3, negative ion mode MALDI-TOF/TOF-MS/MS was used to investigate the major structural features of the French oak virgin released lignin (hardwood). Different classes of compounds and/or derivatives were identified in this

work, such as flavonoids (mainly tricin), lignans, flavonolignans, and several lignin derivatives abundant in syringyl units. In addition, most of these identified compounds were attached to carbohydrates and/or shikimic acid moieties.

In Chapter 4, positive ion mode APPI-QqTOF-MS and MS/MS were used to investigate the structure of the French pine virgin released lignin (softwood). This analysis showed that pine lignin contains different classes of lignin-carbohydrate complexes (LCCs). Fifteen novel LCCs were identified beside the identification of the simple monomer acetylated eugenol. In general, the identified LCCs were composed of one or two carbohydrate moieties attached to one or two or three lignin monomers.

In Chapter 5, positive ion mode APPI and ESI-QqTOF-MS and MS/MS were used to investigate more structural features of the French oak virgin released lignins. This chapter represents more information about the oak lignin structure besides what was concluded from the MALDI-TOF/TOF analysis (Chapter 3). In this study, ten novel lignans were identified in addition to the neolignane cedrusin. These ten novel lignans were eight aryl tetralin lignans, one lignan-dehydroshikimic acid complex, and a novel lignan trimer.

The identification of all these novel lignin derivatives (Chapter 3-5) was not a straightforward task. The overlapping complex isotopic distributions and centroid mass shifts extremely complicated the identification of the unknown lignin ions. Consequently, the use of the powerful tandem mass spectrometry was essential, which allowed the report of all these tentatively assigned lignin formulas and/or structures. It is important to note that some complex product ion scans were called messy MS/MS (Chapters 4 and 5),

indicating that the selected precursor ion for MS/MS is not a single ion but a mixture of isobars and/or isomers, which clearly demonstrates the complexity of the lignin mixture.

The identified novel lignin compounds (Chapters 3-5) gave new perspectives on lignin biosynthesis and its native structure. For example, the identification of tricin and other flavonoids in the oak VRL (Chapter 3) may reveal a relationship between lignin and flavonoids biosynthesis. Also, the detection of the flavonoids and/or lignin – shikimic (Chapter 3) indicates that shikimic acid derivatives may have a different role in the lignin biosynthesis. Especially that shikimic acid was only described to exist as an intermediate in the lignin biosynthesis, and it was never described before to exist in the final lignin network. [3] Moreover, the identification of lignans in the oak VRL (chapters 5) indicates that there may be a relationship between lignans and lignin biosynthesis. Especially lignans are known to be synthesized by proteins, while lignins are known to be formed by random oxidative polymerization.[4] Lastly, identifying various lignin carbohydrate complexes in the pine VRL (Chapter 4) suggests the great association between lignin and carbohydrates in the wood network. All these previous examples show that there is a wealth of scientific information hidden inside the VRLs, which can be lost if these VRLs were subjected to further purifications and/or degradations.

6.2. Future Work

Although, this work offers novel details about the sporopollenin structure that was a mystery for almost a century! However, future projects should be established to find the whole sporopollenin shell structure. One of these projects is to use molecular dynamics simulations to build a hypothetical complete sporopollenin shell using the major building

blocks identified in this thesis. [5] This can help in visualizing how the sporopollenin chains and/or units are connected. Also, it can help in deciphering how the chains interact together to form this extremely stable sporopollenin shell. These interactions are expected to be intramolecular and/or intermolecular hydrogen bonding due to the presence of the beta diketone moiety explained in this work.

Another possible future project is to use the same analytical strategies and/or techniques described in Chapter 2 to investigate different sporopollenin species worldwide. This is necessary to examine if other sporopollenin species have different chemical structures. Moreover, the investigation of the sporopollenin structure with various MALDI matrices may unravel other structural features in sporopollenin.

Regarding lignin, as mentioned before, the complexity of the lignin samples was demonstrated clearly in chapters 4 and 5 by assuming that some selected precursor ions are existing as a mixture of isomers and sometimes isobars. For this reason, future work using ion mobility coupled with high-resolution mass analyzers such as orbitrap or FTICR is essential to help in resolving and/or separating this complex mixture of isomers and isobars. The use of ion mobility is essential to resolve isomass isomers. [6] This isomers separation can help in the simplification of the resulting product ion scans and enhances confidence in the assignment of chemical formulas and/or structures.

The “do-it-by yourself” approach in assigning chemical formulas and/or structures to unknown ions could be sometimes tedious and challenging; however, it opens a way for discovering novel compounds that are not present in any available compound libraries. Despite the effectiveness of the latter approach, we plan to support the complex MS data

interpretation with one-dimensional and/or two-dimensional Kendrick mass defect plots. [7,8] These plots are useful in sorting and classifying closely related lignin ions (same class or family). For example, KMD plots can align all ions (same class or family) that differ in the number of CH_2 units on the same horizontal line. In that case, if the right structure of one of these related ions is identified and supported by CID-MS/MS, all other ions formulas can be deduced without the need to do more MS/MS. [7,8] Therefore, these KMD plots will help in getting general information about the lignin mixture compounds classes and/or types rather than studying each individual ion separately.

References

[1] Joule, J.A.; Mills, K.: *Heterocyclic Chemistry*; John Wiley and Sons: Chichester, West Sussex, UK; **2010**.
[2] Lam, H.Q.; Le Bigot, Y.; Delmas, M.; Avignon, G.: A new procedure for the destructuring of vegetable matter at atmospheric pressure by a catalyst/solvent system of formic acid/acetic acid. Applied to the pulping of triticale straw. *Ind. Crops Prod*. **2001**, *14*, 139-144.
[3] Vanholme, R.; Cesarino, I.; Rataj, K.; Xiao, Y.; Sundin, L.; Goeminne, G.; et al.: Caffeoyl shikimate esterase (CSE) is an enzyme in the lignin biosynthetic pathway in Arabidopsis. *Science*. **2013**, 341(6150), 1103-1106.
[4] Davin, L.B.; Lewis, N.G.: Lignin primary structures and dirigent sites. *Curr. Opin. Biotechnol*. **2005**, *16*(4), 407-415.
[5] Karplus, M.; McCammon, J.A.: Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **2002**, *9*(9), 646-652.
[6] Lanucara, F.; Holman, S.W.; Gray, C.J.; Eyers, C.E.: The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics. *Nat. Chem.* **2014**, *6*(4), 281-294.
[7] Qi, Y.; Fu, P.; Li, S.; Ma, C.; Liu, C.; Volmer, D. A.: Assessment of molecular diversity of lignin products by various ionization techniques and high-resolution mass spectrometry. *Sci. Total Environ.* **2020**, *713*, 136573.
[8] Qi, Y.; Hempelmann, R.; Volmer, D. A.: Two-dimensional mass defect matrix plots for mapping genealogical links in mixtures of lignin depolymerisation products. *Anal. Bioanal. Chem.* **2016**, *408*(18), 4835-4843.

Appendix A. Supporting Information for Chapter 2

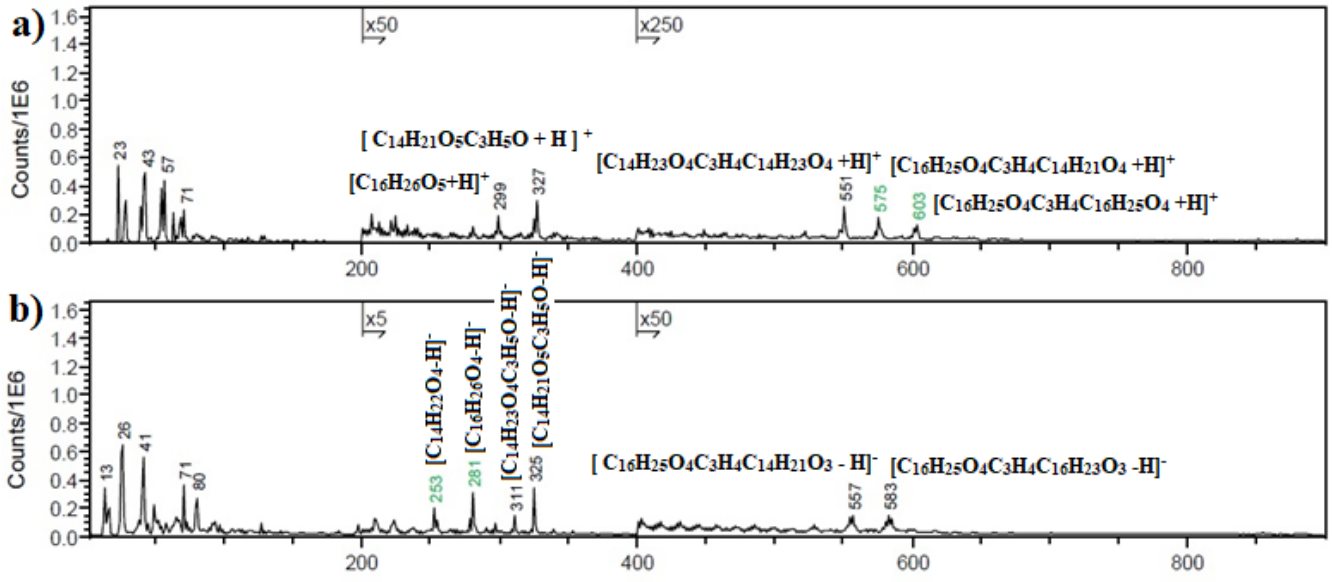


Figure A1.1. TOF-SIMS of *Lycopodium Clavatum* sporopollenin a) positive ion mode b) negative ion mode. The highlighted peaks are identified by parallel imaging MS/MS.

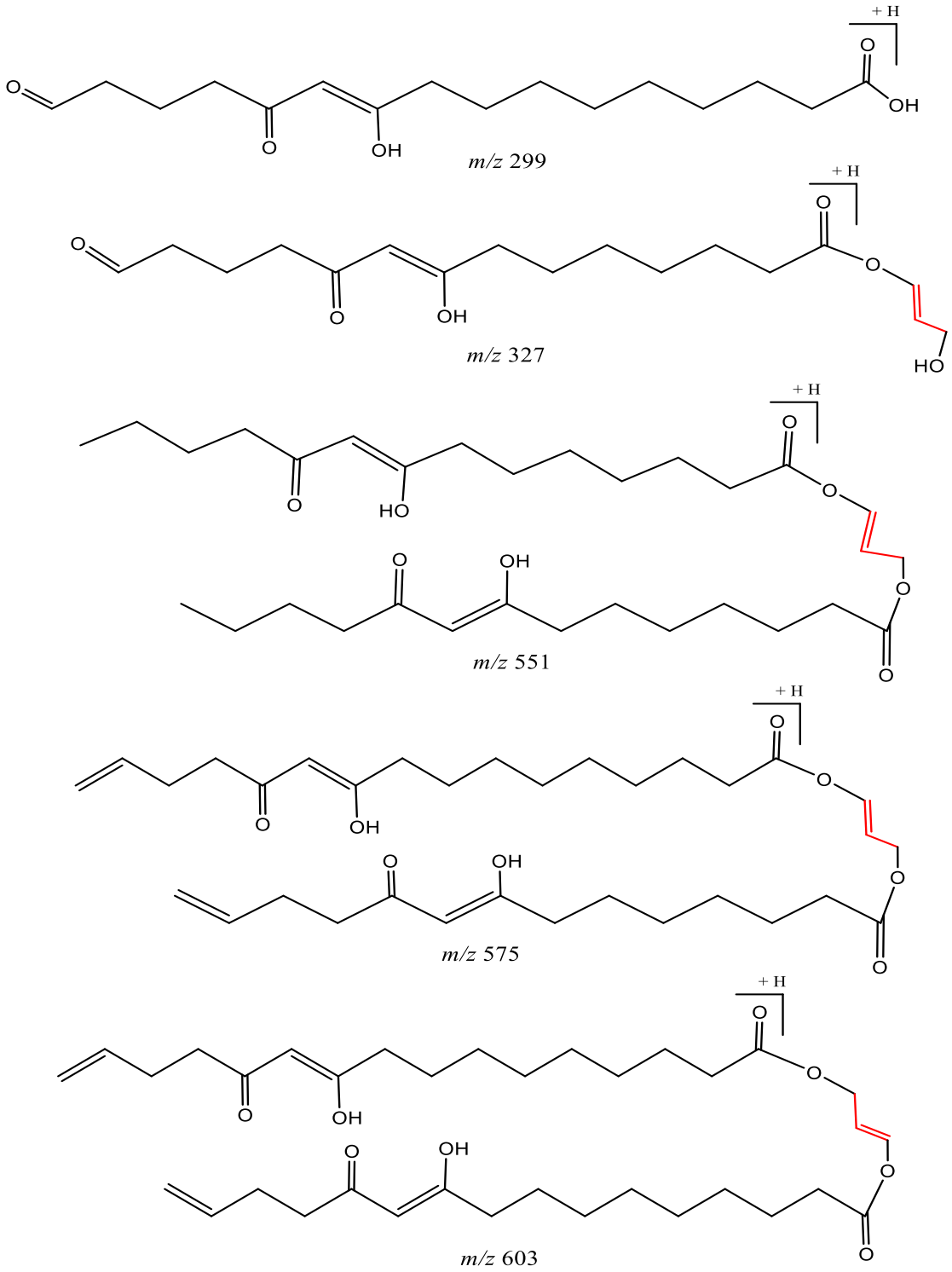


Figure A1.2. The proposed structures of the ions identified in the positive ion mode TOF-SIMS (glycerol carbons are in red colour)

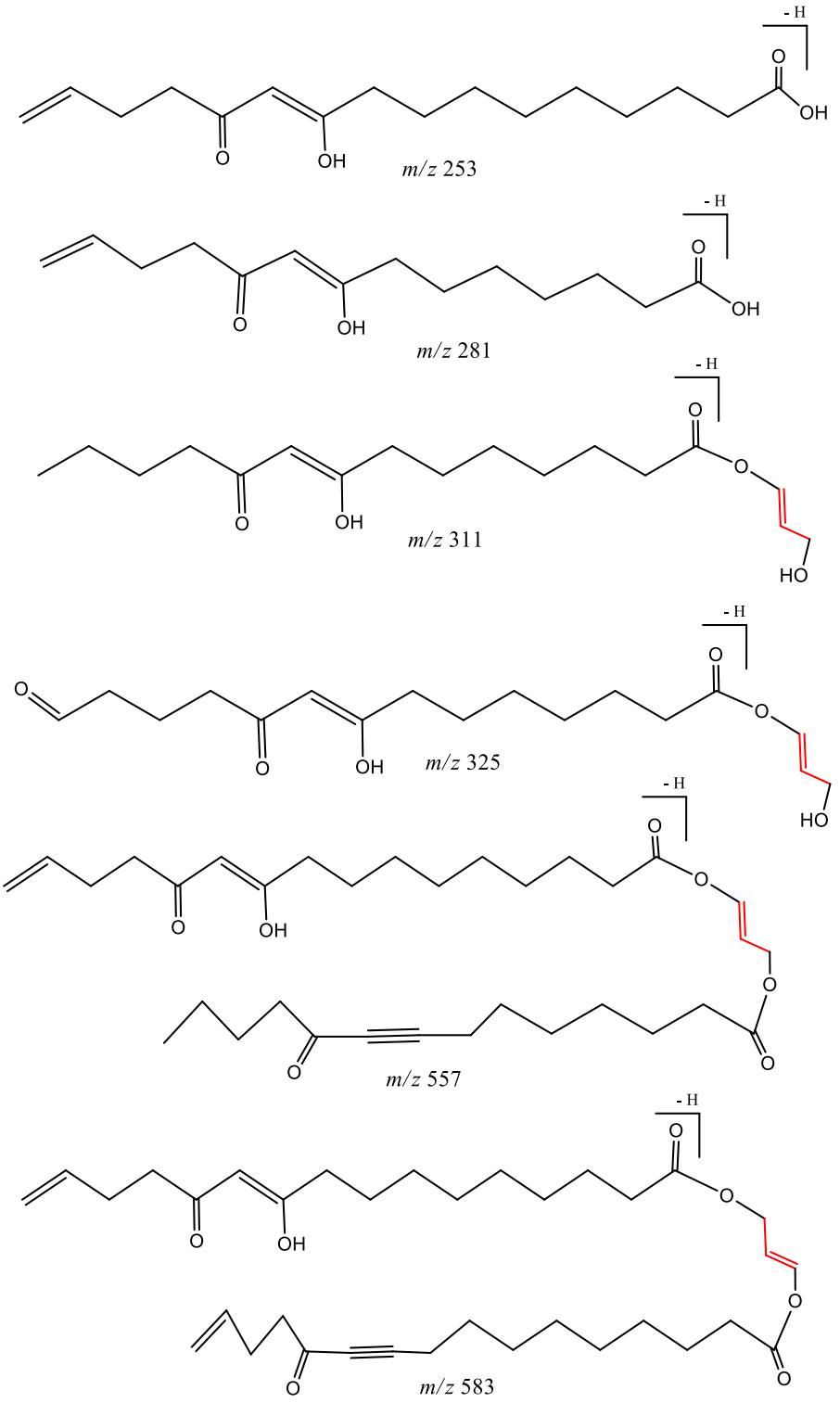


Figure A1.3. The proposed structures of the ions identified in the negative ion mode TOF-SIMS (glycerol carbons are in red colour)

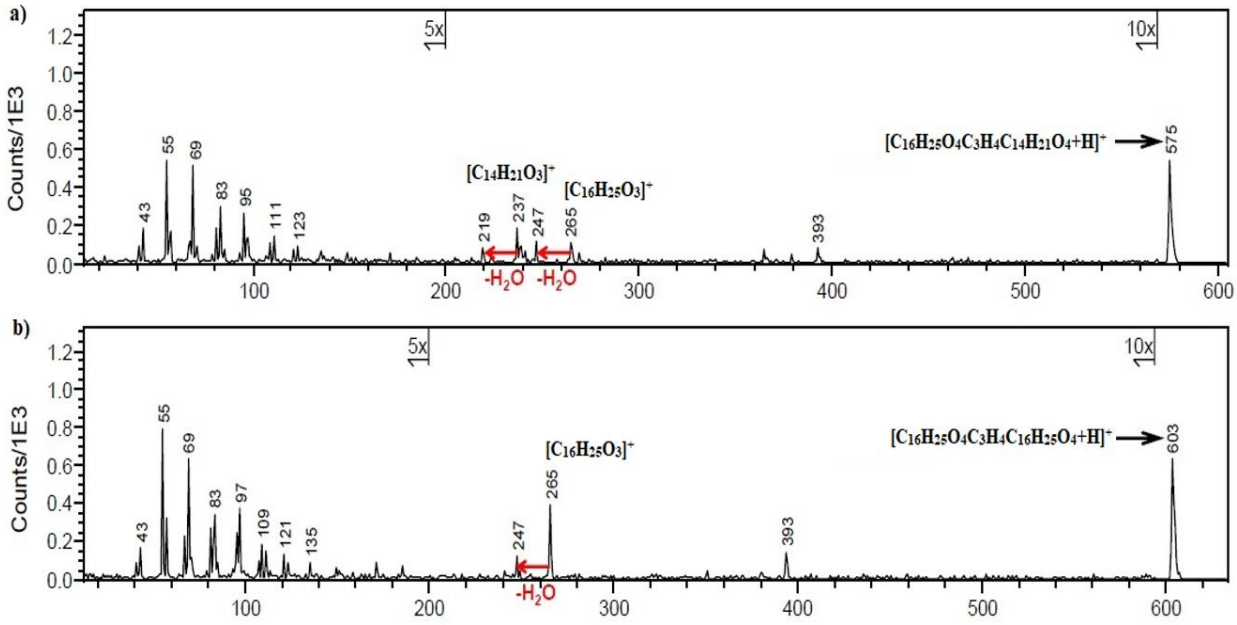
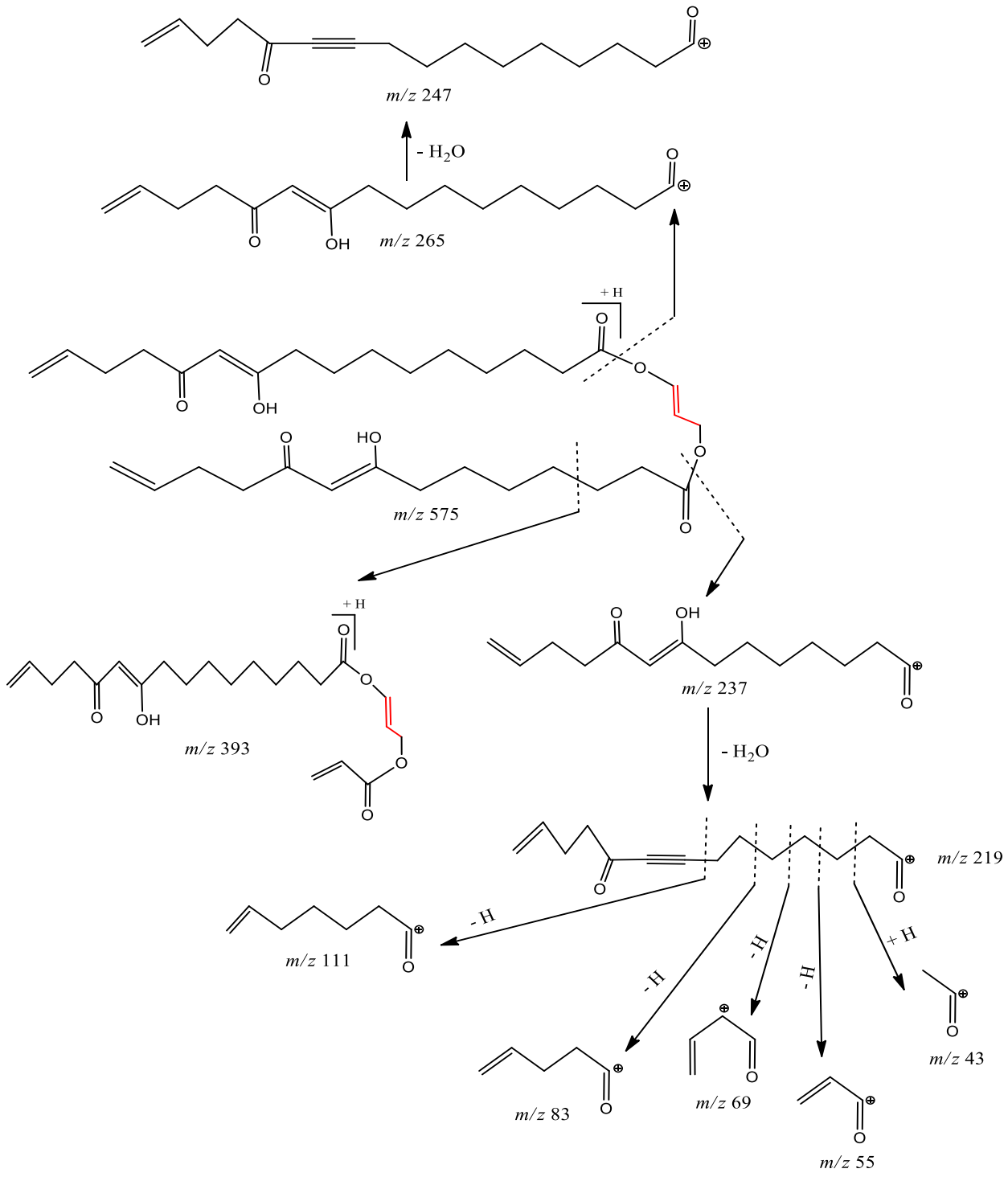
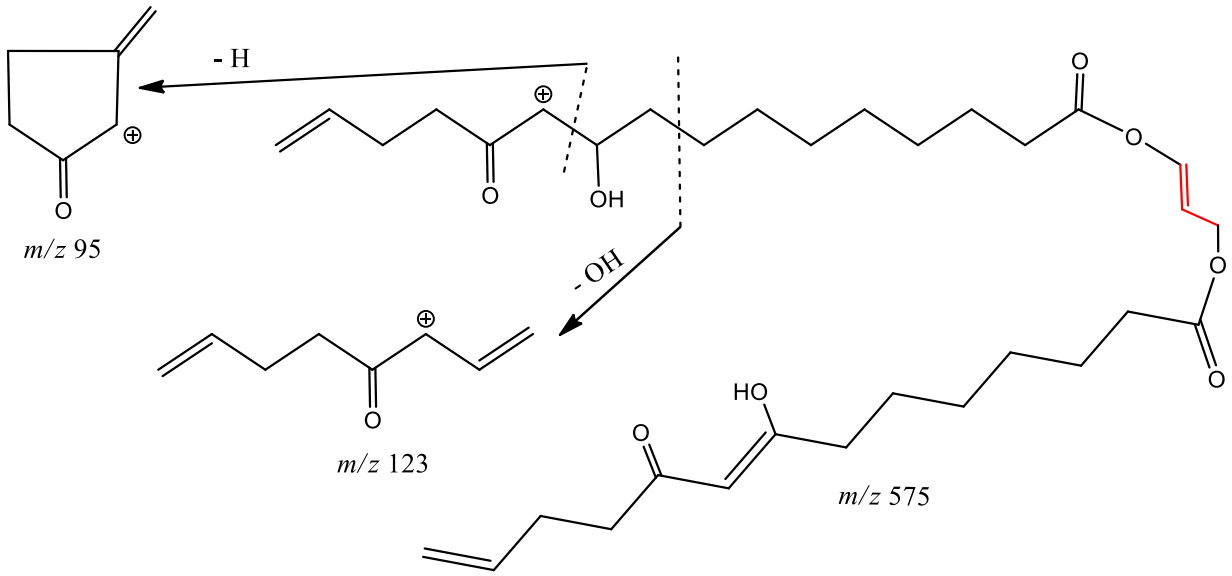


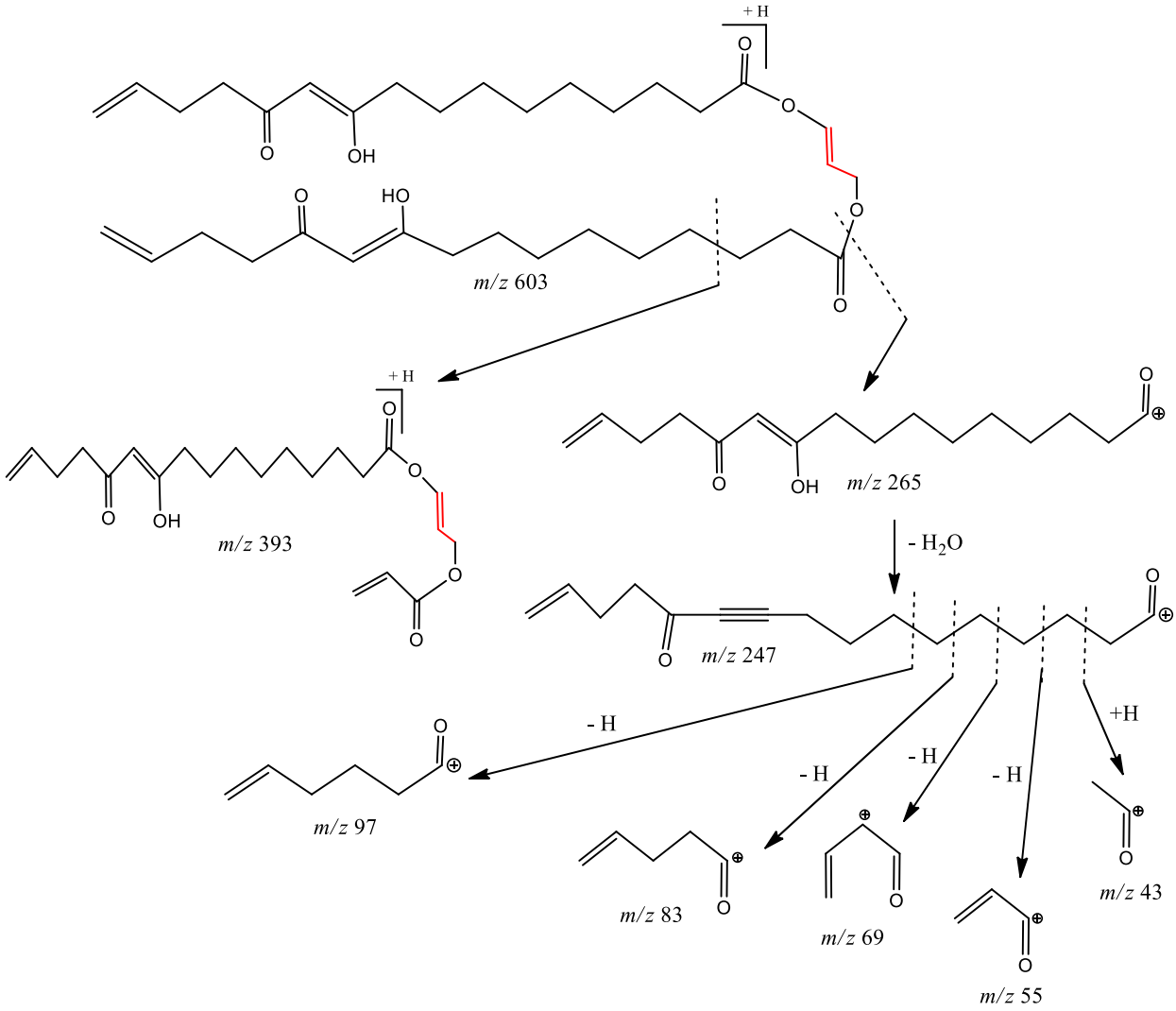
Figure A1.4. a) Product ion scan of the precursor ion at *m/z* 575 identified in positive ion mode TOF-SIMS. b) Product ion scan of the precursor ion at *m/z* 603 identified in positive ion mode TOF-SIMS



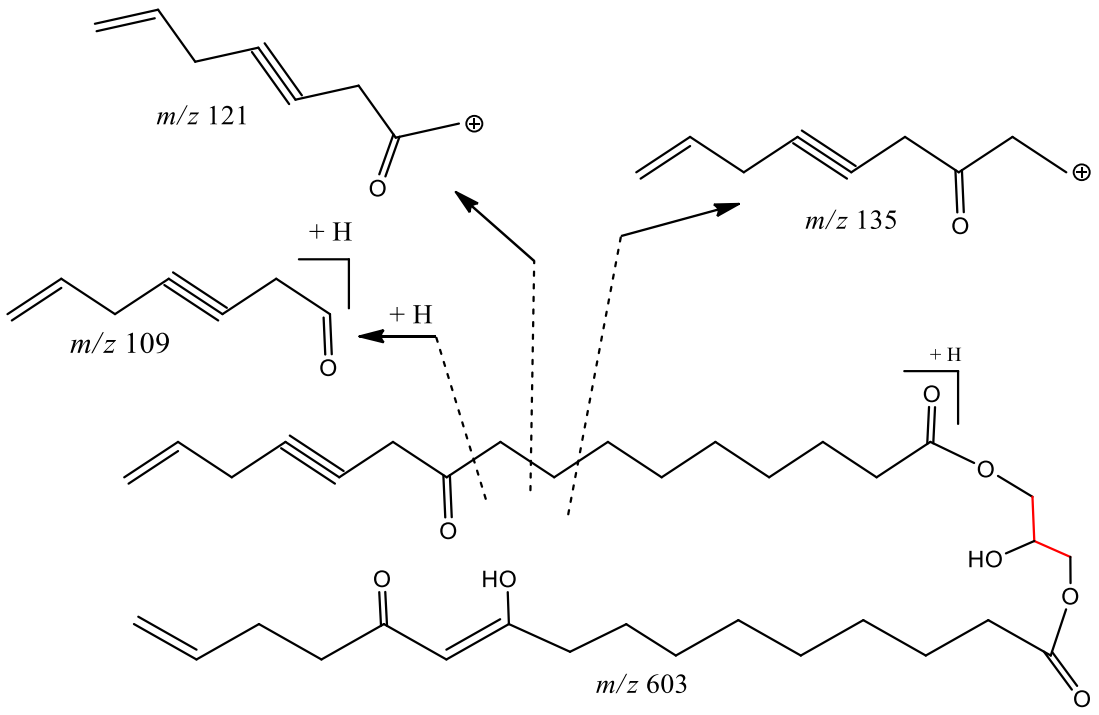
Scheme A1.1a. Proposed MS/MS fragmentation mechanism of the precursor ion at *m/z* 575



Scheme A1.1b. Proposed MS/MS fragmentation mechanism of the precursor ion at *m/z* 575. Note: one of the double bonds in the precursor ion at *m/z* 575 is protonated to give this positively charged ion



Scheme A1.2a. proposed MS/MS fragmentation mechanism of the precursor ion at *m/z* 603



Scheme A1.2b. proposed MS/MS fragmentation mechanism of the precursor ion at *m/z* 603. Note: the precursor ion at *m/z* 603 is represented here by a different isomer to account for the formation of the ions at *m/z* 109, 121 and 135.

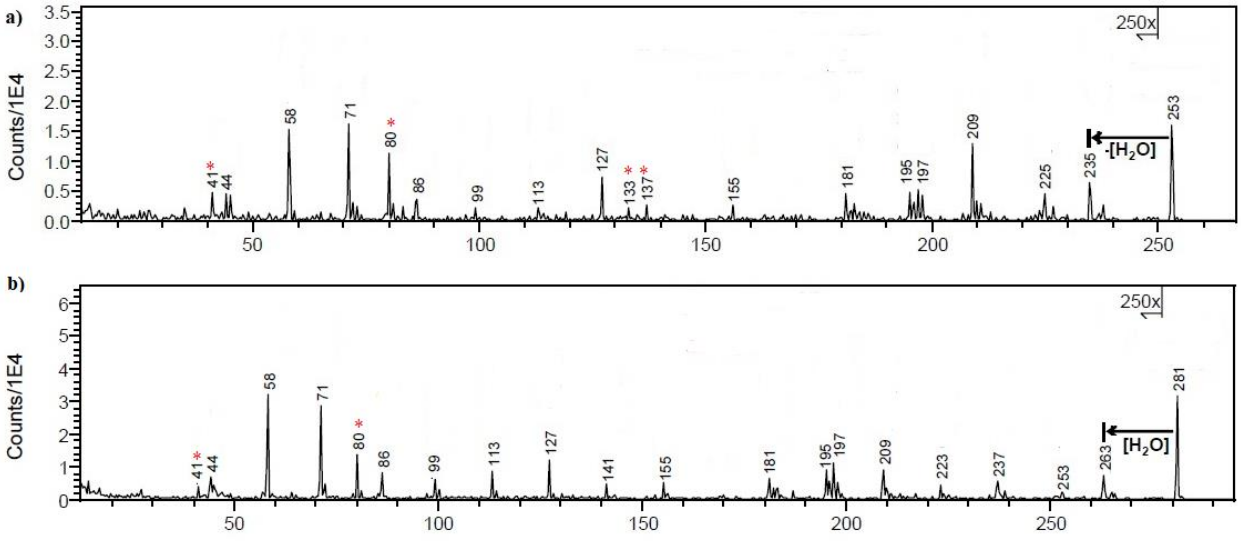
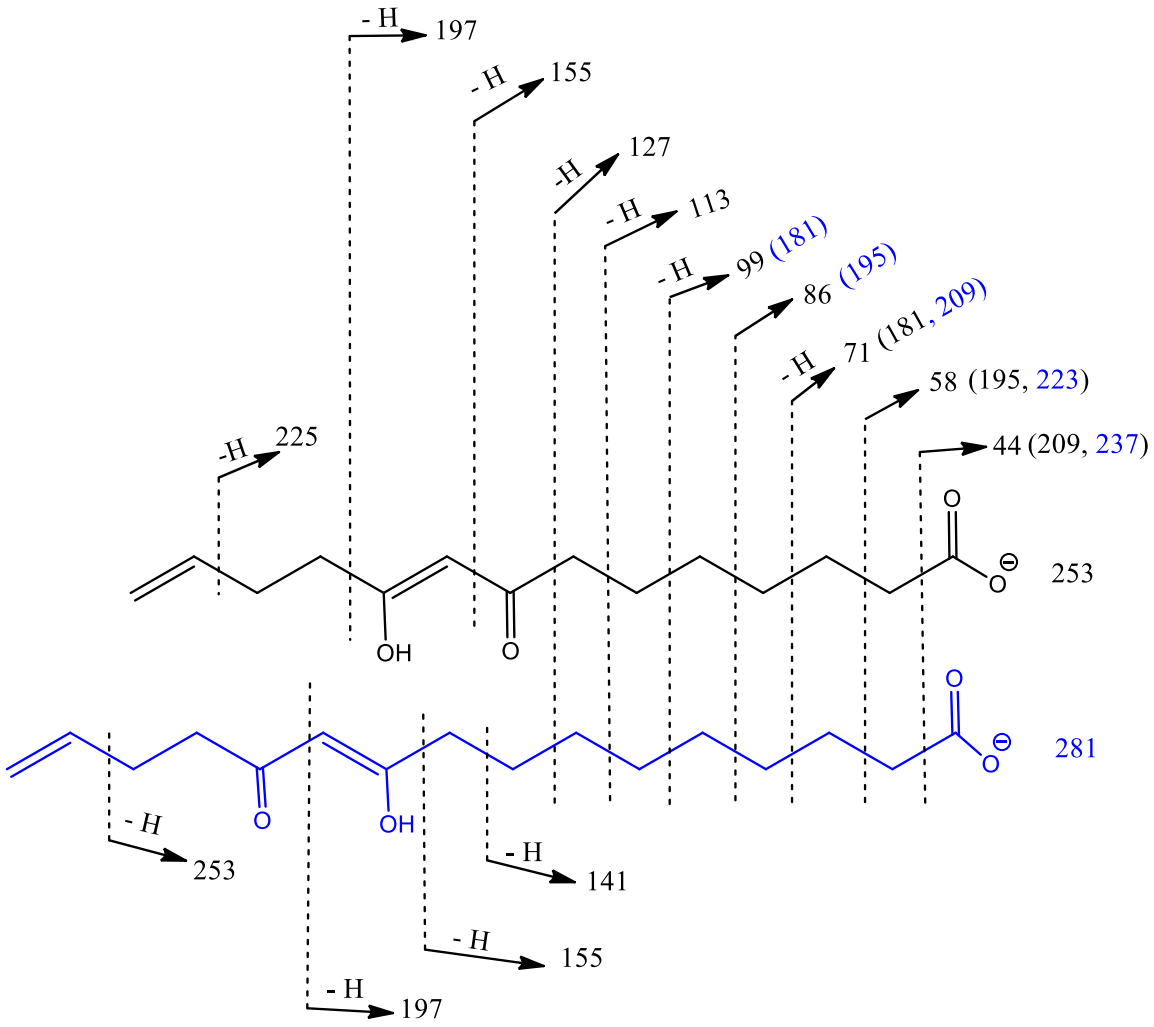
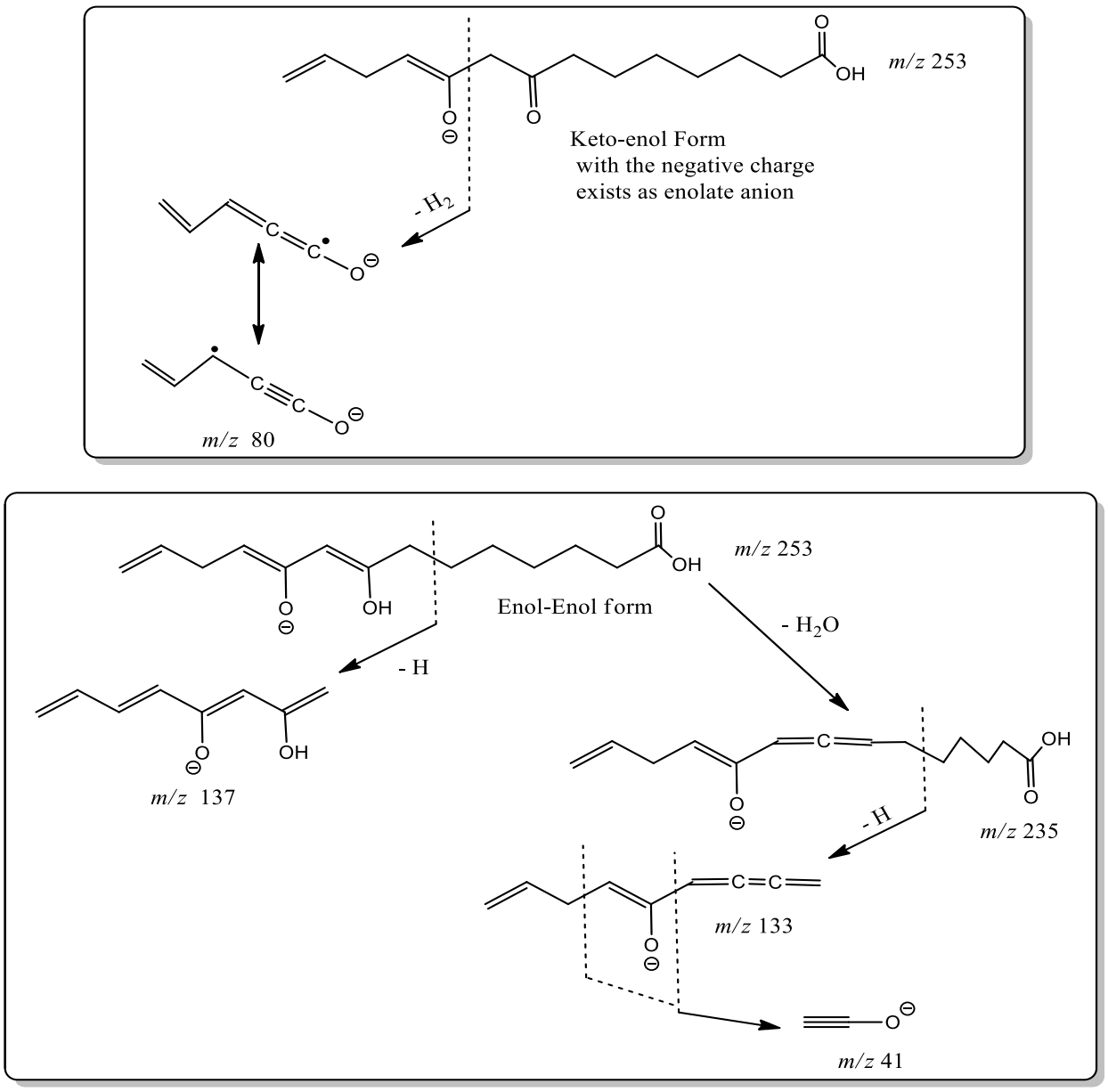


Figure A1.5. a) Product ion scan of the precursor ion at *m/z* 253 identified in Negative ion mode TOF-SIMS. b) Product ion scan of the precursor ion at *m/z* 281 identified in Negative ion mode TOF-SIMS



Scheme A1.3a. The MS/MS Fragmentation pathways of the Ketol fatty acids identified at *m/z* 253 and *m/z* 281. **Note**: The numbers between brackets represent the complementary fragments, for example, ion at *m/z* 44 (209, 237) means that product ion at *m/z* 209 is formed after the loss of 44 from precursor ion at *m/z* 253 and product ion at *m/z* 237 is formed after the loss of 44 from the precursor ion at *m/z* 281.



Scheme A1.3b. The proposed formation of the unusual ions at *m/z* 41, 80, 133, and 137, which are diagnostic and support the proposed structures of the precursor ions at *m/z* 253 and 281. In order to demonstrate the formation of the ion at *m/z* 80, we consider that the negative charge can exist as an enolate anion instead of carboxylate anion. Also, to describe the low intensity ion at *m/z* 133 and 137, we consider that the beta-diketone moiety can exist as enol-enol form (This will be described later in the Figure A1.6 describing the 8 possible forms of the beta-diketone moiety). Lastly, it shows an example of how one water molecule can be lost from the precursor ion at *m/z* 253 supporting that one hydroxyl group exist in the enol form and the other in the keto form.

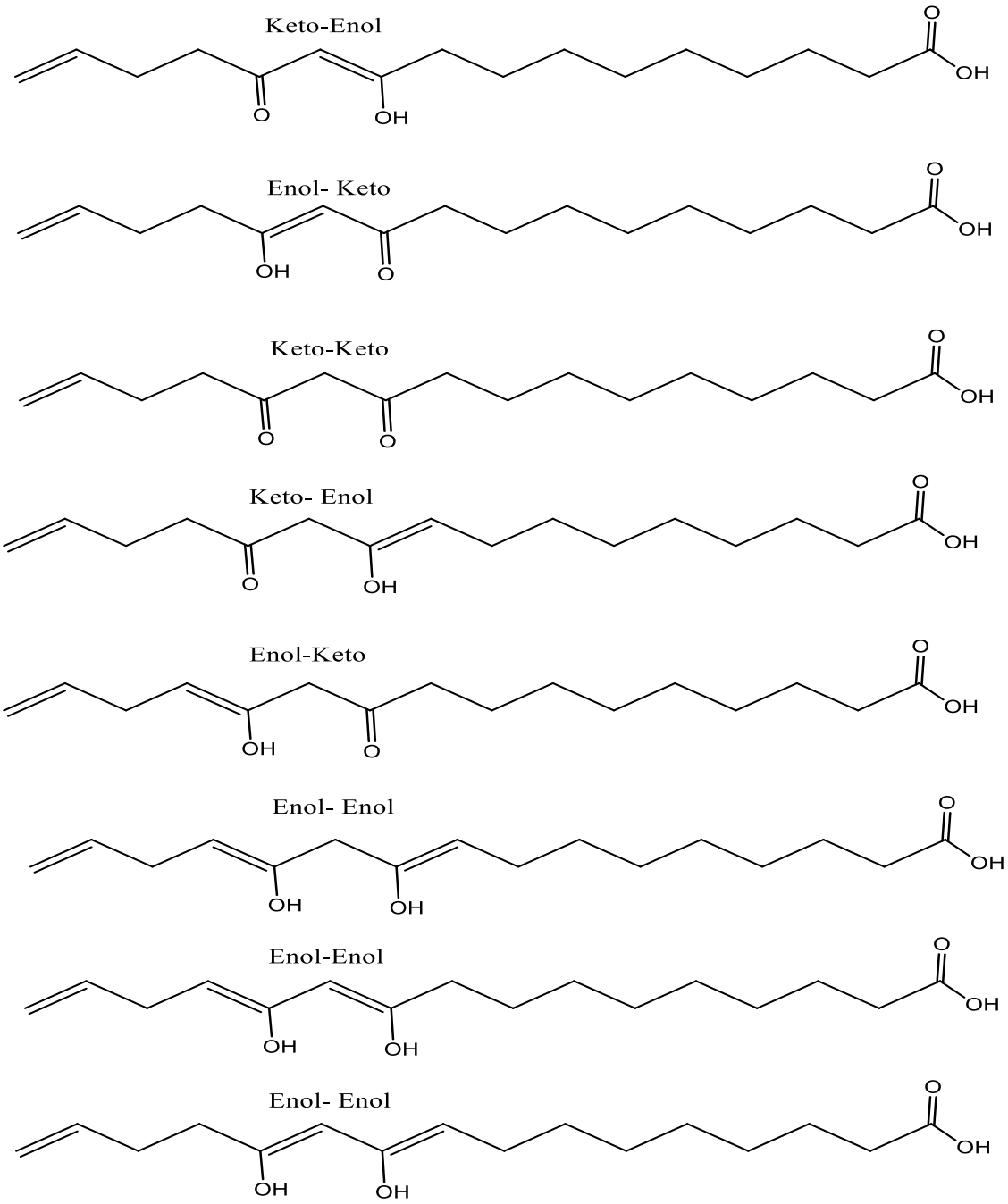


Figure A1.6. Eight different forms of the beta diketone moiety present in the fatty acid monomers identified by using TOF-SIMS-MS. The question that should be addressed in the future. Does sporopollenin contain all these forms? Do these forms have a role in intermolecular hydrogen bonding between sporopollenin chains giving explaining part of the sporopollenin strength?

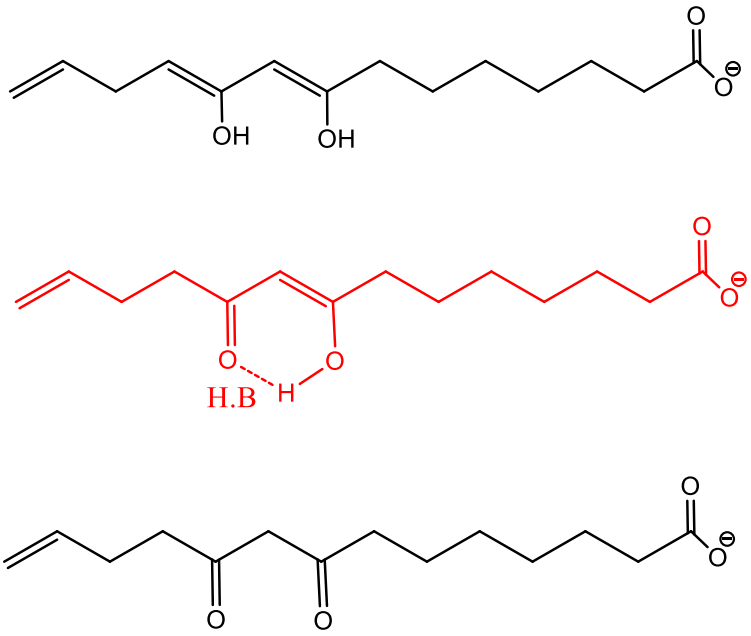


Figure A1.7. Different forms of beta-diketone moiety in the identified carboxylic acid using TOF-SIMS-MS/MS. The middle structure is most probably the dominant form due to the formation of intramolecular hydrogen bond and can be account for why only one water molecule is lost instead of two.

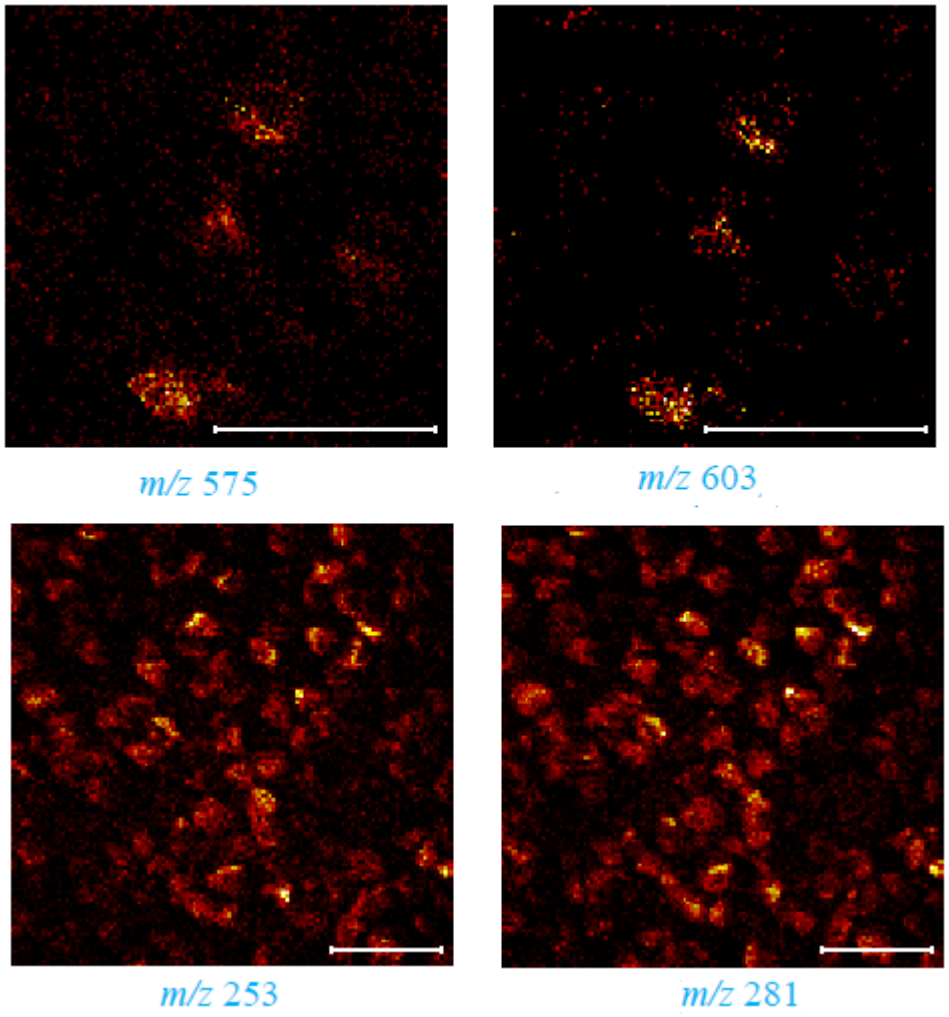


Figure A1.8. Imaging of chemical distributions of ions at *m/z* 253 and 281 (-ve ion mode), and ions at *m/z* 575 and 603 (+ve ion mode). Note: markers are 100 micrometers

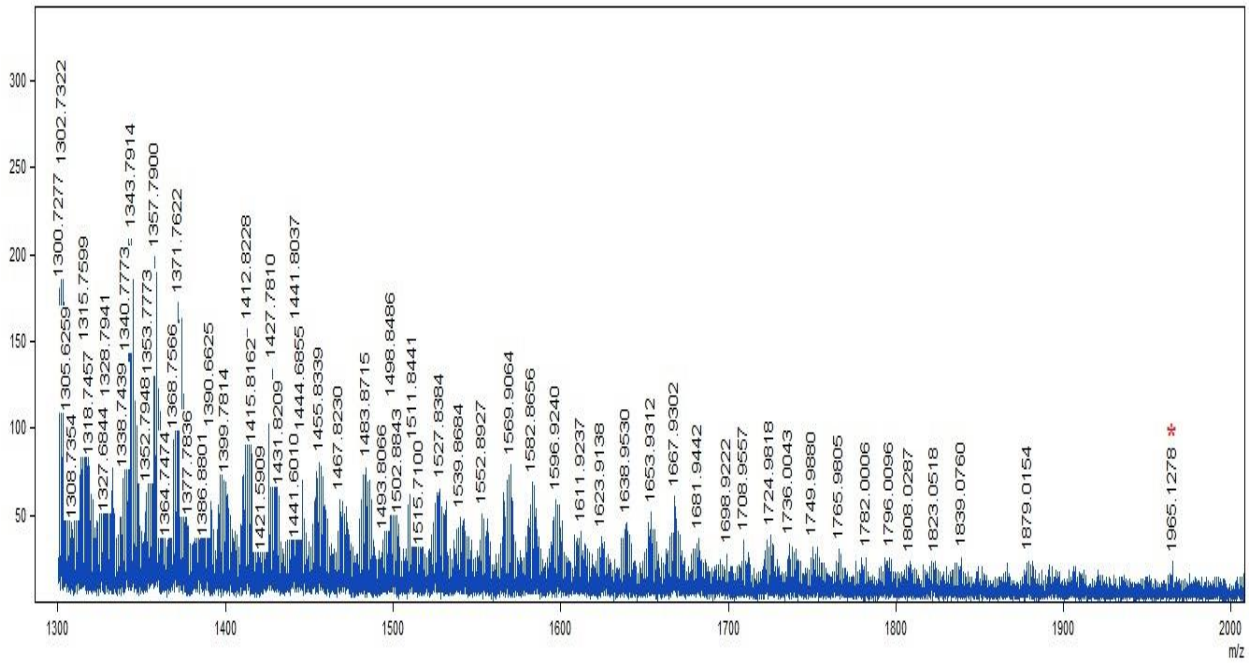


Figure A1.9a. MALDI-MS of sporopollenin using DAN as a matrix in the range of 1300-2000

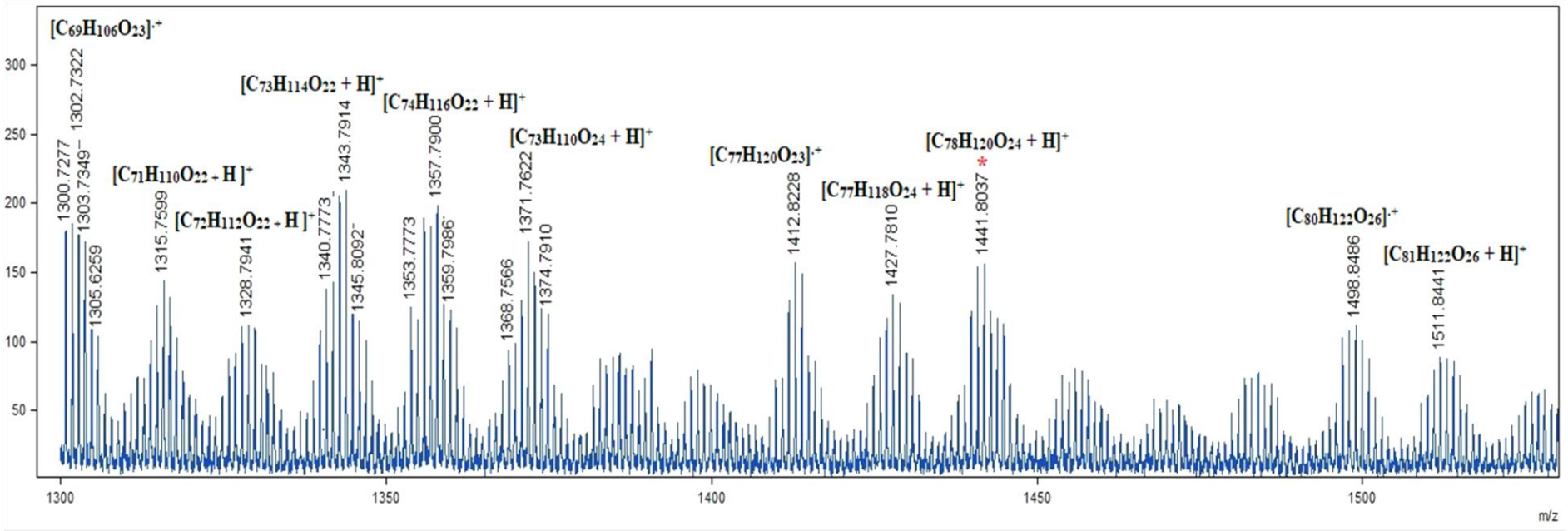


Figure A1.9b. expansion of Sporopollenin MALDI MS using DAN as a matrix in the range of *m/z* 1300-1500

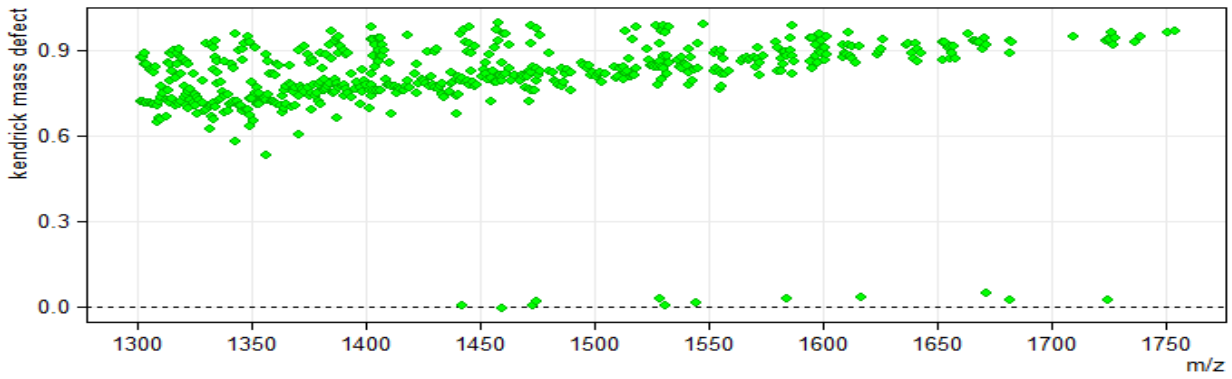


Figure A1.10. Kendrick mass defect plot for better visualization of complex Sporopollenin MALDI-MS using DAN is showing a bundle of ions that are closely related to each other. All ions on the same horizontal ion (same mass defect) differs by the number of CH2 while ions on the same vertical line have the same number of CH_2 but differ in the number of oxygens and/or DBE that created a higher mass defect. Conditions for the creation of this plot (Int. Threshold was set to 10% plus removing isotopes)

Table A1.1. Chemical formulas of some selected ions in the sporopollenin MALDI-MS using DAN as a matrix. All ions almost fit with the C/H (0.63) , C/O (3.33) and H/O (5.26) ratios of the empirical formula of *Lycopodium clavatum* sporopollenin

Experimental *m/z*	Assigned Chemical Formula	Error ppm	C/H (0.63)	C/O (3.33)	H/O (5.26)
1302.7322	$[C_{69}H_{106}O_{23}]^{+\bullet}$	15.2	0.65	3.00	4.60
1315.7599	$[C_{71}H_{110}O_{22} + H]^{+}$	2.4	0.64	3.22	5.00
1328.7941	$[C_{72}H_{112}O_{22}]^{+\bullet}$	22.2	0.64	3.27	5.09
1343.7914	$[C_{73}H_{114}O_{22}+ H]^{+}$	2.5	0.64	3.32	5.18
1357.7900	$[C_{74}H_{116}O_{22}+ H]^{+}$	10.0	0.64	3.36	5.27
1371.7622	$[C_{73}H_{110}O_{24}+ H]^{+}$	11.4	0.66	3.04	4.58
1412.8228	$[C_{77}H_{120}O_{23}]^{+\bullet}$	-19.6	0.65	3.34	5.2
1427.7810	$[C_{77}H_{118}O_{24}+ H]^{+}$	0.5	0.65	3.20	4.92
1441.8037	$[C_{78}H_{120}O_{24}+ H]^{+}$	-14.5	0.65	3.08	5.00
1498.8486	$[C_{80}H_{122}O_{26}]^{+\bullet}$	17.4	0.65	3.08	4.69
1511.8441	$[C_{82}H_{126}O_{25}+ H]^{+}$	-14.8	0.65	3.28	5.04
1965.1278	$[C_{106}H_{164}O_{33}]^{+\bullet}$	6.3	0.64	3.21	4.96

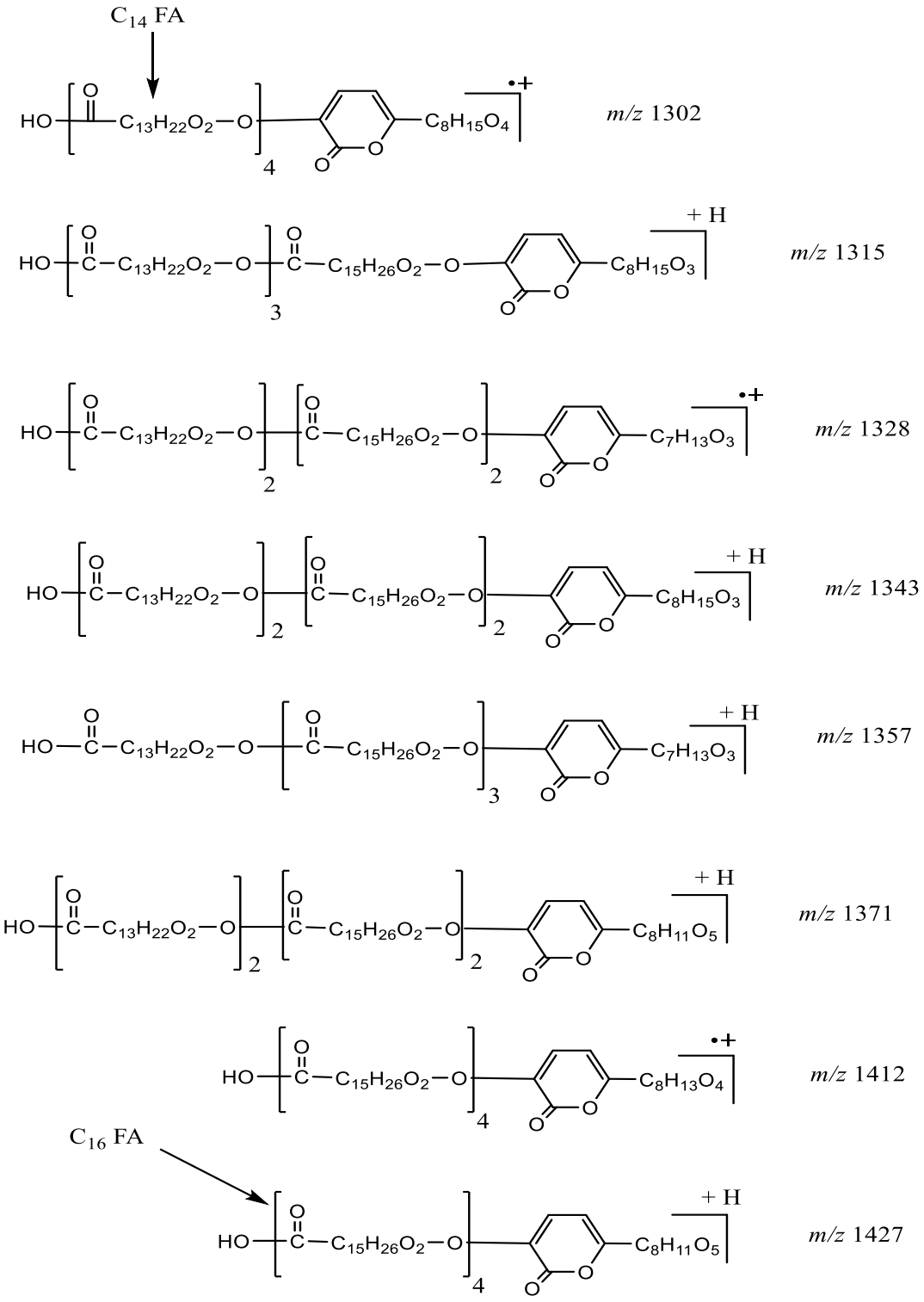


Figure A1.11a. Possible chemical composition of some selected ion in the Sporopollenin MALDI-MS using DAN as a matrix

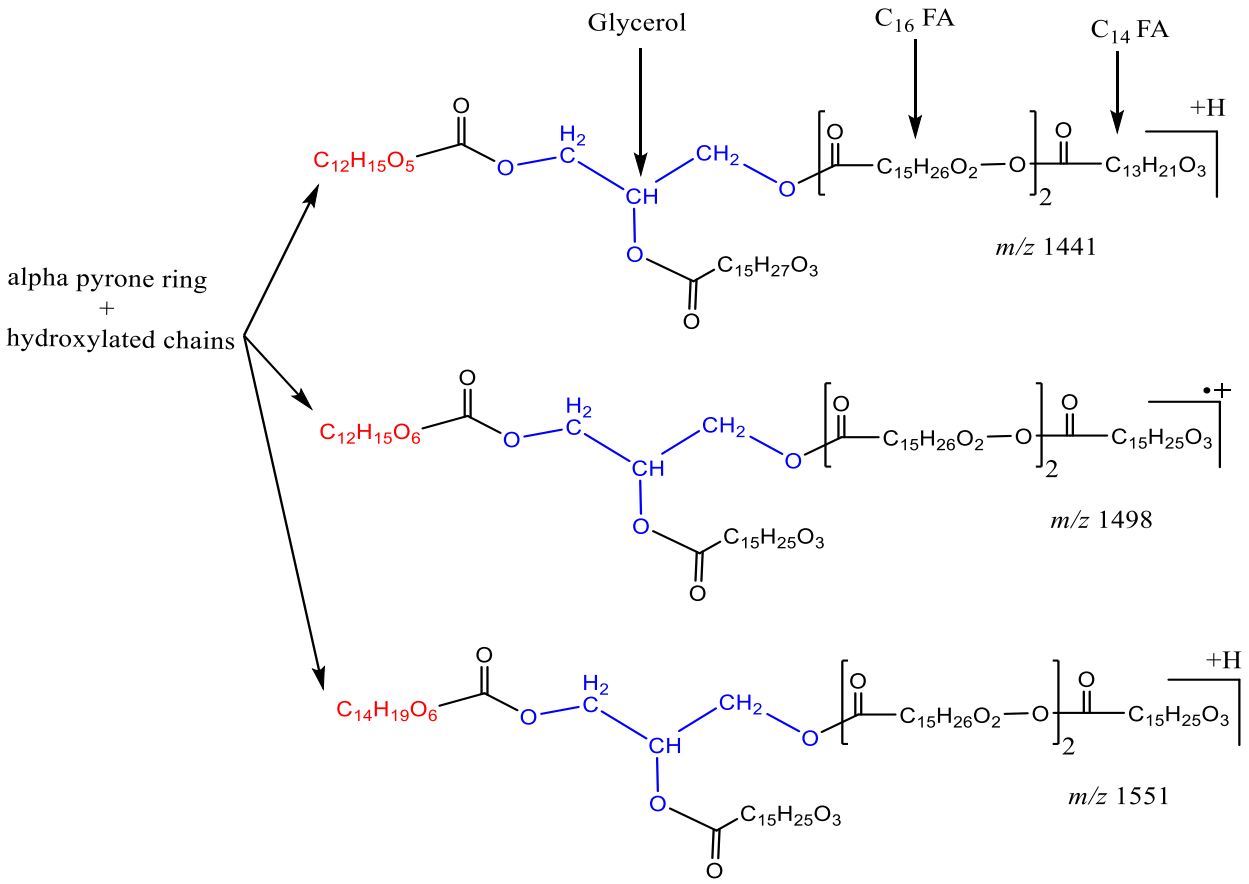


Figure A1.11b. Possible chemical composition of some selected ion in the Sporopollenin MALDI-MS using DAN as a matrix

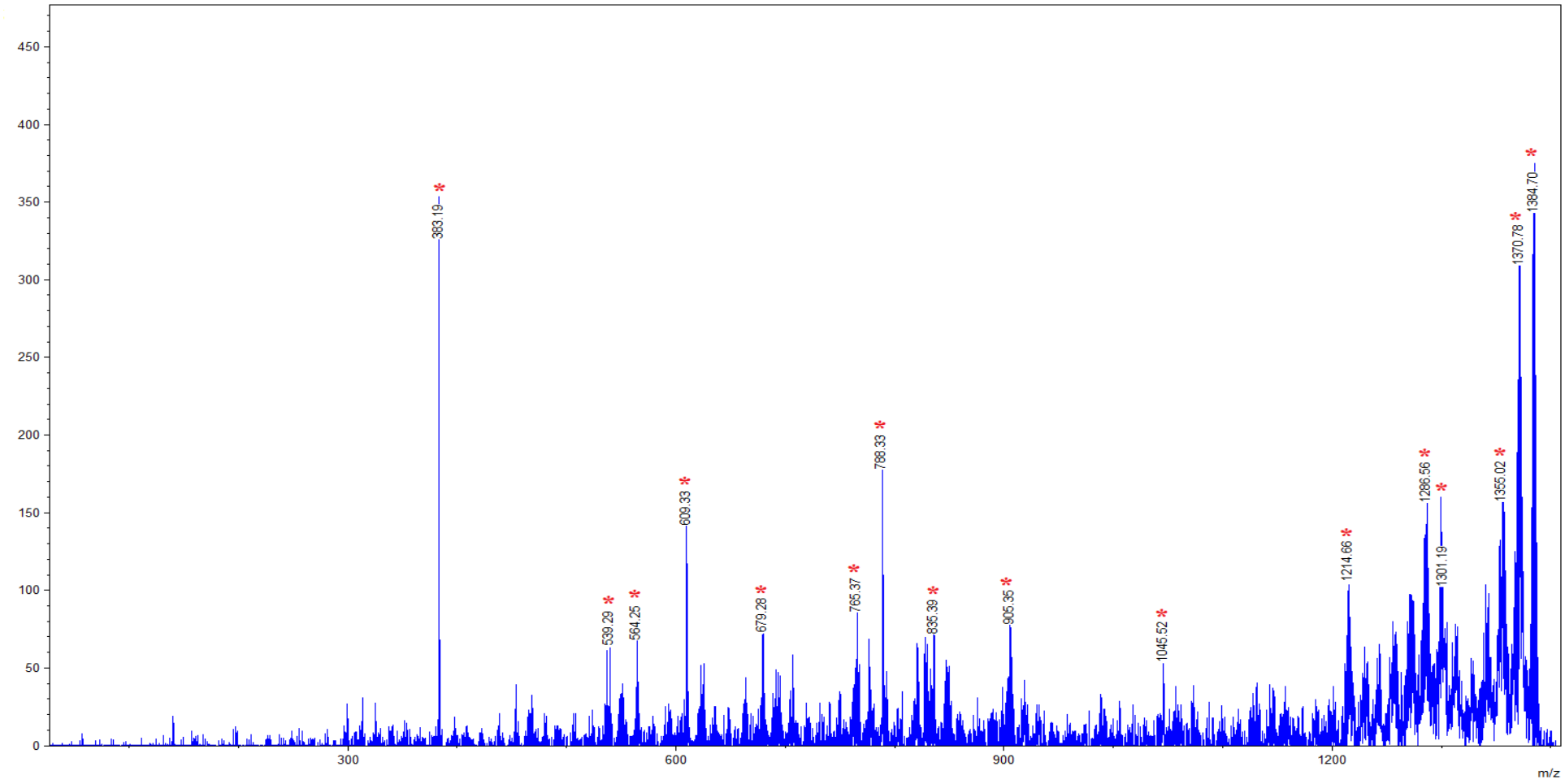
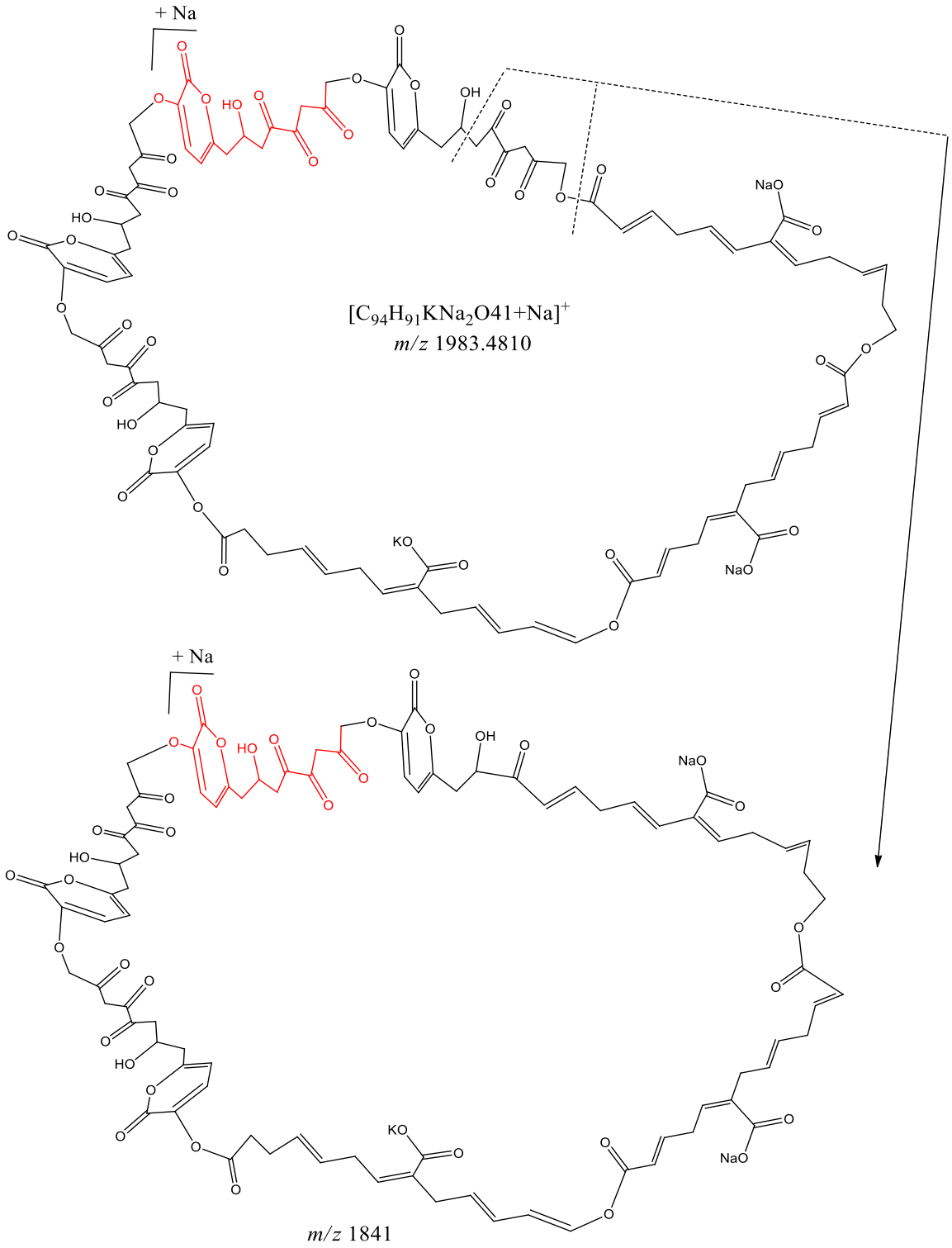
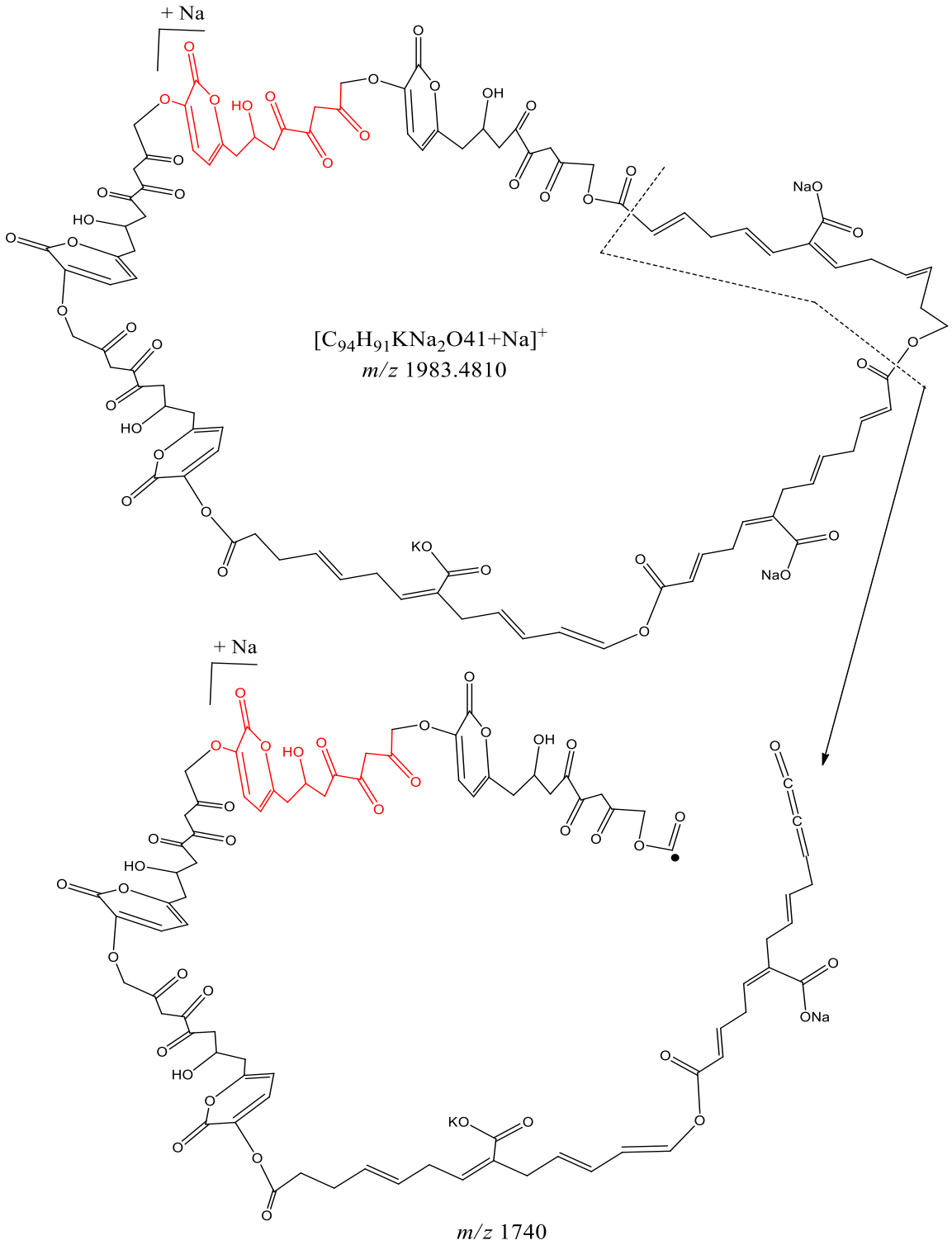


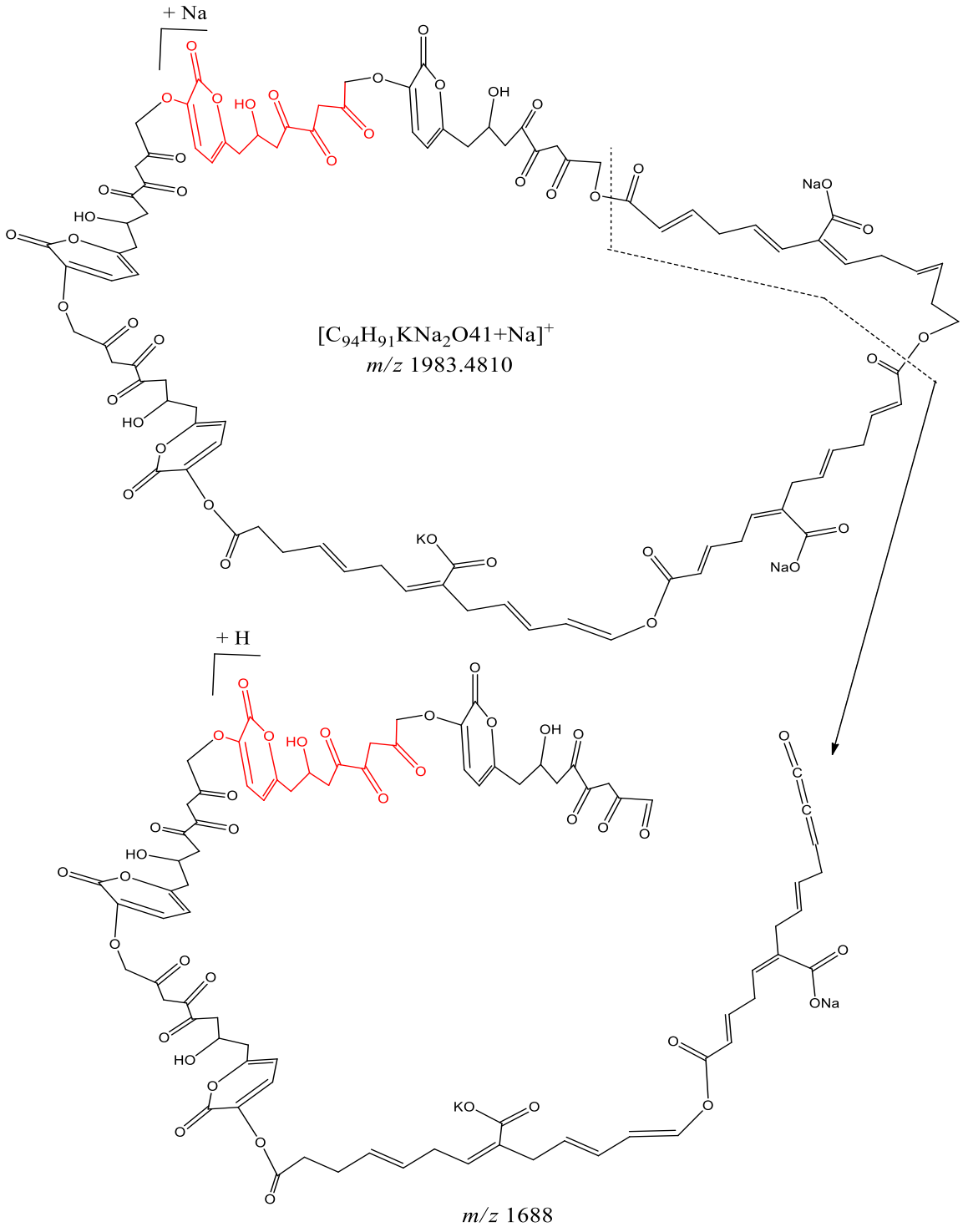
Figure A1.12. Product ion scan of *m/z* 1441.8037 extracted from the MALDI-TOF-MS of the empty sporopollenin using DAN as a matrix



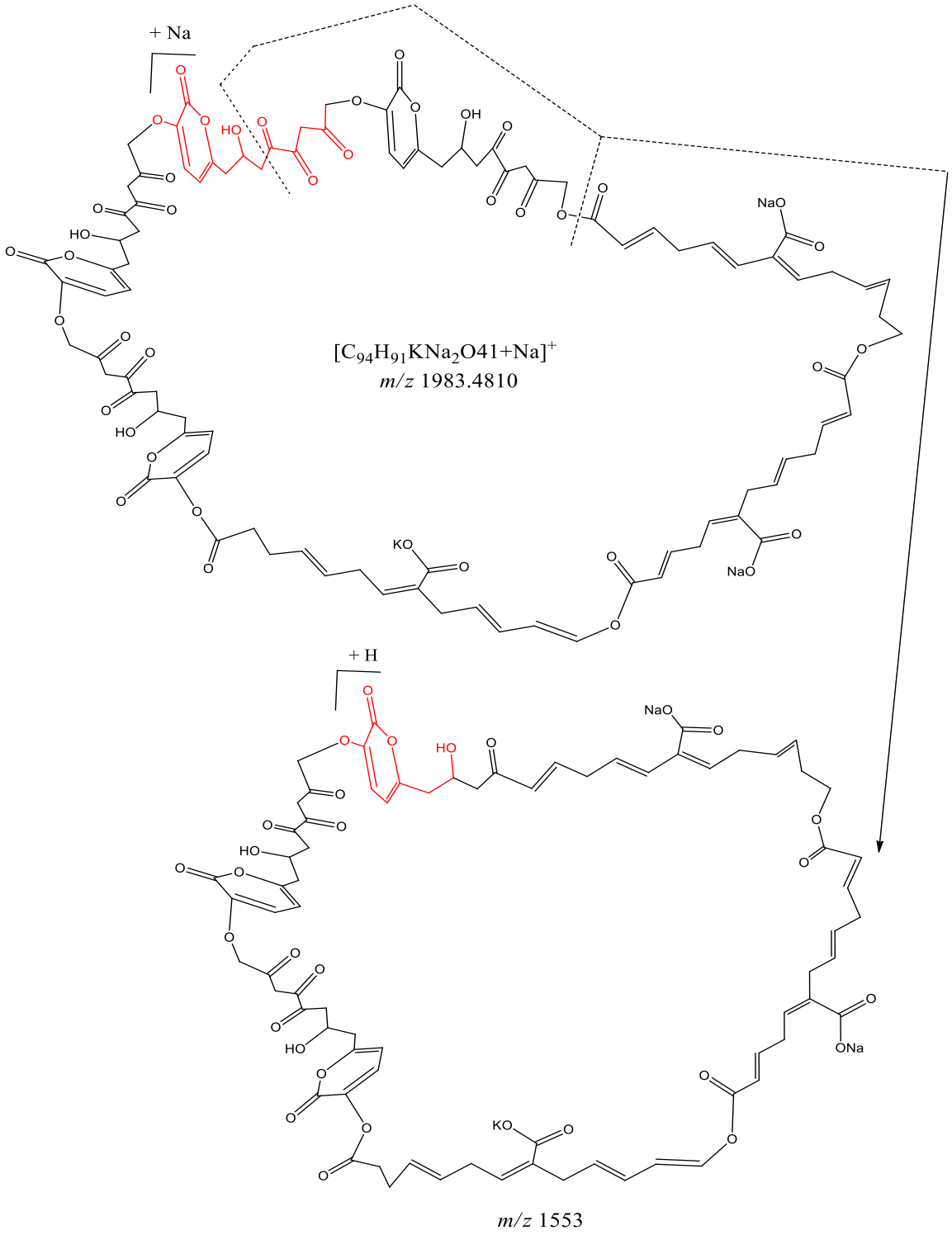
Scheme A1.4a. MS/MS fragmentation pathway of the precursor ion at *m/z* 1983 identified in the sporopollenin MALDI-MS using HABA as a matrix



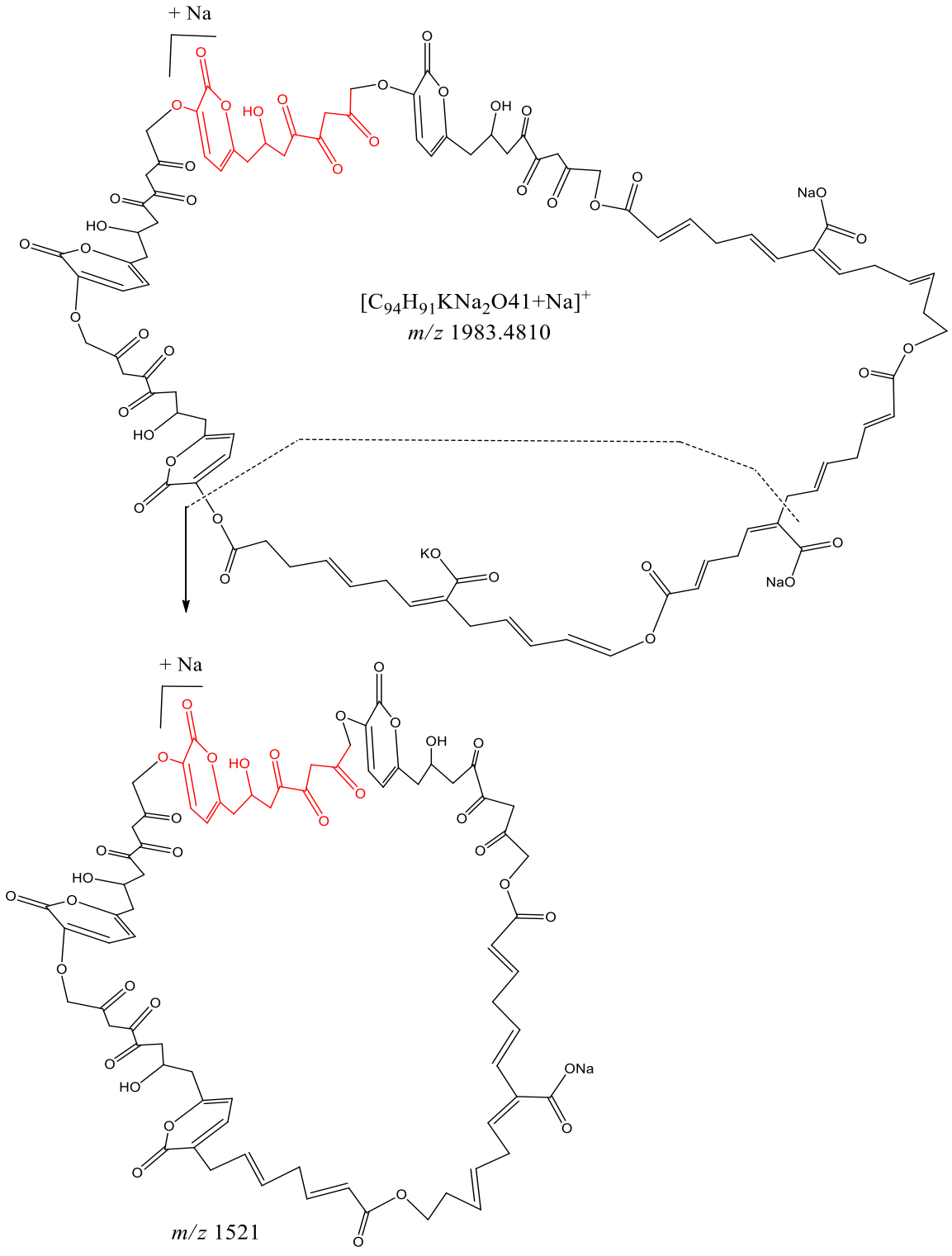
Scheme A1.4b. MS/MS fragmentation pathway of the precursor ion at *m/z* 1983 identified in the sporopollenin MALDI-MS using HABA as a matrix



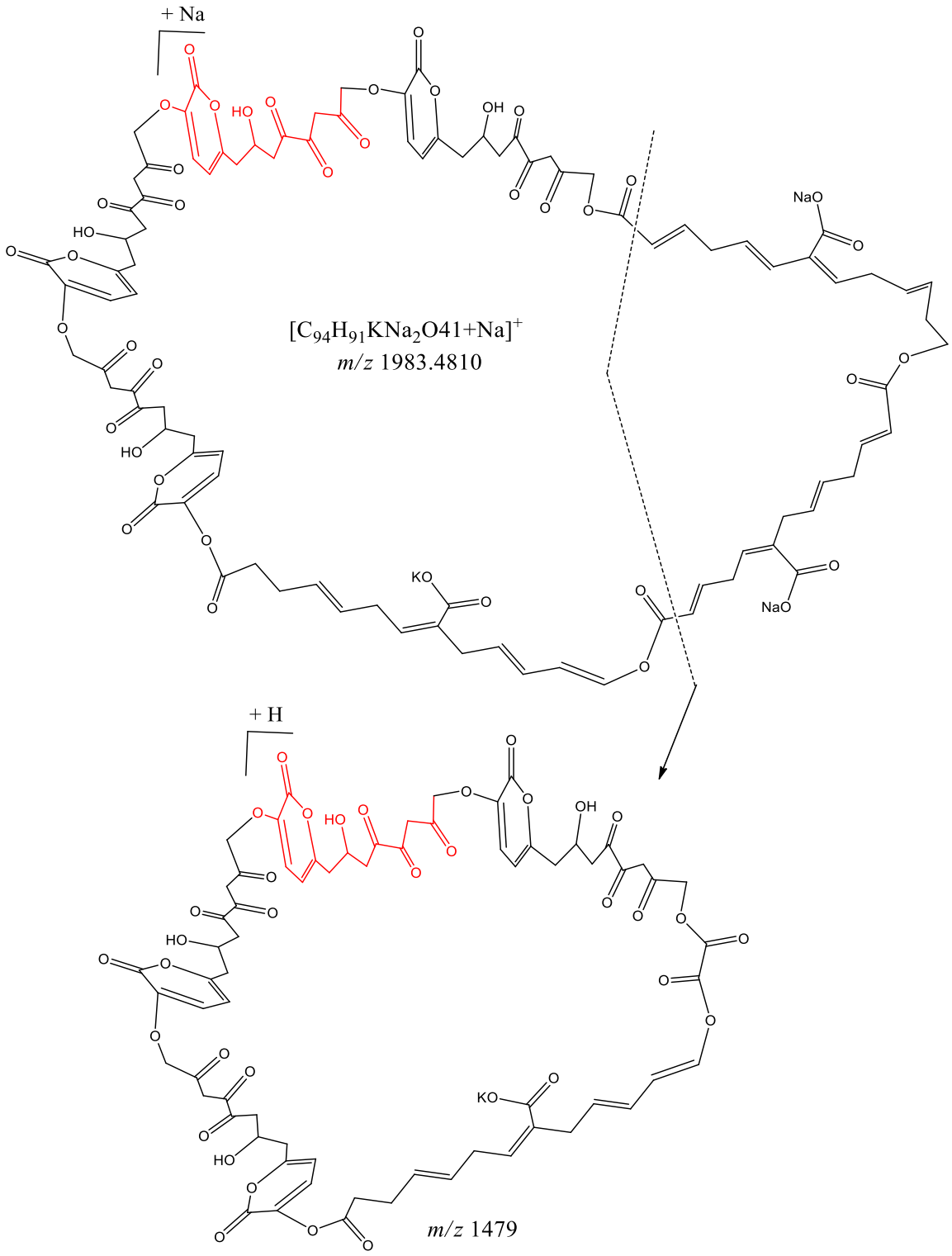
Scheme A1.4c. MS/MS fragmentation pathway of the precursor ion at *m/z* 1983 identified in the sporopollenin MALDI-MS using HABA as a matrix



Scheme A1.4d. MS/MS fragmentation pathway of the precursor ion at *m/z* 1983 identified in the sporopollenin MALDI-MS using HABA as a matrix



Scheme A1.4e. MS/MS fragmentation pathway of the precursor ion at *m/z* 1983 identified in the sporopollenin MALDI-MS using HABA as a matrix



Scheme A1.4f. MS/MS fragmentation pathway of the precursor ion at *m/z* 1983 identified in the sporopollenin MALDI-MS using HABA as a matrix

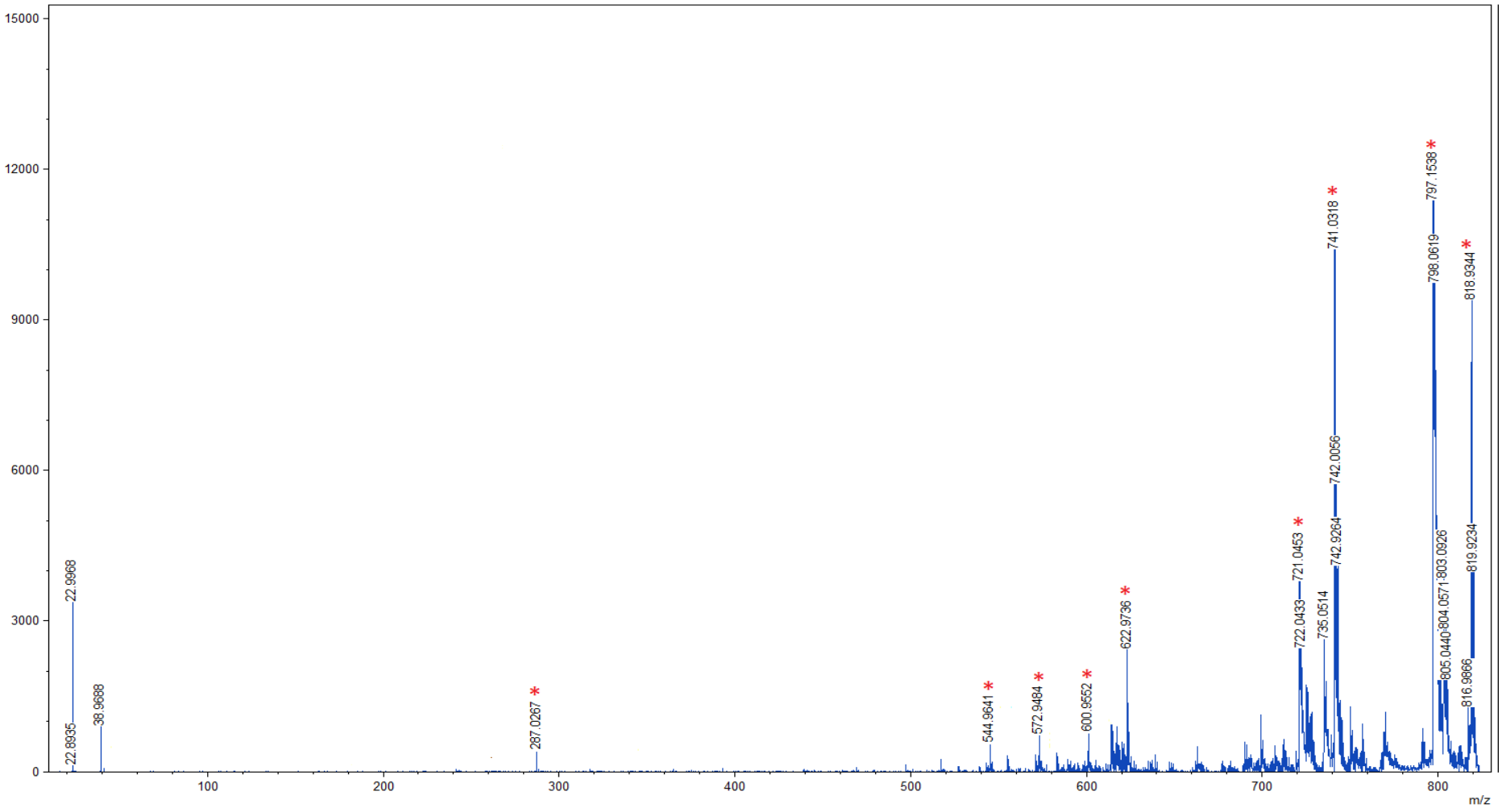


Figure A1.13. *Quasi*-MS^3 of the product ion at *m/z* 863 formed in MS^2 of the precursor ion at *m/z* 1983

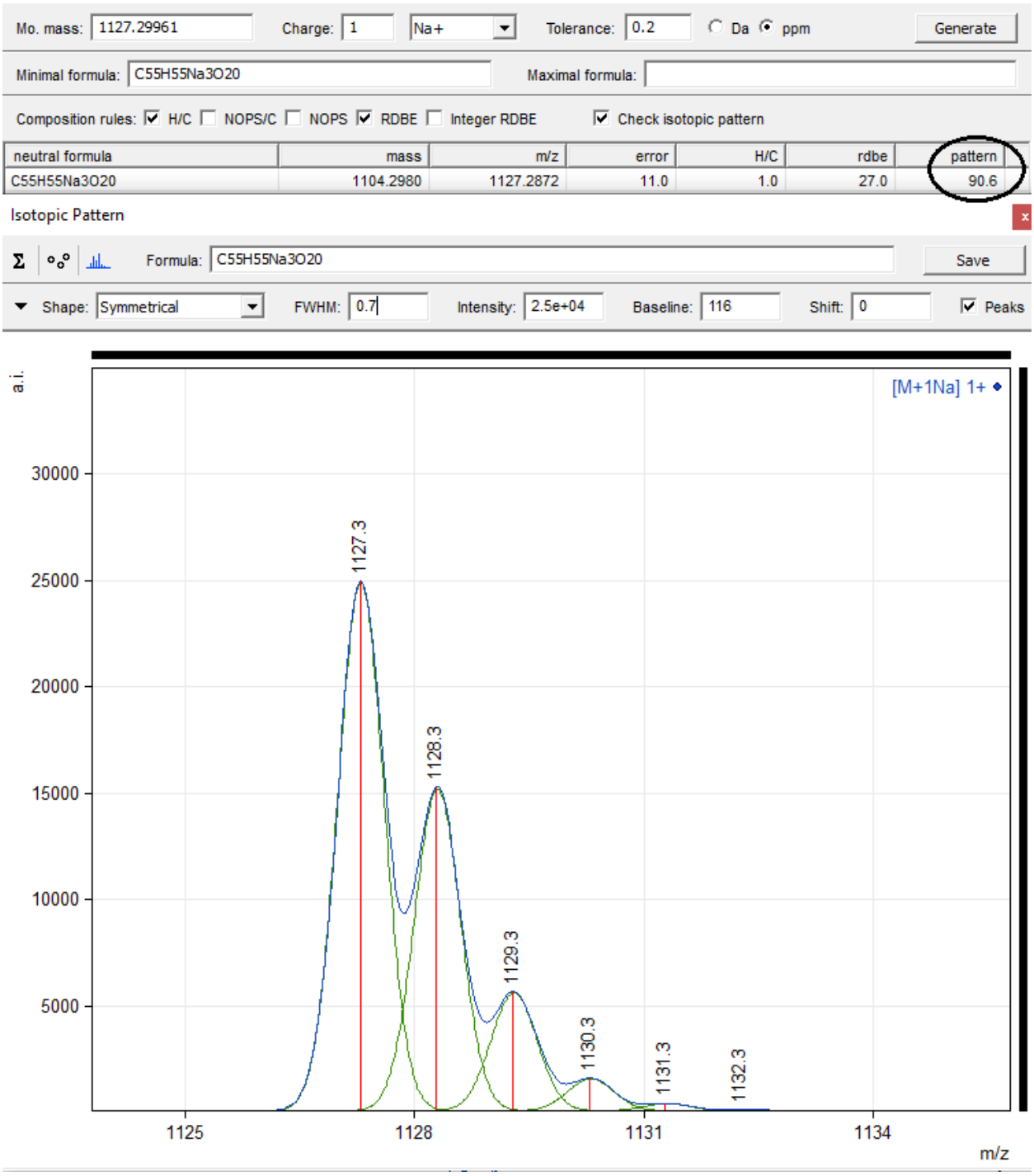


Figure A1.14. Comparison between the actual isotopic distribution of the peak at *m/z* 1127 (Green) with the theoretical isotopic distribution of the assigned formula $[C_{55}H_{55}Na_3O_{20}+Na]^+$ (Blue).

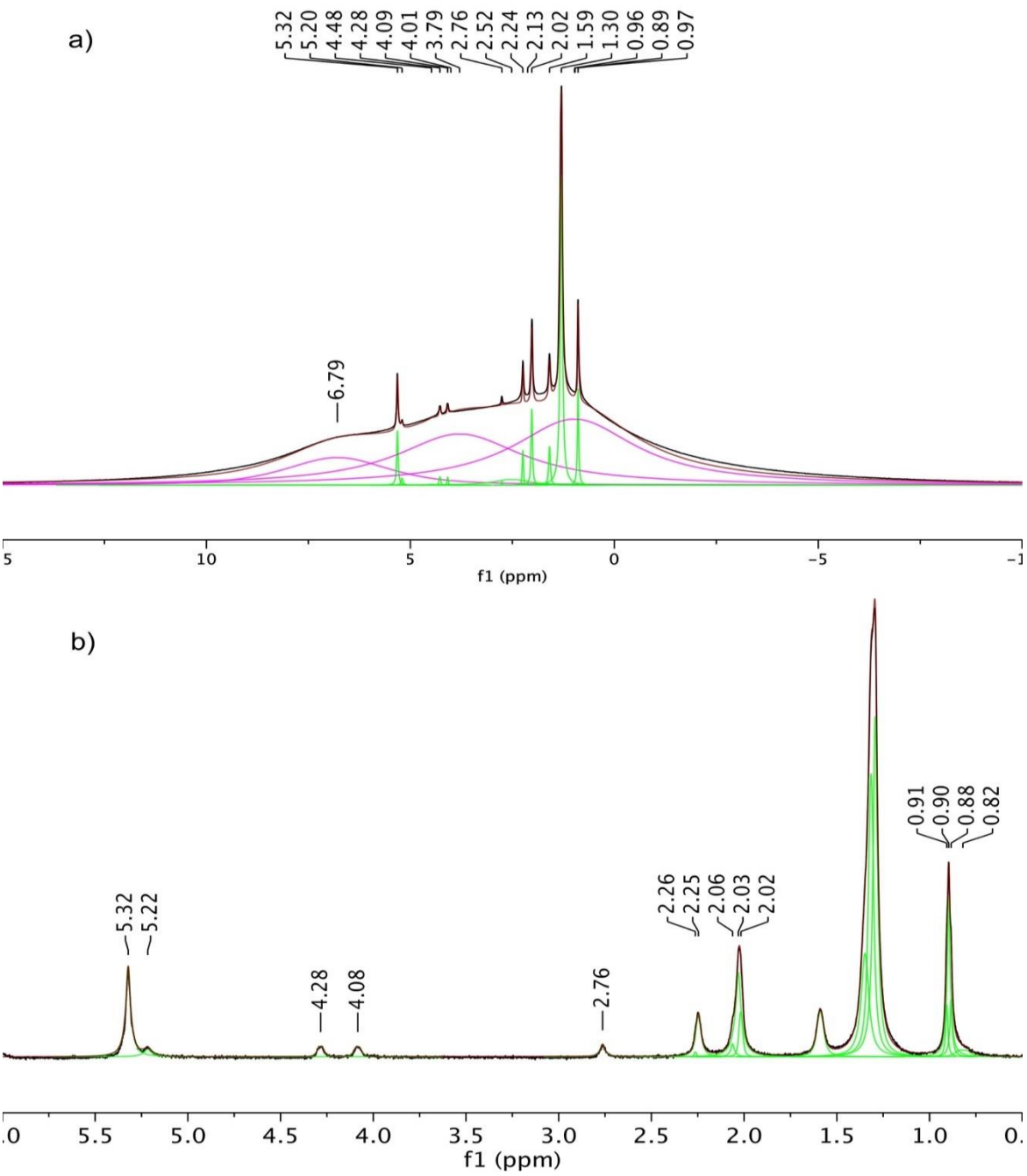


Figure A1.15. ^{1}H spectrum of Sporopollenin at ν_r=20kHz using 90 pulse (a) and using CPMG, with four loops of 1ms, (b). Peak deconvolution for each spectrum is shown in green (liquid-like phase) and pink (amorphous phase).

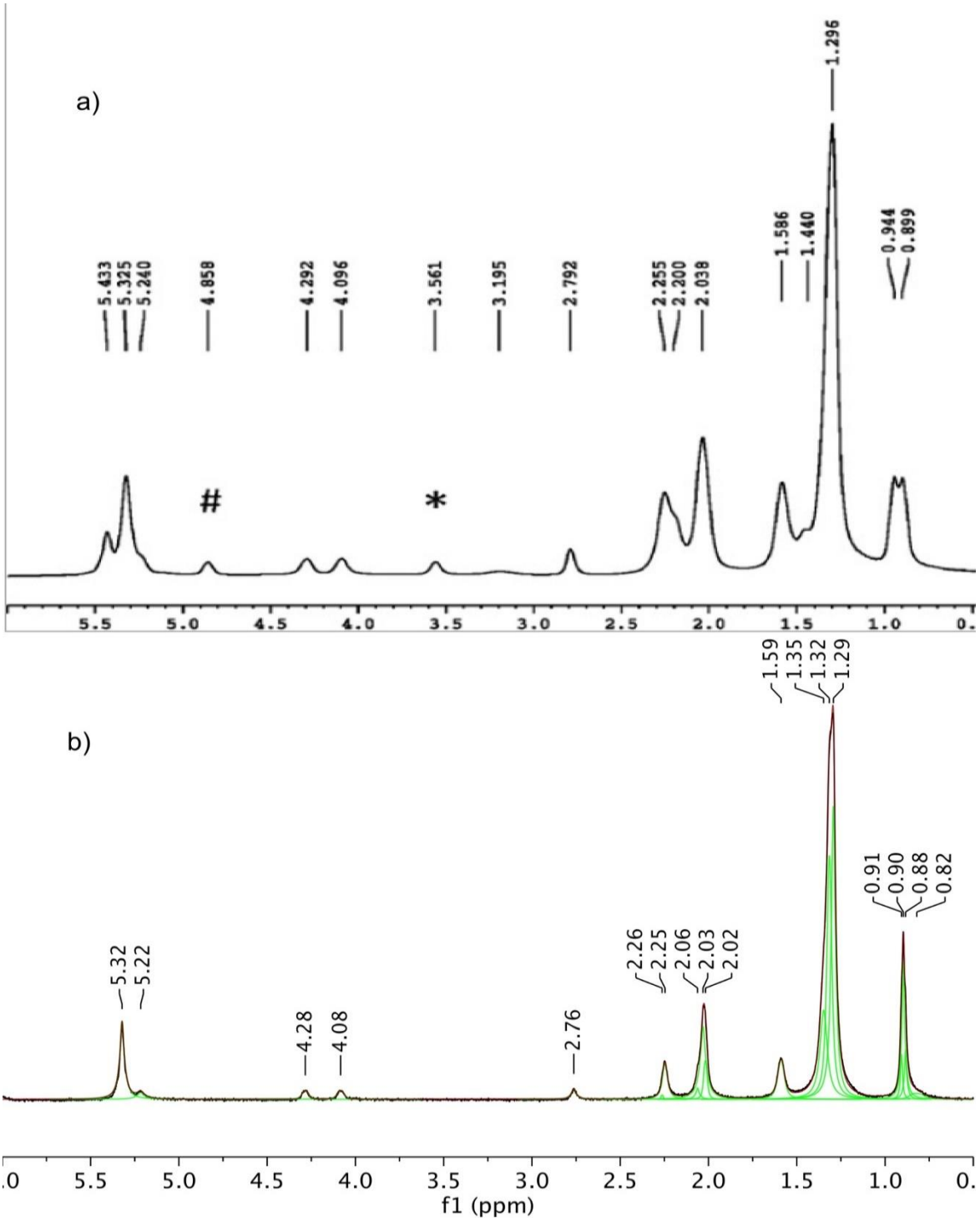


Figure A1.16. Comparison of the ^{1}H NMR spectrum of a) *Lesquerella lyrata* (adapted with permission from *Ind. Crop. Prod*. **2012**, *37*(1), 186-194) b) *Lycopodium clavatum* sporopollenin

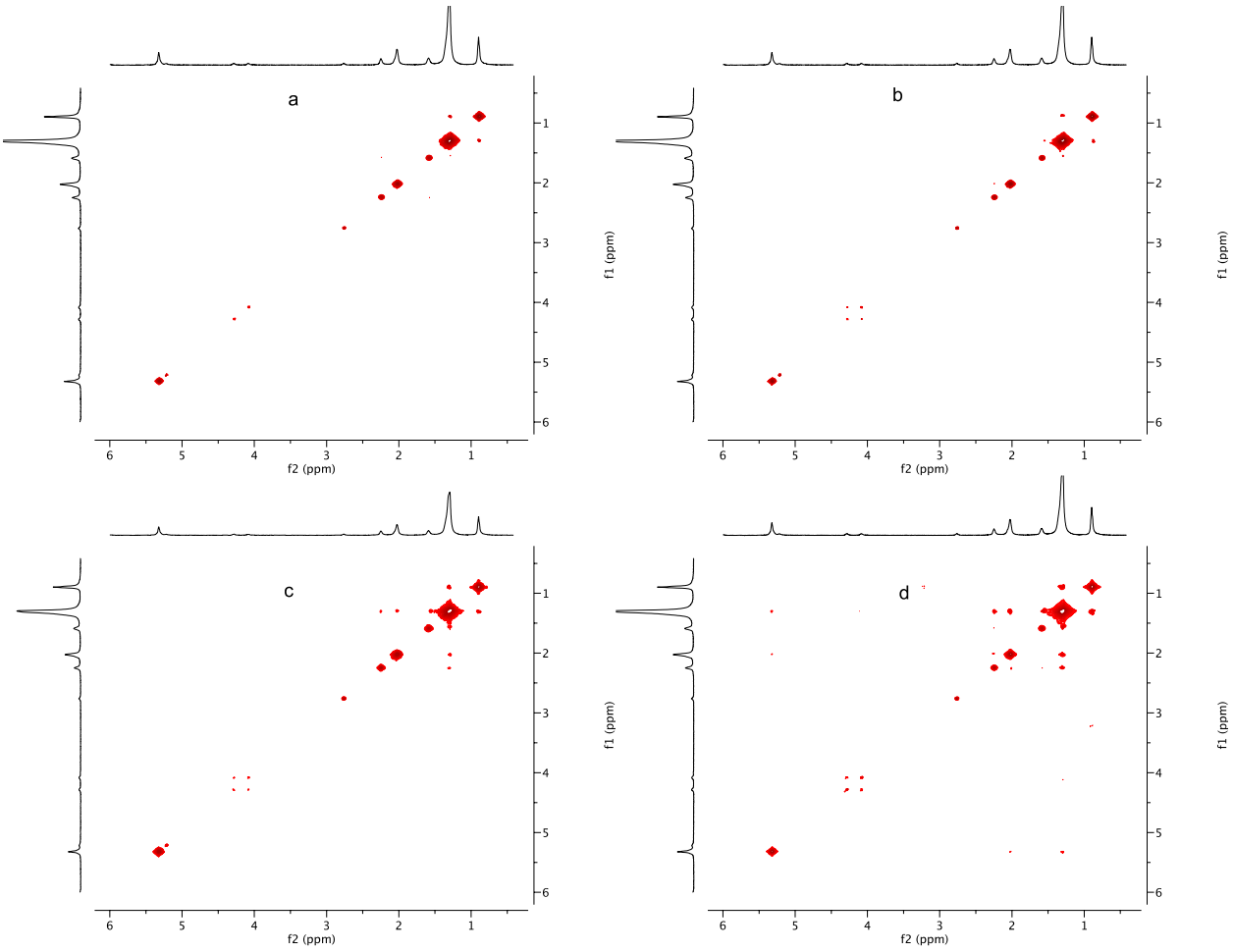


Figure A1.17. ^{1}H-^{1}H NOESY on Sporopollein for various mixing times: a) 10ms, b) 50ms, c) 100ms, d) 200ms

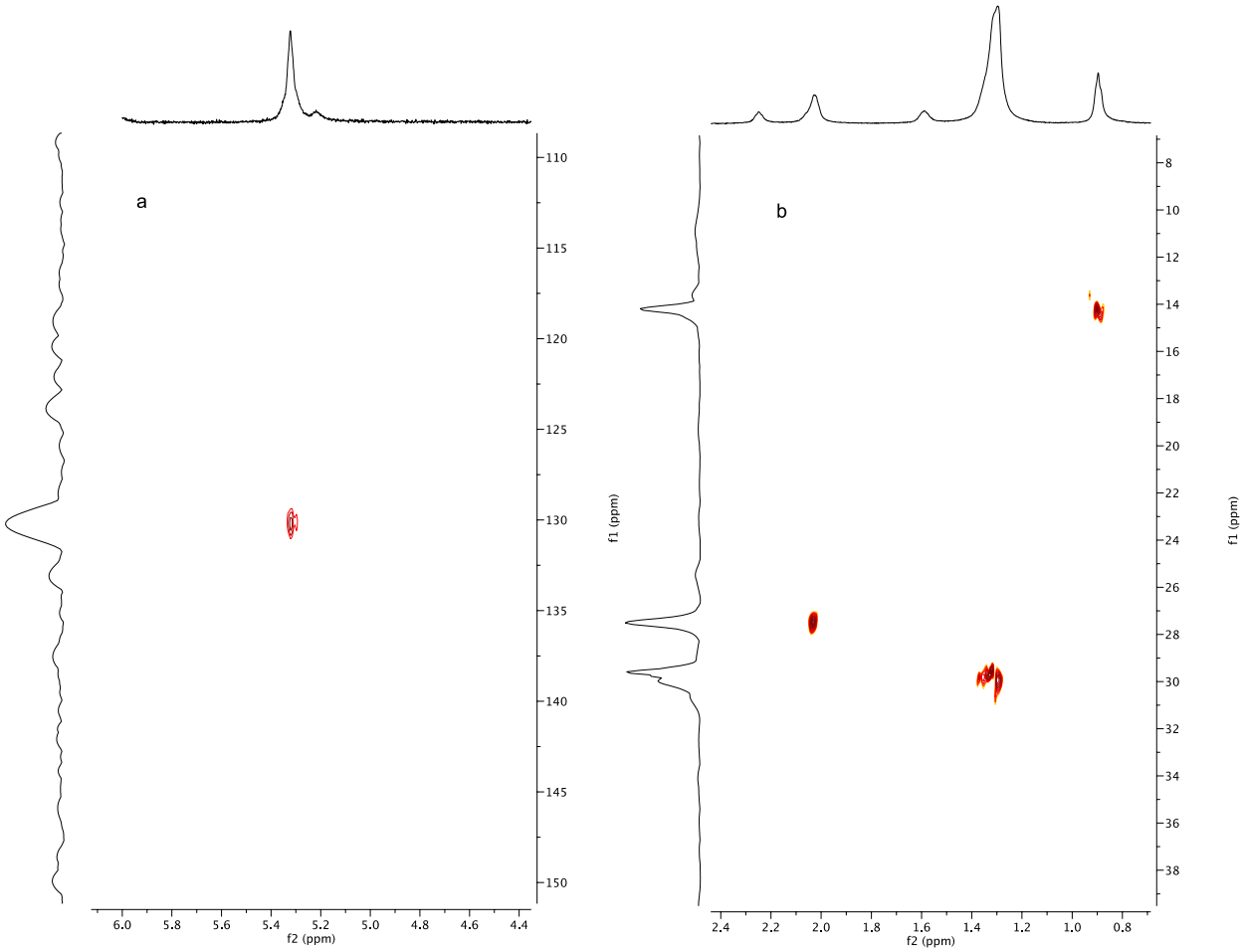


Figure A1.18. Rotor-synchronized $^{13}C\{^{1}H\}$ HSQC on Sporopollein for a) J=170Hz and b) J=125Hz

Table A1.2. 1H and 13C chemical shifts, NOE correlation, and suggested assignment.

#	δ (^{1}H) ppm from deconvolution	δ (^{13}C) ppm from HSQC	NOE **STRONG**, *MEDIUM*, weak	Suggested Groups
A	0.82 0.88 0.90 0.91	14.27	**B**	-CH_3
B	1.29 1.32 1.35	29.96 29.6 29.88	**A, C**, *D, E*, j	-$(CH_2)_n$-
C	1.59		**B**	-**CH_2**-CH_2-COO
D	2.02 2.03 2.06	27.5	*A*, e, j	-**CH_2**-CH=
E	2.25 2.26		*B*, d	-**CH_2**-COO
F	2.76			-CH=CH-**CH2**-CH=CH-
G	4.08		**H**	-**CH_2**-O (diastereotopic H)
H	4.28		**G**	
I	5.22			-C**H**-O- (β-C of glycerol)
J	5.32	130.7	b, d	-CH=CH-

Table A1.3. List of peaks and their assignment revealed in C-13 Solid state NMR

NMR shift	Assignment
14.36	CH_3
23.05	CH_2
25.56	CH_2
27.63	CH_2-CH=CH
29.8	CH_2
36.37	CH_2-C=
43.26	CH_2-Pyrone ring
56.25	CO-CH_2-CO
64.98	CH_2 -O
72.98	CH-O
101.99	Alpha pyrone ring C-5
116.3	Alpha pyrone ring C=C
124.43	Alpha pyrone ring C=C
129.72	Chain C=C
133.54	Chain C=C
143.84	Alpha pyrone ring C-3
159.58	Keto or enol carbon =C-OH or -C=O
172.49	COOH

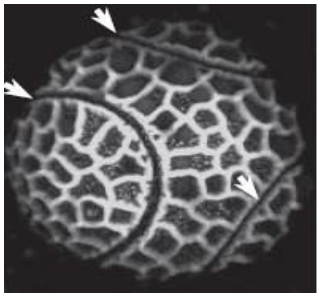


Figure A1.19. *Passiflora sp. (Passifloraceae) sporopollenin* SEM image adapted with permission from *Annu. Plant Rev. online.* **2018**, *1*(2):1-40.

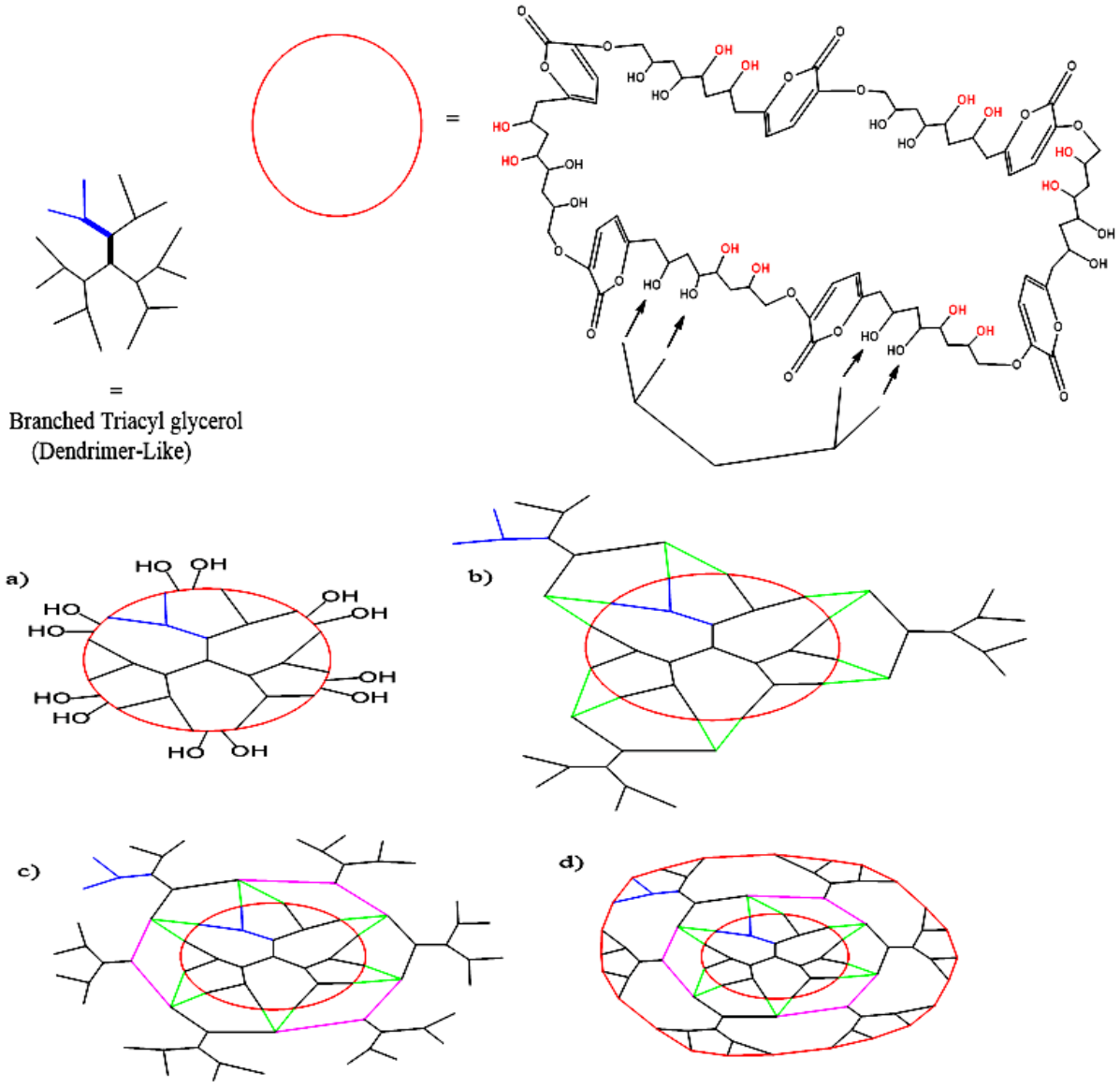


Figure A1.20. Stepwise proposal for the formation of sporopollenin. In the Top of this figure, the two-building blocks used in this model are shown, the first unit is the circular polyhydroxylated tetraketide polymer (Red circle) which is rich with hydroxyl groups and the second unit is the branched triglycerides (dendrimer-like) that can be attached to the hydroxyl groups of the circular polymer. **a)** the branched triacyl glycerides approach the bottom hydroxyl group of the circular polymer leaving the rest of the OH groups that are pointing up. **b)** Branched triglycerides (green) can be attached to the free OH groups that are pointing up in the circular polymer. **c)** The added branched triglycerides (green) in the last step can be connected by more branched triglyceride (purple). **d)** all branches can be added to another circular polymer (red) and so on.

The summary of the manuscript in the form of questions and answers

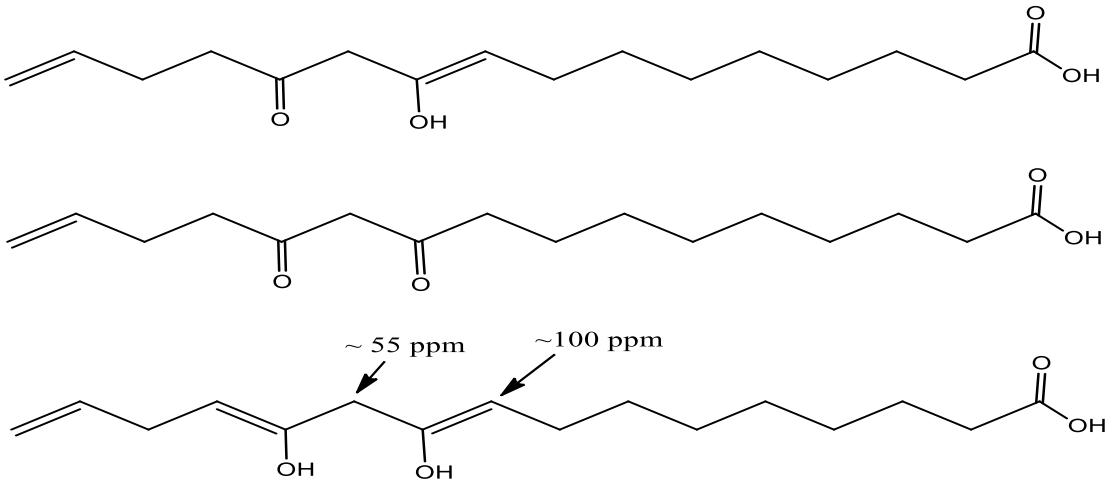
1) Why did we begin our study with TOF-SIMS-MS/MS?

Based on our Knowledge that Sporopollenin is extremely stable and insoluble in any solvent, we decided to use the energetic secondary ions to strike sporopollenin aiming to find any characteristic ions for sporopollenin

2) What did we get from TOF-SIMS-MS/MS?

We find some fatty acids, mono, and diglycerides, which give us a hint about the nature of sporopollenin, then we decided to do MS/MS for some selected peaks in the positive and negative ion mode. As an example, the MS/MS of the carboxylate ions at *m/z* 281 and 253 was characterized by the loss of water and the appearance of unusual ions such as *m/z* 41, 80, 133, and 137. After studying the MS/MS fragmentation very well and after several proposals, we successfully find the right structures of these fatty acids monomers. From this study, we deduced that these fatty acids contain a terminal hydroxyl group that is needed to link the fatty acids together plus two oxygen atoms that exist in their chains as a beta diketone moiety. Interestingly, this beta diketone moiety can exist in three different forms Keto-Keto, Keto-enol, and enol-enol. The MS/MS of these fatty acids supports that they are present in the Keto-enol form, but if we draw the enol-enol form, the CH_2 in between can be assigned to the peak at 55 ppm appeared in the C-13 NMR spectrum. Based on this result, we can raise some questions that should be addressed in the future. Does this beta diketone moiety exist in different forms in the sporopollenin structure? Are they

responsible for Hydrogen bonding between the fatty acids chains giving more strength to this network, and lastly, Are these central OHs used for cross-linking between the chains?



3) After knowing the possible monomers using TOF-SIMS-MS/MS, what was our next step?

Based on sporopollenin's insolubility in any solvent, we decided to use the solid sample to perform the MALDI-MS using some matrices that are good for lipids such as DAN and CHCA. We aimed from this step to use a soft ionization technique to see large fragments compared to the TOF-SIMS or any kinds of polymeric fatty acid or TAGs. After seeing large masses in both DAN and CHCA, we begin to propose structures for these masses using the monomers identified in TOF-SIMS. Surprisingly, we were successful in building chemical structures for these masses using the C-14 and C-16 monomers identified using TOF-SIMS-MS/MS. Especially for an ion in the MALDI-MS using CHCA as a matrix that appeared at *m/z* 1643, which was assigned as the protonated molecule of $C_{89}H_{142}O_{27,}$ a formula which is close enough to the empirical formula of *Lycopodium clavatum* empirical

formula $C_{89}H_{142}O_{27}$. This ion was assigned as Diacylglycerol containing six fatty acids (five C14 and one C16) plus glycerol moiety. Also, all the identified chemical formulas in the MALDI MS using both DAN and CHCA have C/H, C/O, and H/O ratios that fit with the empirical formula of the *Lycopodium clavatum.* For the sake of brevity, we showed in the Manuscript one example for the MS/MS of *m/z* 1441 identified in DAN spectra. It should be mentioned that all other ions in both the DAN and CHCA spectra are subjected to further MS/MS studies that can be published later in a separate paper.

4) What did we decide to do after revealing structures in the MALDI-MS spectrum that correspond to the empirical formula of *Lycopodium clavatum*?

We decided to try other matrices aiming to find more information about sporopollenin structures such as HABA, which is known to be used for the analysis of polymers. Surprisingly, it showed a different kind of spectrum indicating a repeating unit and/or polymeric structure, and so we deduced that this could be different information than what we knew from MALDI-MS using DAN and CHCA. By surveying the literature about the biosynthesis of sporopollenin, we found that many researchers suggest that sporopollenin is composed of fatty acids and/or alcohols plus polyhydroxylated tetraketide monomers, which in general is composed of an alpha pyrone ring plus a hydroxylated aliphatic chain. Actually, we think about these polyhydroxylated tetraketides as a carboxylic acid with many OH groups, but its COOH cyclizes with one of these OH groups to form the cyclic lactone or the alpha pyrone ring. Based on this latter information, we suggest that the

polyhydroxylated tetraketide can form this polymeric pattern in the MALDI-MS using HABA as a matrix. After tedious trials to propose a general structure for these polymeric ions that is close to their masses and fits their MS/MS patterns, we assigned these ions as a circular polyhydroxylated tetraketide polymer. This circular polymer has some opened pyrone rings in the form of carboxylates and some dehydrogenated CH-OH present in the hydroxylated chains (CHOH is in the form of C=O). Also, the C-13 NMR showed peaks that support the structure of the pyrone ring and the hydroxylated chain in these polymeric structures.

5) What did we do after revealing the empirical formula of sporopollenin and the possibility of the presence of the pyrone rings plus the hydroxylated chains?

We start to study all possible NMR techniques to know if our proposed structures using mass spectrometry make sense with the NMR or not. Surprisingly, all NMR data support what we revealed from mass spectrometry. We can summarize the NMR findings in the following points:

a) The solid-state proton NMR spectrum showed two kinds of the overlapped spectrum.

1) The first spectrum was a very sharp spectrum indicating that this sharp spectrum is corresponding to liquid-like structures, which fit with fatty acid chains having mobility. Interestingly, we were lucky to survey the literature and find a remarkable similarity between this sharp spectrum and the spectrum of TAG estolides (branched esters of hydroxy fatty acids) of *Lesquerella lyrate*

seeds. Only two peaks were absent in the sporopollenin. The first one at 3.5 ppm is characteristic of the proton attached to the carbon carrying the OH group. This indicates that OH in sporopollenin is present on a double bond (vinylic carbon with no hydrogens). The second peak at 4.8 ppm is characteristic for TAG Estolides (branched ester bonds), indicating that sporopollenin hydroxycarboxylic acids are connected linearly and not in a branched way.

2) The second spectrum was a broad spectrum indicating an amorphous part of the sporopollenin network, which appears to be a more rigid sub-structure on which the fatty acids network is attached, hindering its movement. This broad-spectrum showed that sporopollenin containing alpha-pyrone rings (broad peak at 6.79 ppm) and hydroxylated chains (broad peaks at 3.79 and 0.98 ppm). These assignments support the presence of polyhydroxylated tetraketide composed of the alpha-pyrone ring and hydroxylated aliphatic chain, like what we revealed from the MALDI-MS using HABA as a matrix.

b) We drew a structure using the identified sporopollenin units, which explains every peak in the C-13 NMR spectrum. Moreover, it fits the C-13 quantitative data.

6) Why did we perform High-resolution X-ray Photoelectron spectroscopy?

After performing the different NMR spectra, we thought that the broad peak at 6.79 ppm in the proton NMR spectra and the peak at 143 ppm in the C-13 NMR spectra indicates the possibility of the presence of aromatics in the sporopollenin structures. So, we performed

XPS after knowing that it can characterize aromatic polymers through some satellite peaks. Interestingly, there were no satellite peaks in the XPS supporting the absence of aromaticity or any kind of aromatic polymer or compounds. Also, we thought that alpha pyrone rings are aromatic; however, after checking one of the heterocyclic books, we found that pyrone rings are not typical aromatics (pseudo aromatic). This is due to their characteristic reactions like ring-opening, Diels-Alder reactions, and electrophilic additions. Fortunately, for this latter reason, it does not contradict the absence of aromaticity in the sporopollenin revealed by the XPS.

7) What are we planning to do in the future?

Our future work will focus on using different matrices in the MALDI-TOF-MS/MS in the positive or the negative ion mode, which may reveal other diagnostic structural details of the *Lycopodium clavatum* sporopollenin. Moreover, we will try to investigate the structure of other sporopollenin species by using the same characterization techniques presented here. Lastly, we are willing to perform some theoretical studies and modeling to build a complete sporopollenin network using the units identified in this manuscript and study the network's stability and the possible interactions between their chains.

Appendix B. Supporting information for Chapter 3

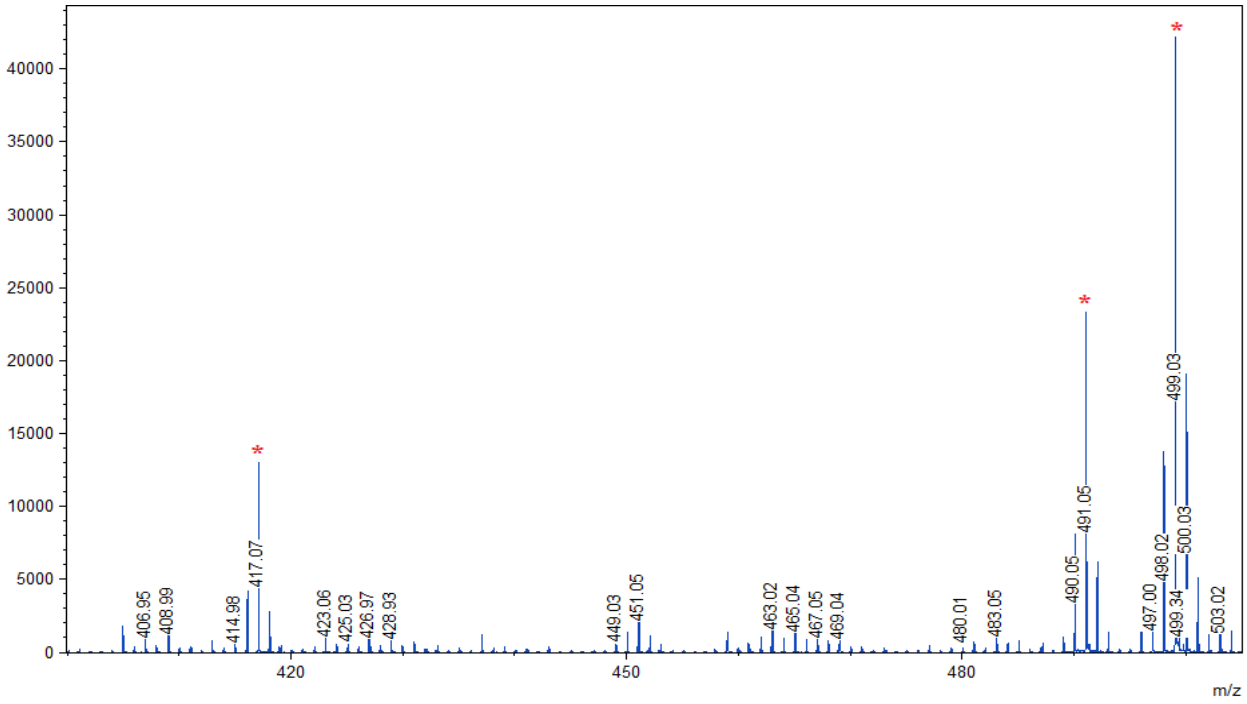


Figure A2.1a. Zooming of the spectrum from *m/z* 400 to 500

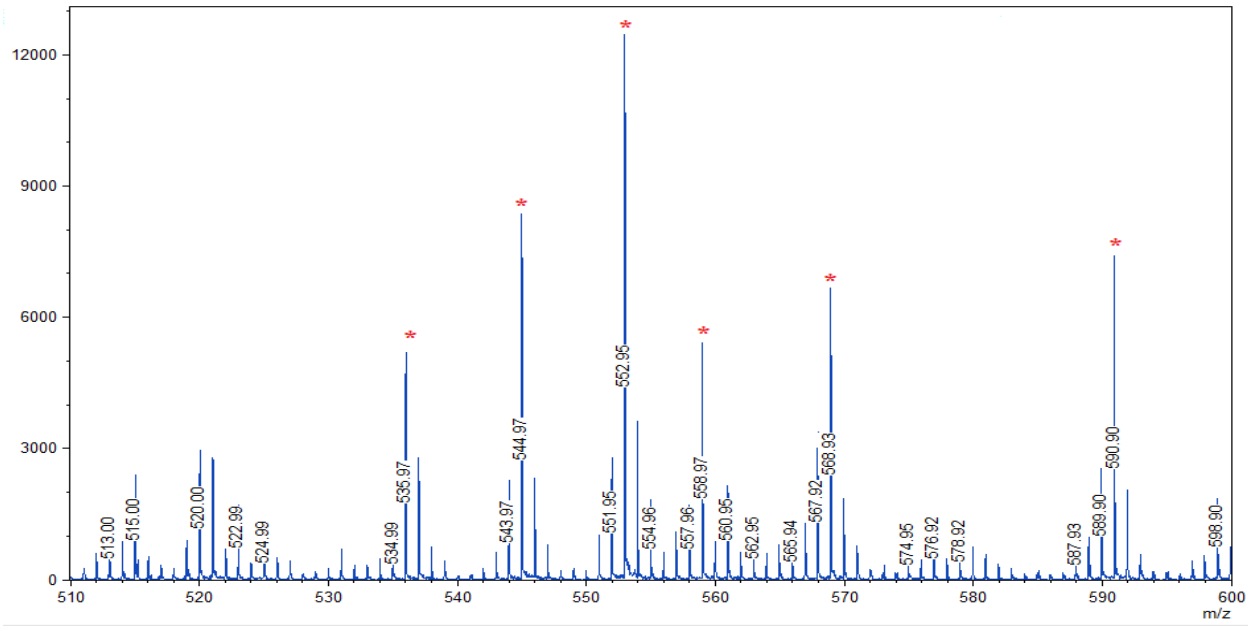


Figure A2.1b. Zooming of the spectrum in the region from 500 to 600

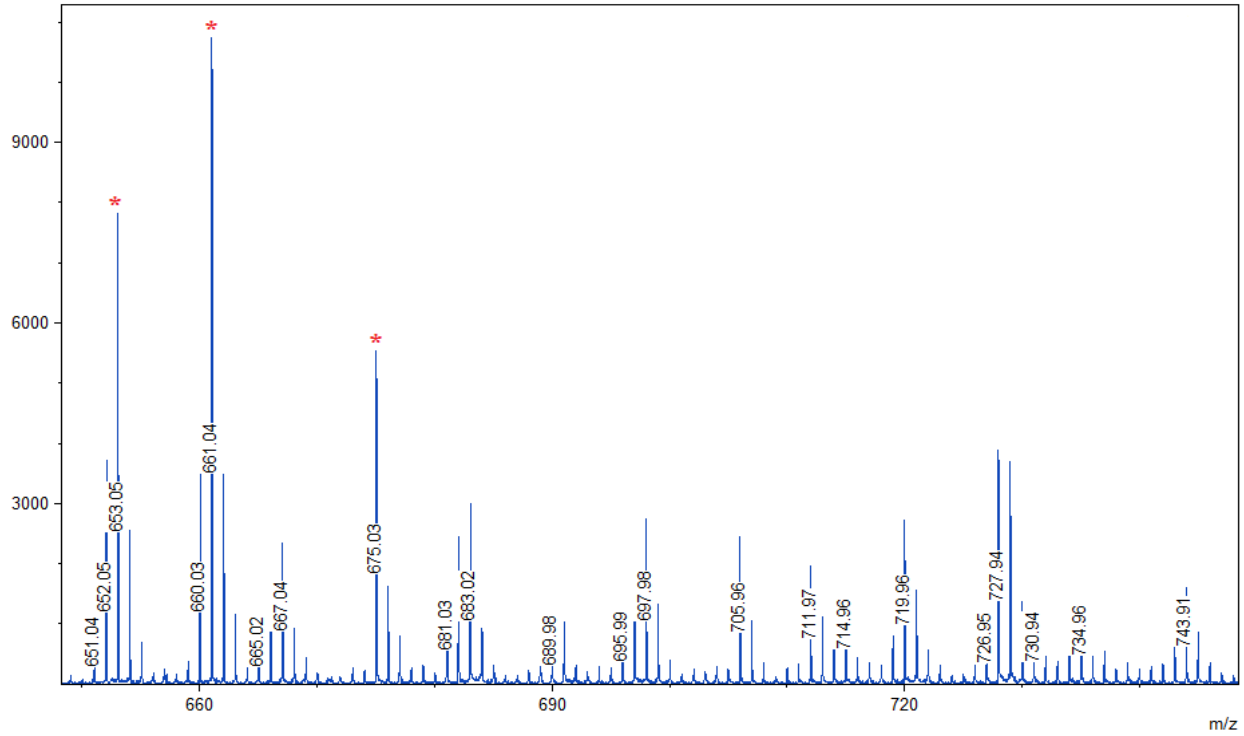


Figure A2.1c. Zooming of the spectrum from *m/z* 650 to 750

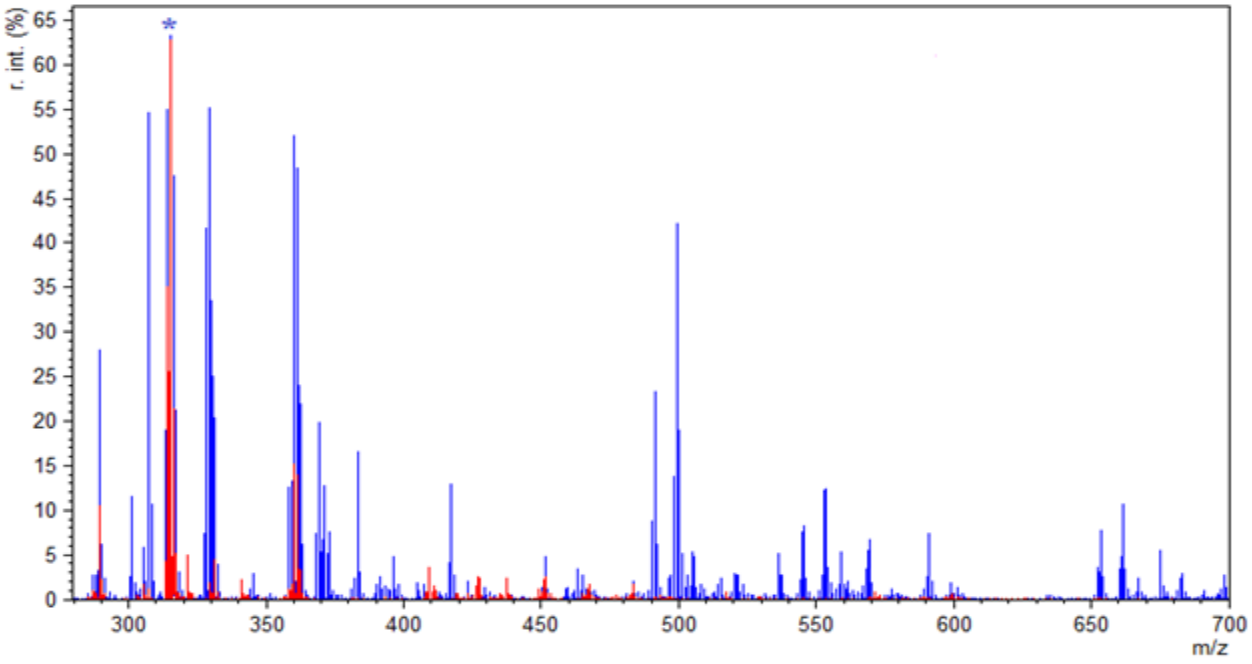
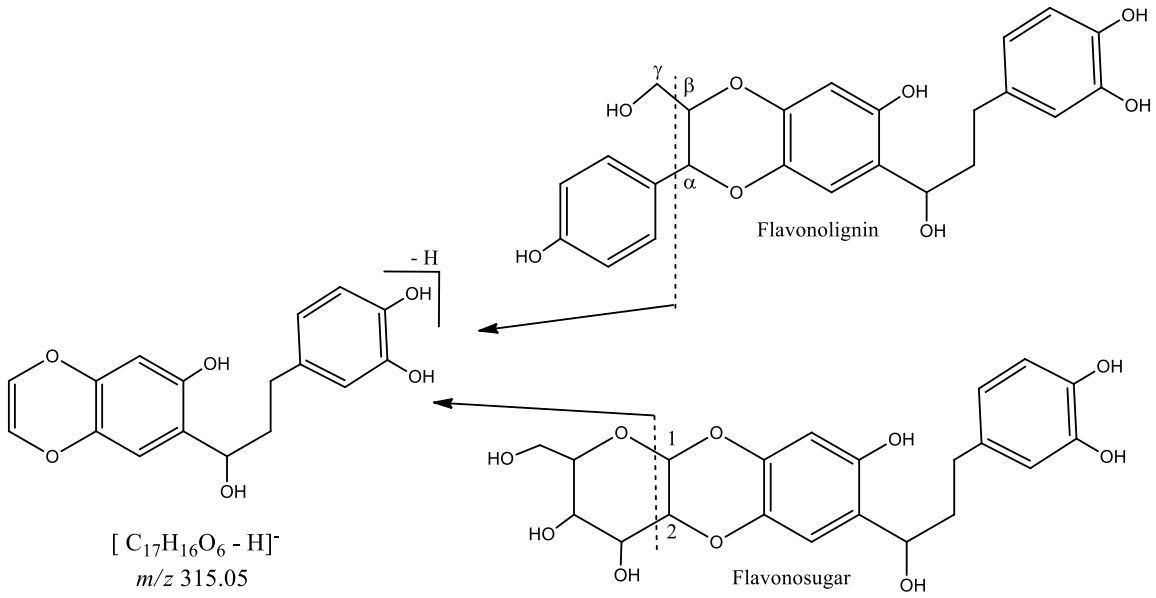


Figure A2.2. Comparison between the Matrix only (RED) and the sample (Blue). There is a strong overlap for the peak at *m/z* 315, as denoted by an asterisk.



Scheme A2.1. Possible origin of the precursor ion *13* at *m/z* 315.05

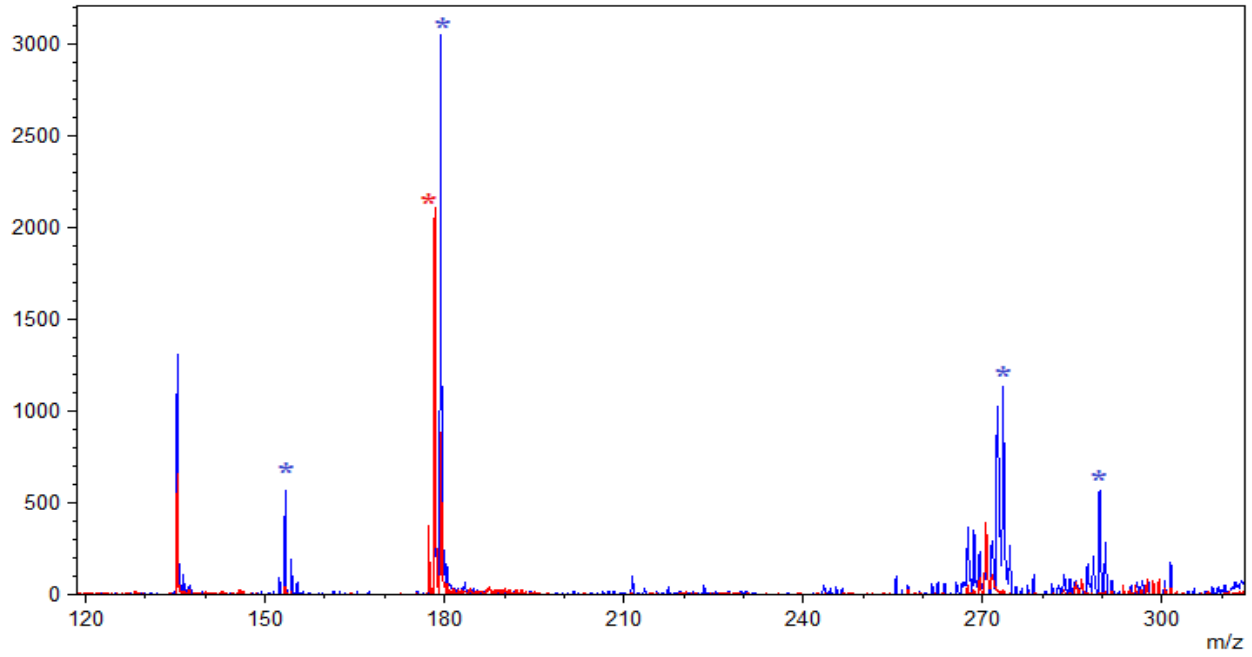
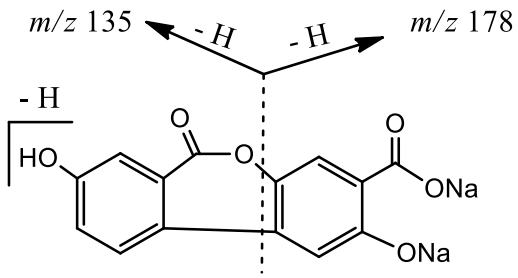


Figure A2.3. The comparison between the MS/MS of *m/z* 315 of the matrix (RED) and m/z 315 of the sample (Blue). The blue asterisk denotes characteristic peak for the sample ion, which is m/z 153, 179, 273, and 289. The red asterisk denotes the peak at m/z 178 that characterize the matrix ion. The peak at m/z 135 is a common peak between the two ions.



[2DHB - $2H_2O$ + 2Na - 3H]$^-$

m/z 315

Scheme A2.2. The fragmentation pathway of the proposed matrix cluster at *m/z* 315

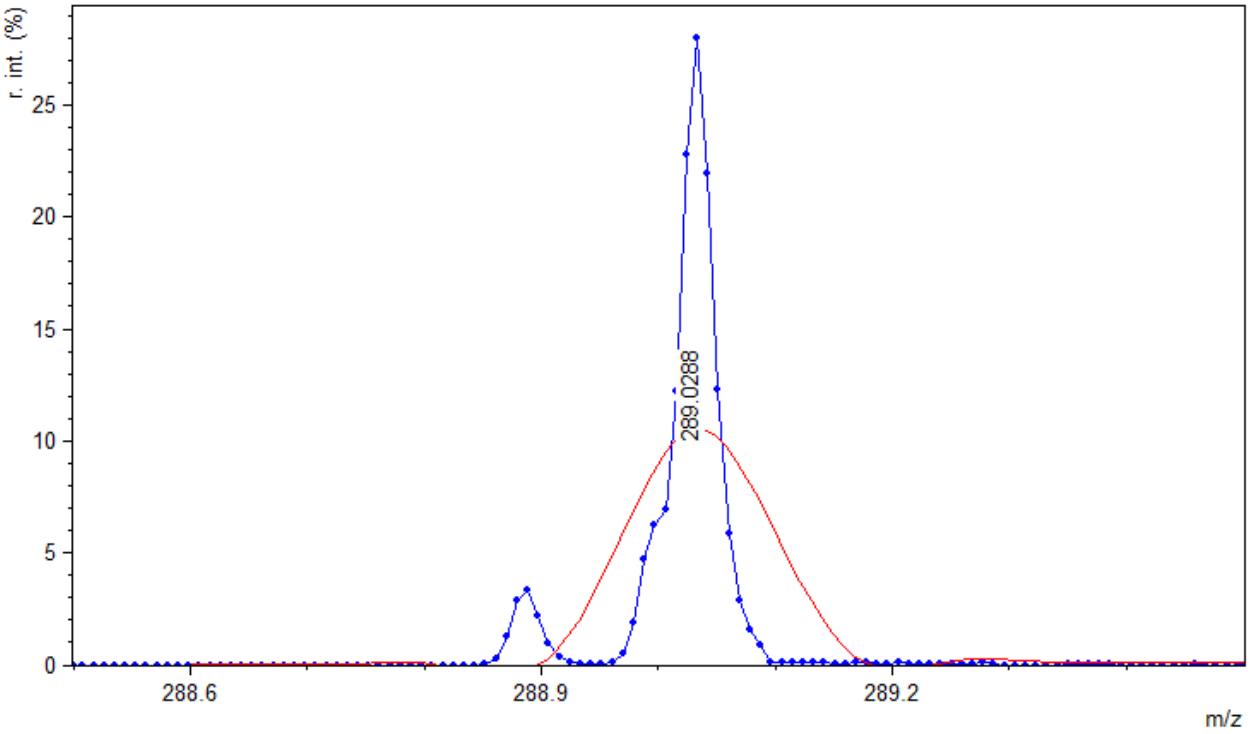


Figure A2.4. Intensity comparison between the ion at m/z 289 in the LDI spectrum of DHB only (Red) and the MALDI spectrum of the sample (Blue). The increase in the intensity in the MALDI spectrum (Blue) with respect to the LDI spectrum of the matrix only (RED) supports the presence of the sample and its contribution to the increase of the peak intensity.

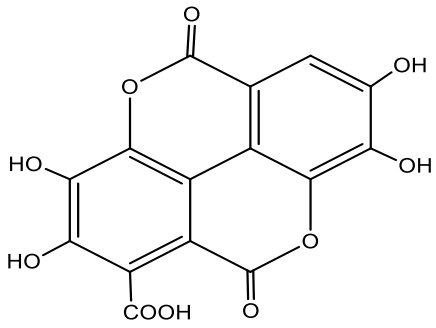
Mo. mass: 345.023951 Charge: 1 H+ Tolerance: 0.2 Da ppm Generate

Minimal formula: CHO Maximal formula:

Composition rules: H/C NOPS/C NOPS RDBE Integer RDBE Check isotopic pattern

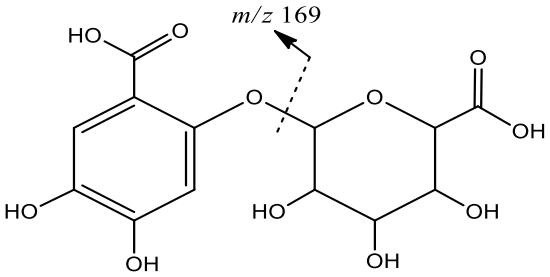
neutral formula	mass	m/z	error	H/C	rdbe	pattern
C5H2N19O ✗	344.07	345.08	-0.05	0.4	14.5	97.6
C15H4O10 ✓	343.98	344.99	0.04	0.3	14.0	89.2
C13H12O11 ✓	344.04	345.05	-0.02	0.9	8.0	89.2
C12H8O12 ✓	344.00	345.01	0.02	0.7	9.0	89.2
C14H2NO10 ✗	343.97	344.98	0.05	0.1	14.5	89.2
C14H16O10 ✓	344.07	345.08	-0.06	1.1	7.0	89.1

Figure A2.5. The search for possible chemical formulas for the precursor ion at *m/z* 345. The list shows the top six chemical formulas suggested for this ion and ranked according to their isotopic distribution pattern. It should be noted that formulas containing nitrogen are excluded from the investigation. The fragmentation pathway comparison between these possible formulas is demonstrated in Scheme A2.3 to decide which is the best formula to be assigned to the precursor ion at m/z 345.



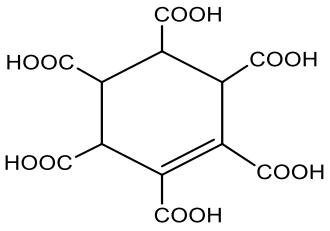
X [$C_{15}H_6O_{10}$ - H]$^-$

Neutral loss of 30 Da is not possible



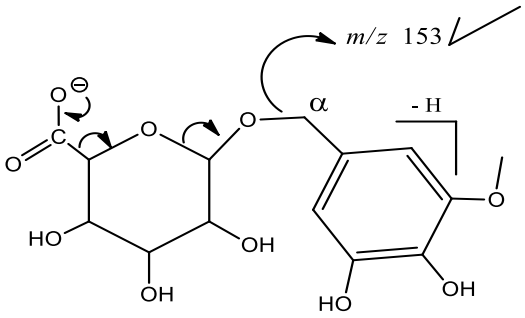
X [$C_{13}H_{14}O_{11}$- H]$^-$

The product ion at *m/z* 169 is more likely to be formed and not the ion at *m/z* 153



X [$C_{12}H_{10}O_{12}$- H]$^-$

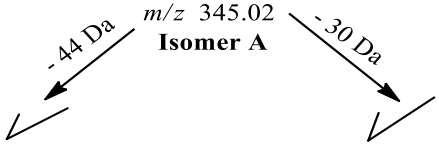
Formation of the product ion at *m/z* 153 is not possible

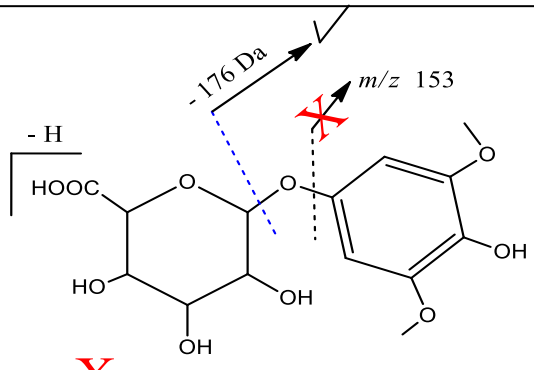


[$C_{14}H_{18}O_{10}$ - H]$^-$

m/z 345.02

Isomer A





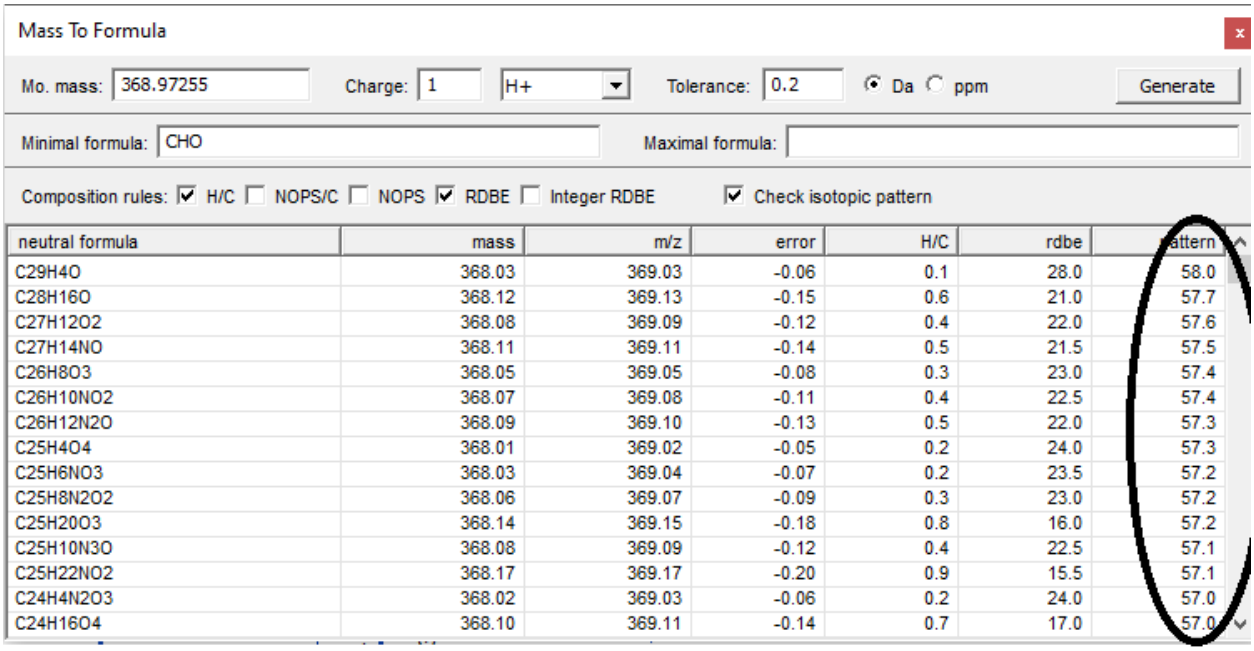
X **Isomer B**

[$C_{14}H_{18}O_{10}$ - H]$^-$

m/z 345.02

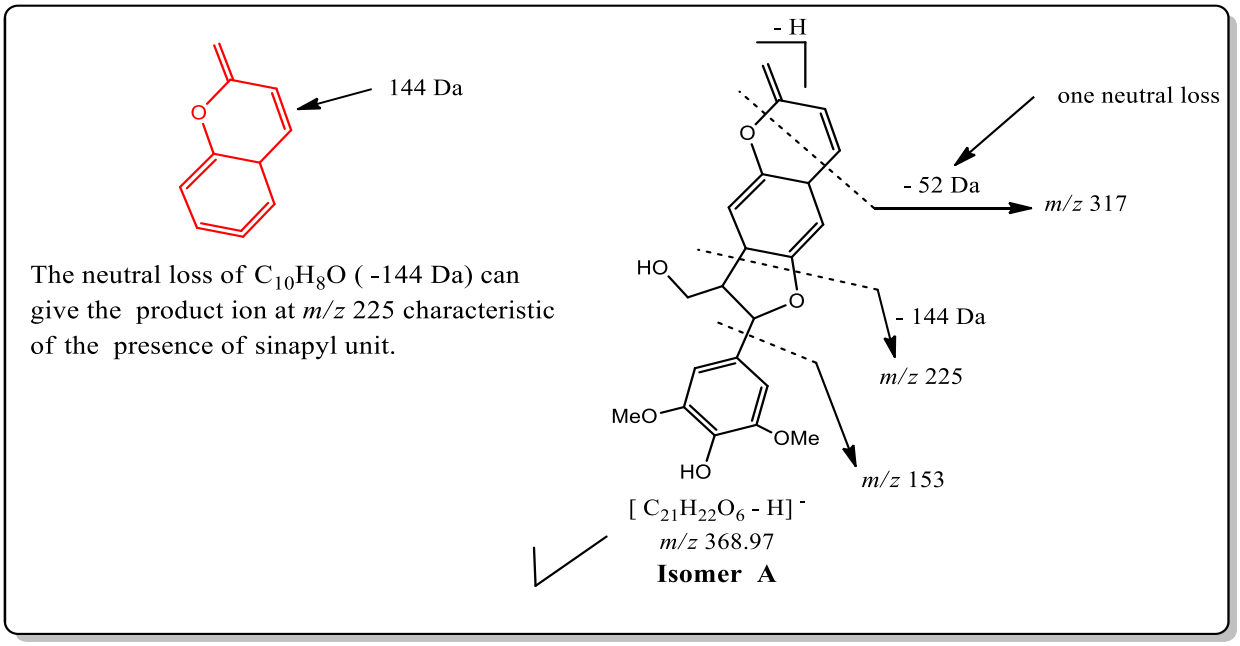
(This kind of glucuronides is characterized by the loss of 176 Da, so it is less likely to form the product ion at *m/z* 153)

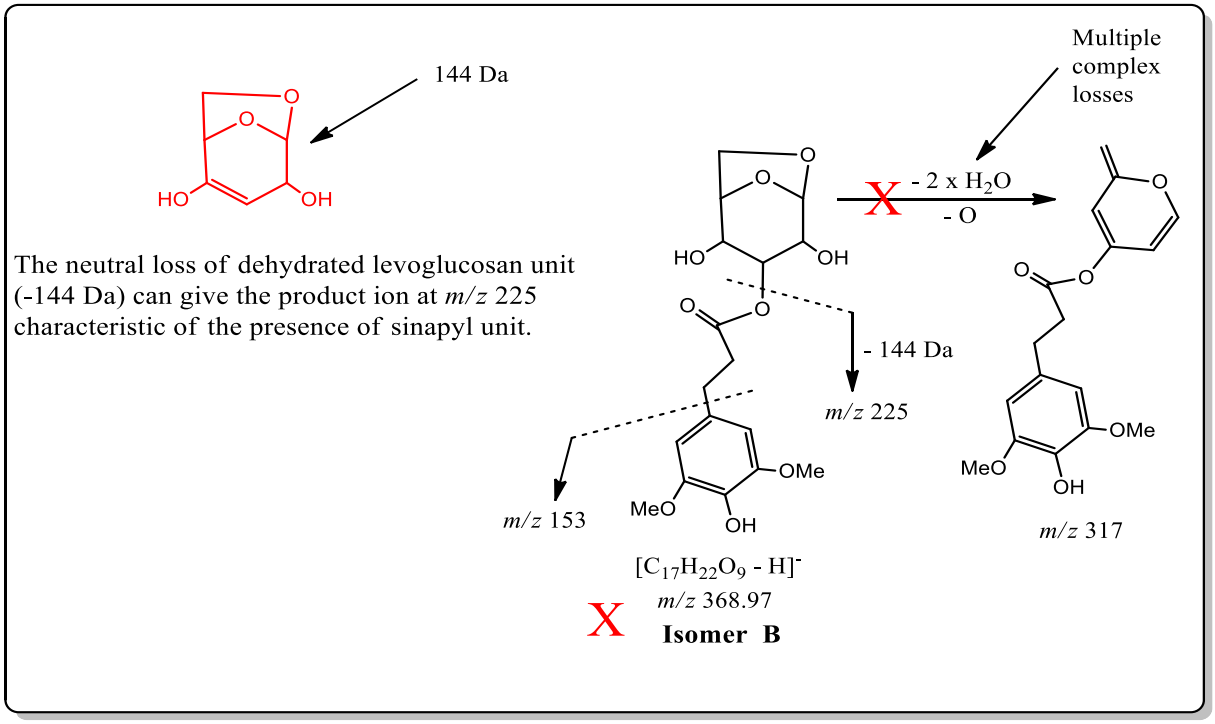
Scheme A2.3. Fragmentation pathway comparison between the top four possible candidates for the precursor ion at m/z 345 and how the best candidate has been chosen.



neutral formula	mass	m/z	error	H/C	rdbe	attern
C29H4O	368.03	369.03	-0.06	0.1	28.0	58.0
C28H16O	368.12	369.13	-0.15	0.6	21.0	57.7
C27H12O2	368.08	369.09	-0.12	0.4	22.0	57.6
C27H14NO	368.11	369.11	-0.14	0.5	21.5	57.5
C26H8O3	368.05	369.05	-0.08	0.3	23.0	57.4
C26H10NO2	368.07	369.08	-0.11	0.4	22.5	57.4
C26H12N2O	368.09	369.10	-0.13	0.5	22.0	57.3
C25H4O4	368.01	369.02	-0.05	0.2	24.0	57.3
C25H6NO3	368.03	369.04	-0.07	0.2	23.5	57.2
C25H8N2O2	368.06	369.07	-0.09	0.3	23.0	57.2
C25H20O3	368.14	369.15	-0.18	0.8	16.0	57.2
C25H10N3O	368.08	369.09	-0.12	0.4	22.5	57.1
C25H22NO2	368.17	369.17	-0.20	0.9	15.5	57.1
C24H4N2O3	368.02	369.03	-0.06	0.2	24.0	57.0
C24H16O4	368.10	369.11	-0.14	0.7	17.0	57.0

Figure A2.6. The search for possible chemical formulas for the precursor ion at *m/z* 368.9. The list shows that the top candidate has an isotopic distribution of 58.0%, which means that this ion has a complex overlapping isotopic distribution pattern. In this case, tandem mass spectrometry is much more needed to investigate the most likely structure for this ion as a low isotopic percentage will not help in assigning the best formulas





Scheme A2.4. Fragmentation pathway comparison between two possible candidates for the precursor ion at m/z 369 and how the best candidate has been chosen. It should be noted that the product ion at *m/z* 225 was the key to propose the structure of the precursor ion at m/z 368.9 as it is known to be characteristic for the presence of syringyl lignin units.

Appendix C. Supporting Information for Chapter 4

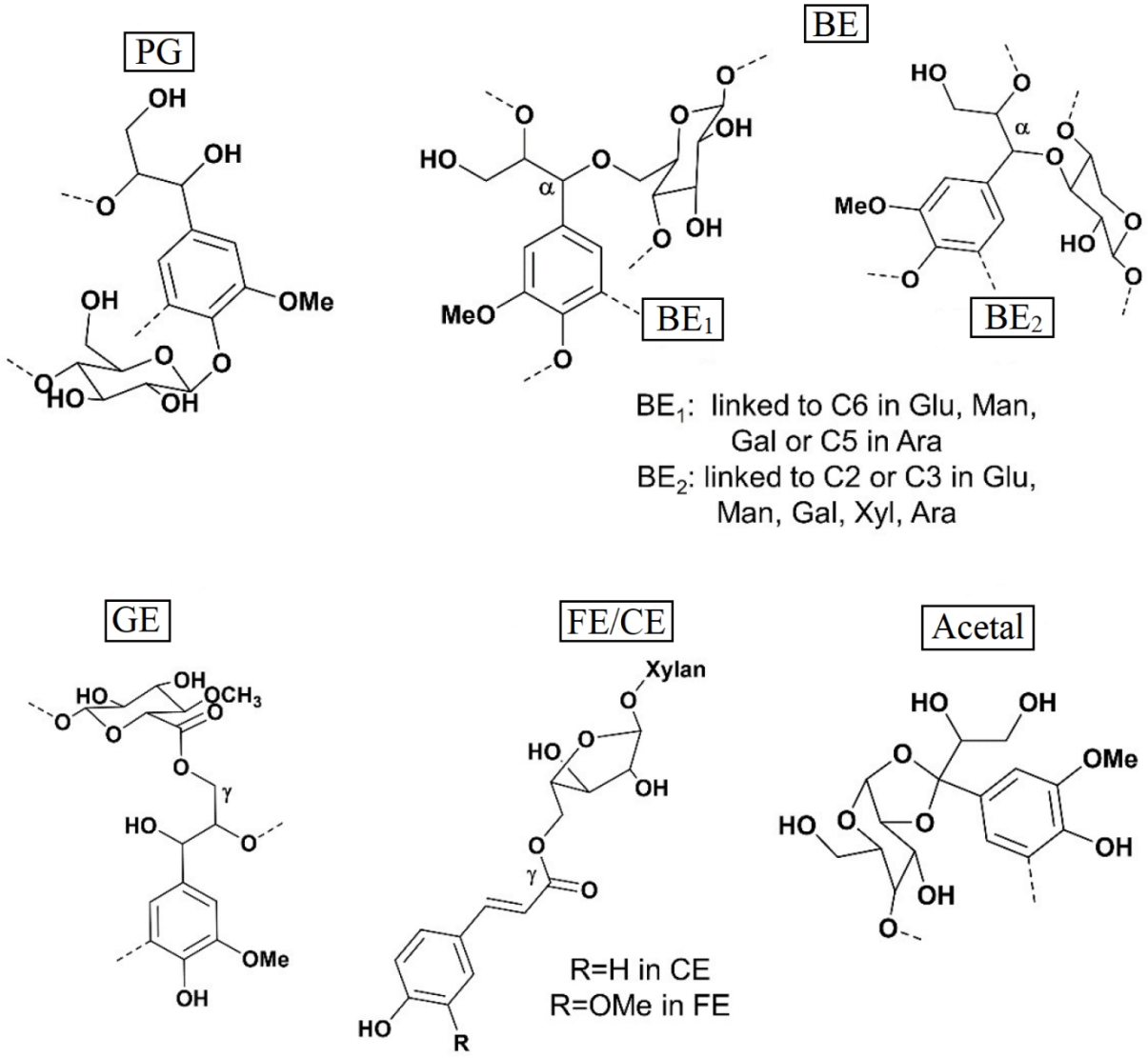


Figure A3.1. Suggested structures of Lignin-Carbohydrate bonds (LC) in wood and grass. In this figure, PG=phenyl glycosides, BE=benzyl ethers; GE= γ-esters; FE=ferulate esters, and CE=coumarate esters. Reprinted with permission from *Green Chem.* **2019**, *21*(7), 1573-1595.

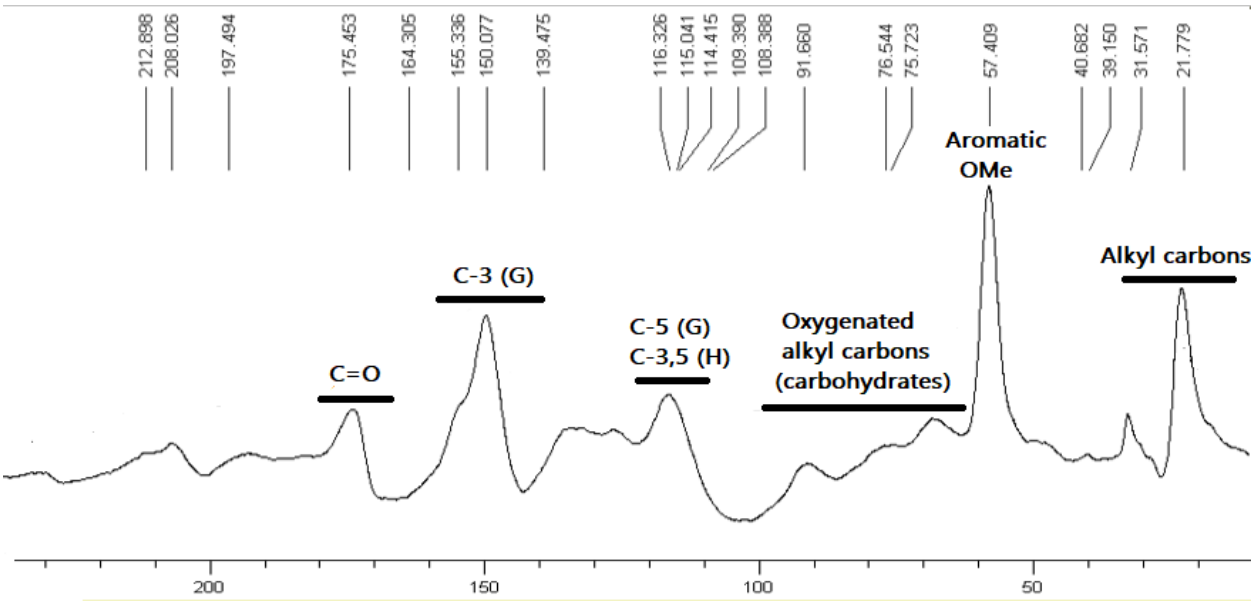


Figure A3.2. Solid-state ^{13}C- NMR of the extracted French Pine VRLs.

Table A3.1. Summary of the solid-state ^{13}C- NMR peaks

C-13 NMR shift regions and/or peaks	Possible Assignment	Reference
20-40 ppm	Alkyl carbons	1
57 ppm	Aromatic OMe	2
75-90 ppm	Aliphatic Carbons-O-Cα or Cβ or Cγ and/or oxygenated alkyl carbons in carbohydrates	3, 4
114-116 ppm	Aromatic C-3 and C-5 of H lignin or Aromatic C5 of G lignin	2
150 ppm	Aromatic C-5 of G unit	2
175 ppm	Aliphatic C=O	1

[1] Bianchi TS. *Biogeochemistry of Estuaries.* New York, NY: Oxford University Press; 2007.
[2] Fox SC, McDonald AG. Chemical and thermal characterization of three industrial lignins and their corresponding lignin esters. *BioResources.*2010; *5*(2): 990-1009.
[3] Baldock JA, Nelson PN. Soil Organic Matter, Section B. In Sumner ME, ed. *Handbook of Soil Science*. Boca Raton, FL: CRC Press; 2000: 25–84
[4] Ravenscrofta N, Bertib F. NMR characterization of bacterial glycans and glycoconjugate. In: Rauter AP, Christensen BE, Somsak L, Kosma P, Adamo R, eds. *Recent Trends in Carbohydrate Chemistry: Synthesis and Biomedical Applications of Glycans and Glycoconjugates.* Radarweg, Amsterdam: Elsevier; 2020: 239-281.

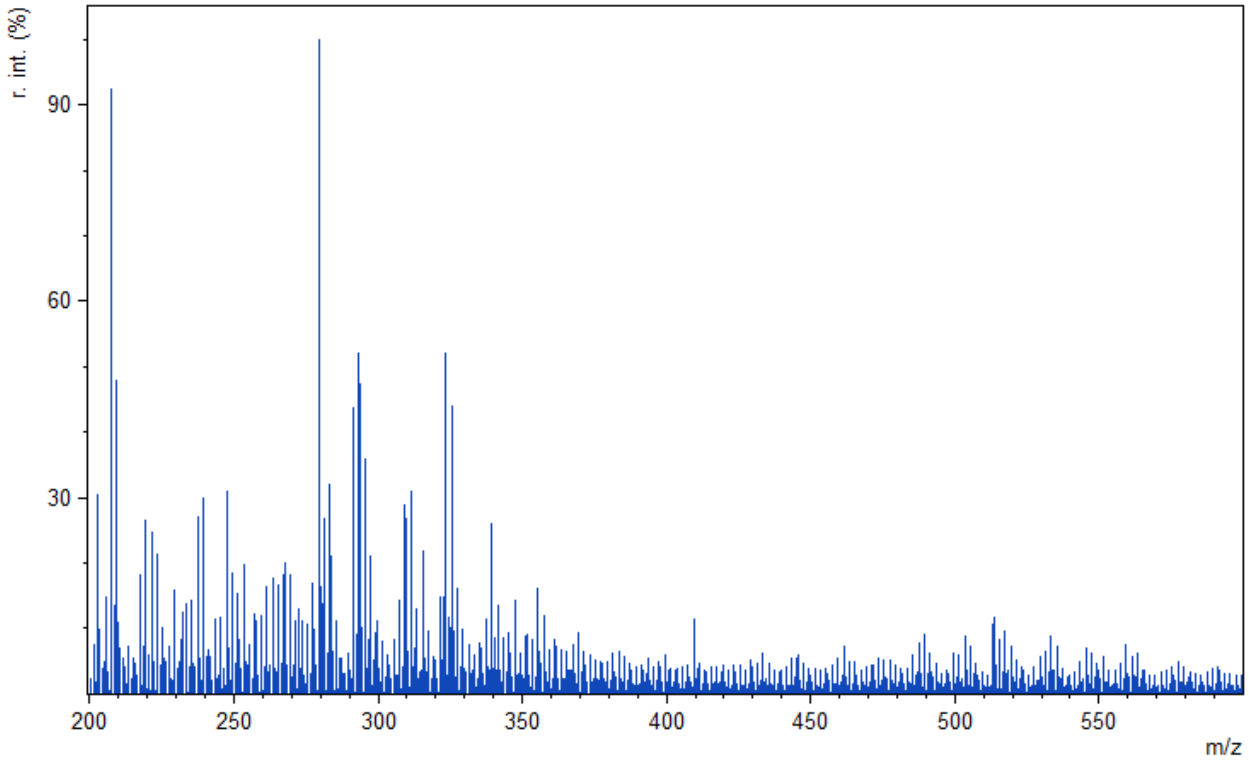


Figure A3.3a APPI-QqTOF-MS of the French pine lignin from *m/z* 200-600

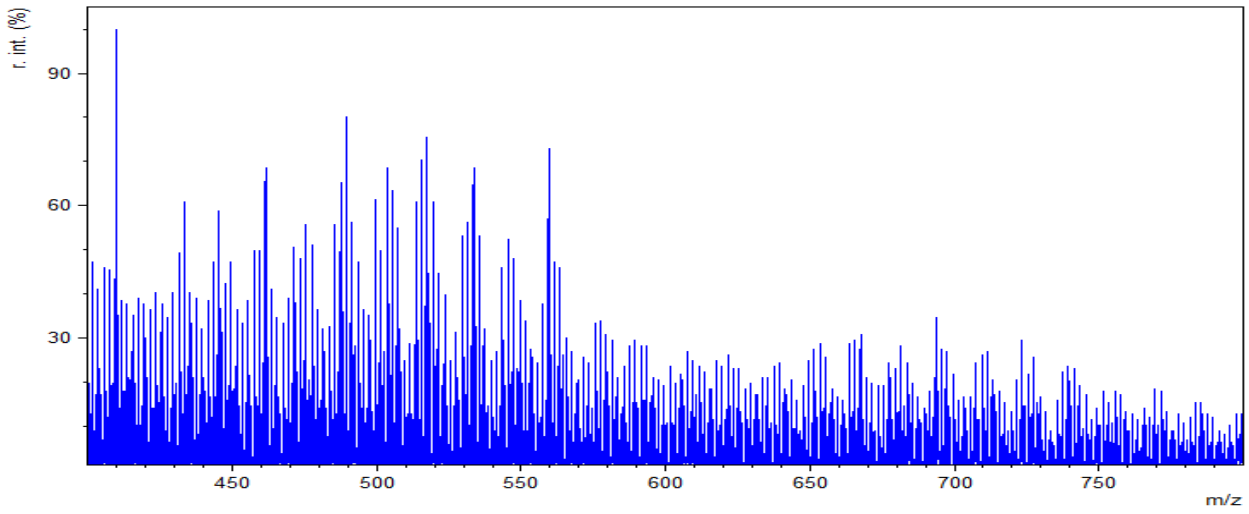


Figure A3.3b. APPI-QqTOF-MS of the French pine lignin recorded from *m/z* 400-800

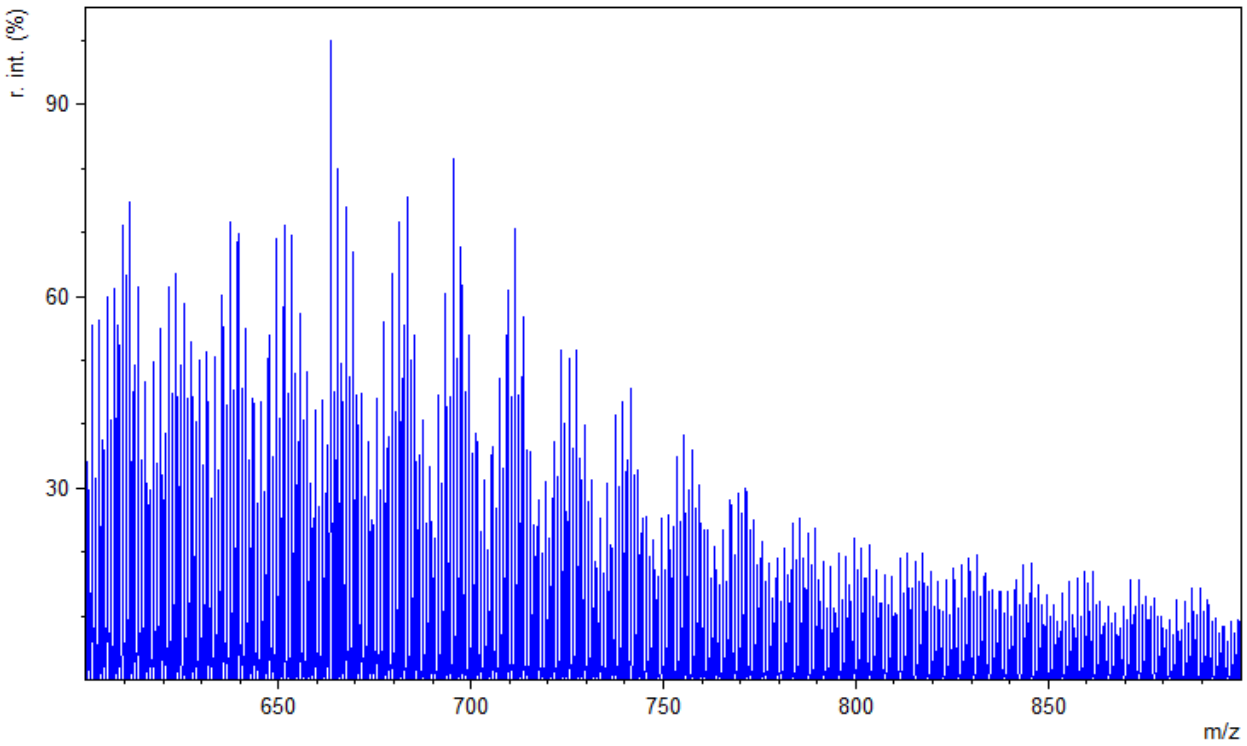


Figure A3.3c. APPI-QqTOF-MS of the French pine lignin recorded from *m/z* 600-900

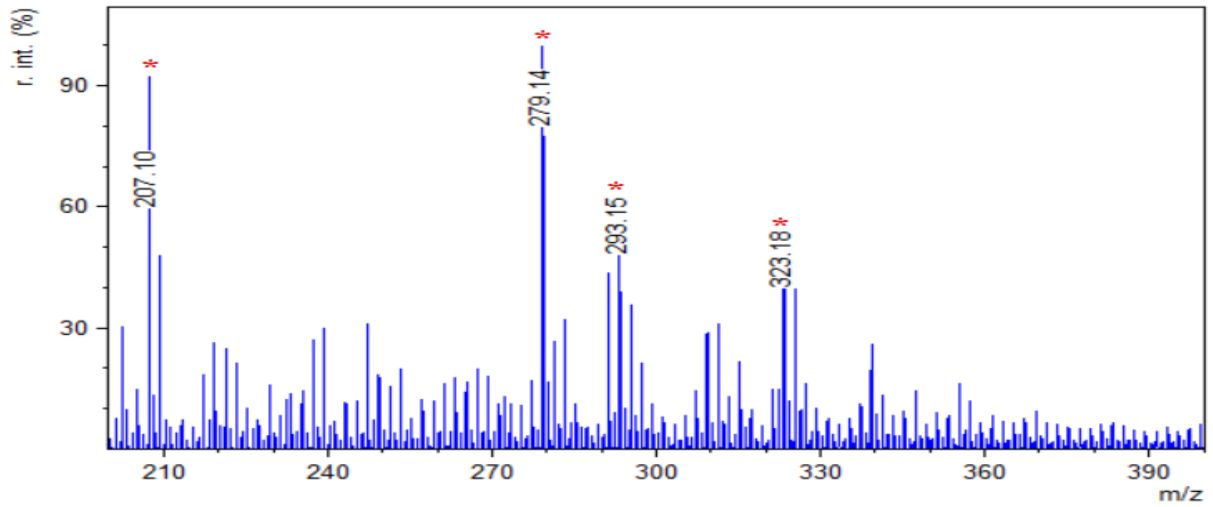


Figure A3.4a. Zooming of the *m/z* 200-400 range of the APPI-QqTOF-MS of the extracted French Pine VRLs.

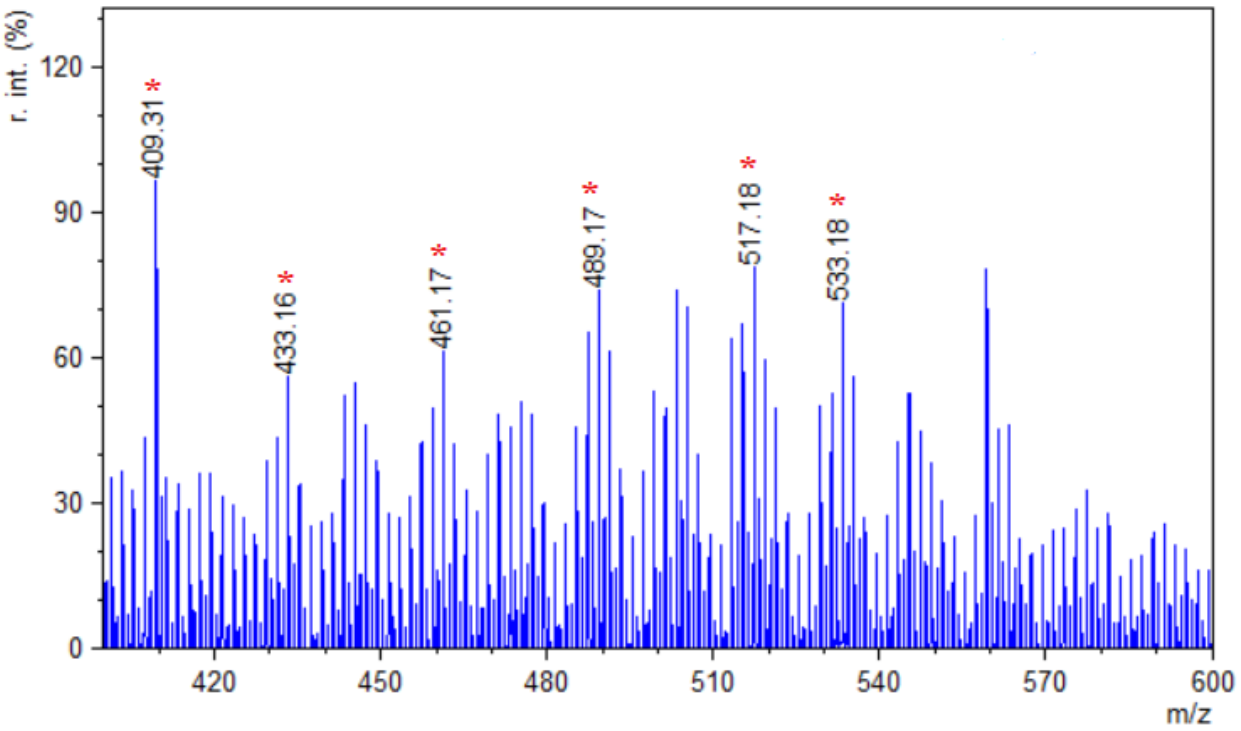


Figure A3.4b. Zooming of the *m/z* 400-600 range of the APPI-QqTOF-MS of the extracted French Pine VRLs.

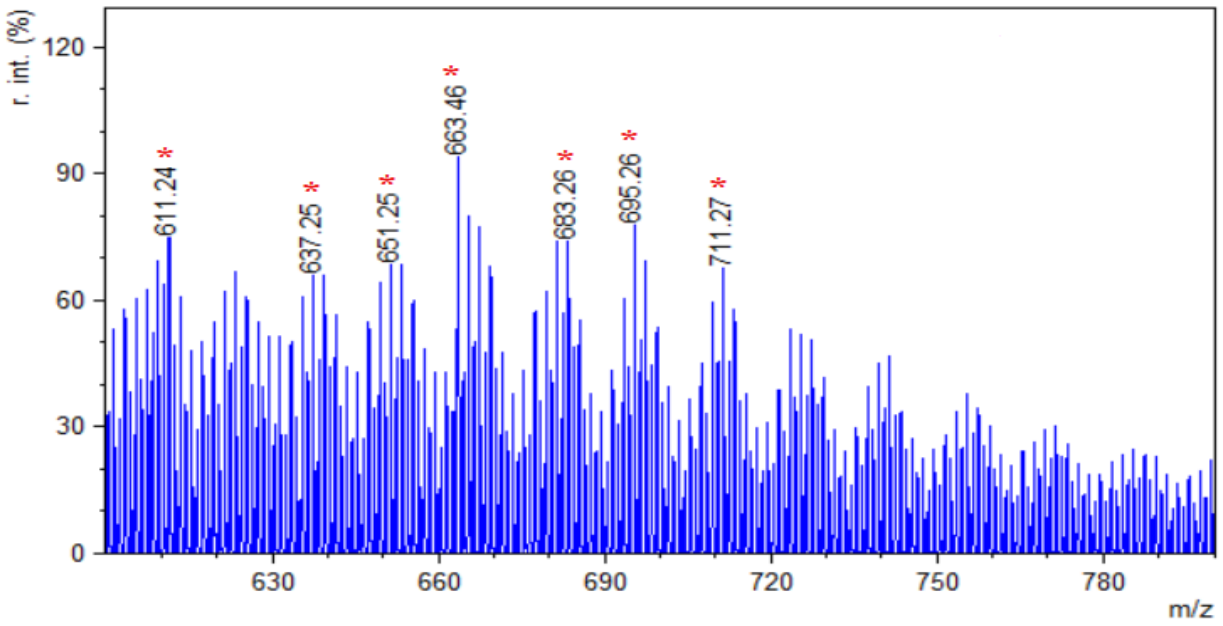


Figure A3.4c. Zooming of the *m/z* 600-800 range of the APPI-QqTOF-MS.

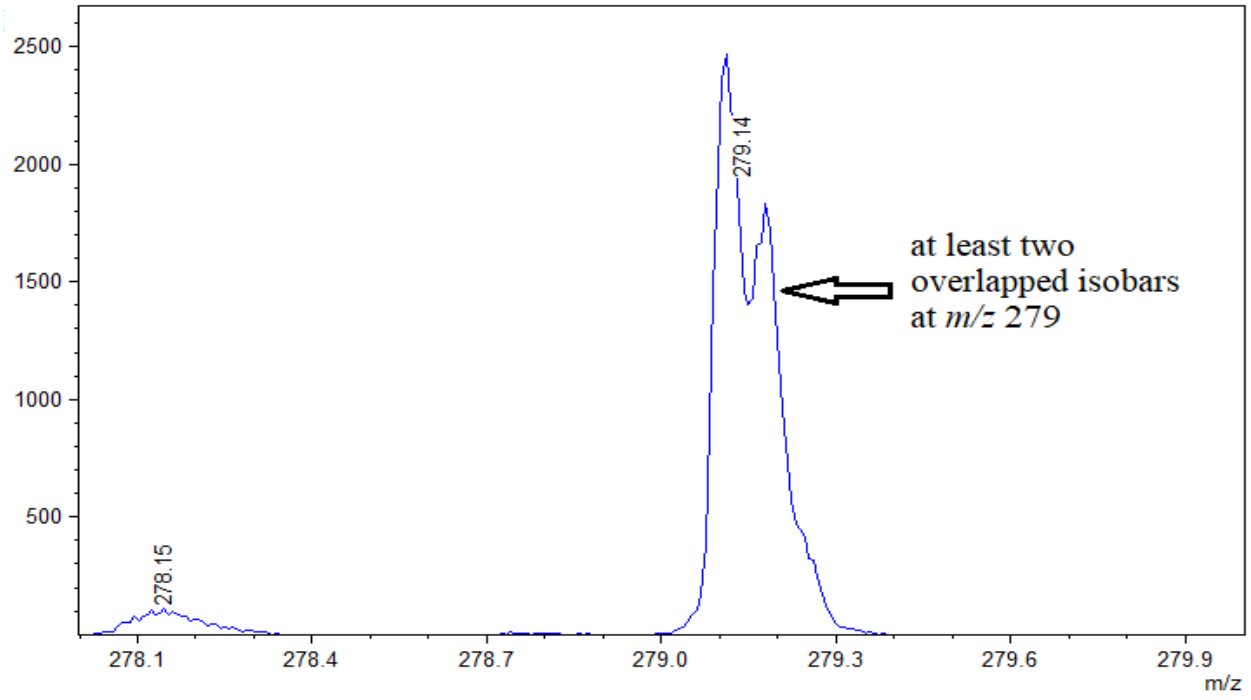
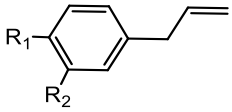
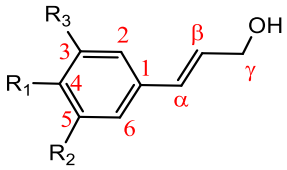


Figure A3.5. Magnification of the mass spectrum at *m/z* 279 as an example to show the complex overlapping between ions at every *m/z* unit.



R_1= $OCOCH_3$, R_2= OCH_3, acetylated eugenol (Ion *1*:*m/z* 207)



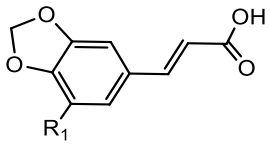
R_1= OH, R_2= H, R_3=H, coumaryl alcohol (Ions *8, 13*: *m/z* 517, 611)

R_1= OH, R_2= OCH_3, R_3=H, coniferyl alcohol (Ions *2, 11, 15*: *m/z* 279, 651, 711)

R_1= OH, R_2= OH, R_3=H, caffeyl alcohol (Ion *13: m/z* 611)

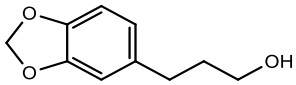
R_1= OH, R_2= $OCOCH_3$, R_3=H, acetylated caffeyl alcohol (Ion *4: m/z* 323)

R_1= OH, R_2= OCH_3, R_3=H, 5-hydroxy coniferyl alcohol (Ions *12*: *m/z* 695)

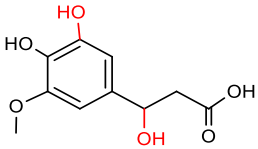


R_1=H, 3,4-(methylenedioxy) cinnamic acid (Ions *8, 9, 11, 12, 15: m/z* 517, 533, 651, 695, 711)

R_1=OH, 5- hydroxy 3,4-(methylenedioxy) cinnamic acid (Ions *15: m/z* 711)



3,4-methylenedioxy cinnamyl alcohol (Ions *12*: *m/z* 695)



3-(3,4-dihydroxy-5-methoxyphenyl)-3-hydroxypropanoic acid (Ion *9*: *m/z* 533)
(dihydroxy dihydroferulic acid)

Figure A3.6a. Summary of the lignin derivatives nomenclature used to describe the lignin moieties of the proposed LCC precursor ions. The general numbering system of lignin units is shown on one of the structures. Also, the precursor ions that contain each of these units is written between brackets

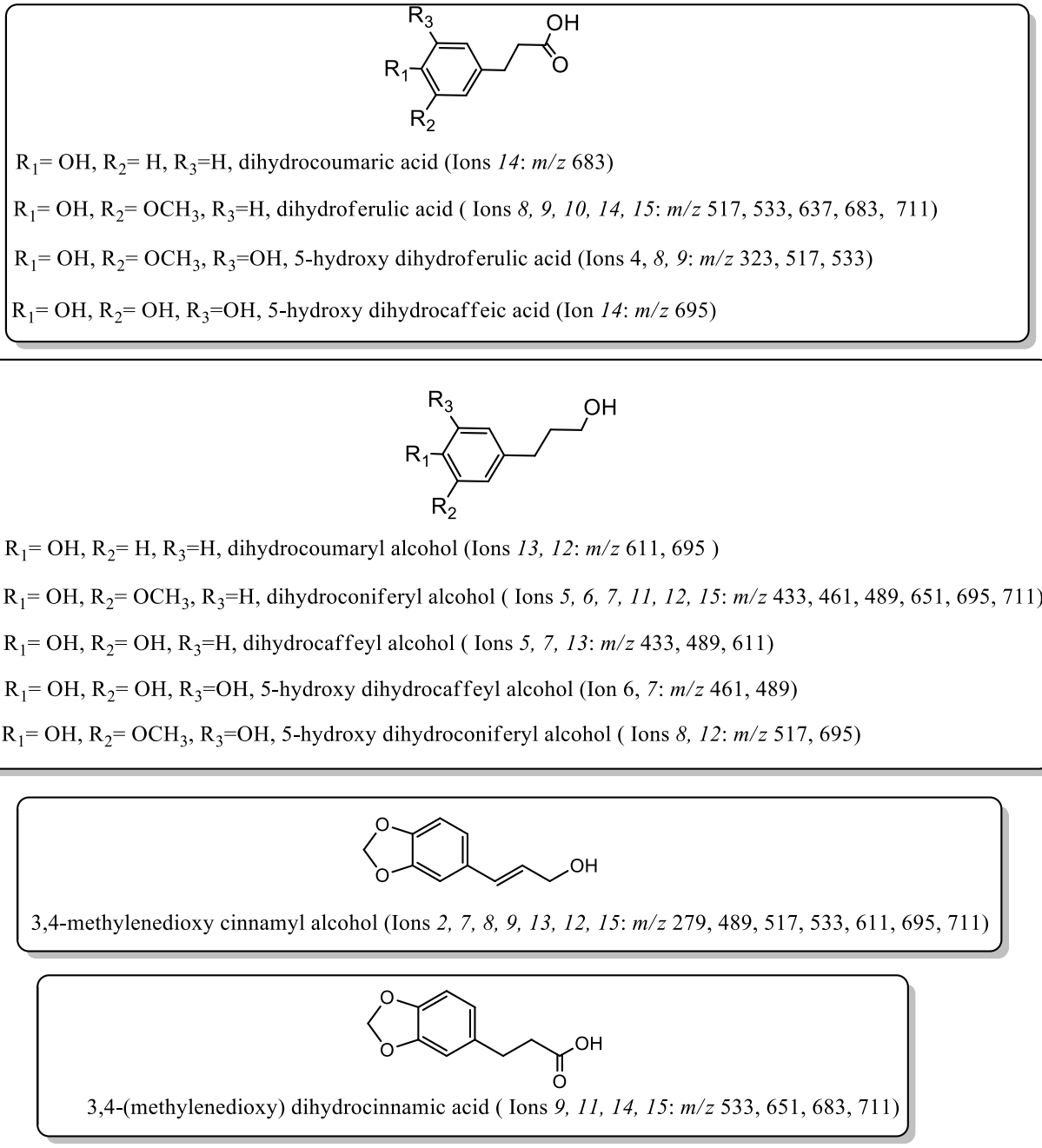


Figure A3.6b. Summary of the lignin derivatives nomenclature used to describe the lignin moieties of the proposed LCC precursor ions. The general numbering system of lignin units is shown on one of the structures. Also, the precursor ions that contain each of these units is written between brackets

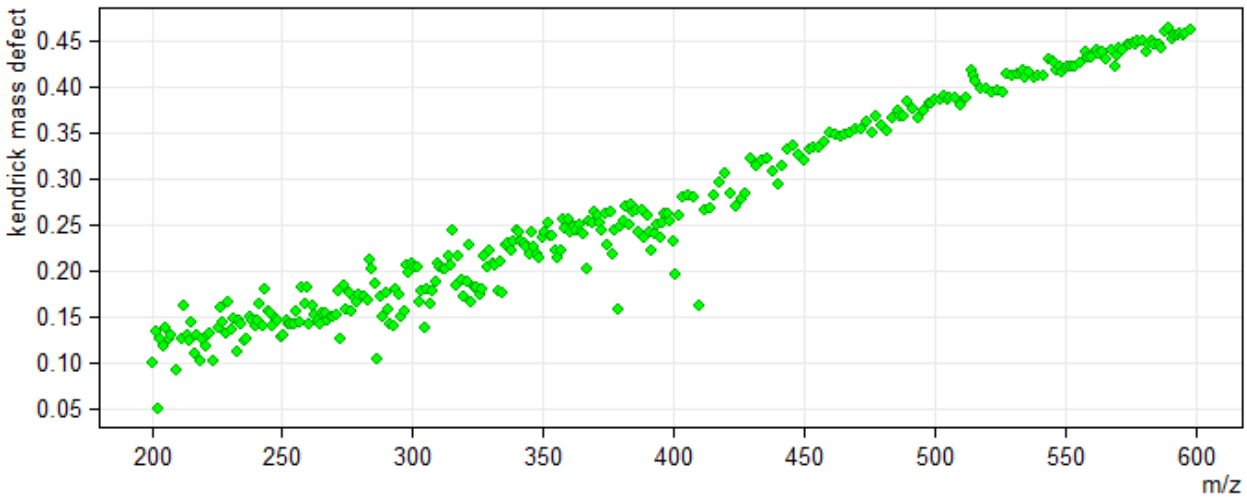


Figure A3.7a. Kendrick mass defect for the spectrum in the *m/z* 200-600 range of the APPI-QqTOF-MS of extracted French Pine VRLs.

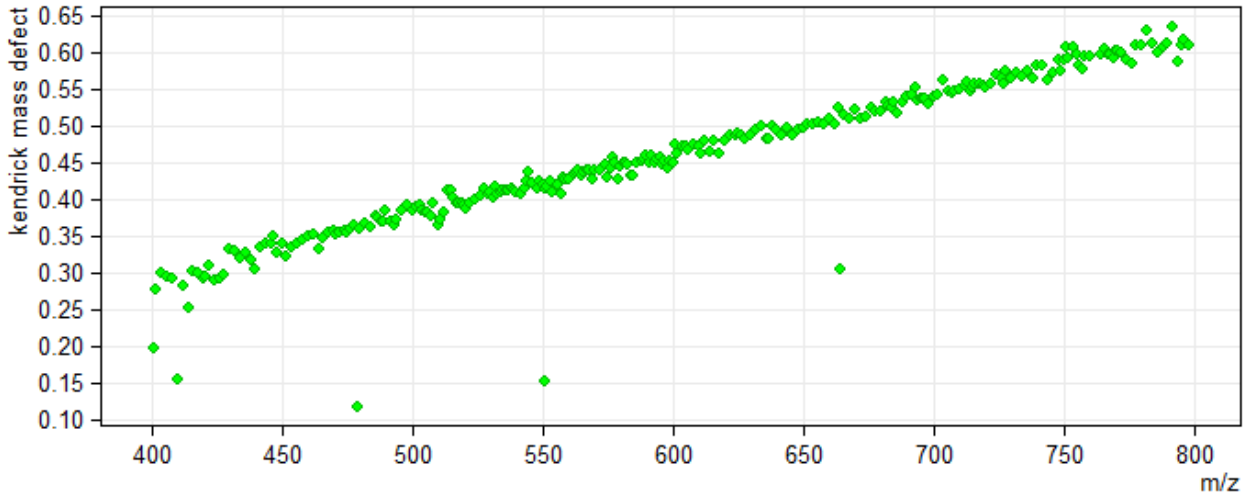


Figure A3.7b. Kendrick mass defect for the spectrum in the *m/z* 400-800 range of the APPI-QqTOF-MS of the extracted French Pine VRLs.

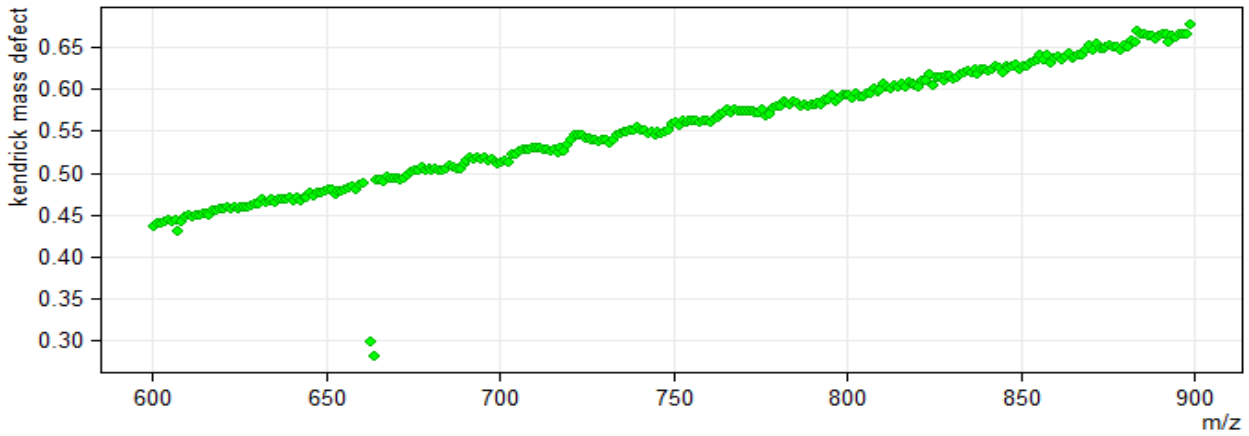


Figure A3.7c. Kendrick mass defect for the spectrum in the *m/z* 600-900 range of the APPI-QqTOF-MS of the extracted French Pine VRLs.

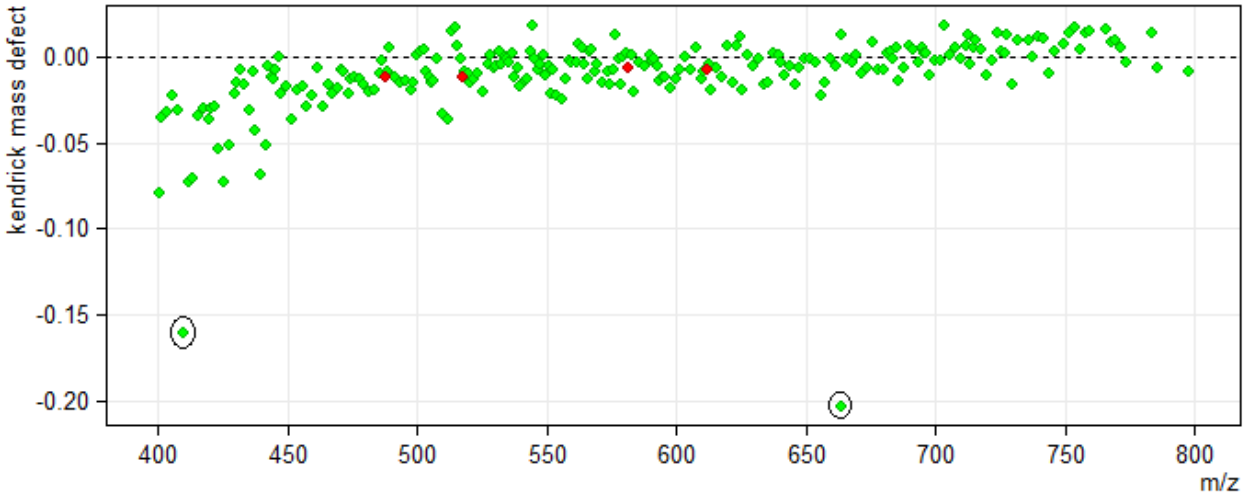


Figure A3.8a. Kendrick mass defect using [OCH2] units for the spectrum in the *m/z* 400-800 range of the APPI-QqTOF-MS of the VRLs French pine. The circled ions are the ions at m/z 409 and 663, which an entirely different mass defect than the mass defect range of the major bundle of ions.

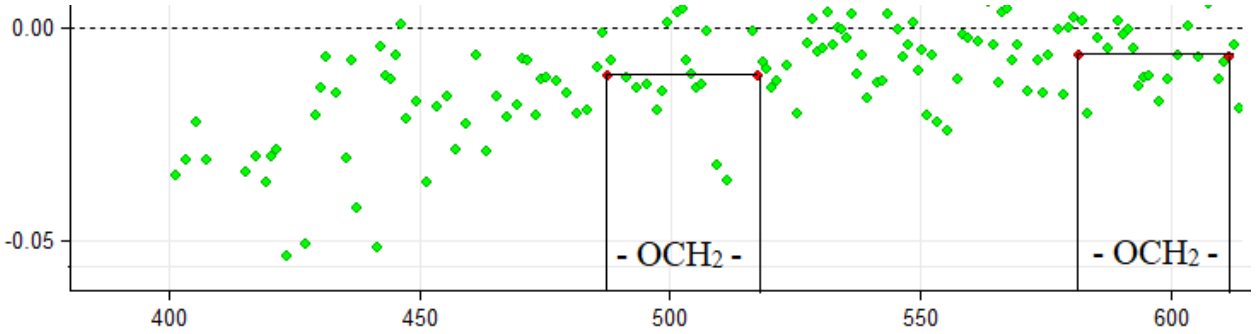


Figure A3.8b. Zooming of the Figure SI-4 to show the alignment of the ions at *m/z* 487 and *m/z* 517, and the ions at *m/z* 581 and *m/z* 611.

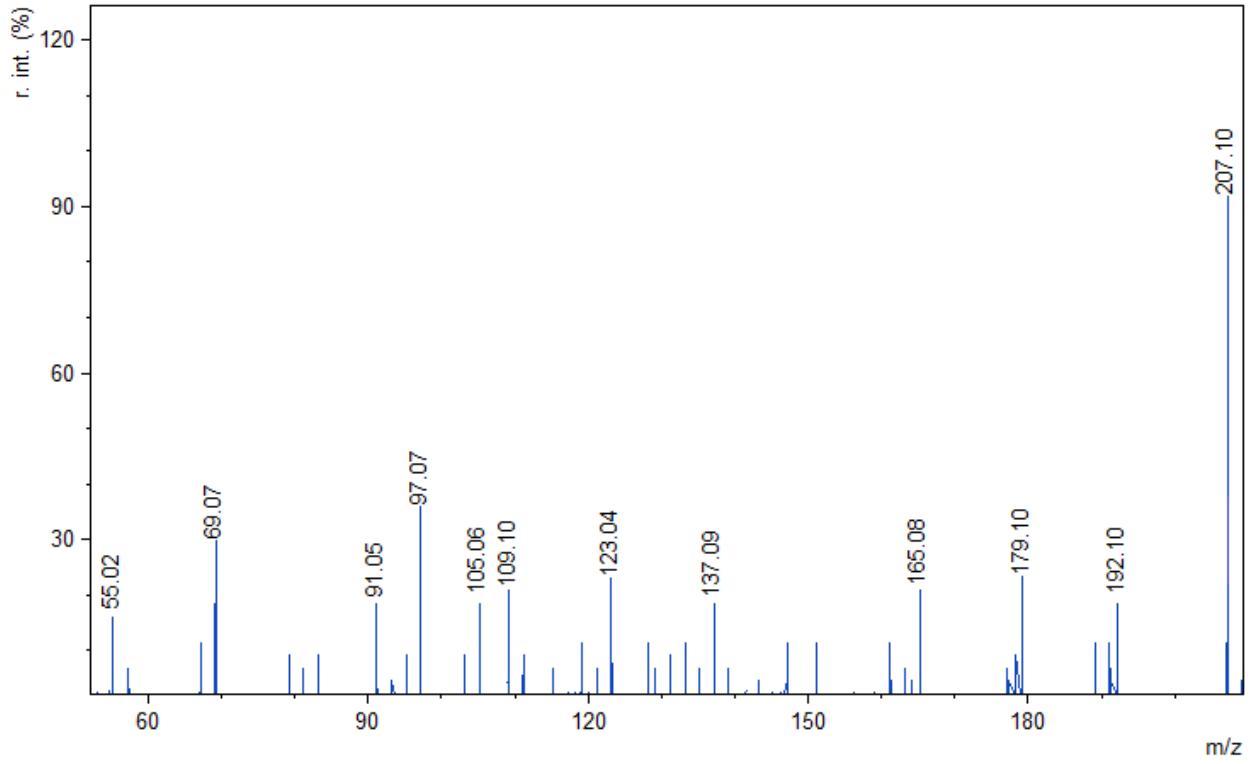
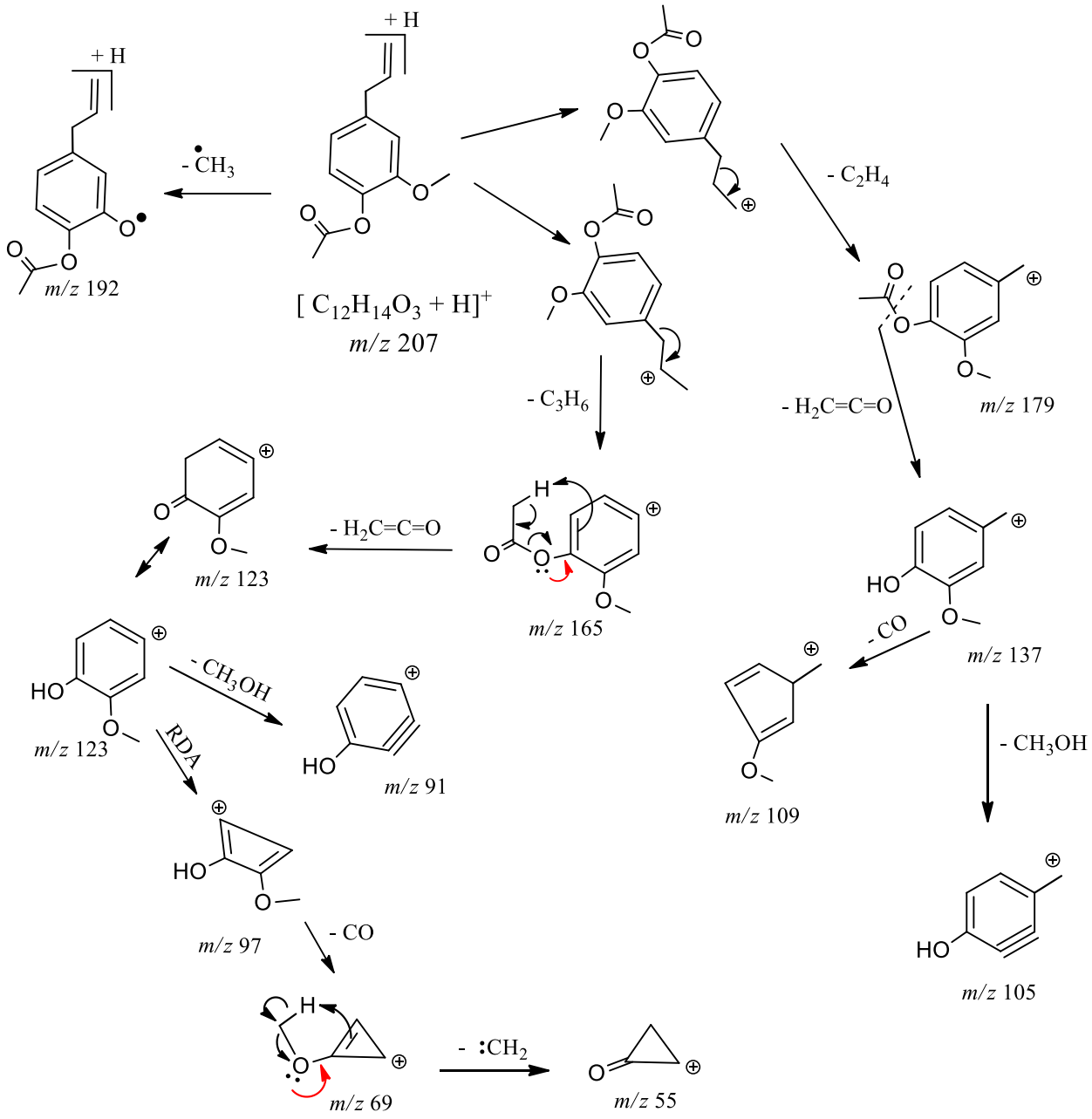


Figure A3.9. Product ion scan of the protonated precursor ion 1 at *m/z* 207.10



Scheme A3.1. Low energy CID-MS/MS fragmentation mechanism of the protonated precursor ion *1* at *m/z* 207.10

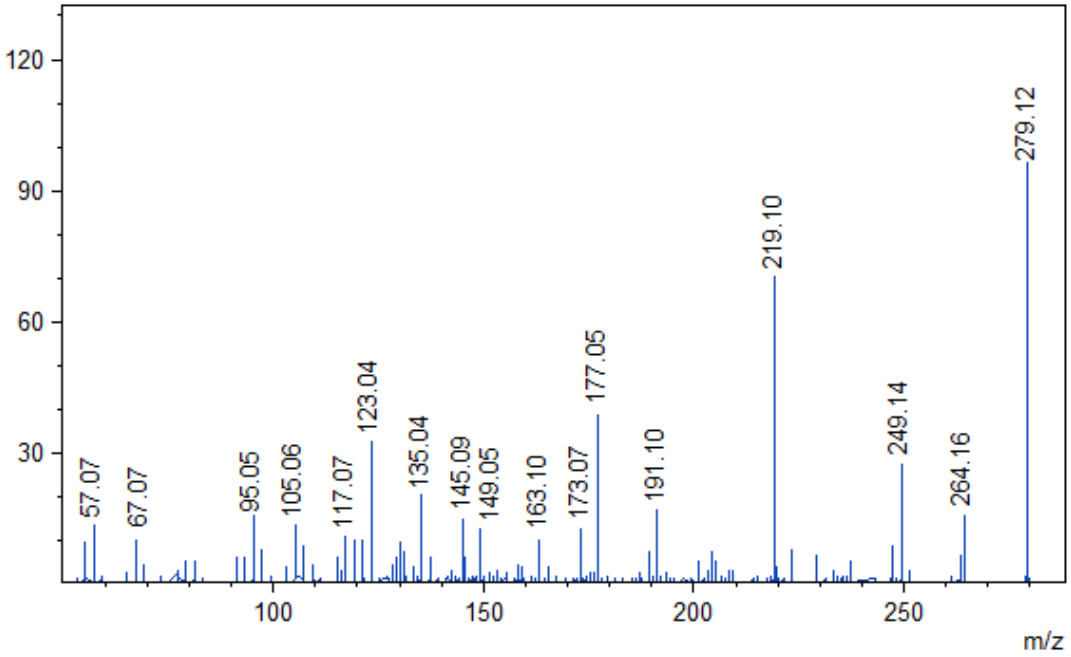
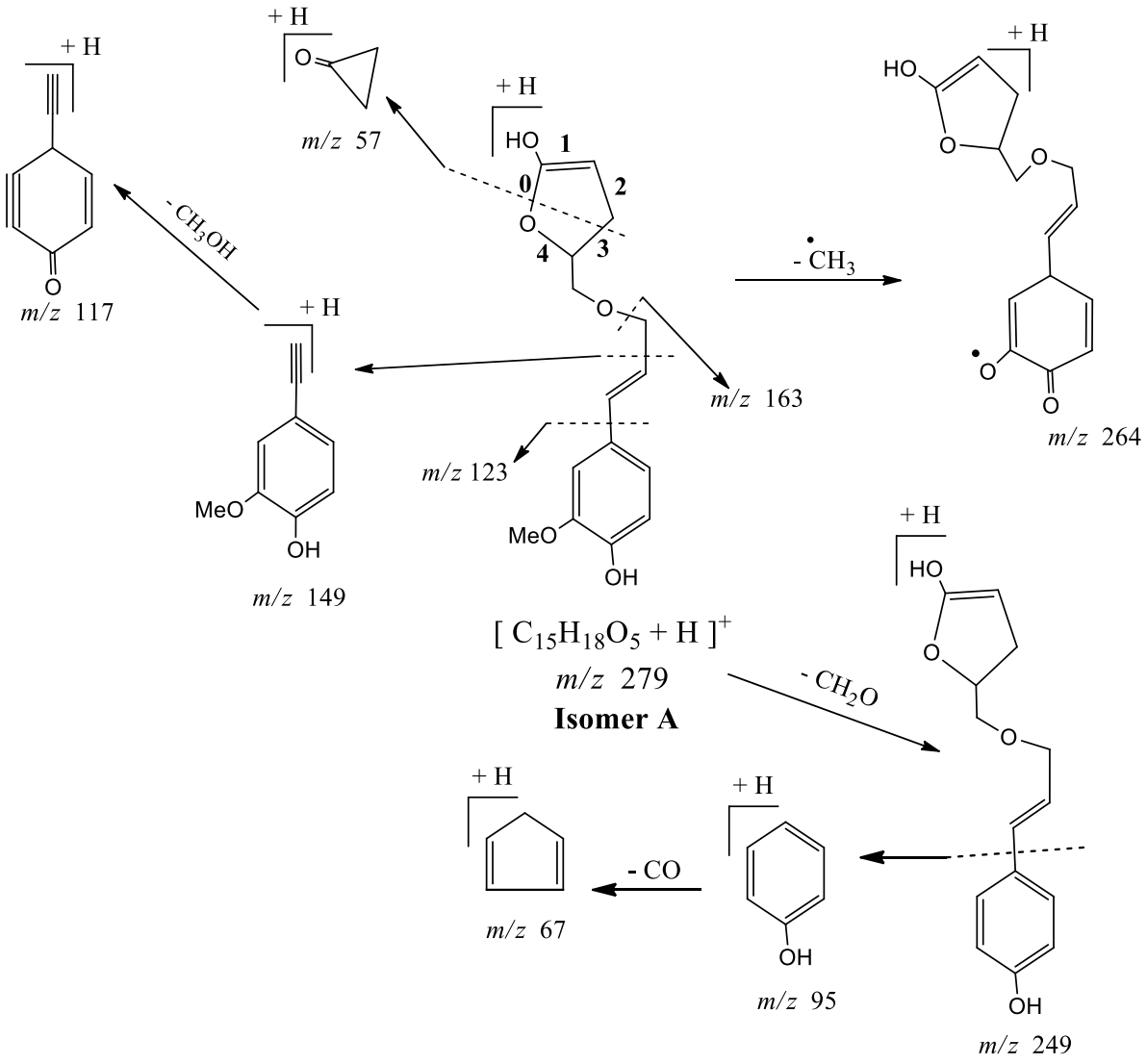
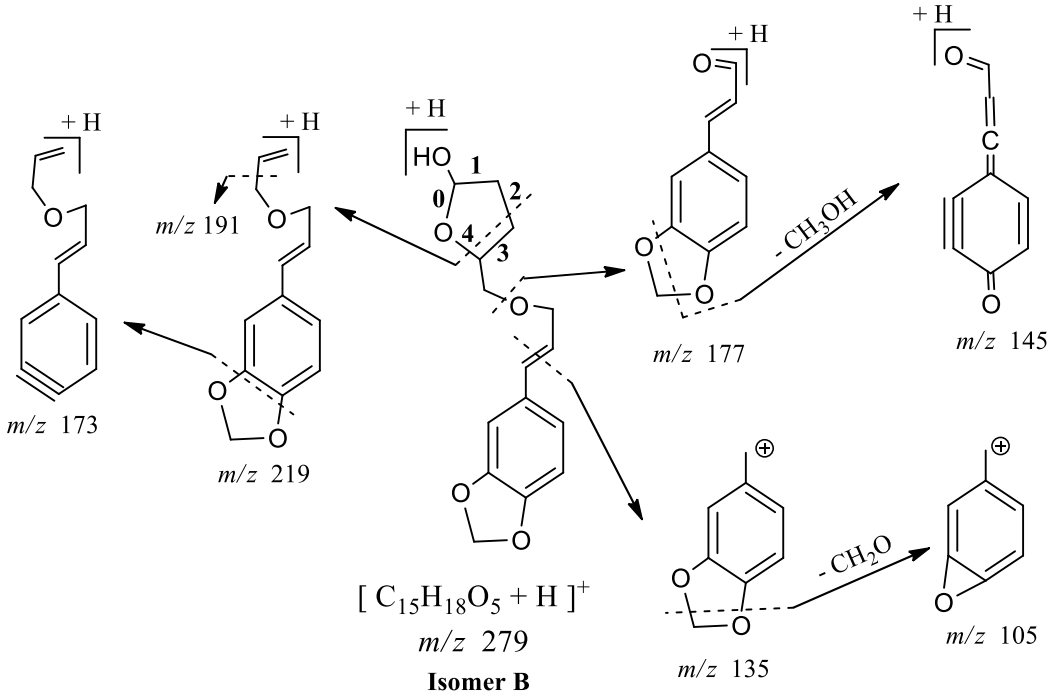


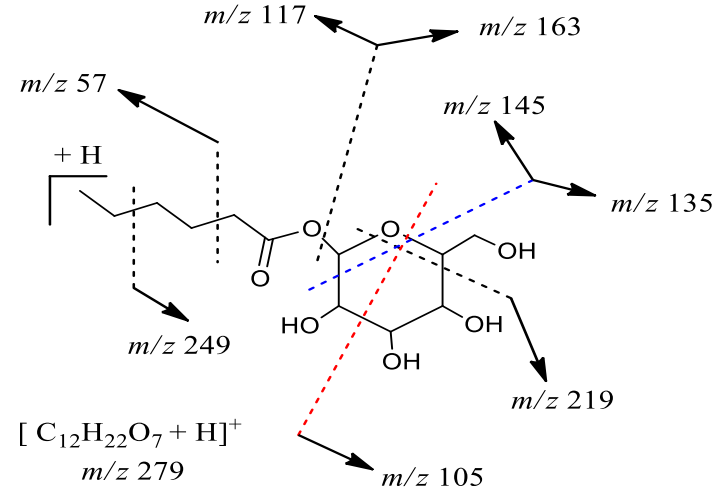
Figure A3.10. Product ion scan of the protonated precursor ion *2* at *m/z* 279.14



Scheme A3.2a. Low energy CID-MS/MS fragmentation mechanism of the protonated precursor ion *2* at *m/z* 279.14 (Isomer A)



Scheme A3.2b. Low energy CID-MS/MS fragmentation mechanism of the protonated precursor ion *2* at *m/z* 279.14 (Isomer B)



Different Isobar

Loss of methyl radical X

Formation of aromatic ions *m/z* 177, *m/z* 149, *m/z* 123 X

Formation of protonated phenol $C_6H_7O^+$ at *m/z* 95 X

Scheme A3.3. Fragmentation mechanism of the possible precursor ion *2* Isobar with the formula $[C_{12}H_{22}O_7 + H]^+$

A3.1- The product ion scan of LCC composed of a glycose residue attached to one lignin unit

A) The product ion scan of the protonated molecule *3* at *m/z* 293.15

The formula $[C_{16}H_{20}O_5 + H]^+$ (DBE=6.5) was assigned to the protonated molecule *3* at *m/z* 293.15 (Figure A3.11 and Scheme A3.4). This novel protonated molecule was found to be composed of O-methyl cymaroside (methyl 2,6-dideoxy-3-O-methyl-ß-D-ribo-hexopyranoside) attached to the aromatic moiety benzofuranol. It is suggested that the benzofuranol moiety resulted from the degradation of two lignin units attached through [β-5(α-O-4)] linkage (phenylcoumaran).

The product ion scan of precursor ion *3* showed the sequential loss of two methanol molecules to yield the product ions at *m/z* 261 and *m/z* 229, respectively. Please note that these sequential methanol losses occurred in the sugar moiety, as shown in Scheme A3.4. The precursor ion 3 showed a sugar ring cleavage (-60 Da) to yield the product ion at *m/z* 233 ($^{1,5}X$). Lastly, the loss of the sugar moiety from the precursor ion *3* leads to the creation of the product ion at *m/z* 135, as demonstrated in Scheme A3.4.

Lastly, another isomer B composed of benzofuranol attached to a different sugar isomer to discuss the formation of the product ion at *m/z* 57 ($C_3H_5O^+$) as shown in scheme A3.4. It should be noted that there is another possibility to assign the product ion at *m/z* 57 as $C_4H_9^+$ that can be formed from another isomer and/or isobar containing a butyl chain.

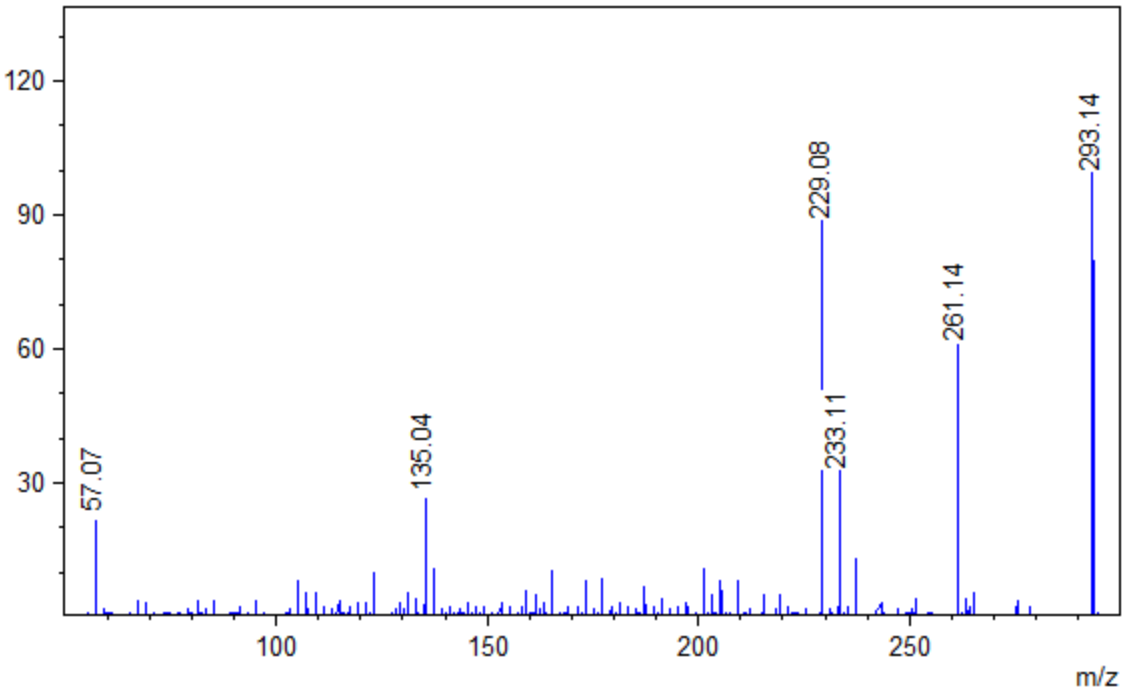
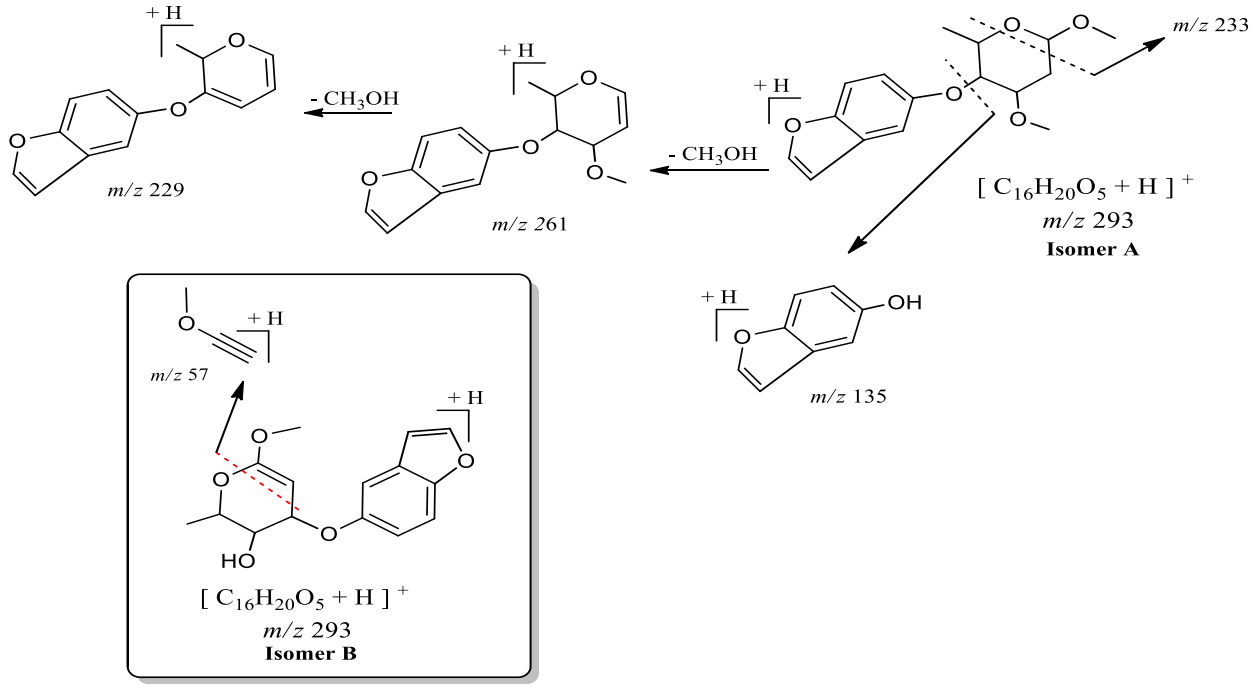


Figure A3.11. Product ion scan of the protonated precursor ion *3* at *m/z* 293.15



Scheme A3.4. Low energy CID-MS/MS fragmentation mechanism of the protonated precursor ion *3* at *m/z* 293.15

b) Product ion scan of the protonated molecule *4* at *m/z* 323.18

The formula $[C_{16}H_{18}O_7 + H]^+$ (DBE=7.5) was assigned to the protonated ion *4* at *m/z* 323.18, and it was tentatively deduced to be composed of a dehydrated pentose moiety attached to an acetylated caffeyl alcohol (Lignin C unit) (Figure A3.12 and Scheme A3.5). Once more, this protonated ion can be represented by another isomer B composed of sugar derivative furaneol attached to 5-hydroxy dihydroferulic acid unit. This was thought to be necessary to explain the messy MS/MS of this precursor ion and to account for the most abundant product ions.

The product ion scan of the precursor ion *4* at *m/z* 323.18 (Isomer A) indicated the loss of one methanol molecule to produce the product ion at *m/z* 291. As well, this product ion scan can occur by two different sugar ring cleavages, to create the product ions at *m/z* 263 ($^{0,3}X$) and *m/z* 249 ($^{0,2}X$).[43] This latter ion at *m/z* 249 loses 60 Da in the form of CH_3COOH to produce the secondary ion at *m/z* 189. This latter ion can show two consecutive loss of a ketene molecule (- CH_2=C=O) and a carbon monoxide molecule to yield the product ions at *m/z* 147 and *m/z* 119, respectively.

It should be noted that the product ion at *m/z* 263 can be formed directly from precursor ion *4* through the loss of one molecule of acetic acid instead of a sugar ring cleavage. This latter ion at *m/z* 263 loses one water molecule from the sugar ring to yield the product ion at *m/z* 245. Moreover, this product ion scan of the precursor ion at *m/z* 323 (Isomer A) also can be initiated by the cleavage of the C1-O bond present in the sugar moiety to yield the aromatic lignin unit (acetylated caffeyl alcohol) at *m/z* 207. Also, Isomer A can experience

the cleavage of the Cα=Cβ in the lignin unit to create the benzylic cation at *m/z* 165, as shown in Scheme A3.5.

The Isomer B of the precursor ion at *m/z* 323.18 was proposed to describe the product ions at *m/z* 128 and *m/z* 57. This isomer can experience the cleavage of the ester bond between the 5-hydroxy dihydroferulic acid unit and furaneol to give the furaneol radical cation at *m/z* 128. As well, Isomer B can be subjected to a sugar ring cleavage to produce the product ion at *m/z* 57. As well, there is another possibility to assign the product ion at *m/z* 57 as $C_4H_9^+$ that can be formed from another possible isomer and/or isobar containing a butyl chain.

It is significant to mention that the gas-phase separation of these two isomeric precursor ions could not be achieved with our QqTof-MS/MS instrument. However, we anticipate that the separation of these two precursor ions could be realized by using ion mobility mass spectrometry.

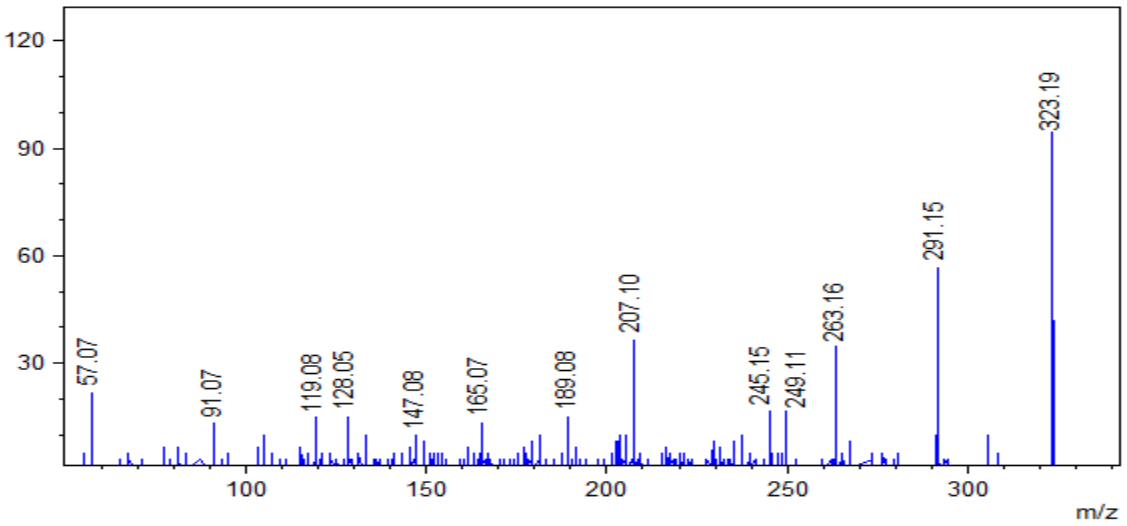
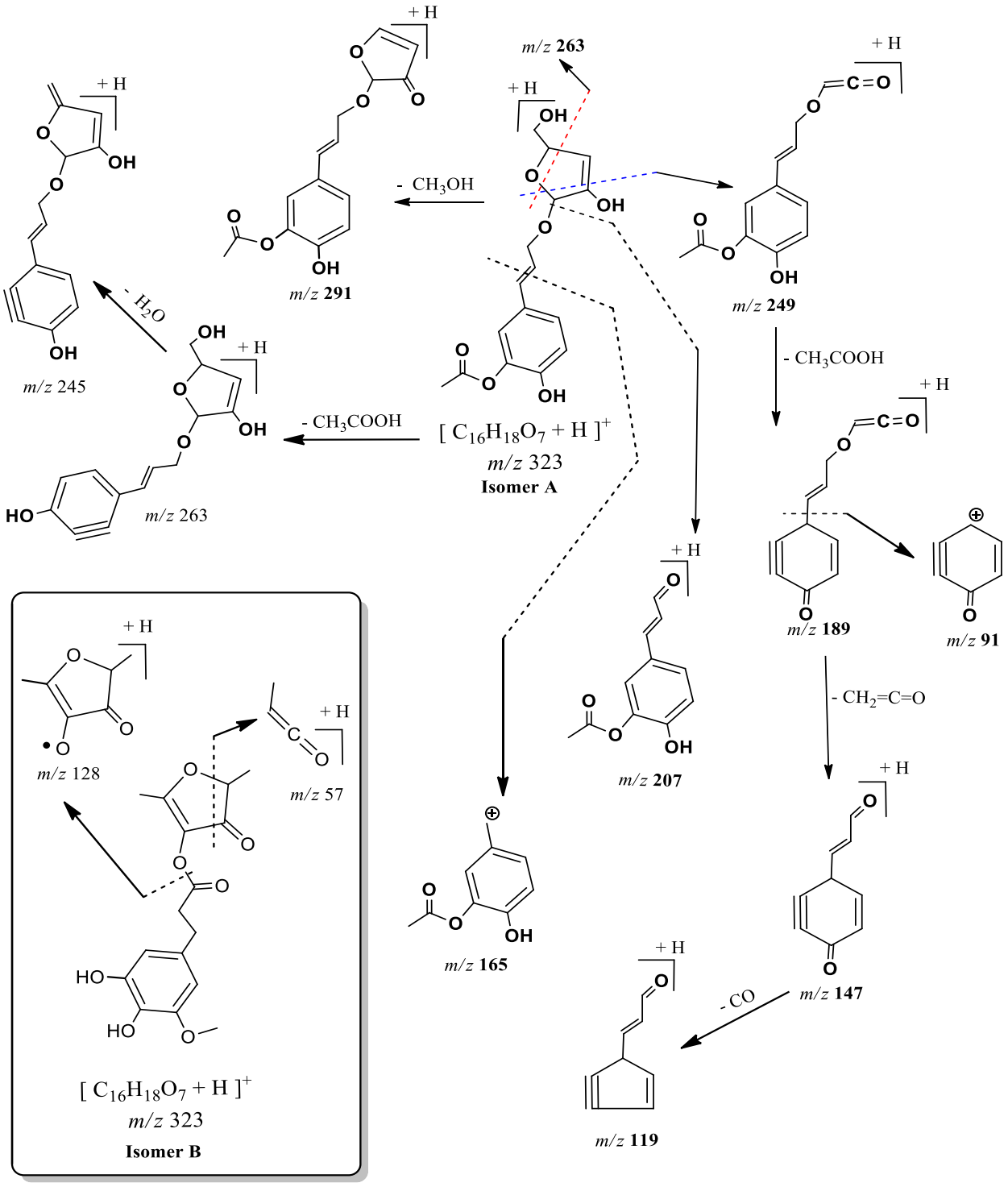


Figure A3.12. Product ion scan of the protonated precursor ion *4* at *m/z* 323.18



Scheme A3.5. Low energy CID-MS/MS fragmentation mechanism of the protonated precursor ion *4* at *m/z* 323.18

A3.2- The product ion scan of LCC composed of two glycoses attached to one lignin unit

A) The protonated molecule *5* at *m/z* 433.16

The formula $[C_{19}H_{28}O_{11} + H]^+$ (DBE=5.5) was assigned to the protonated molecule *5* at *m/z* 433.16. Once more, this protonated molecule appeared to be composed of two isomeric A and B structures (Figure A3.13 and Scheme A3.6). Isomer A was composed of three units, which are tetrose, pentose derivative [2-(hydroxymethyl) oxolane-2,3,4,5-tetrol], and the lignin unit dihydroconiferyl alcohol. In contrast, isomer B was composed of three units, which are hexose, tetrose, and dihydrocaffeyl alcohol.

The product ion scan of the precursor ion at *m/z* 433.16 (Isomer A) was initiated by the loss of one methanol molecule to give the ion at *m/z* 401. This latter ion loses one methyl radical to yield the protonated radical product ion at *m/z* 386. In addition, the ion at *m/z* 401 experience the cleavage of the Cβ-Cγ bond of the lignin unit to yield the ion at *m/z* 295.

Additionally, the precursor ion at *m/z* 433.16 (Isomer A) can also be initiated by a C1-Cα bond cleavage in the lignin unit to create product ion at *m/z* 309. Also, the cleavage of the lignin unit Cα-Cβ bond in the isomer A produces the product ions at *m/z* 137 and 295. This latter product ion at *m/z* 295 loses one water molecule to form the ion at *m/z* 277. Lastly, the presence of the pentose derivative [2-(hydroxymethyl) oxolane-2,3,4,5-tetrol] was supported by the formation of its protonated molecule at *m/z* 167, as demonstrated in Scheme A3.6. This last product ion at *m/z* 167 loses two consecutive water molecules to yield the product ion at *m/z* 149 and 131, respectively.

The product ion scan of the precursor ion at *m/z* 433.16 (isomer B) was subjected to two different hexose sugar ring cleavage to produce the product ions at *m/z* 373 (^{2,4}X) and *m/z* 359 (^{3,5}X). This latter product ion at *m/z* 359 losses one water molecule to yield the ion at *m/z* 341. The presence of the hexose-pentose disaccharide was supported by the formation of the disaccharide protonated molecule at *m/z* 283. This latter disaccharide ion at *m/z* 283 can lose the pentose sugar to yield the hexose moiety at *m/z* 163.

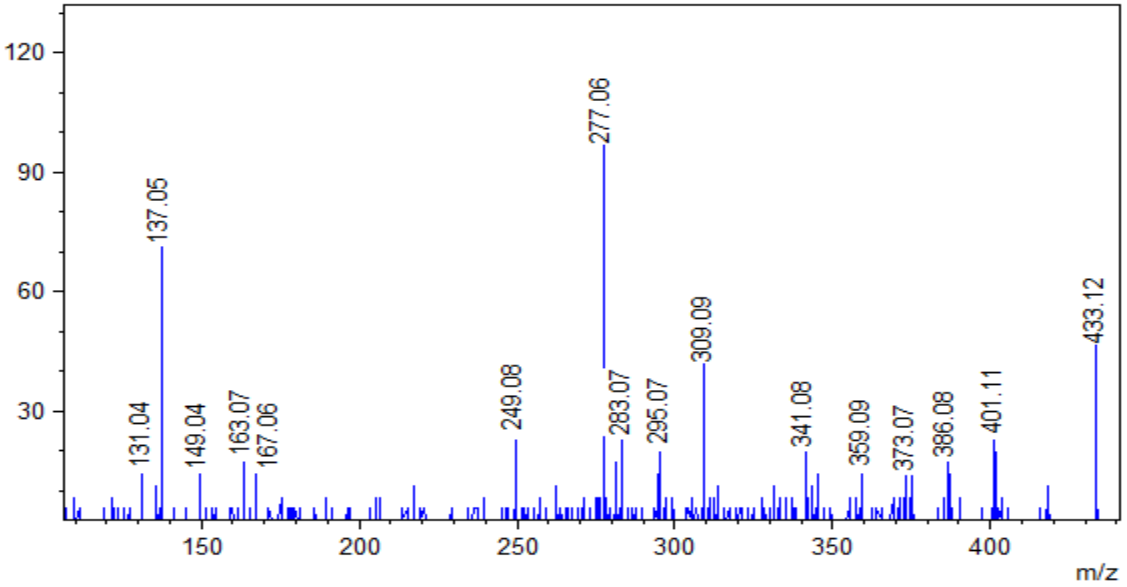
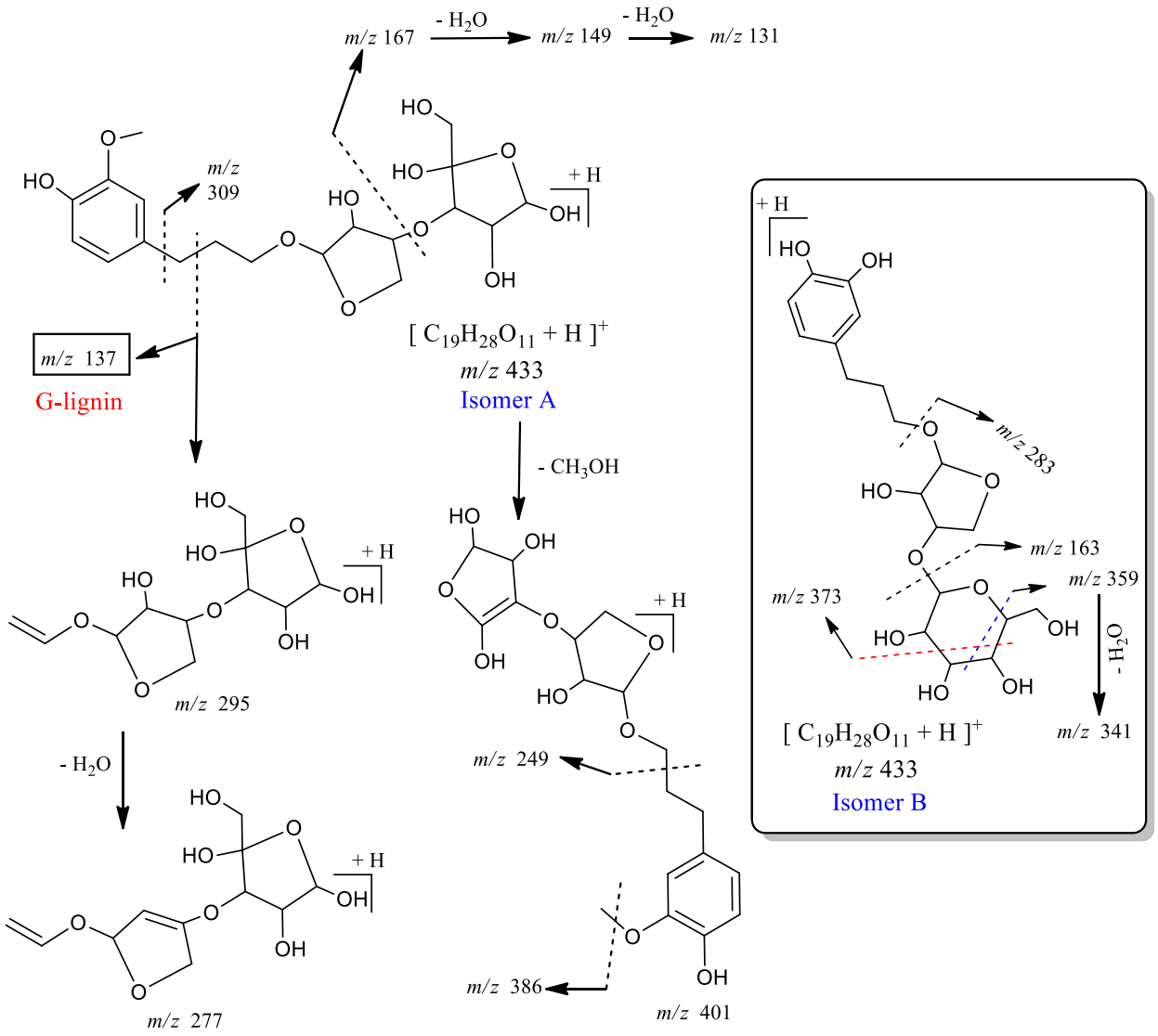


Figure A3.13. Product ion scan of the protonated precursor ion *5* at *m/z* 433.16



Scheme A3.6. Low energy CID-MS/MS fragmentation mechanism of the protonated precursor ion *5* at *m/z* 433.16

A3.3- The product ion scan of LCC composed of a glycose attached to two lignin units

A) The protonated molecule *8* at *m/z* 517.18

The formula $[C_{26}H_{28}O_{11}+ H]^+$ (DBE=12.5) was assigned to the protonated molecule *8* at *m/z* 517.18 (Figure A3.14 and Scheme A3.7a and A3.7b). Once more, we deduced the possible existence of four isomers. The isomer A is composed of a central deoxyhexose unit, which is attached to the lignin derivatives: dihydroferulic acid and 3,4-(methylenedioxy) cinnamic acid. Whereas, isomer B is composed of a dehydrated hexose portion attached to both the 3,4-methylenedioxy cinnamic acid and the 5-hydroxy dihydroconiferyl alcohol residue. The isomer C is a positional isomer of structure B as they differ only in the position of one oxygen atom. Lastly, isomer D is composed of dehydroquinic acid portion attached to both the coumaryl alcohol and the 5-hydroxy dihydroferulic acid moieties. The possible existence of these four isomeric precursor ions is reflected by the complexity of the MS/MS spectrum obtained in Figure A3.14 (Messy MS/MS).

Consequently, the product ion scan of the precursor ion at *m/z* 517.18 (Isomer A) gave the product ion at *m/z* 499 by the elimination of a water molecule (Scheme A3.7a). Likewise, this precursor ion experiences the loss of 46 Da (- CH_2O_2) from the 3,4-(methylenedioxy) cinnamic acid unit to produce the product ions at *m/z* 471. In addition, this precursor ion at *m/z* 517.18 can be subjected to different sugar ring cleavages, to produce the product ions *at m/z* 457 (^{0,4}X), *m/z* 443 (^{3,5}X). The ion at *m/z* 443 loses one methyl radical to produce

the radical ion at *m/z* 428. Also, the product ion at *m/z* 443 can lose one molecule of methanol to yield the product ion at *m/z* 411.

Furthermore, the product ion scan of the precursor ion at *m/z* 517.18 (Isomer A) can proceed by the straightforward loss of a methanol molecule to generate the ion at *m/z* 485. This latter ion is subjected to a sugar ring contraction to produce the ion at *m/z* 277. Also, the product ion at *m/z* 485 experiences the cleavage of the C1-O bond in the sugar ring, which leads to the generation of the ion at *m/z* 295. In addition, the product ion at *m/z* 485 can undergo a C_1-Cα cleavage in 3,4-(methylenedioxy) cinnamic acid unit to yield the product ion at *m/z* 363.

Moreover, the precursor ion *10* (Isomer A) experiences the cleavage of the C1-O bond in the sugar ring, which leads to the generation of the product ion at *m/z* 321. In addition, the cleavage of the Cγ-O of the 3,4-(methylenedioxy) cinnamic acid in the isomer A leads to the formation of the aromatic product ion at *m/z* 175. This later ion loses one methylene carbene from the methylenedioxy moiety to yield the product ion at *m/z* 161, as demonstrated in Scheme A3.7a. Furthermore, the cleavage of the Cβ- Cγ bond of the dihydro ferulic acid unit in the precursor isomer A leads to the formation of the aromatic product ion at *m/z* 151. Lastly, the cleavage of the Cα-Cβ bond of dihydro ferulic acid in isomer A leads to the formation of the two product ions at *m/z* 137 and *m/z* 379.

The proposed isomeric precursor ion B showed the cleavage of the Cγ-O bond in the 5-hydroxy dihydroconiferyl alcohol residue to yield the product ion at *m/z* 337. This later product ion at *m/z* 337 can lose 30 Da (-CH_2O) from the methylenedioxy moiety in the 3,4-methylenedioxy cinnamic acid residue to form the product ion at *m/z* 307. In addition , the

proposed isomeric precursor ion C showed a sugar ring cleavage to generate the ion at *m/z* 203, as shown in Scheme A3.7b. Lastly, The isomeric precursor ion D totally agrees with the formation of the dehydroquinic acid protonated ion at *m/z* 191, as shown in Scheme A3.7b. It should be noted that these last three isomers B, C and D were drawn to account for the four product ions at *m/z* 191, *m/z* 203, *m/z* 307 and *m/z* 337 as there is no suitable fragmentation pathway to demonstrate them from Isomer A.

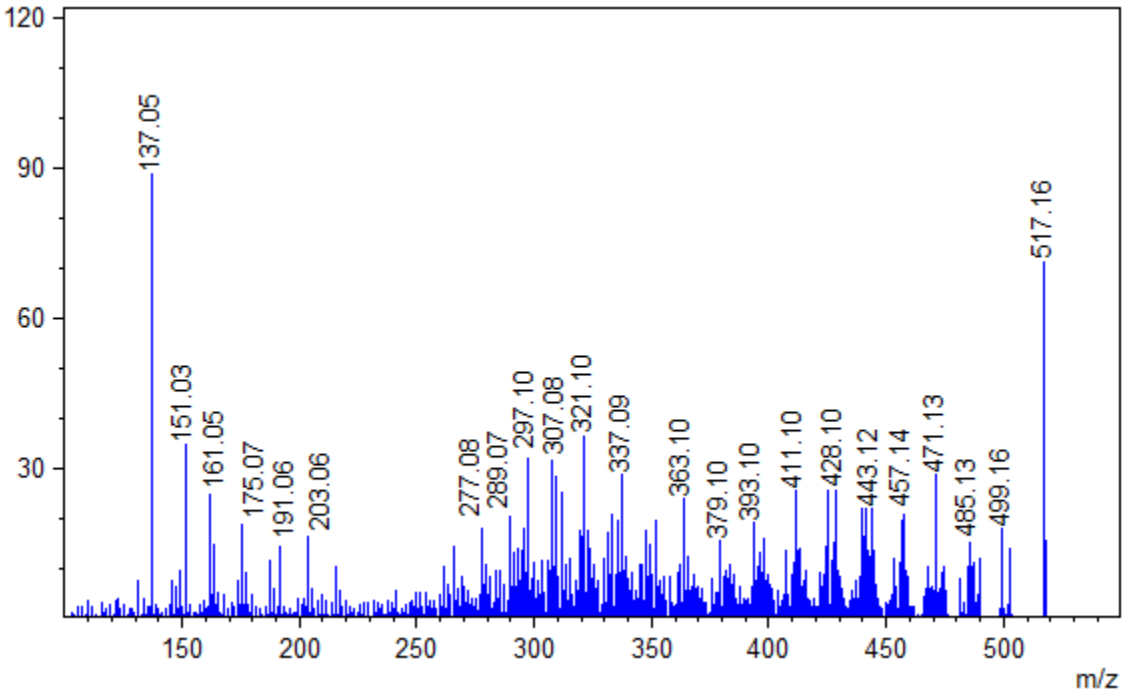
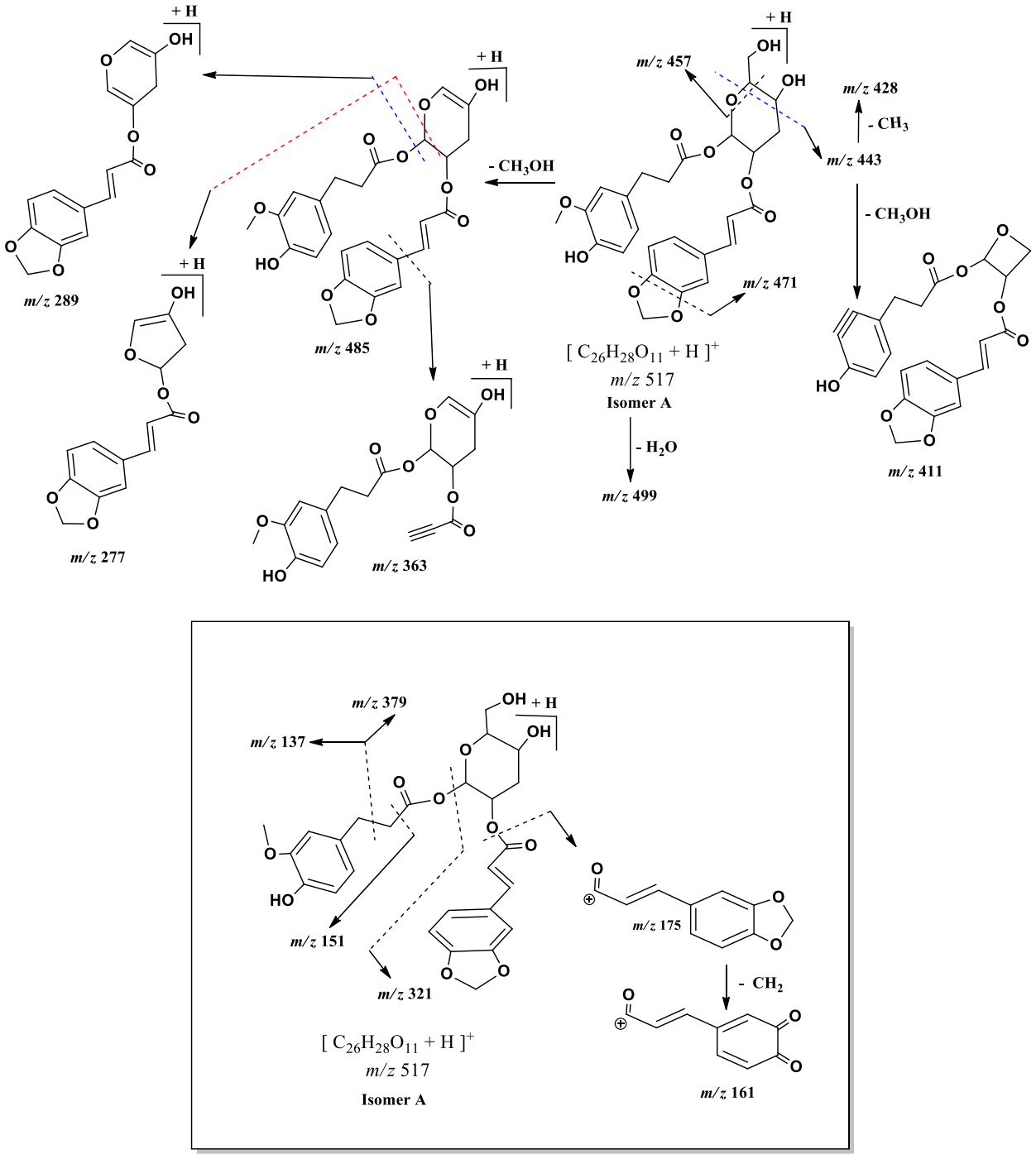
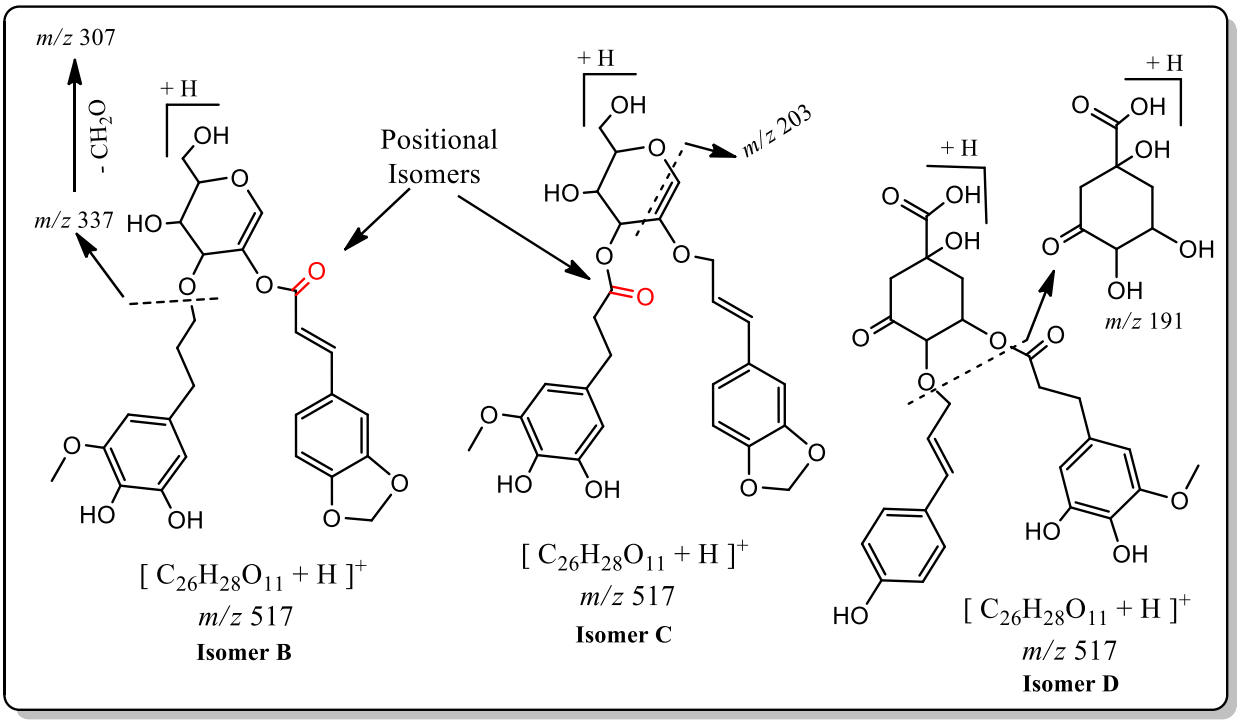


Figure A3.14. Product ion scan of the protonated precursor ion *8* at *m/z* 517.18



Scheme A3.7A. Low energy CID-MS/MS fragmentation mechanism of the protonated precursor ion *8* at *m/z* 517.18



Scheme A3.7B. Low energy CID-MS/MS fragmentation mechanism of the protonated precursor ion *8* at *m/z* 517.18

B) The protonated molecule *9* at *m/z* 533.18

The formula $[C_{26}H_{28}O_{12}+ H]^+$ (DBE=12.5) was assigned to the protonated molecule *9* at *m/z* 533.18, and once more, it could be represented by three isomers (Figure A3.15 and Scheme A3.8a and A3.8B). Isomer A is composed of a hexose sugar connected to dihydroferulic acid and 3,4-(methylenedioxy) cinnamic acid. Isomer B is formed of dehydrated hexose attached 5-hydroxy dihydro ferulic acid and 3,4-(methylenedioxy) dihydrocinnamic acid. Isomer C is composed of dehydrated hexose attached to 3-(3,4-dihydroxy-5-methoxyphenyl)-3-hydroxypropanoic acid (dihydroxy dihydroferulic acid derivative) and 3,4-methylenedioxy cinnamyl alcohol.

The product ion scan of the precursor ion at *m/z* 533.18 (Isomer A) was initiated by the loss of a water molecule to produce the product ion at *m/z* 515 (Scheme A3.8a). This latter ion undergoes a sugar ring cleavage to yield the ion at *m/z* 455 ($^{0,4}X$). In addition, the product ion at *m/z* 515 can lose a carbon monoxide molecule to generate the ion at *m/z* 487. This latter ion, in its turn, can lose 60 Da ($-C_2H_3O_2$) through a sugar ring cleavage to yield the ion at *m/z* 427. In addition, the same product ion at *m/z* 487 can lose 46 Da ($-CH_2O_2$) from the methylenedioxy moiety to create the third-generation product ion at *m/z* 441. Similarly, the same product ion at *m/z* 487 can also lose a methyl radical to generate the radical ion at *m/z* 472. In addition, the product ion at *m/z* 487 can also be subjected to C_1-Cα bond cleavage in the dihydro ferulic acid unit to create the ion at *m/z* 363.

Moreover, the cleavage of the Cβ- Cγ of the dihydro ferulic acid in isomer A leads to the formation of the aromatic product ion at *m/z* 151. Lastly, the cleavage of the Cα -Cβ bond of the dihydro ferulic acid of isomer A leads to the creation of the aromatic product ions at *m/z* 137.

Also, this product ion scan of the precursor ion at *m/z* 533.18 (Isomer A) can be initiated, by a sugar ring cleavage, to produce the product ion at *m/z* 413 ($^{0,2}X$). In addition, the cleavage of the Cγ-O of 3,4-(methylenedioxy) cinnamic acid unit leads to the formation of the aromatic product ion at *m/z* 175. This later product ion loses one methylene carbene from its methylenedioxy moiety to yield the product ion at *m/z* 161.

The product ion scan of isomer B showed a sugar ring cleavage to generate the product ion at *m/z* 297 ($^{1,3}X$). In addition, the cleavage of the C1-O bond in the sugar ring of isomer B lead to the formation of the product ion at *m/z* 321. This latter product ion at *m/z* 321 can

lose a methylene carbene from the methylenedioxy moiety of the 3,4-(methylenedioxy) dihydrocinnamic acid unit to form the product ion at *m/z* 307 (Scheme A3.8b).

Lastly, isomer C showed two different sugar ring cleavage to yield the product ion at *m/z* 203 (^{1,3}X) and *m/z* 277 (^{1,5}X). One more time, these last two isomers B and C were proposed to explain the formation of some product ions that cannot be formed from isomer A. Therefore, the possibility that we are dealing with an isomeric mixture that produces such complicated and messy MS/MS spectra is indeed real.

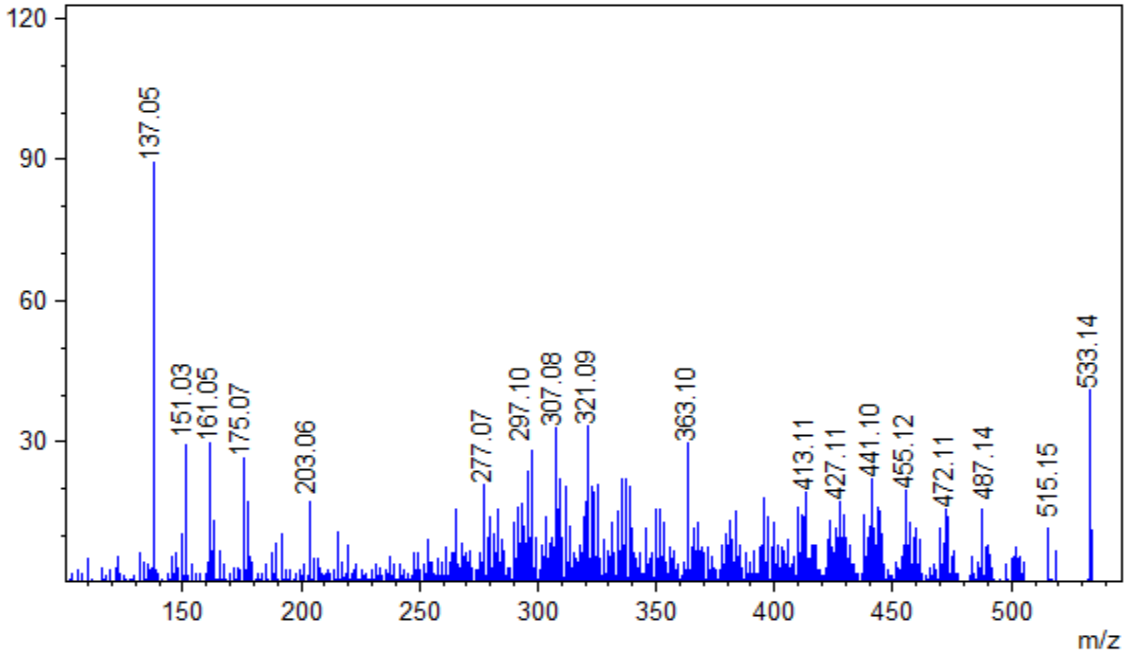
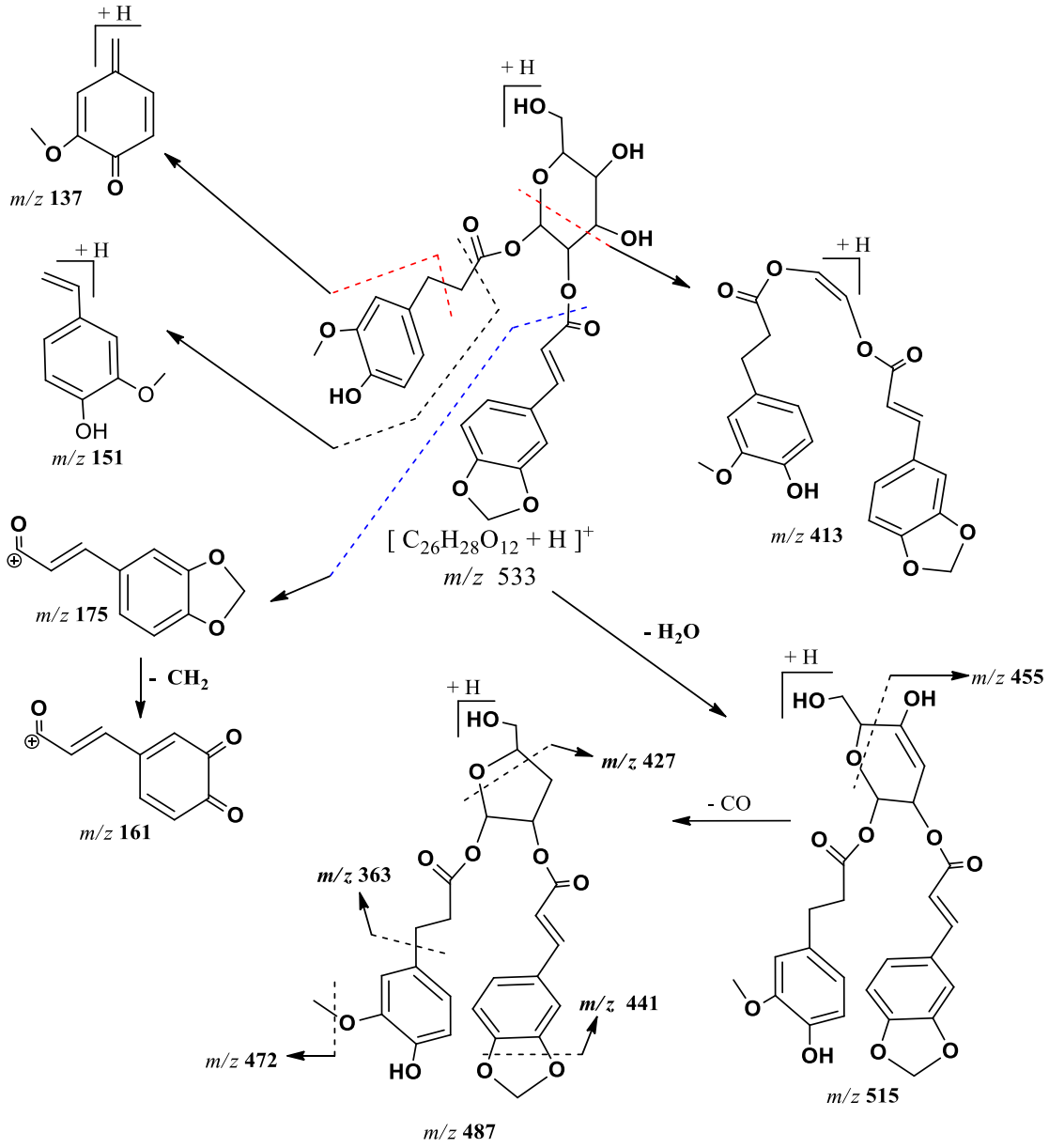
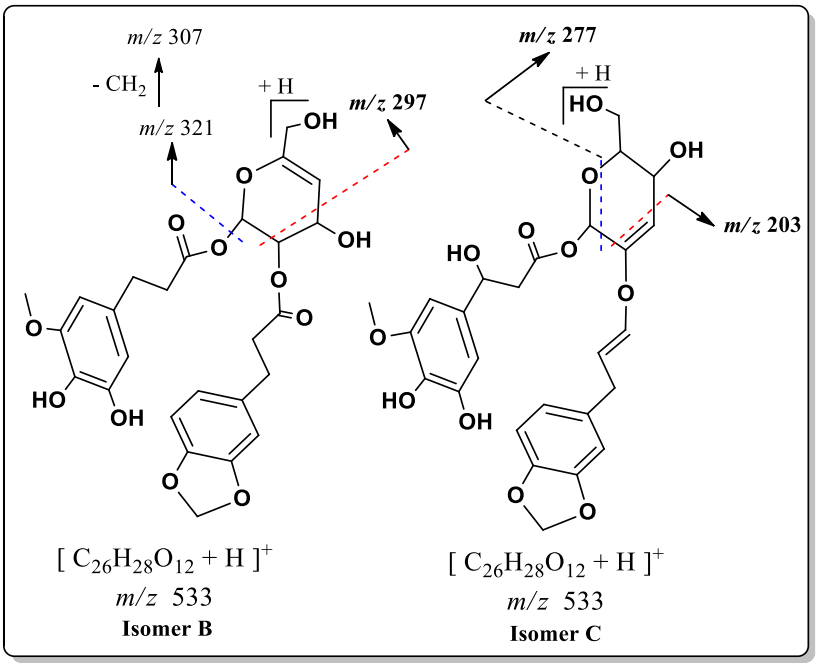


Figure A3.15. Product ion scan of the protonated precursor ion 9 at *m/z* 533.18



Scheme A3.8a. Low energy CID-MS/MS fragmentation mechanism of the protonated precursor ion *9* at *m/z* 533.18



Scheme A3.8b. Low energy CID-MS/MS fragmentation mechanism of the protonated precursor ion *9* at *m/z* 533.18

A3.4- The product ion scan of LCC composed of two glycoses attached to two lignin units

A) The protonated molecule *11* at *m/z* 651.25

The formula $[C_{31}H_{38}O_{15}+ H]+$ (DBE=12.5) was assigned to the protonated molecule *11* at *m/z* 651.25. It was tentatively considered to be a tetramer composed of hexose, pentose, dihydroconiferyl alcohol, and 3,4-(methylenedioxy) cinnamic acid (Figure A3.16 and SchemeA3.9).

Once more, to describe the formation of all major product ions produced by the MS/MS analysis of this precursor ion *11*, it was necessary to consider the presence of isomer B.

This isomer B was composed of hexose, pentose, coniferyl alcohol, and 3,4-(methylenedioxy) dihydrocinnamic acid.

The product ion scan of the precursor ion at *m/z* 651.25 (isomer A) showed the straightforward elimination of a water molecule to produce the product ion at *m/z* 633 (Scheme A3.9). Also, the precursor ion at *m/z* 651.25 experiences loss of the methylenedioxy moiety ($-CH_2O_2$) to give the product ion at *m/z* 605. Moreover, the precursor ion at *m/z* 651.25 showed the loss of formaldehyde (-30 Da) form the aromatic methoxy group of the dihydroconiferyl alcohol unit to create the product ion at *m/z* 621. Moreover, the precursor ion at *m/z* 651.25 experiences two simultaneous bond cleavages, the C3-O and the C4-O bonds in the hexose sugar ring to yield the product ion at *m/z* 309. Furthermore, the precursor ion at *m/z* 651.25 experiences another two simultaneous bond cleavages, the Cγ-O bond dihydroconiferyl alcohol unit and the C4-O bond of the hexose sugar unit to yield the product ion at *m/z* 295.

Additionally, the precursor ion at *m/z* 651.25 is subjected to cleavage of the Cγ-O bond in the 3,4-(methylenedioxy) cinnamic acid unit to form the aromatic product ion at *m/z* 175. Also, this precursor ion can be subjected to the simultaneous cleavages of the Cβ- Cγ and the Cα-Cβ bonds present in the dihydroconiferyl alcohol unit, to create the aromatic product ions at *m/z* 151 and 137, respectively. Lastly, the presence of isomer B was proposed to describe the formation of the product ion at *m/z* 283 and *m/z* 321, as demonstrated in Scheme A3.9.

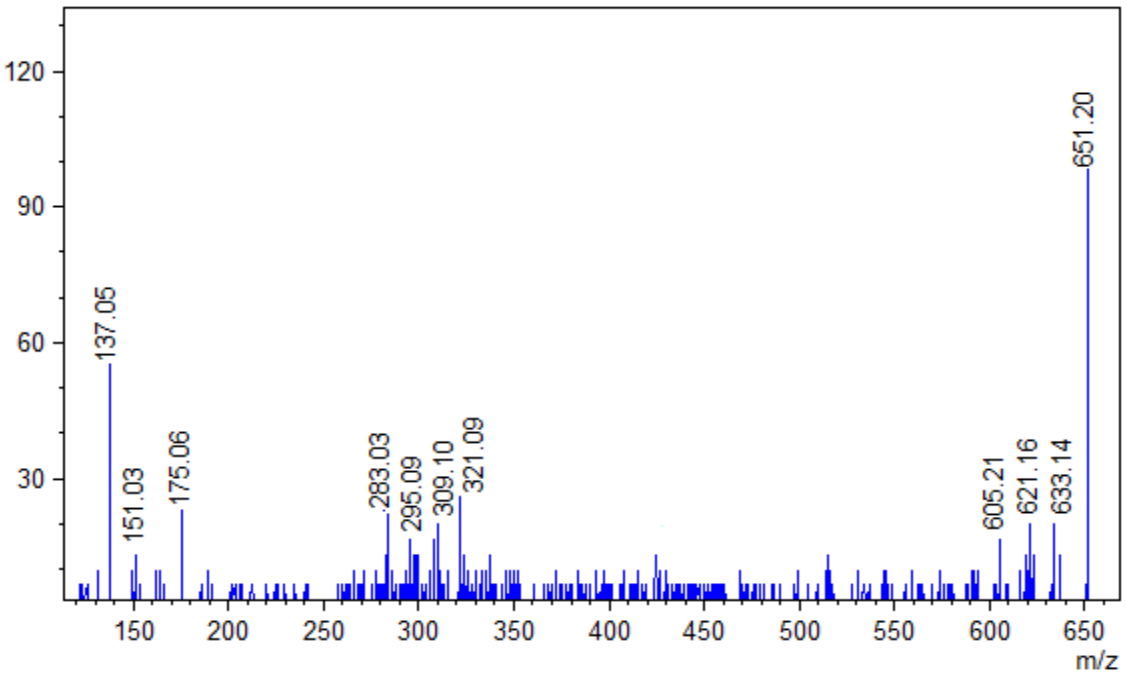
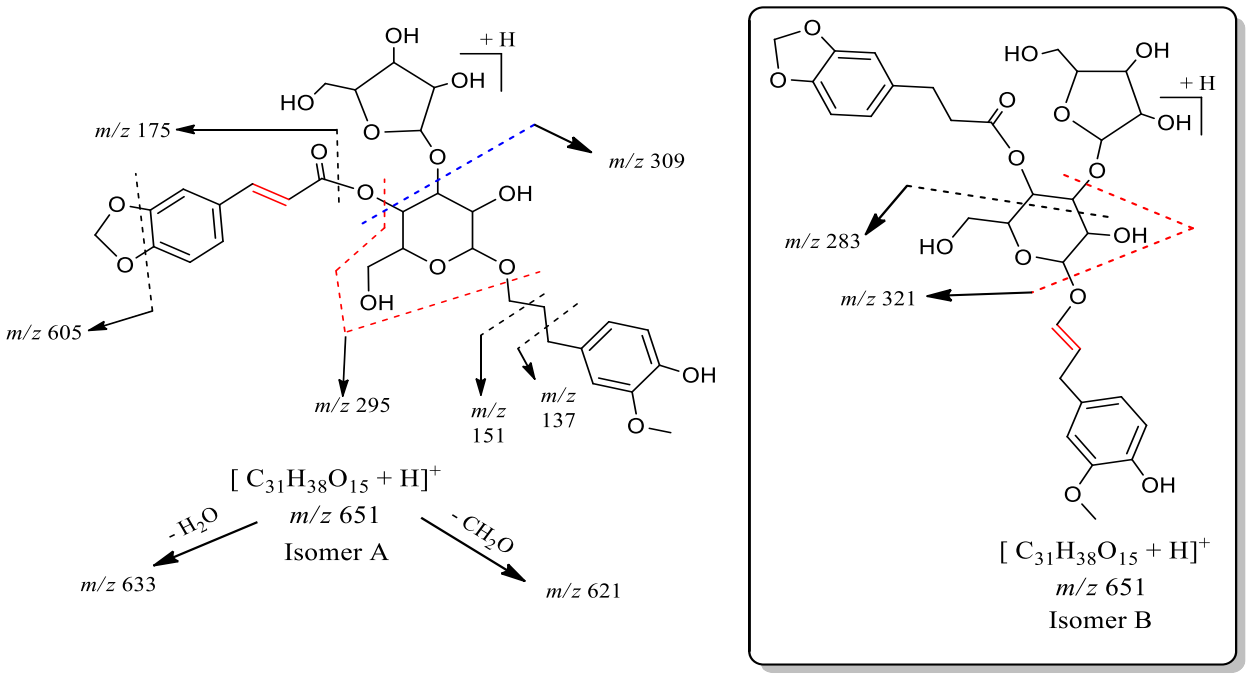


Figure A3.16. Product ion scan of the protonated precursor ion *11* at *m/z* 651.25



Scheme A3.9. Low energy CID-MS/MS fragmentation mechanism of the protonated precursor ion *11* at *m/z* 651.25. Red double bonds indicate the point of difference between the two positional isomers.

B) The protonated molecule *12* at *m/z* 695.26

The formula $[C_{33}H_{42}O_{16}+ H]^+$ formula (DBE=12.5) was assigned to the protonated molecule *12 at m/z* 695.26 (Figure A3.17 and Scheme A3.10).This ion was proposed to be a tetramer composed of methyl hexoside, hexose, dihydroconiferyl alcohol, and 3,4-(methylenedioxy) cinnamic acid (Isomer A). As well, two positional isomers B and C of this formula were proposed. Isomer B is composed of quinic acid, hexose sugar, 3,4-(methylenedioxy) cinnamyl alcohol, and 5-hydroxy dihydroconiferyl alcohol.

The product ion scan of the precursor ion at *m/z* 695.26 (isomer A) was initiated by the loss of 46 Da in the form of CH_2O_2 to create the product ion at *m/z* 649 (Scheme A3.10). Furthermore, the product ion scan of isomer A can also be initiated by the straightforward loss of a water molecule to generate the product ion at *m/z* 677. In addition, the product ion at *m/z* 677 showed a sugar ring cleavage in the methyl hexoside ring to generate the product ion at *m/z* 587 ($^{2,5}X$). Also, the product ion at *m/z* 677 showed two simultaneous cleavages of the C3-O bond in the methyl hexoside ring and the Cγ-O bond in the dihydroconiferyl alcohol unit to yield the product ion at *m/z* 337. This latter ion can lose 30 Da (- CH_2O) from the methylenedioxy unit of the 3,4-(methylenedioxy) cinnamic acid unit to form the product ion at *m/z* 307.

Moreover, the product ion scan of isomer A can be initiated by two different types of sugar ring cleavages. The first cleavage occurs in the methyl hexoside ring to produce the product ions at *m/z* 621 ($^{0,2}X$), while the second cleavage occurs in the hexose ring to yield the product ion at *m/z* 295 ($^{1,3}X$).

Furthermore, the precursor ion at *m/z* 695.26 (Isomer A) is subjected to two simultaneous cleavages of the C2-O and C3-O bonds in the hexose ring to create the ion at *m/z* 321. Additionally, the precursor ion at *m/z* 695.26 is subjected to cleavage of the Cγ-O bond in the 3,4-(methylenedioxy) cinnamic acid unit to form the aromatic product ion at *m/z* 175. This latter ion at *m/z* 175 losses methylene carbene from the methylenedioxy moiety to yield the product ion at *m/z* 161.

Also, this precursor ion at *m/z* 695.26 (Isomer A) can be subjected to the simultaneous cleavages of the Cβ- Cγ and the Cα-Cβ bonds present in the dihydroconiferyl alcohol unit, to create the aromatic product ions at *m/z* 151 and 137, respectively. Lastly, two positional isomers B and C were proposed to describe the formation of the product ion at *m/z* 283 and *m/z* 361, as demonstrated in scheme A3.10b.

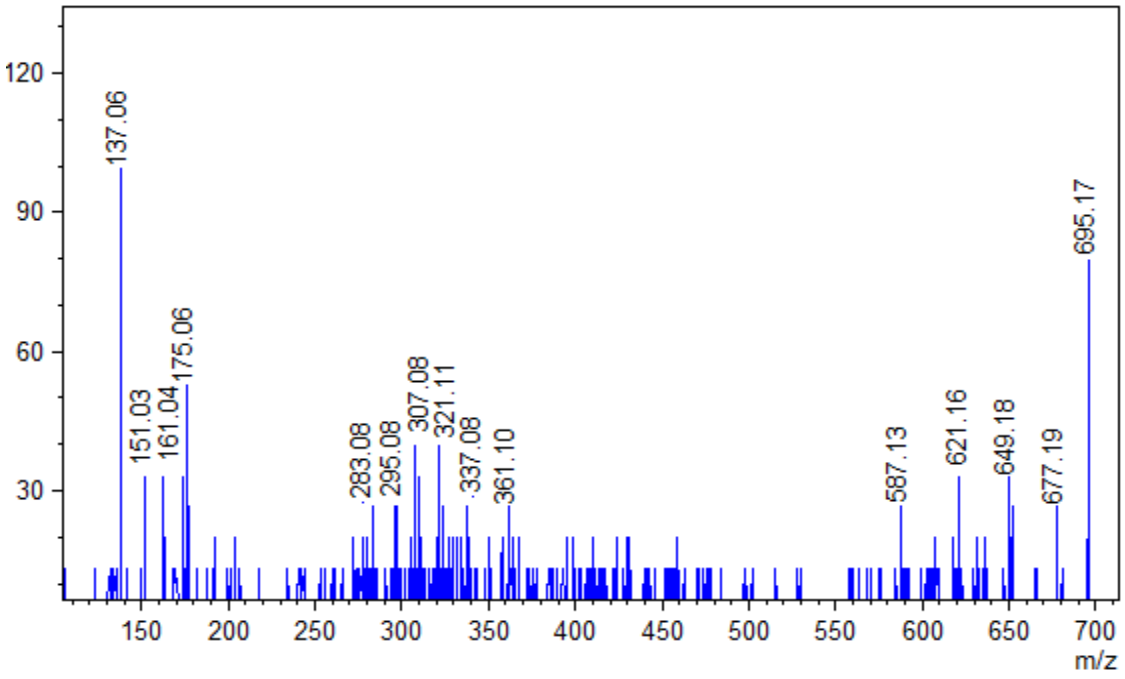
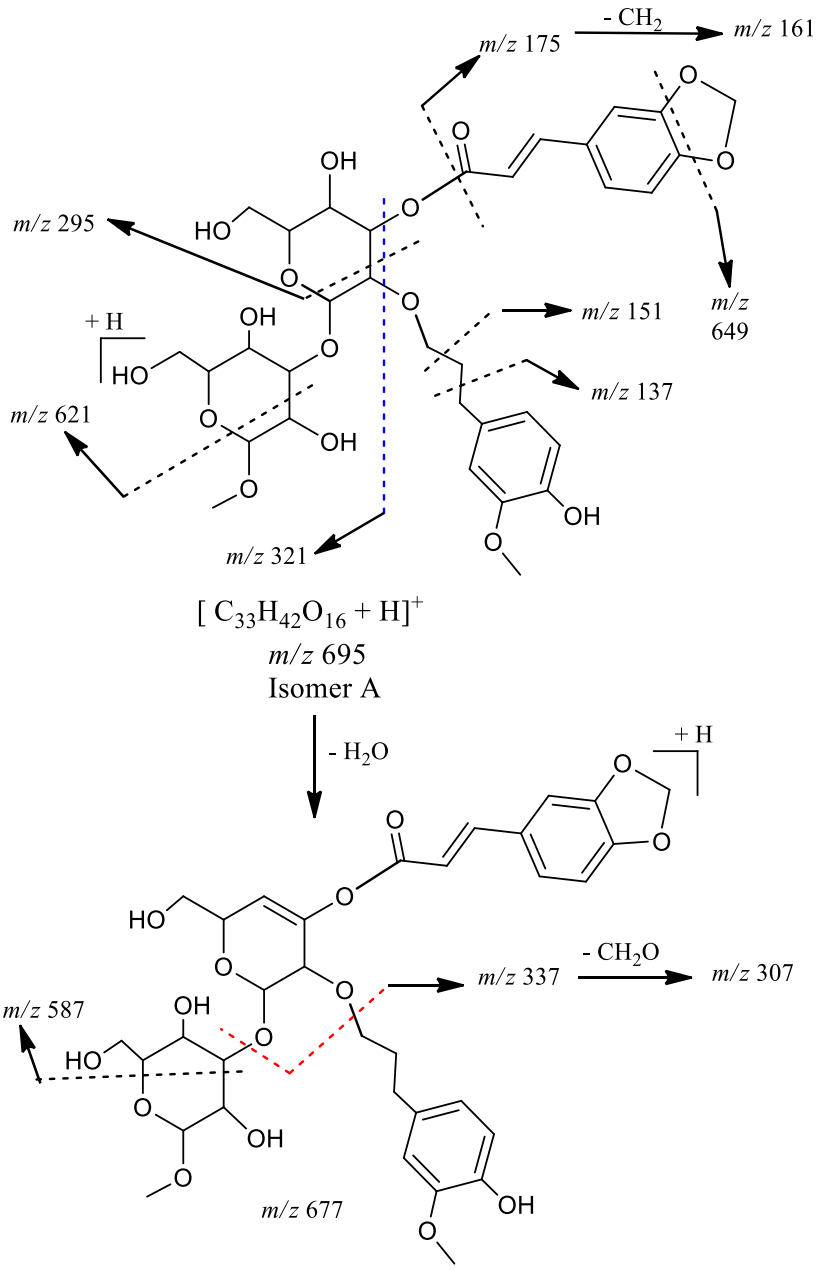
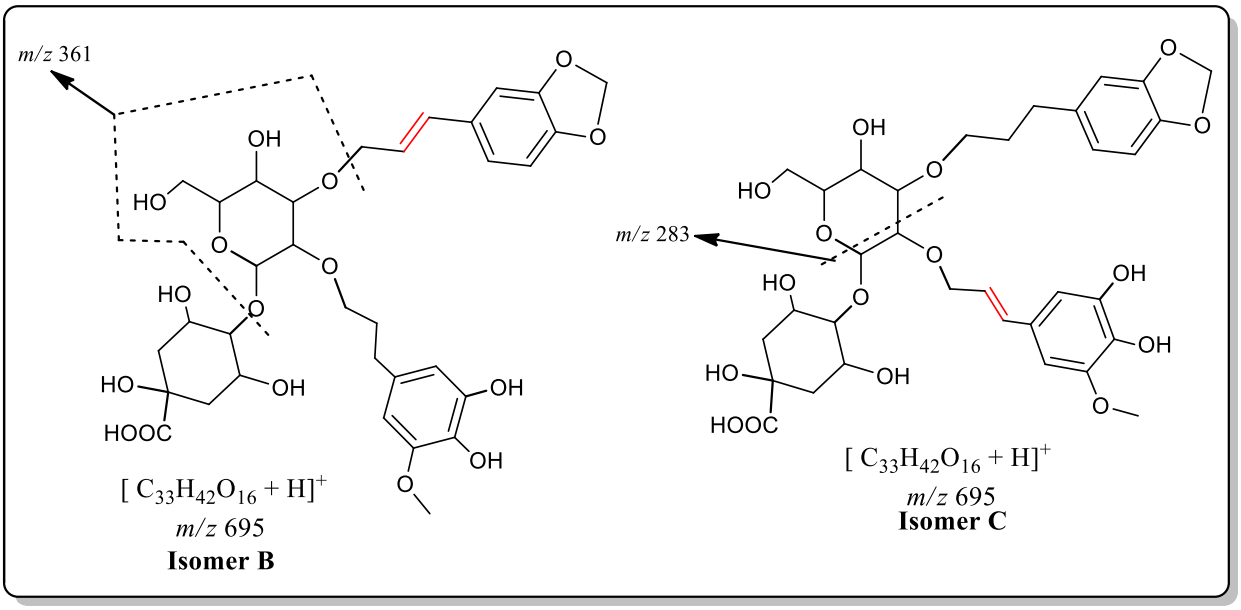


Figure A3.17: Product ion scan of the protonated precursor ion *12* at *m/z* 695.26



Scheme A3.10a. Low energy CID-MS/MS fragmentation mechanism of the protonated precursor ion *12* at *m/z* 695.26



Scheme A3.10b. Low energy CID-MS/MS fragmentation mechanism of the protonated precursor ion *12* at *m/z* 695.26. Red double bonds indicate the point of difference between the two positional isomers.

A3.5- The Product ion scan of LCC composed of one glycose residue attached to three lignol units

A) The protonated molecule *13* at *m/z* 611.24

The $[C_{33}H_{38}O_{11}+ H]^+$ formula (DBE=14.5) was assigned to the protonated molecule *13* at *m/z* 611.24 (Figure A3.18 and Scheme A3.11). This chemical formula could be represented by two isomers A and B. We proposed that isomer A was composed of a pentose sugar attached to three lignin derivatives: two dihydrocaffeyl alcohol, and 3,4-methylenedioxycinnamyl alcohol. The other isomer B is composed of a hexuronic sugar ring attached to two dihydrocoumaryl alcohol and one caffeyl alcohol unit.

The product ion scan of the precursor ion at *m/z* 611.24 (Isomer A) was initiated by the loss of a water molecule to produce the product ion at *m/z* 593 (Scheme A3.11a). This latter ion loses hydroxymethyl radical to yield the radical ion at *m/z* 562. Another possible positional isomer of this product ion at *m/z* 562 was proposed to be formed from another isomeric precursor ion. This was proposed to demonstrate the formation of the product ion at *m/z* 356, as demonstrated in scheme A3.11.

Additionally, the product ion scan of the precursor ion at *m/z* 611.24 can occur by cleavage of the Cα-Cβ bond of one of the two caffeyl units to create the product ion at *m/z* 487. This latter product ion at *m/z* 487 experiences a sugar ring contraction to yield the product ion at *m/z* 297. It should be noted that this latter ion at *m/z* 487 could be depicted by two different positional isomers that can form the secondary product ion at *m/z* 321 (Scheme A3.11a). This latter ion at *m/z* 321 was created by cleavage of the C2-O bond in the pentose ring of the product ion at *m/z* 487.

Furthermore, the precursor ion at *m/z* 611.24 can be subjected to the instantaneous cleavage of the Cγ-O bond in one of the caffeyl units and the C3-O bond in the pentose ring to yield the product ion at *m/z* 283 (Scheme A3.11a). Lastly, as demonstrated before for most of the previous precursor ions, some cleavages lead to the formation of some aromatic product ions characteristic to the presence of the lignin units such as *m/z* 137, 147 and 161 as shown in Scheme A3.11A.

The proposed isomer B demonstrates the formation of the product ion at *m/z* 343 ($^{2,4}X$) through the hexuronic ring cleavages as shown in Scheme A3.11B. Furthermore, this

isomer B can experience two simultaneous cleavage of the C2-O and C3-O bonds in the hexuronic ring to create the product ion at *m/z* 307 (Scheme A3.11b).

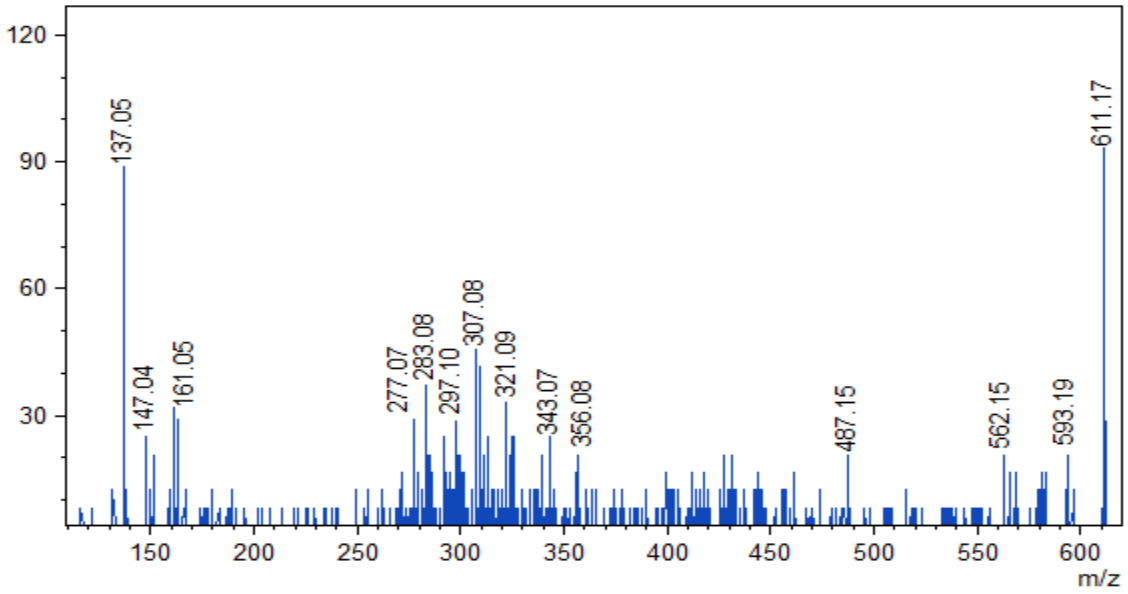
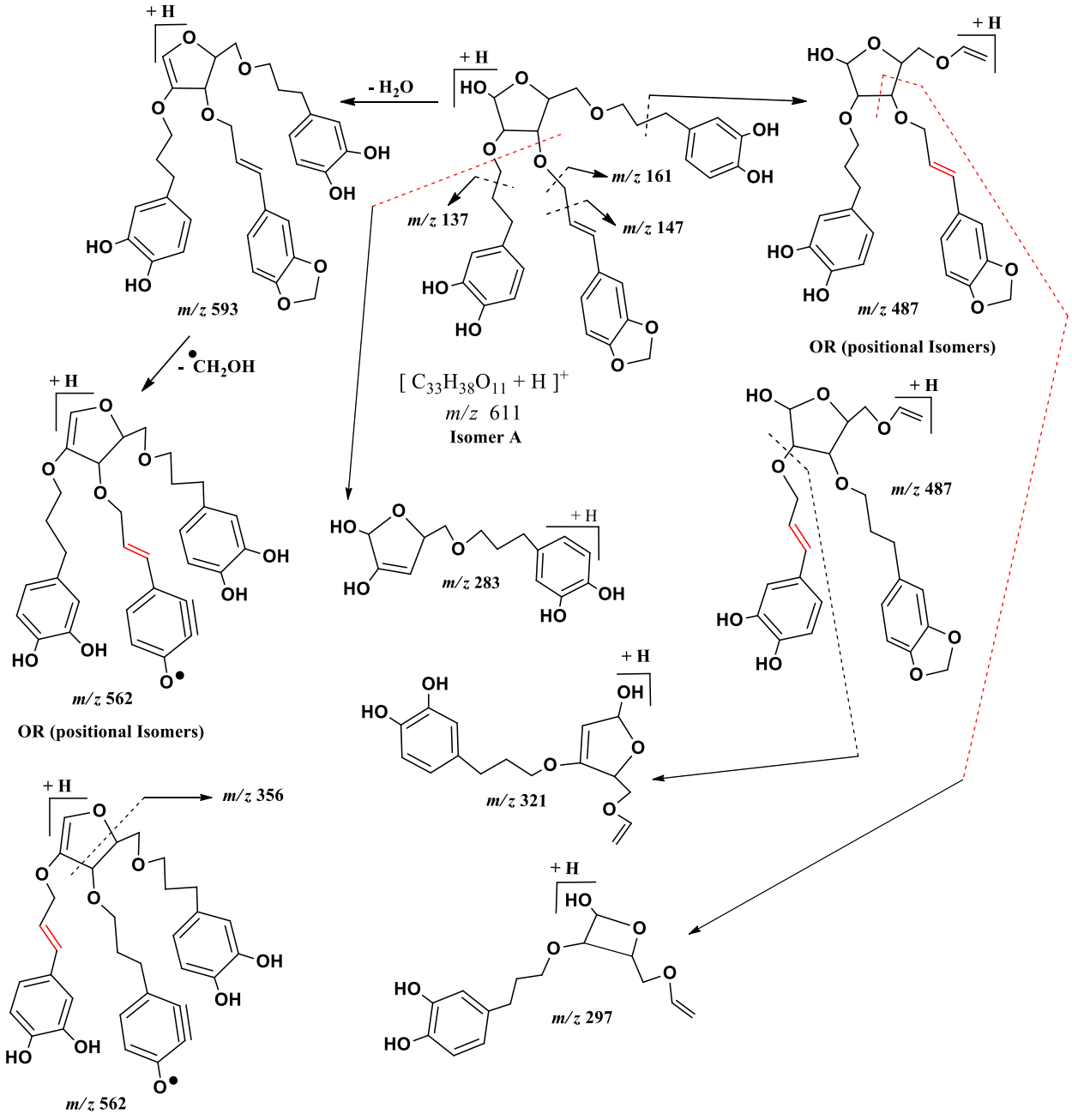
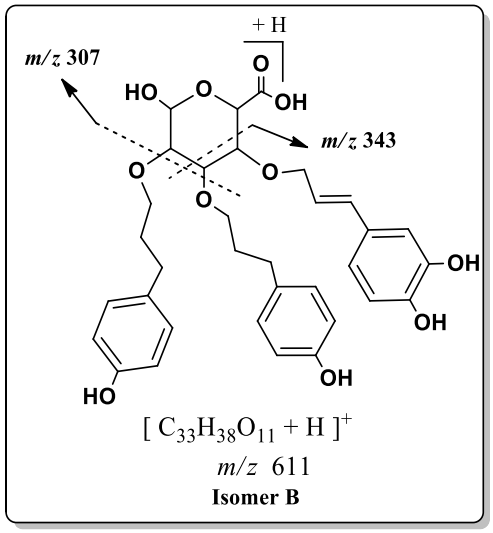


Figure A3.18. Product ion scan of the protonated precursor ion *13* at *m/z* 611.24



Scheme A3.11a. Low energy CID-MS/MS fragmentation mechanism of the protonated precursor ion *13* at *m/z* 611.24. Red double bonds indicate the point of difference between the two positional isomers.



Scheme A3.11b. Low energy CID-MS/MS fragmentation mechanism of the protonated precursor ion *13* at *m/z* 611.24

B) The protonated molecule *15* at *m/z* 711.27

The formula $[C_{36}H_{38}O_{15}+ H]^+$ (DBE=17.5) was assigned to the protonated molecule *15* at *m/z* 711.27 (Figure A3.19 and Scheme A3.12), and it was suggested to be composed of a hexuronic acid moiety attached to three lignin derivatives: two dihydro ferulic acid units and 3,4-methylenedioxy cinnamyl alcohol.

Once more to account for all product ions obtained and to rationalize the complex and messy obtained MS/MS, we had to accept that this precursor ion could be represented by three additional isomers (B, C, and D). Consequently, isomer B was composed of hexuronic acid attached to three lignin units: dihydro coniferyl alcohol, 3,4-(methylenedioxy) cinnamic acid, and dihydro ferulic acid. Whereas, isomer C is a positional isomer of isomer B with only one double bond position changed, as shown in

Scheme A3.12. Lastly, Isomer D is composed of hexuronic acid, dihydro coniferyl alcohol, coniferyl alcohol, and 5-hydroxy-3,4-(methylenedioxy) cinnamic acid.

The product ion scan of the precursor ion at *m/z* 711.27 (isomer A) could be initiated by three different sugar ring cleavage to yield the product ions at *m/z* 249 ($^{3,5}X$), *m/z* 265 ($^{0,3}X$), *m/z* 283 ($^{1,4}X$) (Scheme A3.12).

Furthermore, the product ion scan of Isomer A could also arise by the cleavage of Cβ- Cγ bond and the Cα-Cβ bond present in the dihydroferulic acid unit to yield the aromatic product ions at *m/z* 151 and *m/z* 137. In addition, the precursor ion cleavage of the Cγ-O bond in the 3,4-methylenedioxy cinnamyl alcohol unit leads to the formation of the product ion at *m/z* 161.

Moreover, this product ion scan of isomer A can also be initiated by the simultaneous cleavage of the C_1-O and C_2-O bonds of the sugar ring to yield the product ion at *m/z* 321. Furthermore, the loss of 46 Da in the form of formic acid from isomer A leads to the creation of the product ion at *m/z* 655. This latter product ion fragments by the simultaneous cleavage of the sugar ring C4-O bond and the Cγ-O bond of one of the dihydro ferulic acid units to yield the product ion at *m/z* 309.

The proposed isomeric precursor ion B can fragment by the simultaneous cleavages of C_1-O and C_4-O bonds present in the sugar ring to yield the product ion at *m/z* 333. Lastly, Isomer B experience a sugar ring cleavage to produce the product ion at *m/z* 459 ($^{2,5}X$). The third isomer C was proposed to demonstrate the formation of the product ion at *m/z* 321. Lastly, the formation of the product ion at *m/z* 349 and 297 was demonstrated by proposing the fourth isomeric precursor ion D. It should be noted that the four proposed

isomers are closely related to each other and they differ in the position of double bonds and/or oxygen atoms. Additionally, changing the position of the lignin units on the sugar ring can create more isomers that are not discussed here.

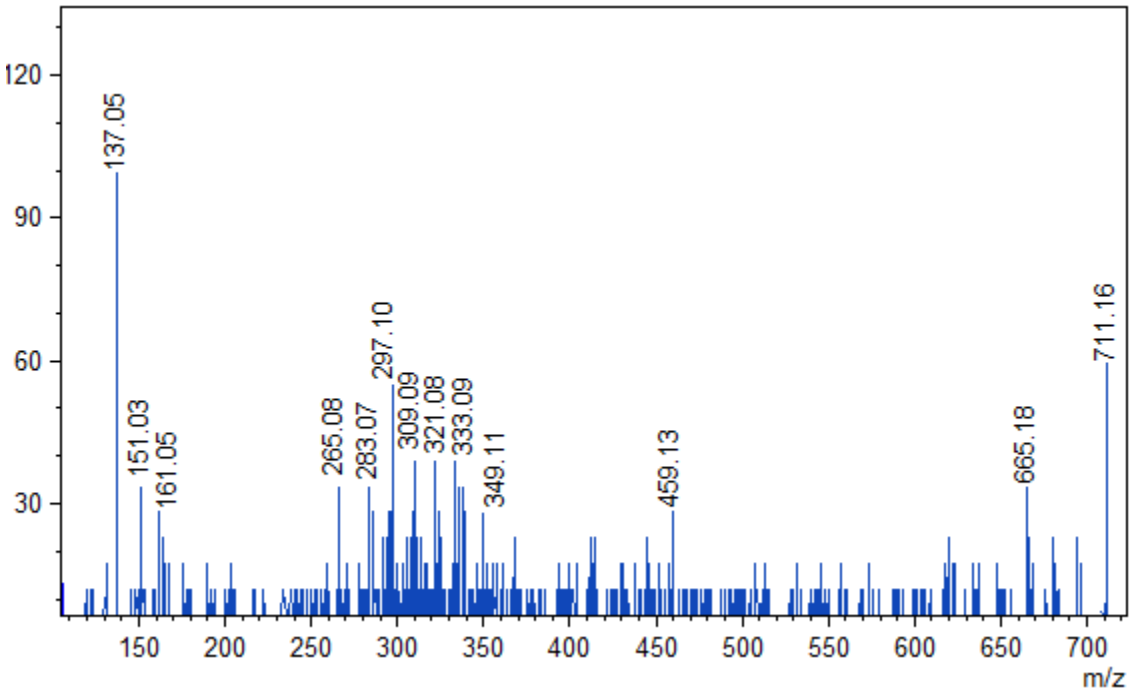
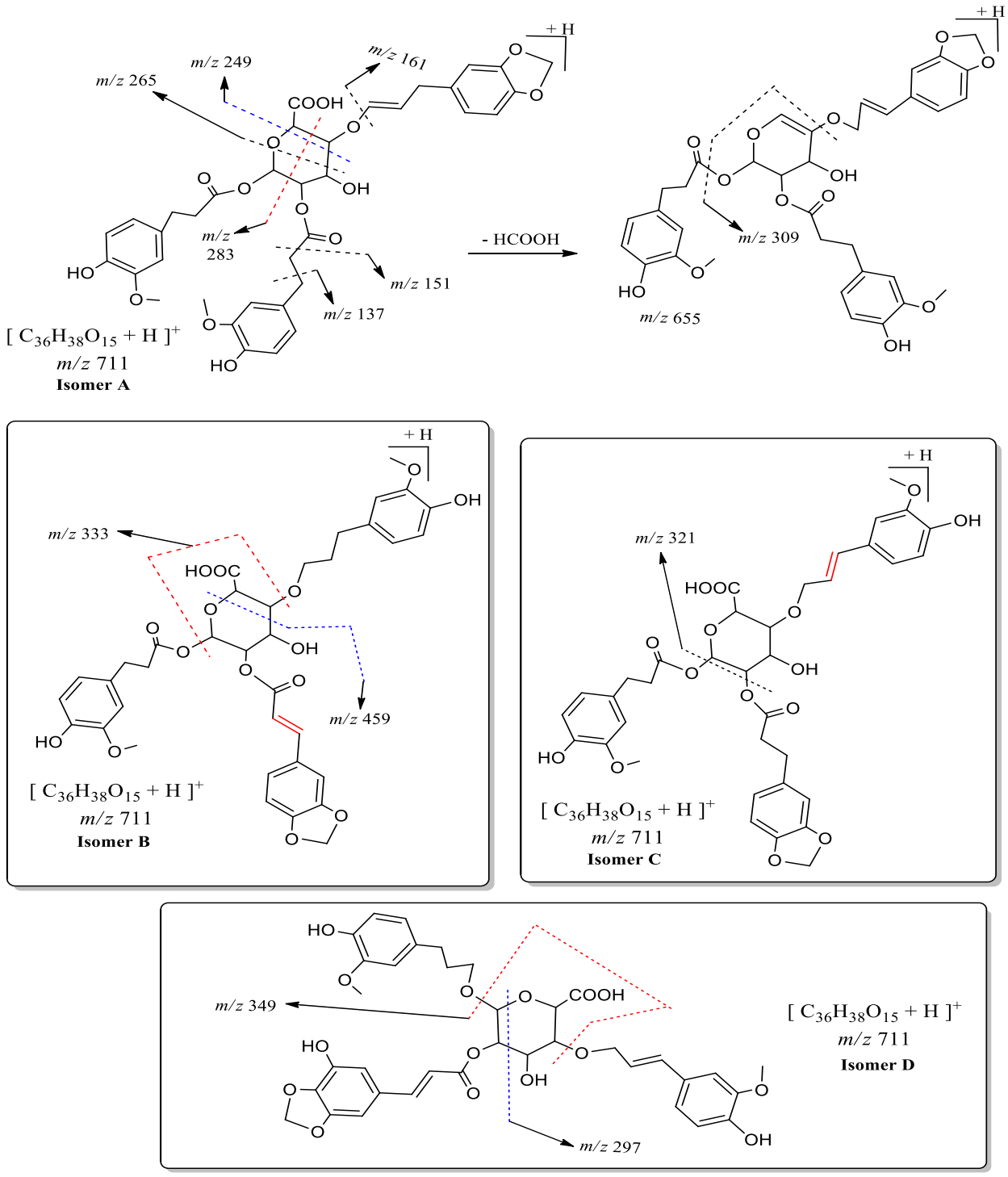


Figure A3.19. Product ion scan of the protonated precursor ion *15* at *m/z* 711.27



Scheme A3.12. Low energy CID-MS/MS fragmentation mechanism of the protonated precursor ion *16* at *m/z* 711.27. Red double bonds indicate the point of difference between the two positional isomers.

Appendix D. Supplementary Information for Chapter 5

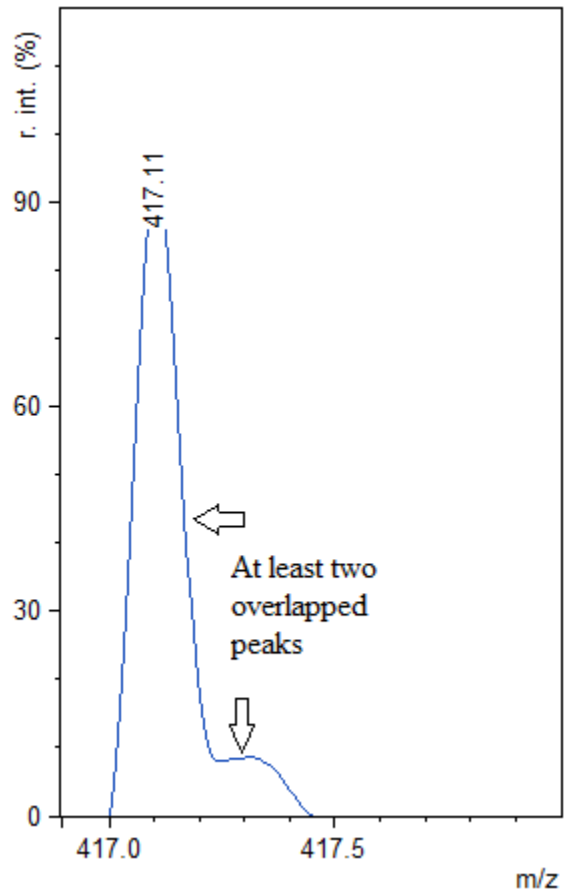


Figure A4.1. Magnification of the mass spectrum at *m/z* 417 as an example to show the complex overlapping between ions at every *m/z* unit

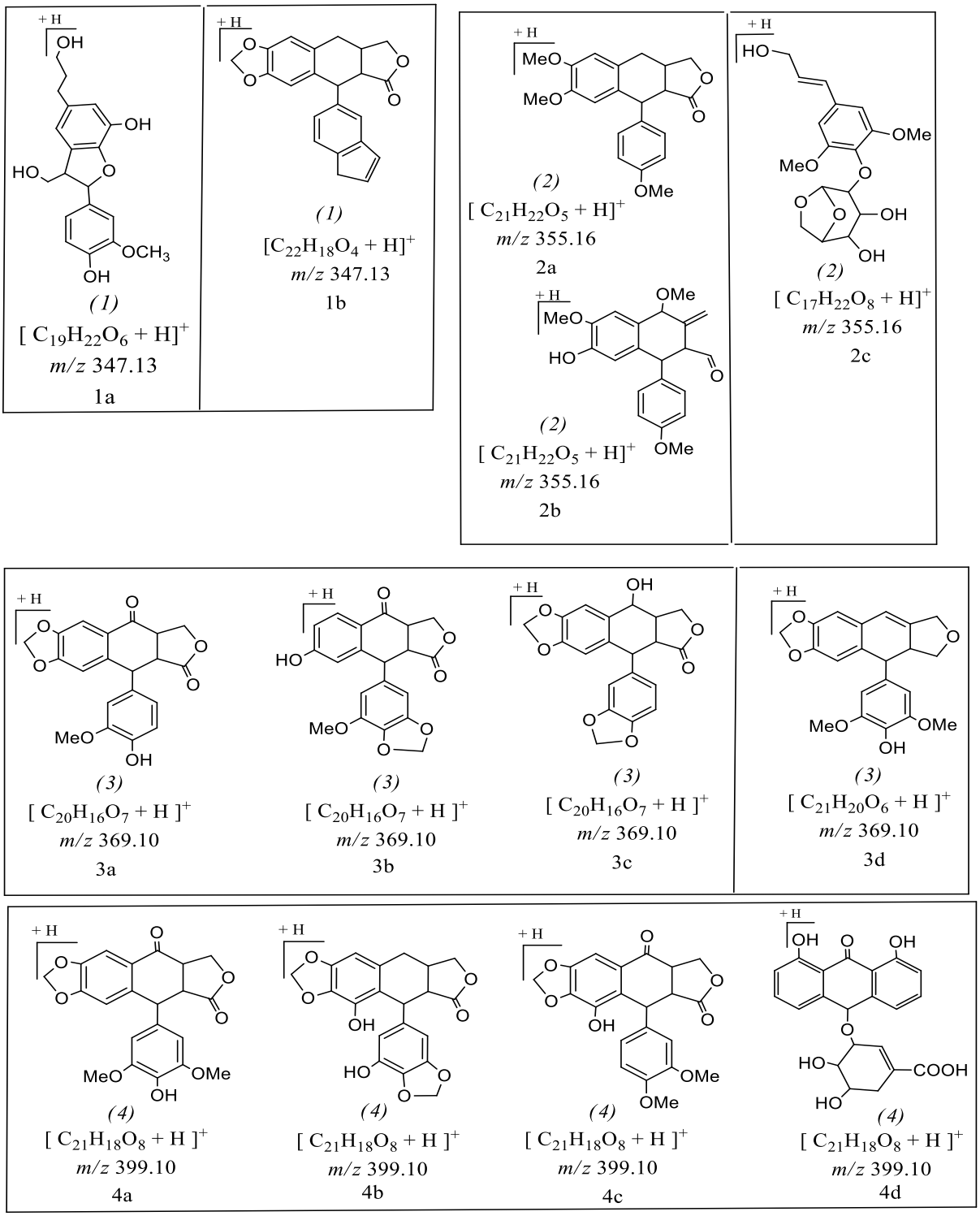
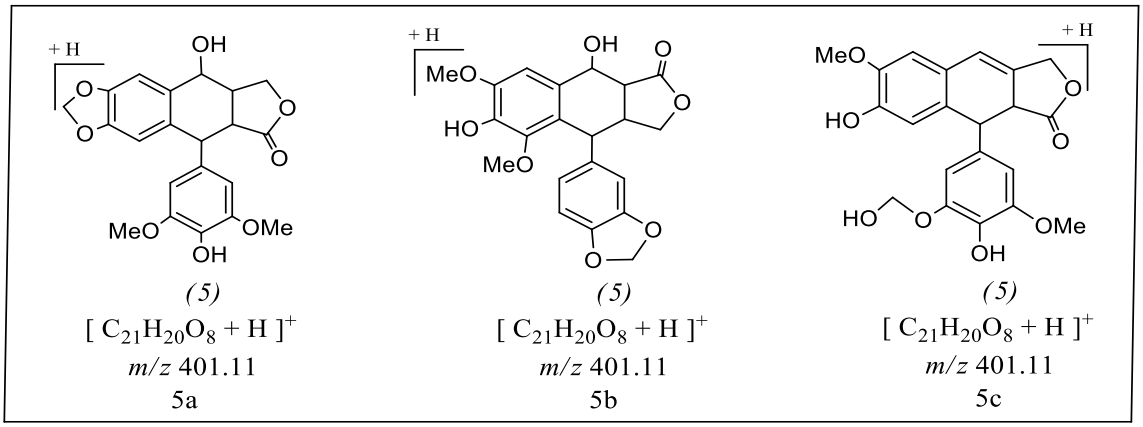
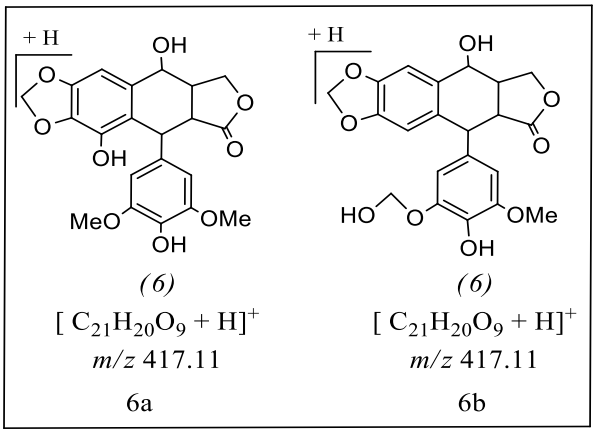
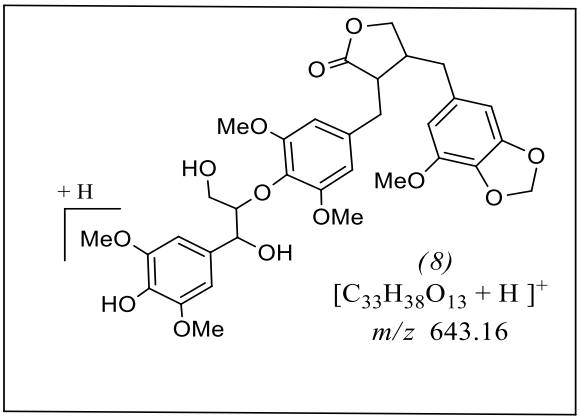


Figure A4.2a. structures of Lignin/Lignan oligomers and/or derivatives identified by APPI-QqTOF-MS/MS







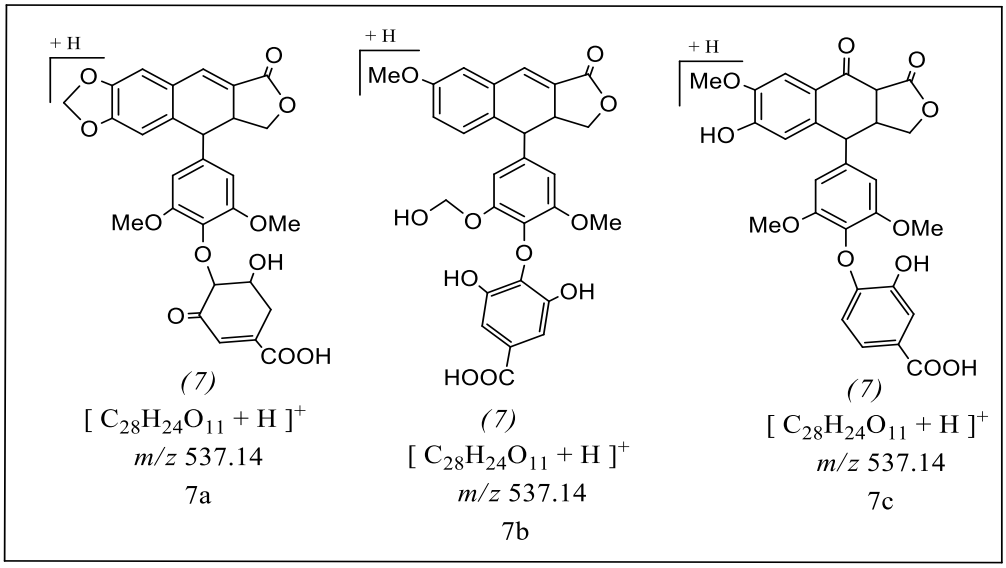


Figure A4.2b. structures of Lignin/Lignan oligomers and/or derivatives identified by APPI-QqTOF-MS/MS

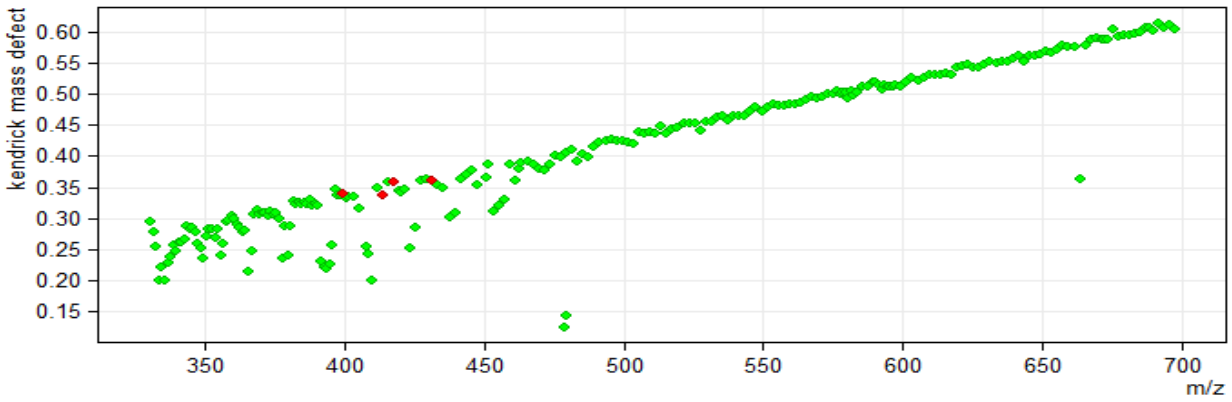


Figure A4.3. Kendrick mass defect (using CH_2 as a base unit) for the spectrum in the *m/z* 340-700 range of the APPI-QqTOF-MS of extracted French Oak VRLs.

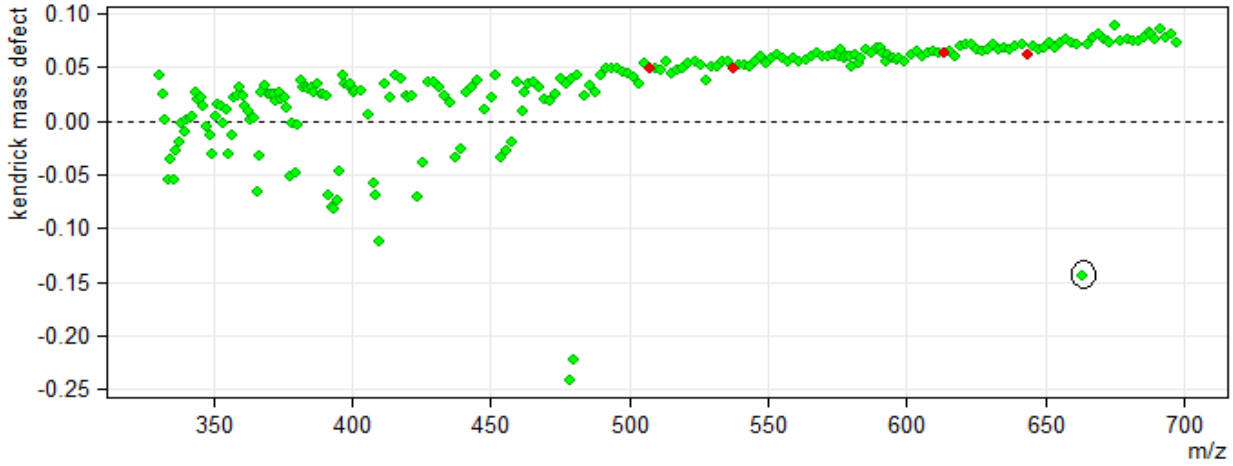
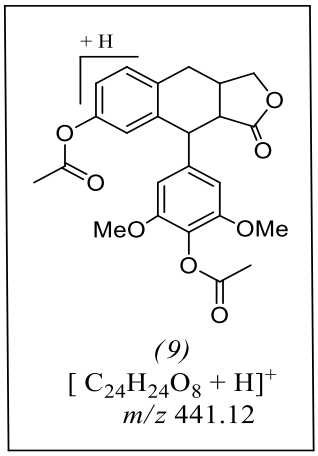
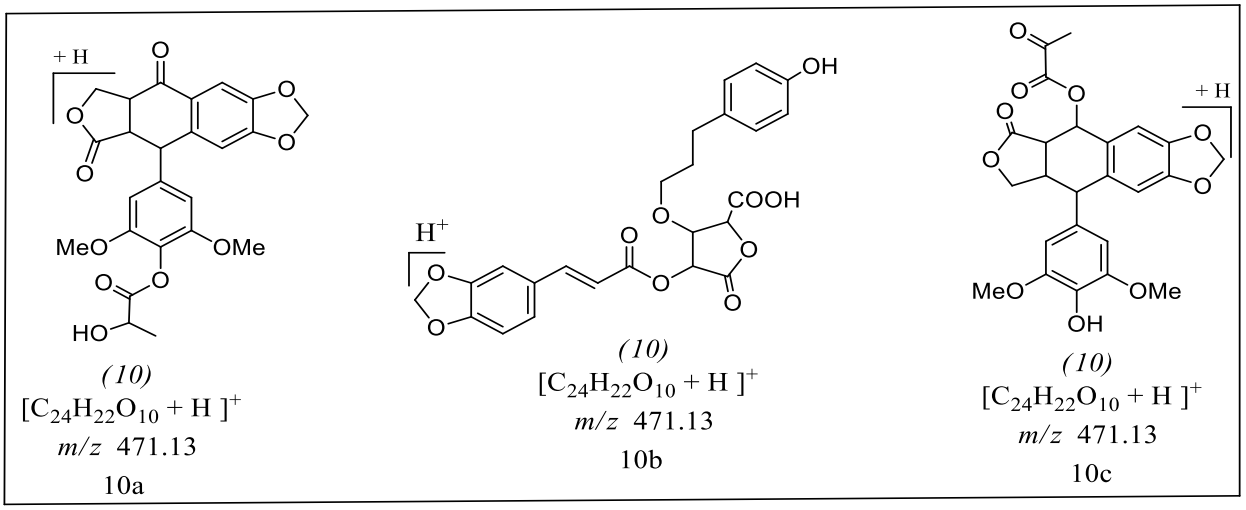


Figure A4.4. Kendrick mass defect (using OCH_2 as a base unit) for the spectrum in the *m/z* 340-700 range of the APPI-QqTOF-MS of extracted French Oak VRLs. The circled ion at *m/z* 663 has an entirely different mass defect than the mass defect range of the major bundle of ions.





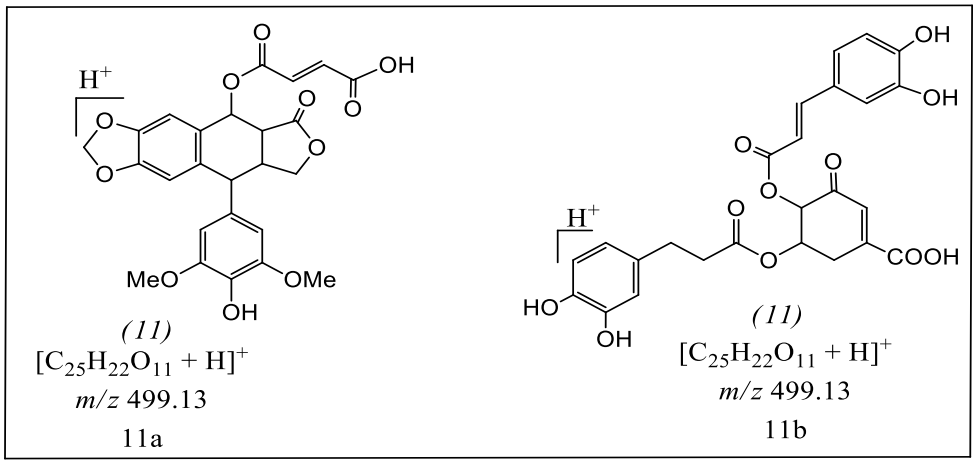
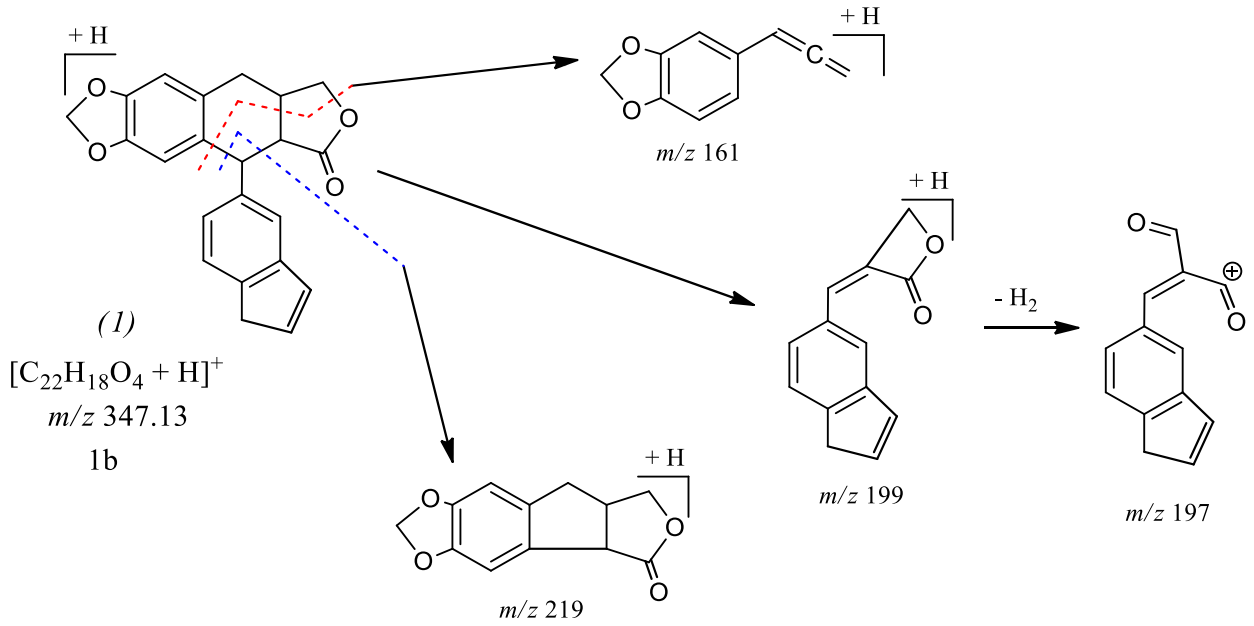
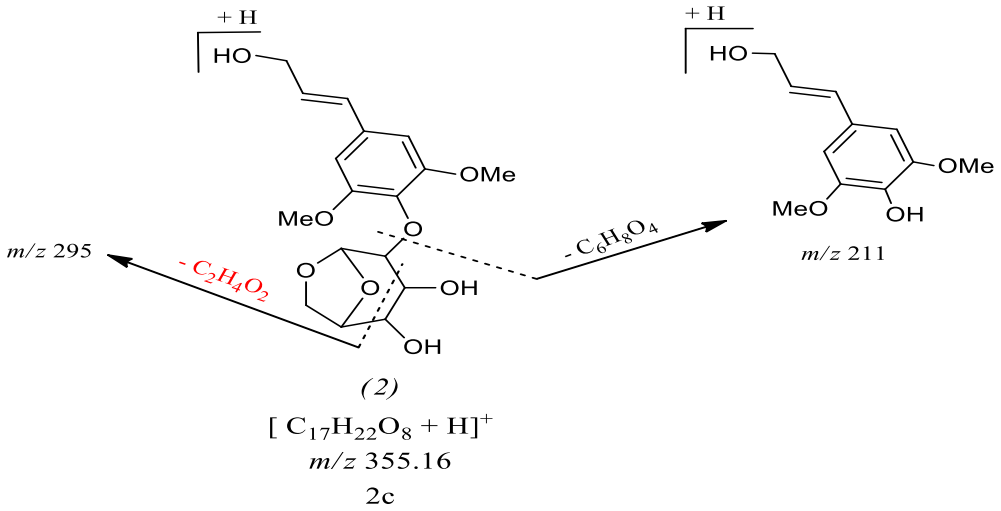


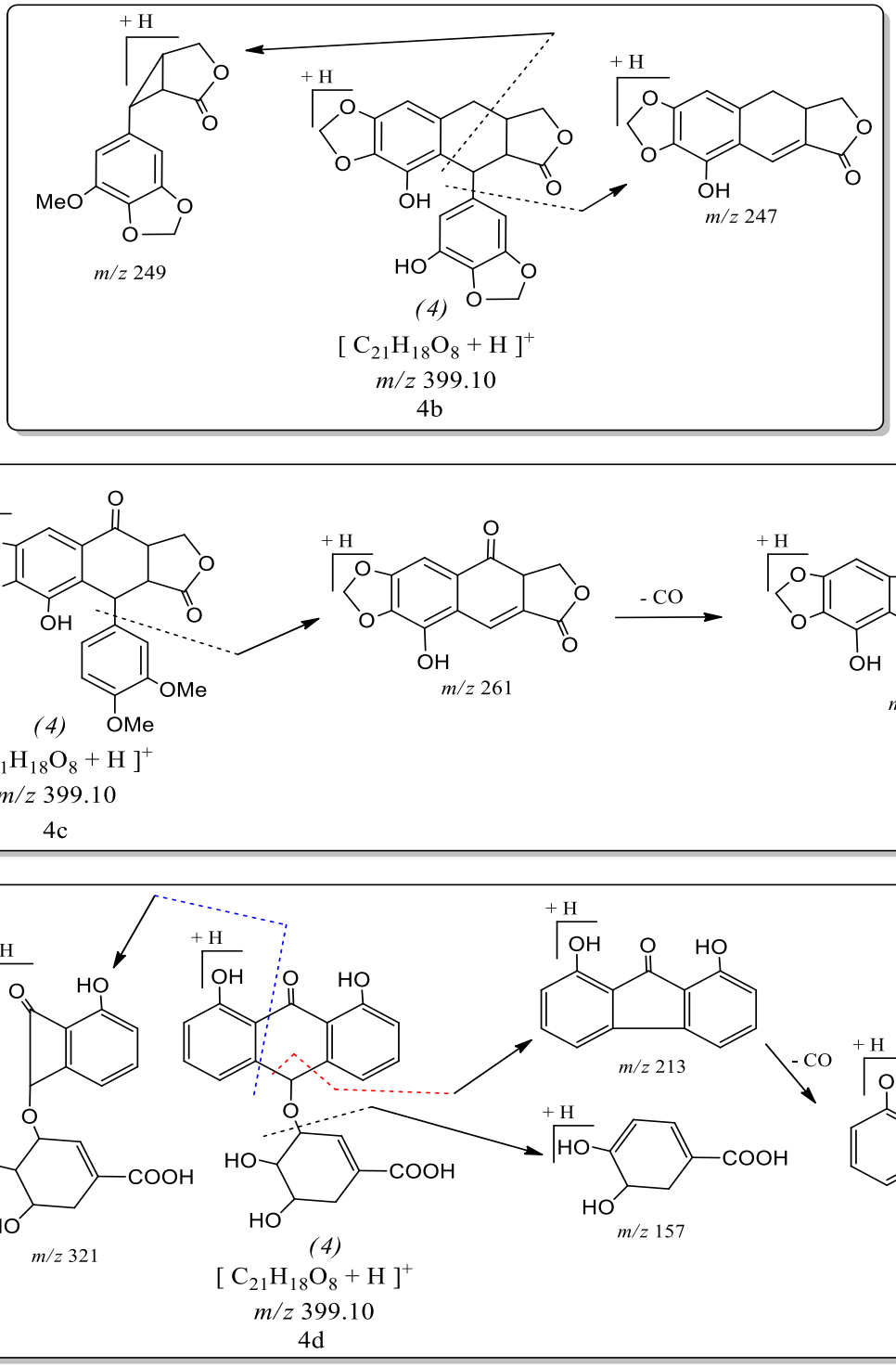
Figure A4.5. structures of Lignin/Lignan oligomers and/or derivatives identified by ESI-QqTOF-MS/MS



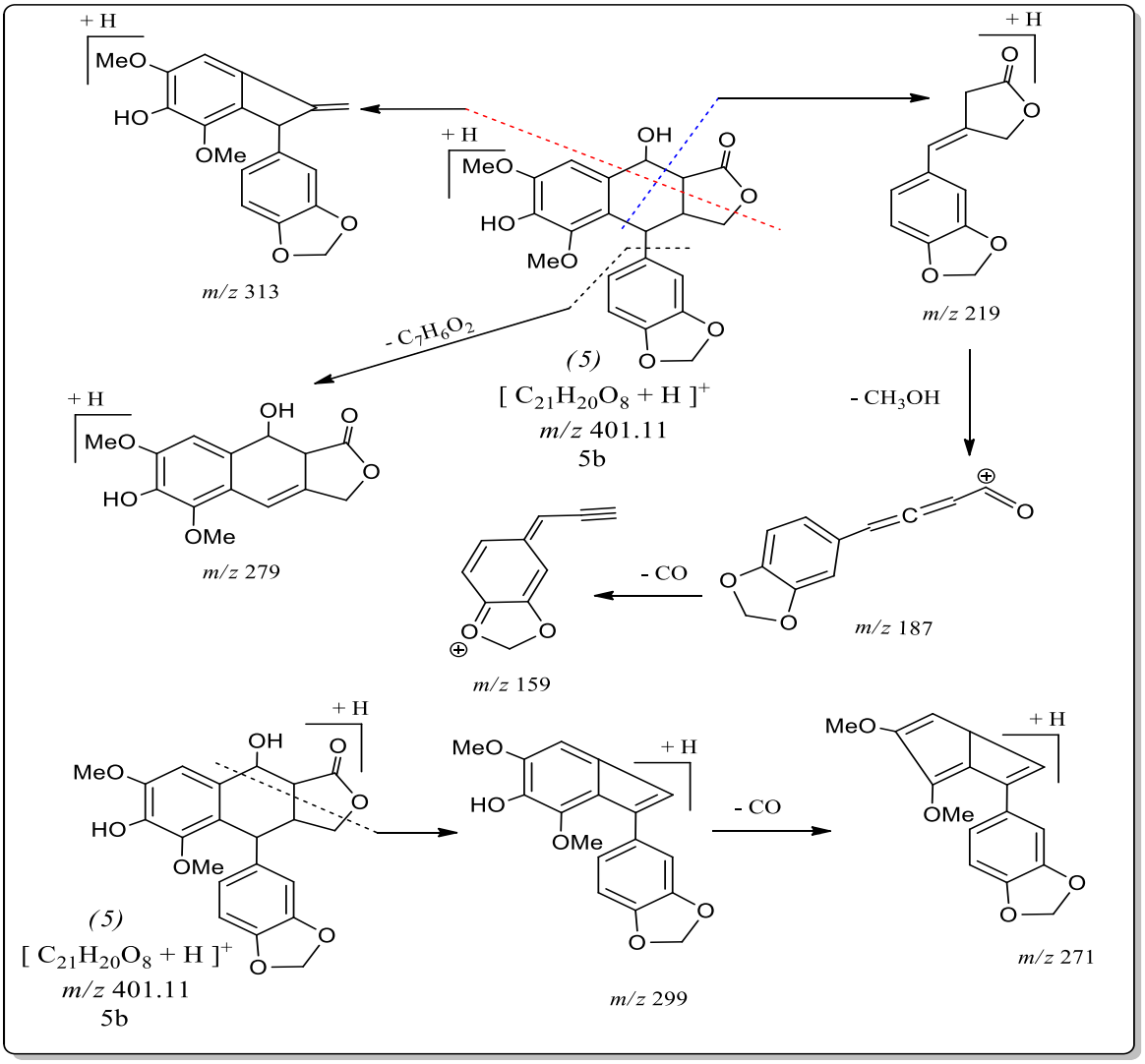
Scheme A4.1. Low energy CID-MS/MS fragmentation mechanism of the protonated precursor ion *1* at *m/z* 347.13

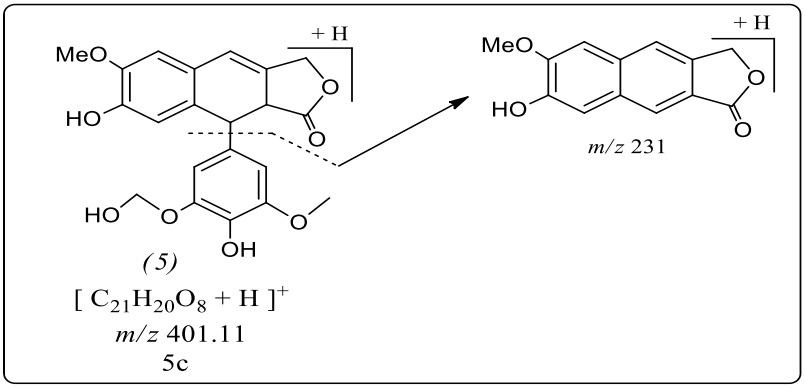


Scheme A4.2. Low energy CID-MS/MS fragmentation mechanism of the protonated precursor ion *2* at *m/z* 355.16

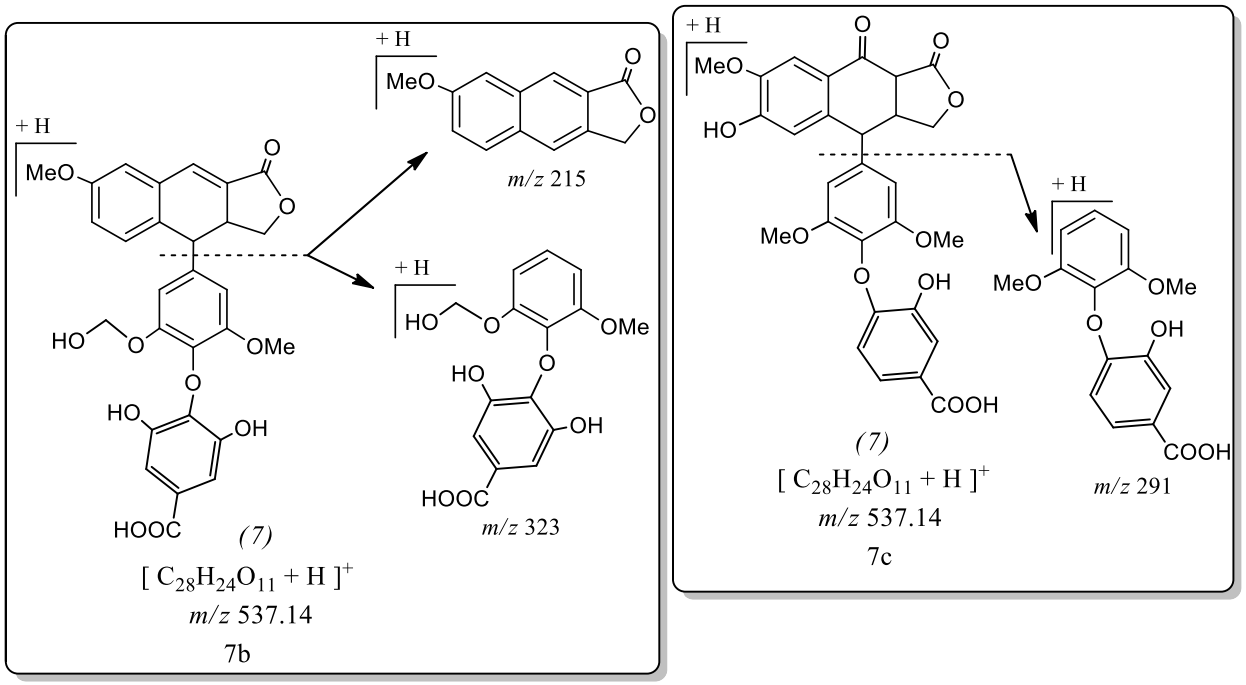


Scheme A4.3. Low energy CID-MS/MS fragmentation mechanism of the protonated precursor ion *4* at *m/z* 399.10

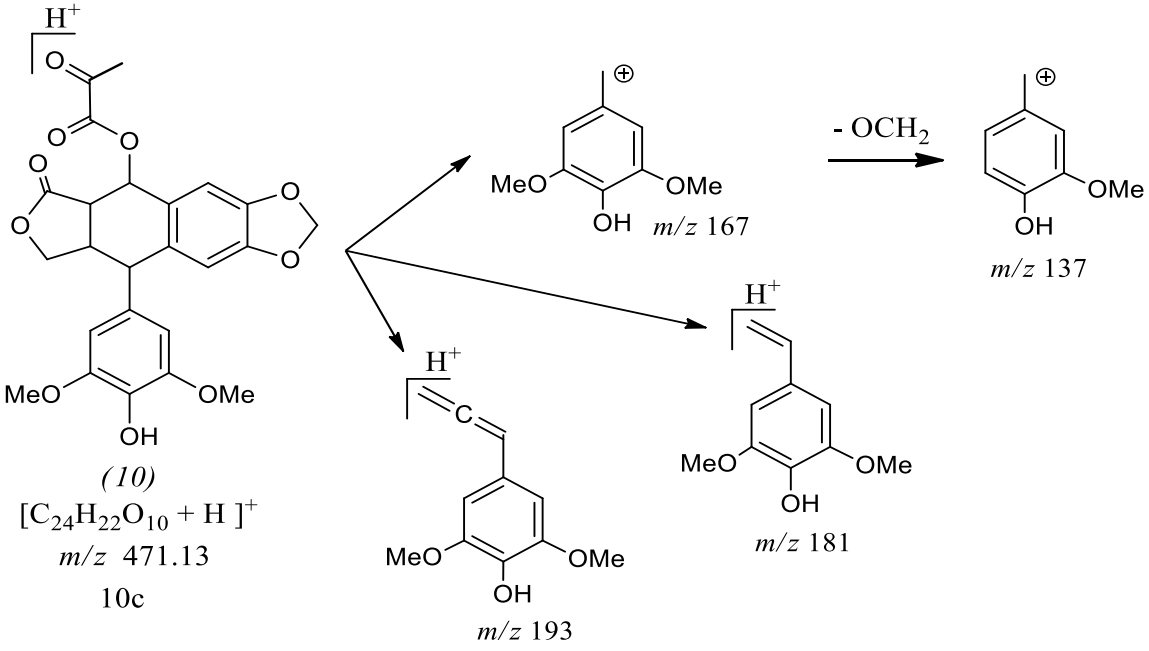




Scheme A4.4. Low energy CID-MS/MS fragmentation mechanism of the protonated precursor ion *5* at *m/z* 401.11



Scheme A4.5. Low energy CID-MS/MS fragmentation mechanism of the protonated precursor ion *7* at *m/z* 537.14.



Scheme A4.6. Low energy CID-MS/MS fragmentation mechanism of the protonated precursor ion *10* at *m/z* 471.13