DEMYSTIFYING AND UNRAVELLING THE FACTUAL MOLECULAR STRUCTURE OF THE BIOPOLYMER SPOROPOLLENIN

Abanoub Mikhael,[1] Kristina Jurcic,[2] Celine Schneider,[3] David Karr,[4] Gregory L. Fisher,[4] Travis D. Fridgen,[1] Alberto Diego-Taboada, [5] Grahame Mackenzie,[5] and Joseph Banoub*[1,6]

[1]Chemistry Department, Memorial University, St John's, 283 Prince Philip Dr, St John's, NL A1B 3X7, Canada

[2] MALDI Mass Spectrometry Facility, Western University, Department of Biochemistry, Medical Sciences Building 392, London, Ontario, Canada, N6A 5C1

[3] NMR Facility, C-CART/CREAIT network, Memorial University, St John's, 283 Prince Philip Dr, St John's, NL, A1B 3X7, Canada

[4]Physical Electronics, Inc. (PHI) 18725 Lake Drive East, Chanhassen, MN 55317, USA

[5]Department of Chemistry, University of Hull, Kingston upon Hull, HU6 7RX, UK

[6]Fisheries and Oceans Canada, Science Branch, Special Projects, 80 East White Hills Road, St John's, NL A1C 5X1, Canada

ABSTRACT

Sporopollenin is a natural highly cross-linked biopolymer composed of carbon, hydrogen, and oxygen which forms the outer wall of pollen grains. Sporopollenin is resilient to chemical degradation. Because of this stability, its exact chemical structure and the biochemical pathways involved in its biosynthesis remains a mystery and unresolved. It is obvious that a well-conceived coherent study of the sporopollenin structure details will help immensely scientists in better understanding the chemistry of their current applications of sporopollenin exines such as drug delivery, peptide synthesis, micro-reactors, and wastewater purification. As well, it may also lead to the discovery of newer biomedical applications in the next coming years.

We have identified and characterized the molecular structure of the clean, intact sporopollenin using mass spectrometric and nuclear magnetic resonance techniques. These analyses showed that sporopollenin is composed of a circular polyhydroxylated tetraketide polymer rigid backbone and a poly(hydroxyacid) branched network. The poly(hydroxyacid) network chains are attached by covalantly ether bonds to the polyhydroxylated tetraketides rigid backbone, forming the scaffold of the spherical sporopollenin.

The shells of the plant spores are natural microcapsules, which have progressed 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 was generally designated as "one of the most exceptionally resistant materials knows in the organic world."[2] Sporopollenin composition was defined to be a highly cross-linked polymer composed of carbon, hydrogen, and oxygen.[2] It is highly resilient to chemical degradation, which is why its exact chemical structure and the biochemical pathways involved in its synthesis are not yet clear.[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.[2] 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] Gordon Shaw *et al.* [5], one of the earlier pioneers studying the structure of sporopollenin, indicated 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 fatty acids and/or alcohols contradicting the proposal that sporopollenin was a carotenoid derivative. In summary, there are three major opinions about the chemical nature of sporopollenin. It can be constructed completely by an aliphatic biopolymer, or it can be completely built as an aromatic biopolymer, and finally it was proposed to exist as a mixture of aliphatic and aromatic biopolymer. [7-9]

It is very important to point out that combustion of different sporopollenins under inert atmosphere (pyrolysis), coupled with 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 material, especially when containing long aliphatic chains that can aromatize and produce aromatic compounds which originally did not exist.[12] Bernard *et al.*[13] studied the thermal degradation of sporopollenin and realized that this occurred in two main stages. The first one below 500 ºC, where Sporopollenin undergoes simultaneous dehydrogenation and deoxygenation. The second stage above 500 ºC, where aromatic products are formed. This study shows that high-

temperature pyrolysis can produce aromatic compounds which lead to blatantly misleading errors in the structure elucidation of sporopollenin. Another main detail from the Bernard study,[13] 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 or alcohols.

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 ketide is one of the important sporopollenin monomers.[14,15] The polyhydroxylated ketides are composed of the α-pyrone ring with a hydroxylated aliphatic chain. It has been proposed that these polyhydroxylated ketides, along with fatty alcohols and or fatty acids, may form the sporopollenin building blocks [16]. 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. [17,18]

Recently, the group of Li *et al.*[19] reported that by using a recent adoption of high-energy ball-milling and a newly developed thioacidolysis degradative method together with state-of-the-art solid-state NMR techniques, they determined the detailed molecular structure of pine sporopollenin. These authors have shown that pine sporopollenin was primarily composed of aliphatic hydroxylated polyketides, aromatic moieties, and acetal groups as a cross-linker.

In this communication, we present a different point of view which does not agree with the work of Li *et al.* [19] (Please refer to SI-1). We present herein, the top-down structural elucidation accomplished on *Lycopodium clavatum.* We have shown by X-ray photoelectron spectroscopy (XPS) the total absence of aromaticity in sporopollenin. In this study, we have identified a series of branched poly(hydroxyacid) with glycerol as a core unit, by using TOF-SIMS tandem MS imaging, and MALDI-TOF-MS followed by high energy collisional induced dissociation CID-TOF/TOF-MS/MS. The poly(hydroxyacids) are composed mainly of various combinations of the C16 and C14 with three oxygens substituted on their chains. These fatty acids contain one terminal hydroxyl group, and the other two hydroxyl groups exist in their chains as a beta-diketone moiety. The structures of these branched poly(hydroxyacid) provides the first-ever report in the literature, showing the existence of a natural "Dendrimer" like molecule.

Furthermore, MALDI-TOF/TOF-MS/MS showed the presence of a cyclic polyhydroxylated tetraketide repeating unit (alpha-pyrone ring plus hydroxylated aliphatic chain) that forms a cyclic polyhydroxylated tetraketide polymer backbone. The hydroxyl groups of this

circular polymer can be attached to the terminal hydroxyl groups of the poly(hydroxyacids) through ether linkages. Please bear in mind that the α-pyrone rings possess weak aromatic characters, as they can undergo some reactions such as ring-opening, Diels-Alder reactions, and electrophilic additions.[20] Fortunately, the α-pyrone ring is not a typical aromatic compound, and for this reason, it does not contradict the absence of aromaticity in the sporopollenin revealed by the XPS. Consequently, this cyclic polymer appears to act like the building stone on which the poly(hydroxyacid) network can be built, forming the scaffold of the spherical sporopollenin. Also, it is possible that sporopollenin wall consists of a dendrimer possessing alternative layers of polyhydroylated tetraketide cyclic polymer and poly(hydroxyacids) network.

Lastly, 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.

Results and Discussion

The *Lycopodium clavatum* sporopollenin used in this investigation was extracted by the classical method of Zetzsche *et al.* [21] method using hot acetone, potassium hydroxide, and phosphoric acid. This series of sequential treatments are required to extract the protein-free hollow intact and clean exine microcapsule. [22] It is essential to note that in this communication, we used sporopollenin identical to the intact, clean exines prepared by the original Zetzsche *et al.* method.[28] Originally, these sporopollenin microcapsule exines were used 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). [23]

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

The HR-XPS C(1s) surface analysis of sporopollenin (SI-figure 1) showed the main types of linked carbons atoms in the in *Lycopodium clavatum* sporopollenin with their percentage in the sample according to the area calculated under each peak. Most importantly, it showed the complete absence of any satellite peak at higher binding energy that results from the π-π^* transition. It should be noted that the absence of this "shake-up line" dismisses the presence of aromatic compounds and support the absence of any extended system of double bonds and/or aromaticity in the sporopollenin. [24]

2. TOF-SIMS and keV-CID MS/MS

The TOF-SIMS tandem MS imaging (+ ion mode) of the sporopollenin showed th*e* ions at *m/z* 575 and 603, these ions are characteristic for diacylglycerol (DAG) derivatives and assigned as $[C_{16}H_{25}O_4C_3H_6OC_{14}H_{21}O_4+H-H_2O]^+$ and $[C_{16}H_{25}O_4C_3H_6OC_{16}H_{25}O_4+H-H_2O]^+$, respectively (SI-figure 2a). The product ion scans of the precursor ions at *m/z* 575 and 603 shows 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 (SI-figure 3 a&b). These acylium product ions contain another two oxygen atoms on their chains, but each product ion eliminates only one water molecule. This indicates 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 from these acylium ions. Moreover, 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 (SI-figure 2b).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 mentioned before (SI-figure 4 a & b). A proposed fragmentation mechanism for these acids is shown in SI-Scheme 1.

Overall, the fragmentation mechanism shown in SI-scheme1 suggests that these carboxylic acids contain a beta diketone moiety in which, one oxygen occurs as enol form (which is lost as a water molecule), whereas the other one exists as keto form (which cannot be lost as a water molecule). This structural pattern is favorable for the presence of a beta diketones structure (SI-Figure 5), as it facilitates the formation of intramolecular hydrogen bond, that gives extra stability to these kinds of structures.[25] This hydrogen bonding resembles the case of intramolecular hydrogen bond in curcumin molecules [26].

It should be noted that these identified fatty acidss contain an extra terminal hydroxyl group in their original structures, that allows them to be attached to another fatty acid through ester bond formation producing the poly(hydroxyacid) network of sporopollenin. This fact was deduced, after performing the MALDI-TOF-MS (+ ion mode using CHCA as a matrix) of sporopollenin, and an ion at *m/z* 1643.9948 was identified and assigned as $[C_{89}H_{142}O_{27}+H]^+$. This ion confirms that the fatty acids forming the poly(hydroxyacid) network contain three oxygens in their chains. This will be discussed again later in this manuscript in section 3.2 (figures 2 a & b, and figure 3).

Moreover, the chemical formulas of all other ions formed in the positive or negative ion mode TOF-SIMS-MS are shown on the spectra (SI-figure 2a&b).

3. MALDI-TOF-MS and HIGH- ENERGY (1 Kev) CID-MS/MS

3.1. MALDI-TOF-MS (+ ion mode) and CID-MS/MS (DAN Matrix)

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.[27] It showed a complex series of ions, that were very closely related; indicating the presence of a very complex heterogeneous mixture (SI-figure 6 a & b). Furthermore, when measuring their Kendrick mass defect plot, we noticed the formation of a bundle of ions, that were strictly related to each other's (SI-figure 7). [28] This Kendrick mass defect plot displayed a series of ions that in general, varied in the number of methylene groups (14 Da). This is clearly demonstrated in the zoomed part of the MS and its Kendrick Plot in the range of *m/z* 1300-1500 Da (SI-figure 6 a & b, and Figure 7).

As an example, we assigned the radical cation at *m/z* 1965.1278 as $[C_{106}H_{164}O_{33}]^{+\bullet}$ composed of two hydroxylated triglycerides (6 x C14) connected by an extra spacer (C16) hydroxylated carboxylic acid (Figure 1) and appears like a branching units in triacylglycerol (TAG) dendrimers.[29] 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. Other ions chemical formulas are shown on the spectrum (SI-figure 6 a & b).

As another example, we have assigned the ion at *m/z* 1441.8037 as $[C_{81}H_{132}O_{21}+H]^{+}$ and its molecular structure was supported by its proposed CID-MS/MS fragmentation pathways, as shown in SI-figure 8 and scheme 1.

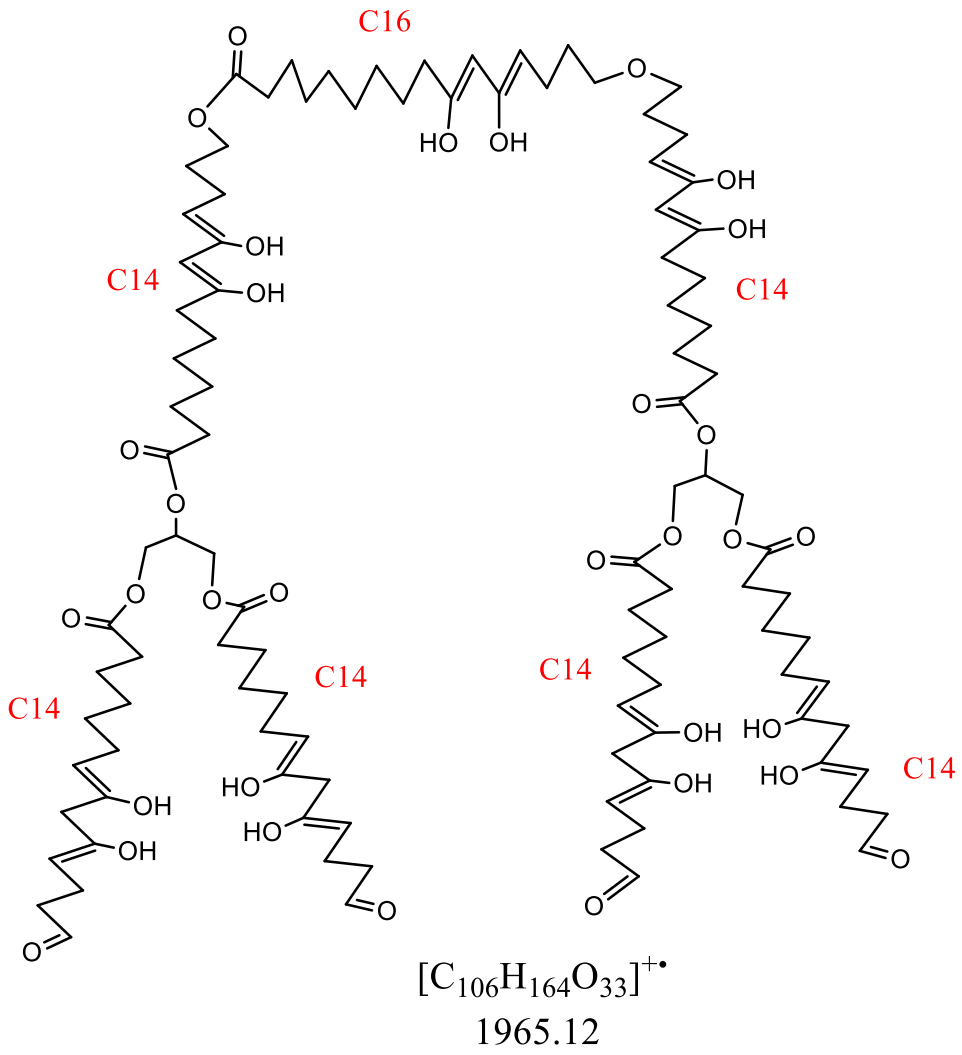
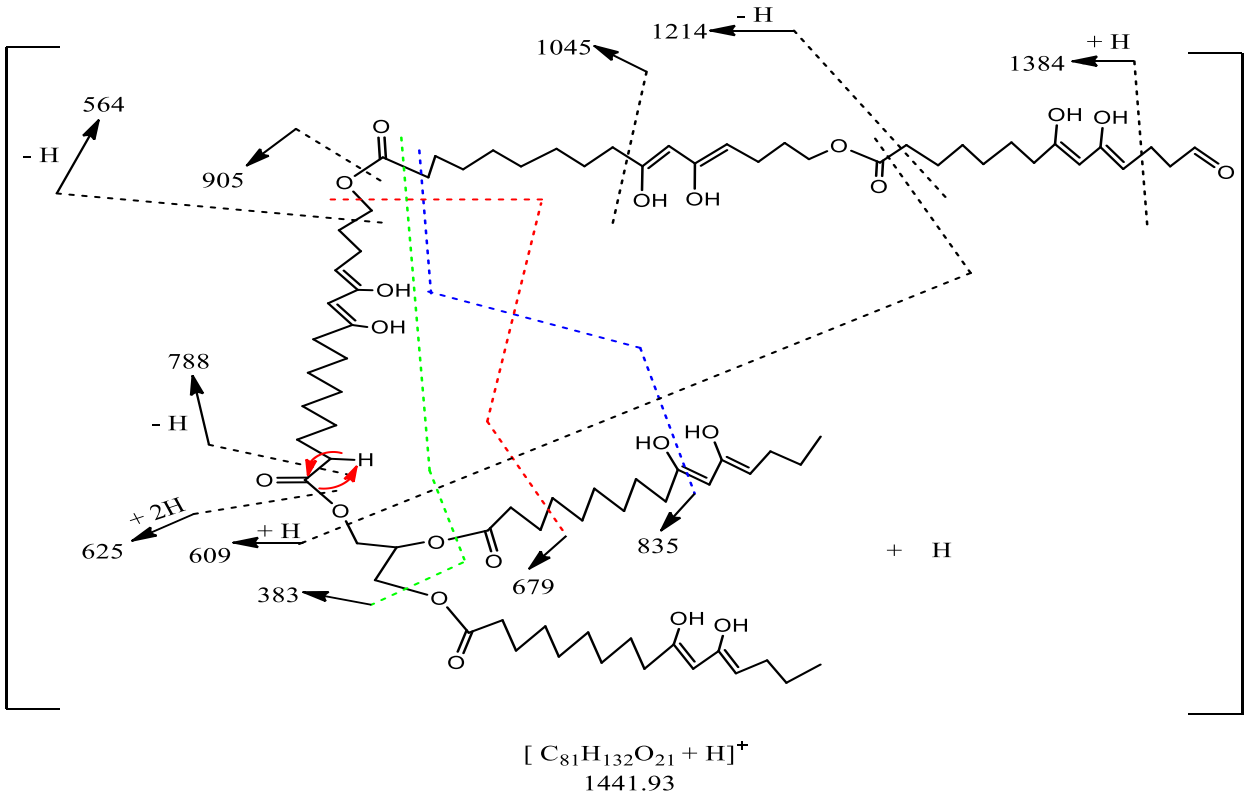


Figure 1. The proposed structure of the poly(hydroxyacid) with glycerol as a core unit at *m/z* 1965 identified in MALDI-MS using DAN as a matrix



Scheme 1. Mechanism of the product ion scan of the precursor ion at *m/z* 1441.93 identified in Sporopollenin MALDI-MS using DAN as a matrix. Note: the product ion at *m/z* 625 is formed through ketene loss indicated by red arrows [30].

3.2. 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) than the ones obtained by DAN (Figure 2a and b). However, it should be noted that the mass region around *m/z* 100-1300, was mainly composed of CHCA matrix peaks,[31] whereas the region from 1500 to 2500 Da was composed of significant analyte ions. In the zoomed region from *m/z* 1500 to 1700, we observed a series of ions that differ by C_2H_4 (2 x CH_2) in their structure (Figure 2a and b). The most important one is the peak at *m/z* 1643.9948 (mentioned before in the TOF-SIMS section) which was assigned as $[C_{89}H_{142}O_{27}+H]^{+\cdot}$ This assignment fair very well with the empirical formula of *Lycopodium clavatum* sporopollenin ($C_{90}H_{142}O_{27}$) calculated by Zetzsche *et al.* (Figure 5) [21]. This old information about the *lycopodium clavatum* sporopollenin empirical formula, was unexpectedly, the most useful information that helped us in the interpretation of the whole data in this manuscript.

This ion was identified with good accuracy (+ 4.9 error ppm), and its chemical formula fits the experimental isotopic distribution pattern with 71.2%.

With this new finding, we finally arrived in obtaining new clear evidence to support and characterize the ions which corresponded to the empirical formula $C_{90}H_{142}O_{27}$ of sporopollenin extracted from *Lycopodium clavatum* (Figure 2a and b).

Another interesting region which is important to discuss is in the range of *m/z* 2300 to 2500, where we observed a series of ions that differed by C_2H_4 units in their structure. As an example from this region, we showed the proposed structure for the peaks at *m/z* 2428.4433 (TAG) assigned as $[C_{133}H_{207}O_{39}]^+$ containing nine fatty acids (seven trihydroxy C14 and two trihydroxy C16) plus glycerol as a core unit (Figure 3). It should be noted that the important ion at *m/z* 1643.9948 (DAG) derived from *m/z* 2428.4433 (TAG), and it is composed of six fatty acids plus glycerol moiety (five trihydroxy C14 and one trihydroxy C16) (Figure 3). These proposed structures provide more support for the presence of poly(hydroxyacid) with glycerol as a core unit and show the structure of the smallest unit (empirical formula) of *lycopodium clavatum* sporopollenin.

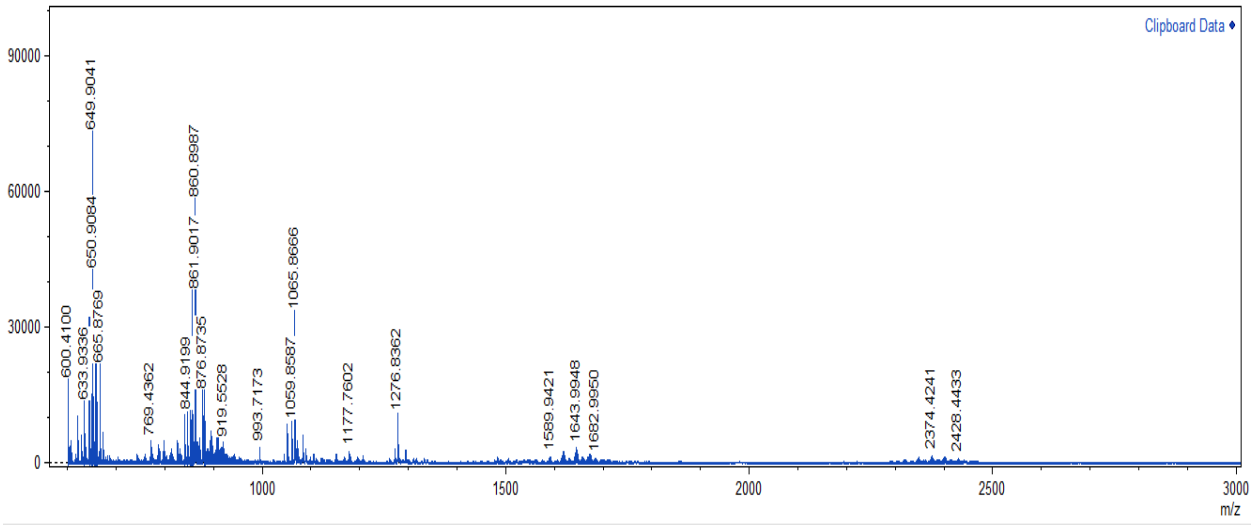


Figure 2a. MALDI-MS of sporopollenin using CHCA as a matrix

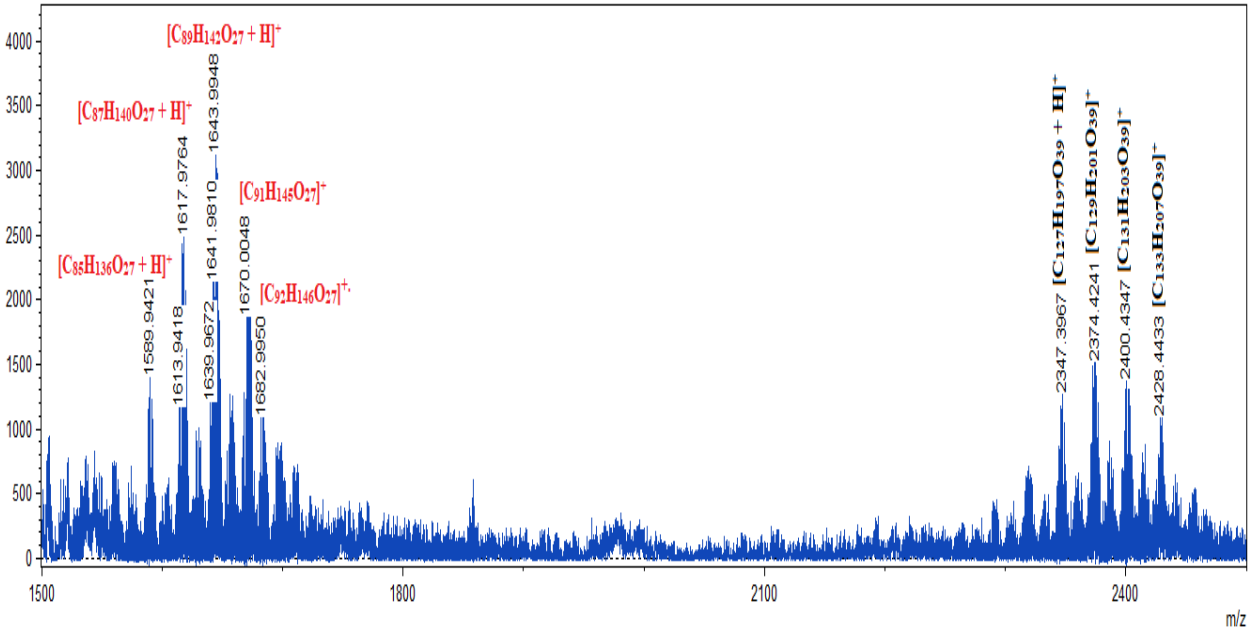


Figure 2b.(continued) Zooming of Sporopollenin MALDI-MS in the high mass region at *m/z* 1500-3000.

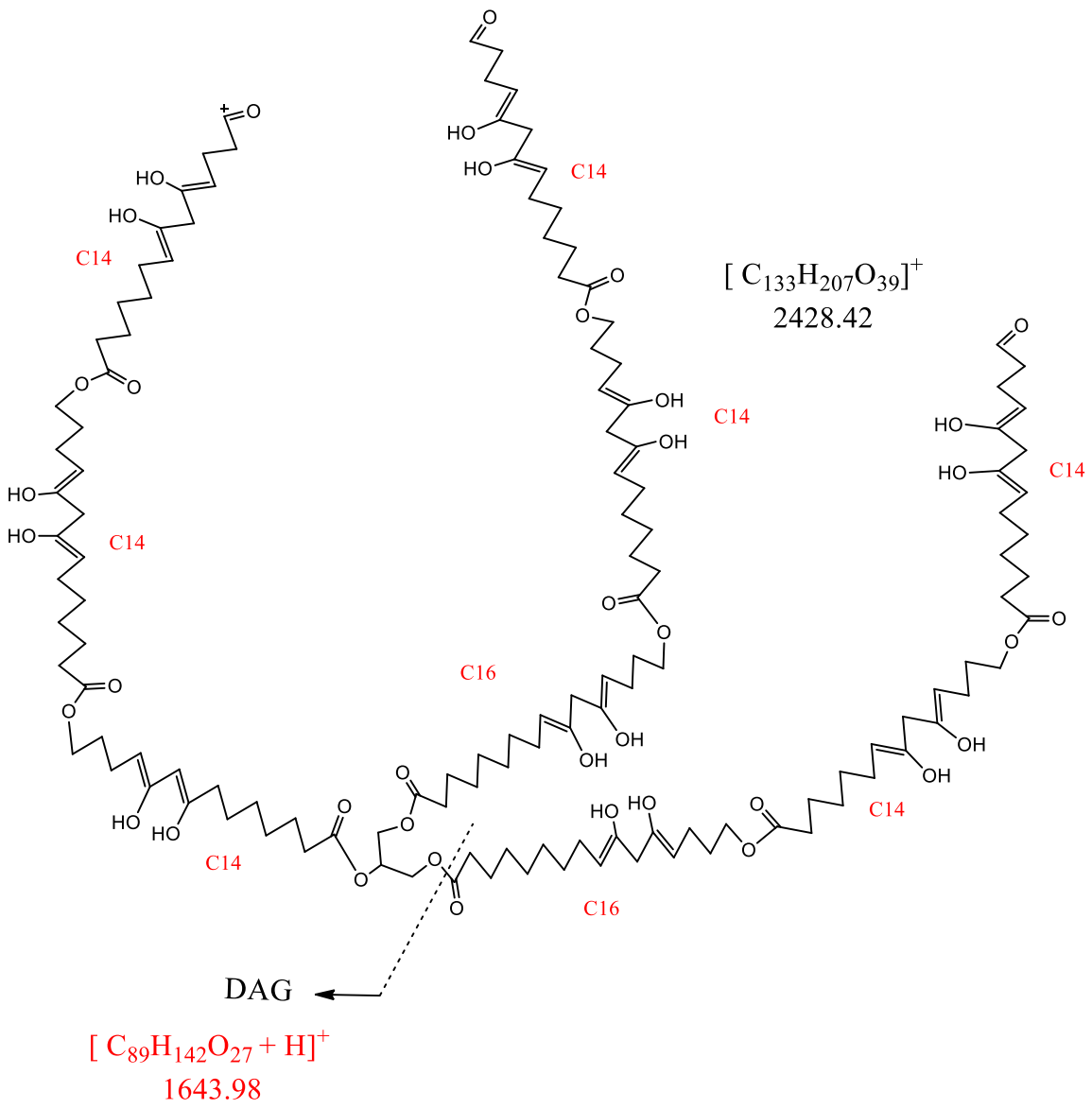


Figure 3. The proposed structure of *m/z* 2428.42 and 1643.98 identified in sporopollenin MALDI-MS using CHCA as a matrix

3.3. MALDI-TOF-MS (+ion mode) and CID-MS/MS (HABA Matrix)

MALDI-TOF-MS (+ ion mode) of the sporopollenin using HABA as a matrix showed a different MS pattern (polymer with a 280 Da repeating unit) than the one observed with either DAN or CHCA . Contrary to these two last matrices, the MALDI-MS with HABA matrix gives new structural information about sporopollenin. A series of sodiated molecular ions containing both Na and K in their molecular formulae was observed using HABA as matrix. 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 4).These four sodiated molecular ions differed from each other by a repeating unit of 280 Da assigned as a polyhydroxylated tetraketide derivative with chemical formula $C_{13}H_{12}O_7$. Please note, that this series of sodiated molecular ions contained both K and Na carboxylates in their respective structures and these were formed during the KOH step in the extraction process of sporopollenin from *Lycopodium clavatum* pollen grains plus the washing step at the end using NaOH as indicated in the experimental section. [32] During this KOH treatment, proteins were removed, and ester bonds in its outer surface became partially hydrolyzed to carboxylic acid salts, and this increases sporopollenin hydrophilicity. [32] Moreover, this

Moreover, we observed another series with almost the same chemical formulae, except, that the K atoms were replaced by Na (-16Da), these ions were 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 4).

Unexpectedly, the high-energy 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 (*m/z* 1983→ m/z 1703→ *m/z* 1423→ *m/z* 1143→ *m/z* 863) (Figure 5 a & b, and scheme 2).

It is evident from the MS/MS of the precursor ion at *m/z* 1983 afforded the product ion at *m/z* 863, which in turn produced the secondary product ion at *m/z* 287. This indicates that the first product ion at *m/z* 863 is composed of three attached units of the secondary product ion at *m/z* 287 (3 x 287 Da + 2H= 863). Furthermore, the product ion scan fragmentation of this precursor ion is initiated by loss of -CH=CH-CH_2-CH=CH- (66 Da), which agrees with the presence of unsaturated fatty acids. Moreover, this product ion scan is not initiated by the loss of 44 Da supporting the absence of free carboxylic acid groups. Other CID-MS/MS fragmentation pathways that support the proposed structure of the precursor ion at *m/z* 1983 are shown in scheme 2.

To support our proposed structure of the precursor ion at *m/z* 1983 and its CID-MS/MS fragmentation patterns presented in scheme 2, we measured the *quasi*-MS^3 spectrum of the product ion at *m/z* 863.[33-35] (SI-figure 9). The *quasi*-MS^3 spectrum of the product ion at *m/z* 863, showed the presence of both sodium and potassium atoms in the chemical composition of this product ion.

It should be noted that the presence of both Na and K atoms in this ion, indeed, 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); (Figure 5 a & b). After examining series of rational structures, we have chosen the most tailored reasonable structure that agreed with the exact mass and the MS/MS of the original precursor ion at *m/z* 1983, $[C_{94}H_{91}KNa_2O_{41}+Na]^+$. Consequently, the product ions at *m/z* 863 can be isobaric either $[C_{42}H_{43}KNa_2O_{13}+Na]^+$ or $[C_{42}H_{43}Na_3O_{13}+K]^+$; and is composed of three tetraenoic C14 carboxylic acid derivatives. Furthermore, the *quasi*-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 (SI-figure 9). This *quasi*-MS^3 fragmentation pattern supports the presence of the free carboxylic acid group in the structure of *m/z* 863 which was created 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 3.

It is important to note that the branched carboxylate salts may exist originally as cyclic esters that were hydrolyzed during the extraction process. Also, the presence of the double bonds may originally exist as shown in the structure of the product ion at *m/z* 863, or it can be formed by the dehydration of a hydroxylated chain, during the extraction process

It should be noted 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). Similarly, we calculated the theoretical isotopic distribution of the chemical formula assigned to the highest intensity peak in the MS (*m/z* 1127). The theoretical isotopic distribution of this peak showed excellent fitting with the experimental isotopic distribution (ca. 90% as shown in SI-figure 10).

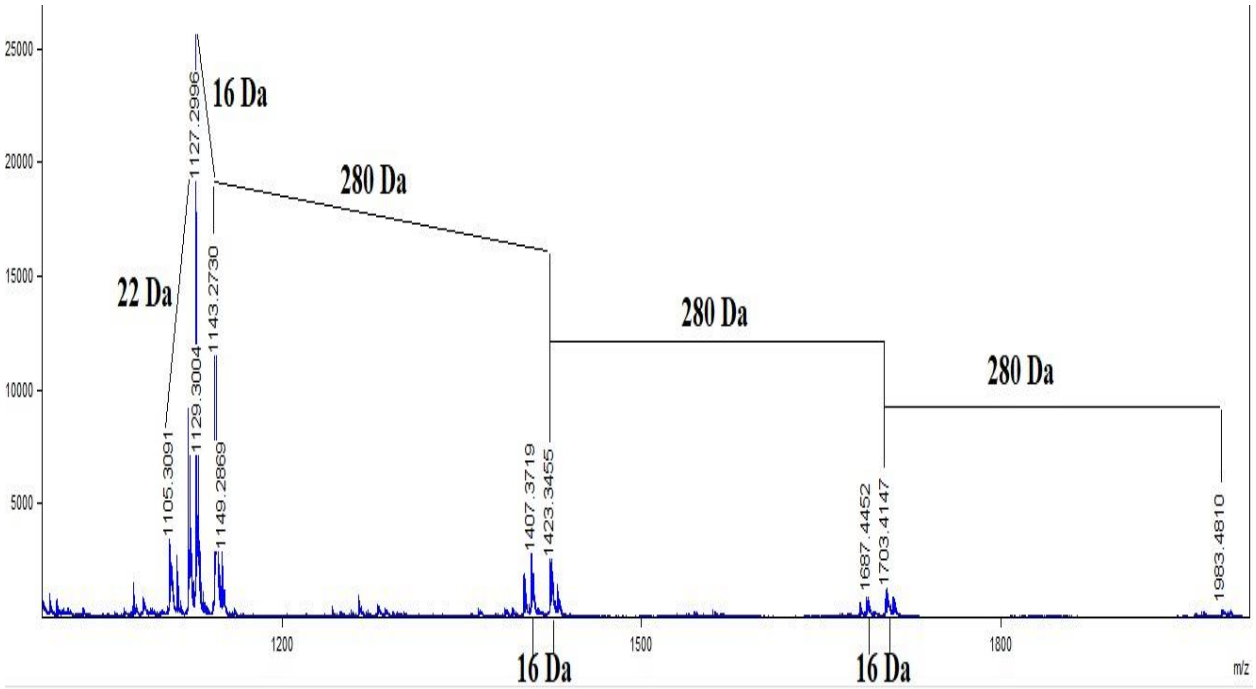


Figure 4: MALDI-TOF-MS (+ ion mode) using HABA as a matrix and displaying the presence of a repeating unit of 280 Da

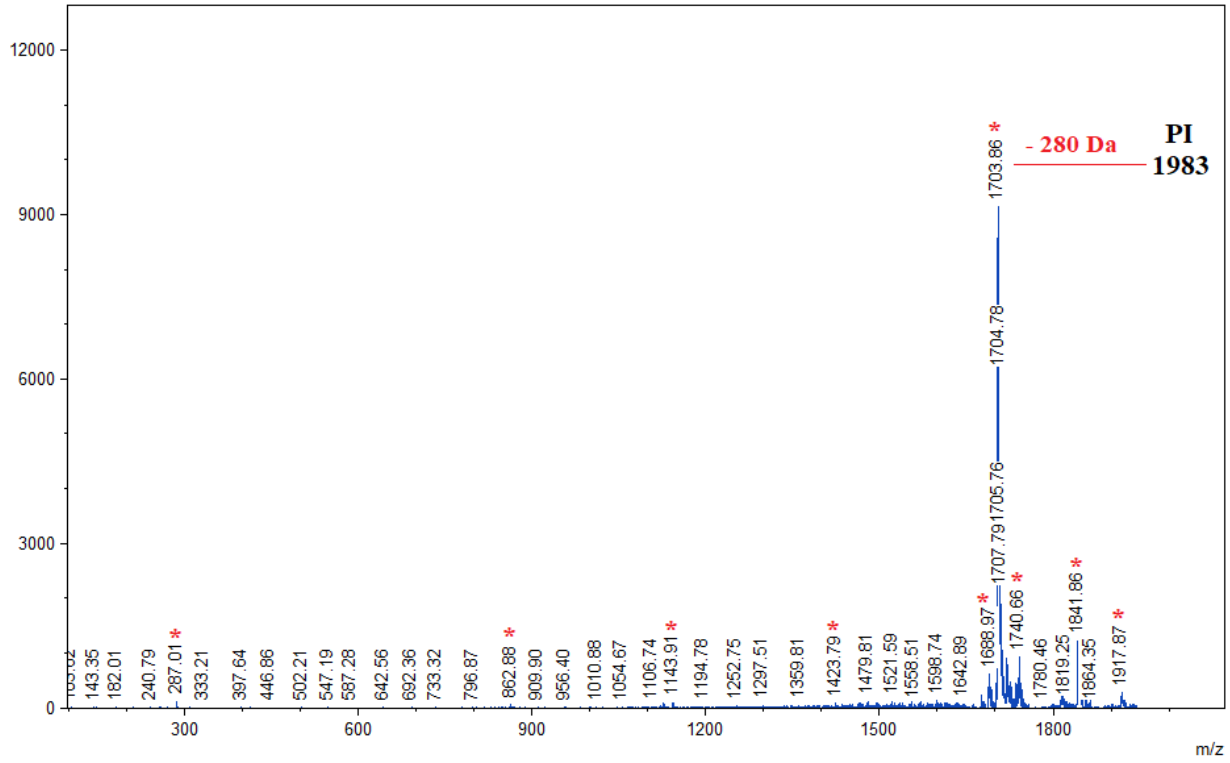


Figure 5a. Product ion scan of the precursor ion at m/z 1983 showing a loss of 280 Da to produce the base peak at m/z 1703

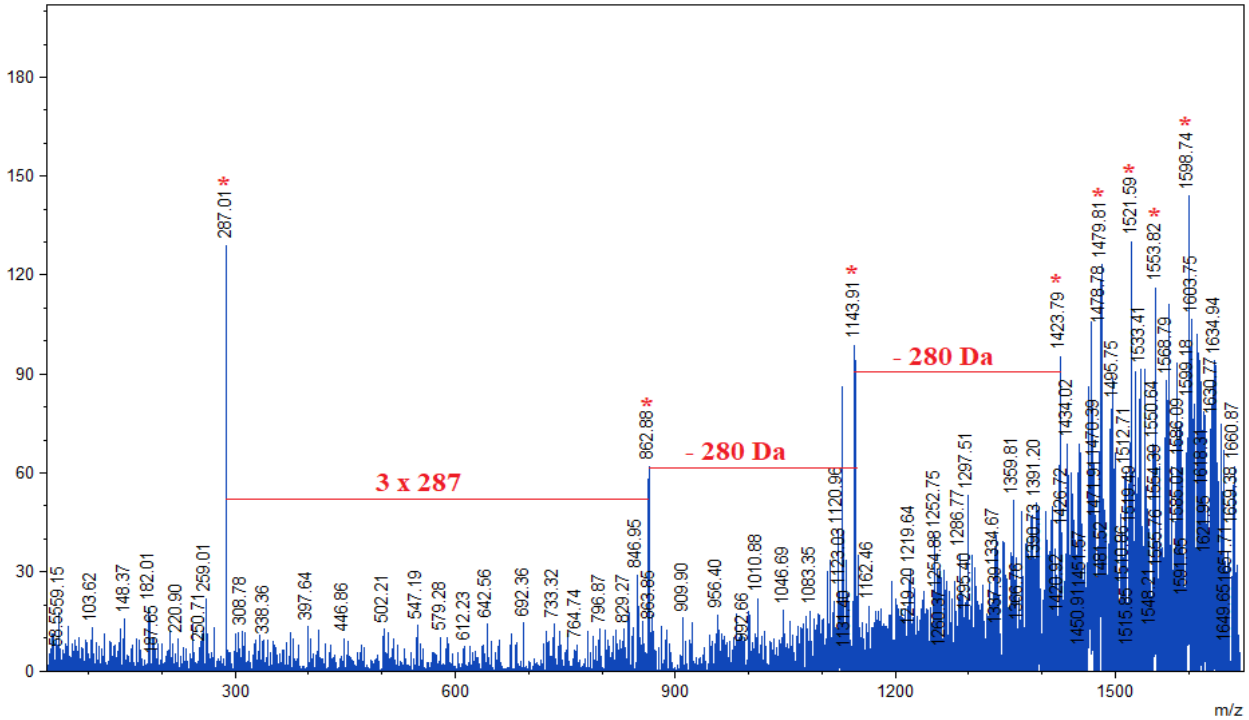
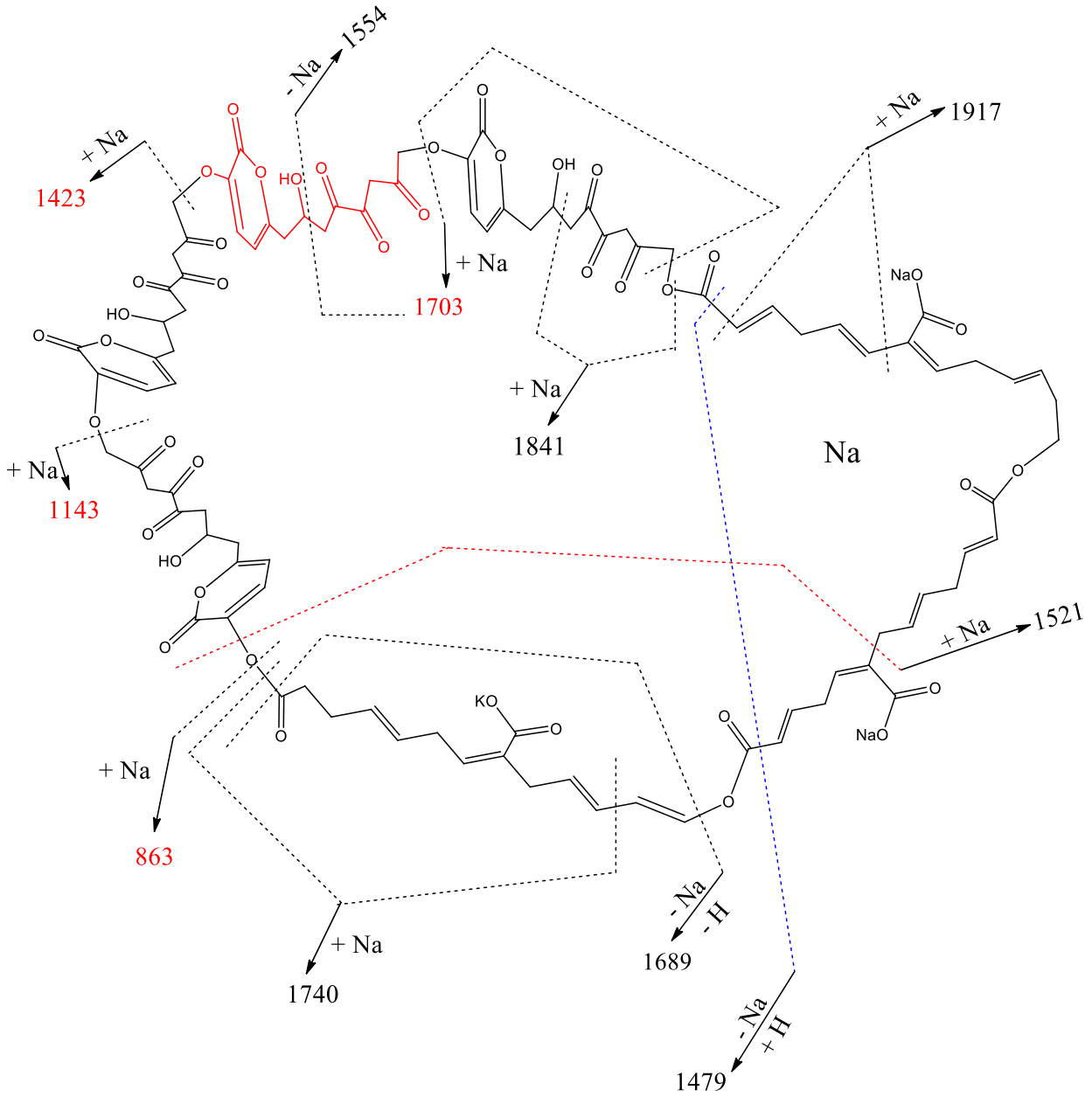
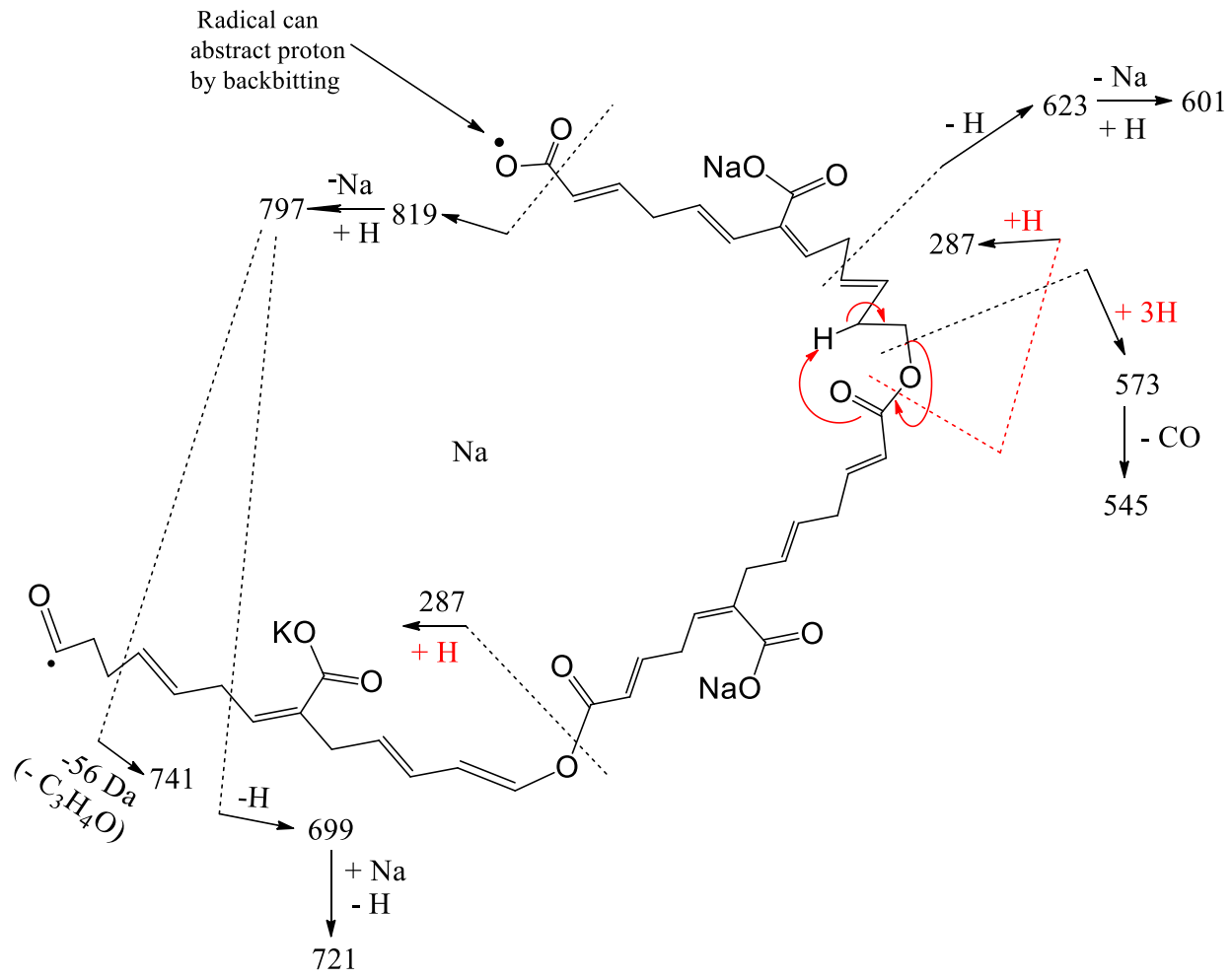


Figure 5b. Zooming of the low mass region in the product ion scan of the precursor ion at m/z 1983



Scheme 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 3: High-Energy CID-MS/MS fragmentation patterns of *Quasi*-MS^3 of the product ion at *m/z* 863. Note: Intramolecular hydrogen transfers in some fragments can occur by backbiting, [36] and Mclafferty rearrangement from the alcohol side as indicated by red arrows. [37 ,38]

4. 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

Before starting this section, we would like the readers to recap that the solid-state ^{1}H and ^{13}C-NMR study of the sporopollenin exine appear like the one obtained by Jing-Ke Weng *et al.* for the pine sporopollenin.[19].

The standard ^{1}H spectrum of sporopollenin shows both sharp and broad peaks indicative of a mixture "liquid-like" or more crystalline phase (green line) and a more amorphous phase (pink line) (SI-Figure 11a). Using a CPMG filter, we can easily remove the amorphous phase and observe only the liquid-like structure (SI-Figure 11b). The "liquid-like phase indicates the presence of a mobile sub-structure, such as long fatty acid chains.

After a vigorous search in literature, we were lucky to find a great similarity between the proton NMR spectrum of the "liquid-like" phase of our sporopollenin with that of the ^{1}H-NMR spectrum of whole seeds of *Lesquerella lyrate* that contain lipids (SI-Figure 12 a & b).[39] 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 is present on a double bond (vinylic), and the peak at 4.8 ppm characteristic for TAG Estolides (branched ester bond). This supports that our hydroxycarboxylic acids 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 composed of the alpha-pyrone ring and hydroxylated aliphatic chain.

Ten cross signals corresponding to different functional groups were identified from the 2D H^{1}-H^{1} NOESY solid state experiments (SI-Figure 13 a & b & c & d). This allowed us to get an idea of the proximity of the proton groups. Whereas, the HSQC (SI-Figure 14 a & b) indicates the ^{13}C correlation, at least with the most populated groups revealed by 1D H-NMR and 2D H^{1}-H^{1} NOESY. The results are listed in SI-Table 1.

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 figure 6 a & b & c, Table 1 and SI-Table 2. Using the building block obtained from TOF-SIMS and MALDI-MS, we were able to

generate a model that is very close to the empirical formula of *Lycopodium clavatum* sporopollenin and fitted with the quantitative data obtained from C-13 NMR to a large extent (Figure 6 a & b & c, and Table1). 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 Jing-Ke Weng *et al.* [19].

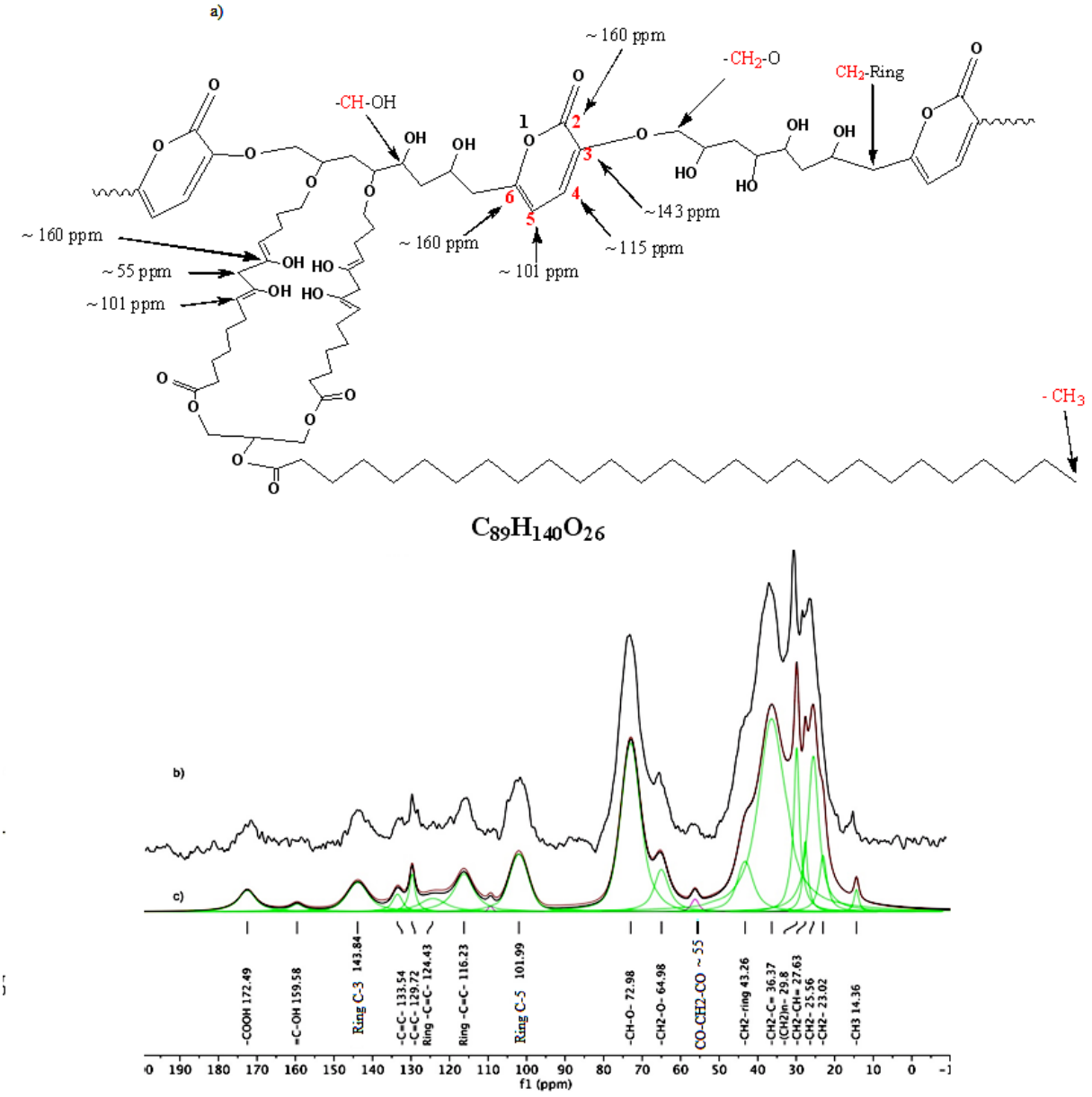


Figure 6: $^{13}C\{^{1}H\}$ multiCP MAS NMR experimental (b) and deconvoluted spectra (c), (a) empirical formula model

% of C	ppm range for integration	Deconvolution	Model
CH_3 (Chain)	0-20	1	2
CH_2 (Chain or ring)	20-50	56	58
O=C-CH_2-C=O	50-60	2	3
CH_2-O or CH-O (Chain)	60-90	19	16
Ring C_5 or Chain C=C-O	90-110	6	3
Ring C_4 or 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	2

Table 1: Distribution of Carbons in the structure according to deconvolution of the experimental spectrum and *Lycopodium clavatum* empirical formula model designed according to moieties identified using MS.

5. Sporopollenin Model

In order to correlate this data to the real sporopollenin scanning electron microscope images (SEM), We use the identified building blocks in this manuscript to build a model for the formation of the sporopollenin exine that looks like its various network SEM images. For example, *Passiflora sp. (Passifloraceae) sporopollenin* SEM image is shown in SI-figure 13, showing a circular aperture with a cross-linked network built on it.[40]

We have shown in our model that sporopollenin consists of a circular polymer composed of the polyhydroxylated tetraketide monomers, which contain hydroxyl groups. The circular polymer hydroxyl groups can be attached to poly(hydroxyacid) network chains (dendrimer-like network) through ether bonds. (SI-figure 14).

CONCLUSION

Although there have been a lot of contradictions about the structural constituents and the molecular identity of the biopolymer sporopollenin, many scientists around the world focused only on the biomedical applications of sporopollenin, as a drug delivery system, without caring or giving credence to its molecular structure.

Until recently, Sporopollenin was described as a highly resistant biopolymer, which was thought to be composed of aromatics, phenolics, and long-chain aliphatic acids. [41,42]

In this communication, we have proven by XPS analysis, beyond any doubts, the total absence of aromaticity in the sporopollenin exine. We also have confirmed for the first time the presence of sporopollenin repeating unit 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 which indicated the presence of the circular polyhydroxylated tetraketides polymer that represent the main circular rigid backbone of the sporopollenin biopolymer. This can be covalently attached by ether links to the poly(hydroxyacid) chain network forming the main framework of sporopollenin. Also, as mentioned in the introduction section, sporopollenin exine could exist as a spherical dendrimer which is a typical type of microcapsule used for drug delivery applications.

It is quite interesting to mention that it was due to the SIMS-TOF and CID-MS/MS analyses by which we initially discovered the chemistries present on the outermost surfaces of the sporopollenin exine through the discovery of ions characteristic for diacylglycerol (DAG). This led to MALDI-TOF-MS and CID-MS/MS analyses that allowed us to cover each individual constituents of the sporopollenin repeating units. In this context, it should be noted that the use of different matrices helped. 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). Finally, the use of the 2-(4-Hydroxyphenylazo)benzoic acid (HABA) allowed us to establish the presence of the circular polyhydroxylated tetraketide polymer (rigid backbone).

In this manuscript, we were capable of revealing the novel molecular structure of the main framework of the spherical sporpollenin exine. Nevertheless, the exact biosynthesis of sporopollenin is now left to the expert biologists, which hopefully could decipher the exact

biosynthesis of sporopollenin, especially that today, the regulation genes of sporopollenin biosynthesis have been discovered.[43]

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.

Experimental

Sporopollenin extraction

The sporopollenin exine used in this manuscript was provided by Prof. Mackenzie 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 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 [44]

High-Resolution X-Ray Photoelectron Spectrometry (HR-XPS).

The experimental details of XPS are shown in SI-2.

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

The experimental details are shown in SI-3.

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

Prior to MALDI mass spectrometric analysis, 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 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 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.

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

The 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,[45] 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.

Acknowledgment

Mr. Abanoub Mikhael would like to acknowledge the Chemistry Department, Memorial University for his Ph.D. fellowship, the authors appreciate the financial help obtained from Fisheries and Oceans Canada, Science Branch and for availing and allowing free access to their laboratories.

The authors are 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 analyses.**

REFERENCES

[1] Mackenzie, G., Beckett, S., Atkin, S., Diego-Taboada, A., Pollen and Spore Shells-Nature's Microcapsules. In Microencapsulation in the Food Industry, Academic Press, 2014, 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.

[3] Ahlers, F., Bubert, H., Steuernagel, S., and Wiermann, R., The nature of oxygen in sporopollenin from the pollen of Typha angustifolia L., *Zeitschrift für Naturforschung 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 Fuels*, 1988, 2(6), 843-847.

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

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

[9] Kim, S.S., Douglas, C.J., 2013. Sporopollenin monomer biosynthesis in *Arabidopsis*. J. *Plant Biol.* 56, 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. (2006). 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. Devon, UK Torquay.

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

[13] 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), e00034.

[14] Kim, S. S., Grienenberger, E., Lallemand, B., Colpitts, C. C., Kim, S. Y., de Azevedo Souza, C., Kombrink, E., 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.

[15] Grienenberger, E., Kim, S. S., Lallemand, B., Geoffroy, P., Heintz, D., de Azevedo Souza, C., Legrand, M., 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.

[16] Nick, P., Opatrny, Z., (Eds.), Applied plant cell biology: cellular tools and approaches for plant biotechnology, *springer science and business media*, 2014, Vol.22.

[17] Fraser, W. T., Lomax, B. H., Jardine, P. E., Gosling, W. D., & Sephton, M. A. (2014). Pollen and spores as a passive monitor of ultraviolet radiation. Frontiers in Ecology and Evolution, 2, 12.

[18] Chu, M., Mierzwa, R., Xu, L., He, L., Terracciano, J., Patel, M., ... & Chan, T. M. (2002). Structure of Sch 419560, a novel α-pyrone antibiotic produced by Pseudomonas fluorescens. The Journal of antibiotics, 55(2), 215-218.

[19] 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.

[20] Joule, J. A., & Mills, K., Heterocyclic Chemistry, *John Wiley and Sons*, 2010

[21] 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.

[22] Gonzalez-Cruz, P., Uddin, M. J., Atwe, S. U., Abidi, N., & Gill, H. S. (2018). Chemical treatment method for obtaining clean and intact pollen shells of different species. *ACS biomaterials science & engineering*, *4*(7), 2319-2329.

[23] Barrier, S. (2008). *Physical and chemical properties of sporopollenin exine particles* (Doctoral dissertation, University of Hull).

[24] Moulder, J. F. (1992). Handbook of X-ray photoelectron spectroscopy. Physical electronics

[25] Gilli, G., Bertolasi, V., Ferretti, V., & Gilli, P. (1993). Resonance-assisted hydrogen bonding. III. Formation of intermolecular hydrogen-bonded chains in crystals of β-diketone enols and its relevance to molecular association. *Acta Crystallographica Section B: Structural Science*, *49*(3), 564-576.

[26] Pan, M. H., Huang, T. M., & Lin, J. K. (1999). Biotransformation of curcumin through reduction and glucuronidation in mice. *Drug metabolism and disposition*, *27*(4), 486-494.

[27] Wesdemiotis, C., Solak, N., Polce, M. J., Dabney, D. E., Chaicharoen, K., & Katzenmeyer, B. C. (2011). Fragmentation pathways of polymer ions. *Mass spectrometry reviews*, *30*(4), 523-559.

[28] Francke, W., Lübke, G., Schröder, W., Reckziegel, A., Imperatriz-Fonseca, V., Kleinert, A., Engels, W. (2000). Identification of oxygen containing volatiles in cephalic secretions of workers of Brazilian stingless bees. *Journal of the Brazilian Chemical Society*, *11*(6), 562-571.

[29] Tengö, J., Groth, I., Bergström, G., Schröder, W., Krohn, S., & Francke, W. (1985). Volatile secretions in three species of Dufourea (Hymenoptera: Halictidae) bees: chemical composition and phylogeny. *Zeitschrift für Naturforschung C*, *40*(9-10), 657-660.

[30] Thomas, A., Charbonneau, J. L., Fournaise, E., & Chaurand, P. (2012). 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. Analytical chemistry, 84(4), 2048-2054.

[31] Hughey, C. A., Hendrickson, C. L., Rodgers, R. P., Marshall, A. G., & Qian, K. (2001). Kendrick mass defect spectrum: a compact visual analysis for ultrahigh-resolution broadband mass spectra. *Analytical Chemistry*, *73*(19), 4676-4681.

[32] Kretzmann, J. A., Evans, C. W., Norret, M., & Iyer, K. S. (2017). Supramolecular assemblies of dendrimers and dendritic polymers in nanomedicine. In Comprehensive Supramolecular Chemistry II (pp. 237-256). Academic Press (Elsevier Inc.)

[33] Murphy, R. C. (2014). *Tandem mass spectrometry of lipids: molecular analysis of complex lipids*. Royal Society of Chemistry.

[34] Neubert, H., Halket, J. M., Ocana, M. F., & Patel, R. K. (2004). MALDI post-source decay and LIFT-TOF/TOF investigation of α-cyano-4-hydroxycinnamic acid cluster interferences. Journal of the American Society for Mass Spectrometry, 15(3), 336-343.

[35] Uddin, M. J., Abidi, N., Warzywoda, J., & Gill, H. S. (2019). Investigation of the fate of proteins and hydrophilicity/hydrophobicity of Lycopodium clavatum spores after organic solvent-base-acid treatment. ACS applied materials & interfaces.

[36] Raska, C. S., Parker, C. E., Huang, C., Han, J., Glish, G. L., Pope, M., & Borchers, C. H. (2002). Pseudo-MS^3 in a MALDI orthogonal quadrupole-time of flight mass spectrometer. Journal of the American Society for Mass Spectrometry, 13(9), 1034-1041.

[37] Banoub, J. H., Demian, W. L., Piazzetta, P., Sarkis, G., Kanawati, B., Lafont, D., Fridgen, T. D. (2015). 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 Communications in Mass Spectrometry, 29(19), 1717-1732.

[38] Joly, N., Aneed, A. E., Martin, P., Cecchelli, R., & Banoub, J. (2005). Structural determination of the novel fragmentation routes of morphine opiate receptor antagonists using electrospray ionization quadrupole time-of-flight tandem mass spectrometry. Rapid Communications in Mass Spectrometry: An International Journal Devoted to the Rapid Dissemination of Up-to-the-Minute Research in Mass Spectrometry, 19(21), 3119-3130.

[39] Zhang, H., Olson, D. J., Van, D., Purves, R. W., & Smith, M. A. (2012). Rapid identification of triacylglycerol-estolides in plant and fungal oils. Industrial Crops and Products, 37(1), 186-194.

[40] Wang, R., & Dobritsa, A. A. (2018). Exine and aperture patterns on the pollen surface: Their formation and roles in plant reproduction. Annual Plant Reviews, JA Roberts, ed (John Wiley & Sons), 1-40.

[41] Piffanelli P, Ross JHE, Murphy DJ (1998) Biogenesis and function of the lipidic structures of
pollen grains. Sex Plant Reprod 11:65–80

[42] Ariizumi T, Toriyama K (2011) Genetic regulation of sporopollenin synthesis and pollen exine
development. Annu Rev Plant Biol 62:437–460

[43] Wang, K., Guo, Z. L., Zhou, W. T., Zhang, C., Zhang, Z. Y., Lou, Y., Yang, Z. N. (2018). The regulation of sporopollenin biosynthesis genes for rapid pollen wall formation. Plant physiology, 178(1), 283-294.

[44] Barrier, S., Diego-Taboada, A., Thomasson, M. J., Madden, L., Pointon, J. C., Wadhawan, J. D., ... & Mackenzie, G. (2011). Viability of plant spore exine capsules for microencapsulation. *Journal of Materials Chemistry*, *21*(4), 975-981.

[45] Johnson, R. L., & Schmidt-Rohr, K. (2014). Quantitative solid-state 13C NMR with signal enhancement by multiple cross polarization. *Journal of Magnetic Resonance*, *239*, 44-49.