DEMystifying and Unraveling the Factual Molecular Structure of the Biopolymer Sporopollenin

Journal: Rapid Communications in Mass Spectrometry

Manuscript ID: RCM-19-0447.R1

Wiley - Manuscript type: Research Article

Date Submitted by the Author: n/a

Complete List of Authors: Mikhael, Abanoub; Memorial University of Newfoundland, Chemistry
Jurcic, Kristina; University of Western Ontario
Schneider, Celine; Memorial University of Newfoundland
Carr, David; Physical Electronics Inc
Fisher, Gregory; Physical Electronics, FRIDGEN, TRAVIS; MEMORIAL UNIVERSITY, Chemistry
Diego-Taboada, Alberto; University of Hull
Georghiou, Paris; Memorial University of Newfoundland
Mackenzie, Grahame; University of Hull
Banoub, Joe; Memorial University of Newfoundland Faculty of Humanities and Social Sciences, Chemistry; Government of Canada, Fisheries and Oceans Canada

Keywords: Lycopodium clavatum sporopollenin exine, TOF-SIMS and 1Kev MS/MS, MALDI-TOF-MS and CID-TOF/TOF-MS/MS, 1D and 2D Solid State 1H- and 13C-NMR, HR-XPS

Abstract:
Rationale
We report the unsolved molecular structure of the complex biopolymer sporopollenin exine extracted from Lycopodium clavatum.

Methods
We have used TOF-SIMS and CID-MS/MS, MALDI-TOF-MS and CID-TOF/TOF-MS/MS for the analysis of this complex biopolymer sporopollenin exine extracted from Lycopodium clavatum. Solid-state 1H- and 13C-NMR, 2D 1H-1H NOESY, Rotor-synchronized 13C{1H} HSQC, and 13C{1H} multi CP-MAS NMR experiments were used to confirm the structural assignments revealed by MS and MS/MS studies.

Results
We have identified and characterized the unsolved molecular structure of the clean intact sporopollenin of Lycopodium clavatum. The compiled MS and NMR analyses showed that sporopollenin contained a poly(hydroxyacid) dendrimer-like network, which accounted for the sporopollenin empirical formula. In addition, the identified hydroxy acid monomers contained a beta diketone moiety. Moreover, MALDI-TOF-MS and MS/MS allowed us to identify a unique circular polyhydroxylated tetraketide polymer. This polymer acted as the rigid backbone on which

http://mc.manuscriptcentral.com/rcm
the poly(hydroxyacid) network can be built, forming the scaffold of the spherical sporopollenin exine

Conclusion
We have shown herein the first-ever report in the literature showing the presence of naturally occurring branched poly(hydroxyacid) natural "Dendrimer" like molecules. Furthermore, we have also have confirmed for the first time the presence of two main sporopollenin building units using solid-state 1H- and 13C-NMR, 2D 1H-1H, and 2D 13C-1H NMR experiments, SIMS-TOF-MS, MALDI-TOF-MS, and CID-MS/MS. These analyses indicated the presence of the circular polyhydroxylated tetraketides polymer that represent the main circular rigid backbone of the sporopollenin biopolymer. This latter polymer can be covalently attached by ether links to the poly(hydroxyacid) chain network to form the sporopollenin biopolymer to form a spherical dendrimer, which is a typical type of microcapsule used for drug delivery application. Finally, High-Resolution X-ray Photoelectron Spectroscopy (HR-XPS) indicated the total absence of aromaticity in the sporopollenin exine
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 Paris E. Georghiou,1 Grahame Mackenzie5 and Joseph Banoub*1,6

1Department of Chemistry, Memorial University, 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
4Physical Electronics, Inc. (PHI) 18725 Lake Drive East, Chanhassen, MN 55317, USA
5Department of Chemistry, University of Hull, Kingston upon Hull, HU6 7RX, UK
6Fisheries and Oceans Canada, Science Branch, Special Projects, 80 East White Hills Road, St John’s, NL A1C 5X1, Canada

ABSTRACT

Rationale

We report the unsolved molecular structure of the complex biopolymer sporopollenin exine extracted from Lycopodium clavatum pollen grains.

Methods

TOF-SIMS and CID-MS/MS, MALDI-TOF-MS and CID-TOF/TOF-MS/MS were used for the analysis of this complex biopolymer sporopollenin exine extracted from Lycopodium clavatum. Solid-state 1H- and 13C-NMR, 2D 1H-1H NOESY, Rotor-synchronized 13C{1H} HSQC, and 13C{1H} multi CP-MAS NMR experiments were used to confirm the structural assignments revealed by MS and MS/MS studies. Finally, high-resolution XPS was used to check for the presence of aromatic components in sporopollenin.

Results

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.

**Conclusion**

We have reported for the first time in the literature, the two main building units that form the *Lycopodium clavatum* sporopollenin exine. The first building unit is a macrocyclic oligomer and/or polymer composed of polyhydroxylated tetraketide-like monomeric units. This macrocyclic unit represents the main rigid backbone of the sporopollenin biopolymer. The second building unit is the poly(hydroxyacid) network upon which their hydroxyl end groups can be covalently attached by ether links to the hydroxylated macrocyclic backbone to form the sporopollenin biopolymer, a spherical dendrimer. Such spherical dendrimers are a typical type of microcapsule that has been used for drug delivery applications. Finally, HR-XPS (high-resolution X-ray photoelectron spectroscopy) indicated the total absence of aromaticity in the sporopollenin exine.

**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. To simplify, 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 °C, where sporopollenin undergoes simultaneous dehydrogenation and deoxygenation. The second stage at temperatures above 500 °C is where aromatic products are formed. This study showed that heating substances at high-temperature 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 (Please refer to SI-1). 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
Results and Discussion

The *Lycopodium clavatum* sporopollenin used in this investigation was extracted by the classical method of Zetzsche *et al.*[^24] 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 whilst the role of the phosphoric acid is to degrade and remove polysaccharide materials, thus leaving only the sporopollenin remaining.[^24] 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. [^25] 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). [^26]

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 strike sporopollenin aiming to discover any characteristic ions for the sporopollenin exine.

1. IDENTIFICATION OF THE POLY(HYDROXY ACID NETWORK BUILDING UNIT REPRESENTING THE EMPIRICAL FORMULA OF LYCOPODIUM CLAVATUM SPOROPOLLENIN EXINE.

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.[^27]
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 $[C_{16}H_{25}O_{4}C_3H_4C_{14}H_{21}O_4+H]^+$ and $[C_{16}H_{25}O_4C_3H_4C_{16}H_{25}O_4+H-]^-$, respectively (SI-Figure 1A and SI-Figure 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 (SI-Figures 4 and SI-Scheme 1 and 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 $[C_{14}H_{22}O_4-H]^-$ and $[C_{16}H_{26}O_4-H]^-$, respectively (SI-Figure 1B and SI-Figure 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 (SI-Figures 5 and SI-Scheme 3).

A proposed fragmentation mechanism for these carboxylate anions at $m/z$ 253 and 281 is shown in SI-Scheme 3. Overall, the fragmentation mechanism shown in SI-Scheme 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 (SI-Figure 6 and 7), as it facilitates the formation of an intramolecular hydrogen bond that gives extra stability to these types of structures.\cite{28} This hydrogen bonding resembles the intramolecular hydrogen bond seen with the β-diketone moiety in tetrahydrocurcumin (THC) molecules.\cite{29} Also, the loss of one water molecule in the product ion scans of the precursor ions $[C_{14}H_{22}O_4-H]^-$ and $[C_{16}H_{26}O_4-H]^-$ at $m/z$ 253 and 281 resemble the fragmentation pattern reported for the diketone THC.\cite{29} 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 SI-Figure 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, and 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 1.3, Figure 2, Figure 3 and Table 1). The proposed chemical structures of all other ions formed in the positive or negative ion mode TOF-SIMS-MS are shown in SI-Figure 2 and SI-Figure 3).

1.2. 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. 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 (SI-Figure 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 (SI-Figure 10). 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 (SI-Figure 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}]^{+•}, 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 1). This latter ion at m/z 1965.1278 has a structure that appears like a branching unit in triacylglycerol (TAG) dendrimers. 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 SI-Table 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 SI-Figure 12 and Scheme 1.
In order to propose a chemical formula with reasonable mass accuracy for this ion, we proposed that this ion contained one alpha-pyrene ring plus a hydroxylated chain in its chemical structure (Scheme 1). This latter unit composed of an alpha pyrene ring and a hydroxylated chain will be explained later in section 2.1.

Also, we proposed the presence of the pyrene 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. (SI-Figure 11 and SI-Table 1).

Needless to say, further MS/MS studies are needed to confirm the assigned chemical formulas 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. \[ 33 \]

**Figure 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 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 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 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 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.

### 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). However, it should be noted that the mass region around \( m/z \) 100-1300 was mainly composed of CHCA matrix peaks,\[ 33 \] 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$_2$H$_4$ (2 x CH$_2$) in their structures (Figure 2 and Table 1) can be seen. The most important one is the ion at m/z 1643.9948 (previously described in the TOF-SIMS Section 1.1), which was assigned as [C$_{89}$H$_{142}$O$_{27}$+H]$^+$. This latter ion at m/z 1643.9948 is a diglyceride which is composed of six fatty acids (five C14 and one C16) plus the glycerol moiety (Figure 3 and Table 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.* [24] 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$_2$H$_4$ 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 C14 and two C16 fatty acids) plus glycerol as a core unit (Figure 3 and Table 1).

The chemical formulas of other selected ions in the MALDI-TOF-MS of sporopollenin using CHCA as a matrix are shown in Table 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 (Table 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 and Table 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 1), which is not surprising based on the complexity of the sporopollenin structure. [35]

To summarize Sections 1.1, 1.2, and 1.3, a series of branched poly(hydroxyacids) with glycerol as a core unit, have 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. [2,36] 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 2A. MALDI-TOF-MS of sporopollenin exine using CHCA as a matrix

Figure 2B. Expansion of the sporopollenin exine MALDI-TOF-MS in the high mass region at m/z 1500-3000

Table 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.

Figure 3A. The proposed structure of m/z 1643.9948 identified in sporopollenin MALDI-TOF-MS using CHCA as a matrix

Figure 3B. The proposed structure of m/z 2428.4438 identified in sporopollenin MALDI-TOF-MS using CHCA as a matrix

2. IDENTIFICATION OF THE PRESENCE OF A POLYHYDROXYLATED TETRAKETIDE-LIKE REPEATING UNIT IN THE SPOROPOLLENIN EXINE STRUCTURE.

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

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_{2}O_{41}+Na]^+, [C_{81}H_{79}KNa_{2}O_{34}+Na]^+, [C_{68}H_{67}KNa_{2}O_{27}+Na]^+ and [C_{55}H_{53}KNa_{2}O_{20}+Na]^+, respectively (Figure 4 and Table 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 $\text{C}_{13}\text{H}_{12}\text{O}_7$. This series of sodiated molecular ions containing both K and Na carboxylates in their respective structures. These carboxylate salts most probably were 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.\[37\]

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.\[37\] 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\(_3\)O\(_{34}\)]\(^+\), [C\(_{68}\)H\(_{67}\)Na\(_3\)O\(_{27}\)]\(^+\) and [C\(_{55}\)H\(_{55}\)Na\(_3\)O\(_{20}\)]\(^+\) (Figure 4 and Table 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 ($m/z$ 1983 → $m/z$ 1703 → $m/z$ 1423 → $m/z$ 1143 → $m/z$ 863) (Figures 5, and Scheme 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 the resulting product ion at $m/z$ 287 (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 SI-Scheme 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 and SI-Scheme 4, the quasi-MS\(^3\) spectrum of the product ion at \( m/z \) 863 was measured.\(^{[38-40]}\) (SI-Figure 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 into three ions of \( m/z \) 287); (Figures 5A&B).

After examining a series of rational potential structures, we have chosen the most reasonable structure for the product ion at \( m/z \) 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. Consequently, the product ion at \( m/z \) 863 was assigned as to 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 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 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 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. 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 SI-Figure 14).
In summary, the MALDI-TOF/TOF-MS/MS using HABA as a matrix showed the presence of a polyhydroxylated tetraketide-like repeating unit (α-pyrone ring plus hydroxylated aliphatic chain) that forms a macrocyclic backbone composed of polyhydroxylated tetraketide-like monomers. These macrocyclic oligomers and/or its related polymers are rich in hydroxyl groups to which the lipid network of the poly(hydroxyacid) can be anchored 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 polyhydroxylated tetraketide-like monomers and the poly(hydroxyacid) network.

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.\cite{41,42} 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.\cite{43,44} On the other hand, HABA has been suggested as an efficient matrix for analyte types such as glycans, steroids, glycoproteins, glycolipids, and synthetic polymers.\cite{30,45,46} Therefore, the selection of the most suitable matrix for a particular analyte, is often a “trial and error” process.\cite{47} 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.

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

**Table 2**: The chemical formulas of the ions identified in the sporopollenin MS using HABA as a matrix

**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**: Expansion 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 3A** *Quasi*-MS$^3$ fragmentation patterns of the product ion at m/z 863.

**Scheme 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, \cite{48}, and McLafferty rearrangement from the alcohol side, as indicated by red arrows. \cite{49,50}

### 3. VERIFICATION OF THE PROPOSED STRUCTURE OF THE SPOROPOLLENIN EXINE USING SOLID STATE NMR

#### 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.\cite{15} 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 15A). Using a CPMG filter, we can easily remove the amorphous phase and observe only the liquid-like structure (SI-Figure 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 (SI-Figures 16A&B).\cite{51} 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 our evidence that the 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 decomposed of alpha-pyrone ring and hydroxylated aliphatic chains. Based on the absence of Estolides (branched esters), we proposed that the hydroxyl end groups of the poly(hydroxy) 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 (SI-Figures 17A-D). This allowed us to get an idea of the proximity of the proton groups.

Whereas, the HSQC (SI-Figures 18 A&B ) 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 SI-Table 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 6A-C, Table 3. and SI-Table 3.

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 (Figures 6A-C, and Table 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 SI-2.

**Figure 6.** *Top:* Empirical formula model; *bottom:* experimental $^{13}$C{$^{1}$H} multi-CP MAS and (C) deconvoluted NMR spectra.

**Table 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.

**4. VERIFICATION OF THE TOTAL ABSENCE OF AROMTICS IN THE SPOROPOLLENIN EXINE**

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

The HR-XPS C(1s) surface analysis of sporopollenin (Figure 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 $\pi-\pi^*$ transition. It should be noted that the absence of this “shake-up line” is consistent with the lack of aromaticity in the sporopollenin. [52]

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 some weak aromatic character, as they can undergo some reactions such as electrophilic additions, ring-opening and Diels-Alder reactions.

**Figure 7.** *Lycopodium clavatum* sporopollenin X-ray photoelectron spectroscopy of carbon (1s).

### 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 shown in SI-Figure 19, reveals a circular aperture with a cross-linked network built on it. 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 oligomeric and dendrimeric-like network) and helps to better visualize how these units may be linked together.

This may provide a clearer view of our model for the sporopollenin exine consisting of a macrocyclic oligo- or polymer composed of polyhydroxylated tetraketide monomers with, hydroxyl groups which can be ether bond-linked to the poly(hydroxyacid) network chains via dendrimer-like network(s). (SI-Figure 20).

### 6. DFT COMPUTATIONAL MODEL USING GAUSSIAN09 OF A HYPOTHETICAL MACROCYCLIC OLIGOMER

A DFT computational model using *Gaussian09* of a hypothetical macrocyclic oligomer based upon the structure shown in Scheme 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 8B. The structure was optimized using the B3LYP/6-31G basis set. The macrocycle clearly shown the “bowl”-shape of a possible segment of the *Lycopodium clavatum* sporopollenin exine (*see* also SI Figures 19 and 20).

**Figure 8.** Left: Hypothetical structure based upon Scheme 2 showing an example of the dendrimer linking points, and right: B3LYP/6-31G DFT-optimized computed molecular stucture.
CONCLUSIONS

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.[56] 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.[57, 58]

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 cyclic and rigid backbone of the sporopollenin biopolymer. This oligomer 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 assistance. The use of 1,5-diaminonaphthalene (DAN) and $\alpha$-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 observations 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.\[^{[59]}\]

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 building units of the spherical sporopollenin exine and the summary of this well laborious and complicated study is presented in Figure 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.\[^{[60]}\]

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.

**Figure 9A:** Summary of the identification of the poly(hydroxyacid) network of sporopollenin  
**Figure 9B:** Summary of the identification of the macrocyclic backbone of sporopollenin

### Experimental

**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 x 300 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 [61]

**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 Kα x-ray source (hv = 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° 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.
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. The elements and qualities of the tandem MS spectra for composition analysis and structure elucidation have been further elaborated in other reports. 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, 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 × 10^{12} Bi3+/cm^2) 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 × 10^{13} Bi3+/cm^2 and 1.99 × 10^{13} Bi3+/cm^2, 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.

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 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.

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 maximal fragmentation coverage allowed by the mass spectrometer which was used.

**Solid-State $^1$H- and $^{13}$C-NMR; $^1$H-$^1$H NOESY 2D and 2D $^1$H-$^{13}$C 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,$^{[6]6}$ 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 \( v_r = 5 \text{kHz} \) with ns=8 scans. 2D \(^1\text{H-}^{13}\text{C}\) HSQC were recorded at \( v_r = 23 \text{kHz} \) (sw=152ppm) for \( J=125\text{Hz} \) and \( J=170\text{Hz} \) with ns=32 scans. Processing, peaks deconvolution, and integration, as well as spectrum prediction, were all performed using MestReNova.

**Acknowledgments**

The authors are extremely grateful to the referees for the excellent effort for refereeing this manuscript. We have really appreciated their comments and wonderful wisdom that helped us to produce a more comprehensible and logic manuscript. Mr. Abanoub Mikhael would like to acknowledge the Chemistry Department, Memorial University, for his Ph.D. fellowship. 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. Finally, we would like to thank Dr. Shofiu Rahman, Chemistry Department MUN for his help in the molecular modelling.

**REFERENCES**


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


Table 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.

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

Table 2. The chemical formulas of the ions identified in the sporopollenin MS using HABA as a matrix.

<table>
<thead>
<tr>
<th>Experimental m/z</th>
<th>Assigned chemical formula</th>
<th>Error ppm</th>
</tr>
</thead>
<tbody>
<tr>
<td>1983.4810</td>
<td>[C_{94}H_{91}KNa_{2}O_{41} + Na]^+</td>
<td>22.3</td>
</tr>
<tr>
<td>1703.4147</td>
<td>[C_{81}H_{79}KNa_{2}O_{34} + Na]^+</td>
<td>21.3</td>
</tr>
<tr>
<td>1687.4452</td>
<td>[C_{81}H_{79}Na_{3}O_{34} + Na]^+</td>
<td>24.2</td>
</tr>
<tr>
<td>1423.3455</td>
<td>[C_{68}H_{67}KNa_{2}O_{27} + Na]^+</td>
<td>17.9</td>
</tr>
<tr>
<td>1407.3719</td>
<td>[C_{68}H_{67}Na_{3}O_{27} + Na]^+</td>
<td>18.4</td>
</tr>
<tr>
<td>1143.2730</td>
<td>[C_{55}H_{55}KNa_{2}O_{20} + Na]^+</td>
<td>9.9</td>
</tr>
<tr>
<td>1127.2996</td>
<td>[C_{55}H_{55}Na_{3}O_{20} + Na]^+</td>
<td>10.5</td>
</tr>
<tr>
<td>1105.3091</td>
<td>[C_{55}H_{55}Na_{3}O_{20} + H]^+</td>
<td>2.9</td>
</tr>
</tbody>
</table>
Table 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.

<table>
<thead>
<tr>
<th>% C</th>
<th>Integration ppm range</th>
<th>Deconvolution</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH$_3$ (Chain)</td>
<td>0-20</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>CH$_2$ (Chain or ring)</td>
<td>20-50</td>
<td>56</td>
<td>58</td>
</tr>
<tr>
<td>CO -CH$_2$-CO (in the enol-enol form)</td>
<td>50-60</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>CH$_2$-O or CH-O (Chain)</td>
<td>60-90</td>
<td>19</td>
<td>16</td>
</tr>
<tr>
<td>Ring C$_5$ or Chain C=C-O</td>
<td>90-110</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>Ring C$_4$ or C=C-CH$_2$-C=C</td>
<td>110-130</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>C=C (Chain)</td>
<td>130-140</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Ring C$_3$</td>
<td>140-150</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>=C-OH or Ring C$_2$ or Ring C$_6$</td>
<td>150-165</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>COOH(Chain)</td>
<td>165-185</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>
LIST OF LEGENDS

Figure 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 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 1B. 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 1C. 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 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.

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

Figure 2B. Expansion of the sporopollenin exine MALDI-TOF-MS in the high mass region at m/z 1500-3000.

Figure 3A. The proposed structure of m/z 1643.9948 identified in sporopollenin MALDI-TOF-MS using CHCA as a matrix

Figure 3B. The proposed structure of m/z 2428.4438 identified in sporopollenin MALDI-TOF-MS using CHCA as a matrix

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. Expansion 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 3A Quasi-MS³ fragmentation patterns of the product ion at m/z 863.

Scheme 3B: Quasi-MS³ fragmentation patterns of the product ion at m/z 863. Note: Intramolecular hydrogen transfers in some fragments can occur by backbiting, [48], and McLafferty rearrangement from the alcohol side, as indicated by red arrows. [49,50]

Figure 6. Top: Empirical formula model; bottom: experimental $^{13}$C{$^1$H} multi-CP MAS and (C) deconvoluted NMR spectra.

Figure 7. Lycopodium clavatum sporopollenin X-ray photoelectron spectroscopy of carbon (1s).

Figure 8. Left: Hypothetical structure based upon Scheme 2 showing an example of the dendrimer linking points, and right: B3LYP/6-31G DFT-optimized computed molecular structure.

Figure 9A: Summary of the identification of the poly(hydroxyacid) network of sporopollenin

Figure 9B: Summary of the identification of the macrocyclic backbone of sporopollenin
Figure 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 1. A. 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 1B. Fragmentation mechanism for the product ion scan of the precursor ion at \textit{m/z} 1441.93 identified in sporopollenin exine MALDI-TOF-MS/MS using DAN as a matrix.
Scheme 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 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.
Figure 2A. MALDI-TOF-MS of sporopollenin exine using CHCA as a matrix

Figure 2B. Expansion of the sporopollenin exine MALDI-TOF-MS in the high mass region at m/z 1500-3000.
Figure 3A. The proposed structure of $m/z$ 1643.9948 identified in sporopollenin MALDI-TOF-MS using CHCA as a matrix.
Figure 3B. The proposed structure of $m/z$ 2428.4438 identified in sporopollenin MALDI-TOF-MS using CHCA as a matrix.
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. Expansion 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 3A Quasi-MS$^3$ fragmentation patterns of the product ion at $m/z$ 863.
Scheme 3B: Quasi-MS³ fragmentation patterns of the product ion at \( m/z \) 863. Note: Intramolecular hydrogen transfers in some fragments can occur by backbiting, \(^{[48]}\) and McLafferty rearrangement from the alcohol side, as indicated by red arrows. \(^{[49,50]}\)
form of questions and answers in SI-2.

Figure 6. Top: Empirical formula model; bottom: experimental $^{13}$C$^{1}$H multi-CP MAS and (C) deconvoluted NMR spectra.
**Figure 7.** *Lycopodium clavatum* sporopollenin X-ray photoelectron spectroscopy of carbon (1s).

<table>
<thead>
<tr>
<th>C 1s functionality</th>
<th>peak BE (eV)</th>
<th>rel. abundance (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C-C &amp; C-H</td>
<td>284.8</td>
<td>71</td>
</tr>
<tr>
<td>C-O</td>
<td>286.0</td>
<td>23</td>
</tr>
<tr>
<td>C=O</td>
<td>287.3</td>
<td>4</td>
</tr>
<tr>
<td>O-C=O</td>
<td>288.0</td>
<td>2</td>
</tr>
<tr>
<td>π→π*</td>
<td>N/A</td>
<td>0</td>
</tr>
</tbody>
</table>

**Figure 8.** *Left:* Hypothetical structure based upon Scheme 2 showing an example of the dendrimer linking points, and *right:* B3LYP/6-31G DFT-optimized computed molecular structure other van Waals interactions.
The identified monomers were used to construct ions in the MALDI-TOF-MS of sporopollenin using DAN or CHCA as a matrix.

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 building block of the sporopollenin exine.

3) A unique peak at 55 ppm in the $^{13}$C-NMR is characteristic 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 presence of betadiketone moiety (No CH-OH) in the fatty acid monomers of sporopollenin and that the hydroxy carboxylic acid monomers are connected linearly (No Estolides).

**Figure 9A:** Summary of the identification of the poly(hydroxyacid) network of sporopollenin
MALDI-TOF-MS/MS of sporopollenin using HABA as a matrix

Macroyclic oligomeric seies with 280 Da repeating unit

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

Hydroxylated chain supported by the quantitative C-13 NMR data, C=O may be formed after the cleavage of the fatty acids attached to this macrocylic oligomer

1) From this study, we deduced the presence 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 presence of this rigid backbone is supported by the broad H-1 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 H-1 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 C-13 NMR.

Figure 9B: Summary of the identification of the macrocyclic backbone of sporopollenin