



**HAL**  
open science

**New insights into the Van Krevelen diagram:  
Automated molecular formula determination from  
HRMS for a large chemical profiling of lichen extracts**

Simon Ollivier, Philippe Jehan, Damien Olivier-Jimenez, Fabian Lambert,  
Joël Boustie, Françoise Lohezic-Le Devehat, Nicolas Le Yondre

► **To cite this version:**

Simon Ollivier, Philippe Jehan, Damien Olivier-Jimenez, Fabian Lambert, Joël Boustie, et al.. New insights into the Van Krevelen diagram: Automated molecular formula determination from HRMS for a large chemical profiling of lichen extracts. *Phytochemical Analysis*, 2022, 33 (7), pp.1111-1120. 10.1002/pca.3163 . hal-03772625

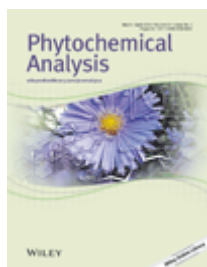
**HAL Id: hal-03772625**

**<https://hal.science/hal-03772625>**

Submitted on 9 Sep 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**New Insights into the Van Krevelen Diagram: Automated Molecular Formula Determination from HR-MS for a Large Chemical Profiling of Lichen Extracts**

Journal:	<i>Phytochemical Analysis</i>
Manuscript ID	PCA-22-0063.R1
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Ollivier, Simon; Univ Rennes, UAR 2025 - ScanMAT - CRMPO (Centre Régional de Mesures Physiques de l'Ouest); Univ Rennes, CNRS - ISCR (Institut des Sciences Chimiques de Rennes) - UMR 6226 Jéhan, Phillipe; Univ Rennes, UAR 2025 - ScanMAT - CRMPO (Centre Régional de Mesures Physiques de l'Ouest) Ollivier-Jimenez, Damien; Univ Rennes, CNRS - ISCR (Institut des Sciences Chimiques de Rennes) - UMR 6226 Lambert, Fabian; Univ Rennes, UAR 2025 - ScanMAT - CRMPO (Centre Régional de Mesures Physiques de l'Ouest) BOUSTIE, Joël; Univ Rennes, CNRS - ISCR (Institut des Sciences Chimiques de Rennes) - UMR 6226 LE DEVEHAT, Françoise; Univ Rennes, CNRS - ISCR (Institut des Sciences Chimiques de Rennes) - UMR 6226 LE YONDRE, Nicolas; Univ Rennes, UAR 2025 - ScanMAT - CRMPO (Centre Régional de Mesures Physiques de l'Ouest)
Keywords:	Lichens, Metabolomics, Structural Classification, Extract Comparison

SCHOLARONE™  
Manuscripts

1           **New Insights into the Van Krevelen Diagram: Automated Molecular Formula**  
2           **Determination from HR-MS for a Large Chemical Profiling of Lichen Extracts**

3  
4   Simon OLLIVIER<sup>1,2,3</sup>, Philippe JÉHAN<sup>1</sup>, Damien OLIVIER-JIMENEZ<sup>2</sup>, Fabian LAMBERT<sup>1</sup>,  
5   Joël BOUSTIE<sup>2</sup>, Françoise LOHÉZIC – LE DÉVÉHAT<sup>2\*</sup>, Nicolas LE YONDRE<sup>1\*</sup>

6  
7   <sup>1</sup> Univ Rennes, CNRS, ScanMAT UAR 2025, CRMPO (Centre Régional de Mesures Physiques de  
8   l'Ouest), F-35000 Rennes, France

9   <sup>2</sup> Univ Rennes, CNRS, ISCR (Institut des Sciences Chimiques de Rennes) - UMR 6226, F-35000  
10   Rennes, France

11   <sup>3</sup> *INRAE, UR BIA, F-44300 Nantes, France and INRAE, PROBE research infrastructure, BIBS facility, F-44300*  
12   *Nantes, France*

13  
14   **\*Correspondence**

15   **Dr. Nicolas Le Yondre**

16   Phone: +33 (0)2 23 23 73 01

17   Email: [nicolas.leyondre@univ-rennes1.fr](mailto:nicolas.leyondre@univ-rennes1.fr)

18   Fax:   +33 (0)2 23 23 67 43

19   **Dr. Françoise Lohéziec-Le Dévéhat**

20   Phone: +33 (0)2 23 23 48 16

21   Email: [francoise.le-devehat@univ-rennes1.fr](mailto:francoise.le-devehat@univ-rennes1.fr)

22

23 Short title (< 70 char): **Chemical Profiling of Lichen Extracts using the Van Krevelen Diagram**

24 Short abstract (< 80 words):

25 Because most metabolomics tools are designed for LC-MS, it is often difficult to obtain structural  
26 information when straying from this gold standard. The Van Krevelen diagram is adapted for this  
27 purpose. Therefore, a new classification method is benchmarked against the classification tool  
28 ClassyFire, and various extracts from four lichens are analyzed with high resolution mass  
29 spectrometry (HRMS). Lichen metabolites are efficiently classified and the obtained results are  
30 consistent with the extraction protocols. This approach provides structural insight in lichen  
31 metabolites from HRMS.

32

### 33 **ABSTRACT**

34 **Introduction:** In recent years, LC-MS has become the gold standard for metabolomic studies.  
35 Indeed, liquid chromatography is relatively easy to couple with the soft electrospray ionization. As  
36 a consequence, many tools have been developed for the structural annotation of tandem mass  
37 spectra. However, it is sometimes difficult to do Data Dependent Acquisition (DDA), especially  
38 when developing new methods that stray from the classical LC-MS workflow.

39 **Objective:** An old tool from petroleomics that has recently gained popularity in metabolomics, the  
40 Van Krevelen (VK) diagram, is adapted for an overview of the molecular diversity profile in  
41 lichens through HR-MS.

42 **Methods:** A new method is benchmarked against the state-of-the-art classification tool ClassyFire  
43 using a database containing most known lichen metabolites ( $n \approx 2,000$ ). Four lichens known for  
44 their contrasted chemical composition were selected, and extractions with apolar, aprotic polar and  
45 protic polar solvents were performed to cover a wide range of polarities. Extracts were analyzed  
46 with Direct Infusion ElectroSpray Ionization Mass Spectrometry (DI-ESI-MS) and Atmospheric  
47 Solids Analysis Probe Mass Spectrometry (ASAP-MS) techniques to be compared with the  
48 chemical composition described in the literature.

49 **Results:** The most common lichen metabolites were efficiently classified, with more than 90% of  
50 the molecules in some classes being matched with ClassyFire. Results from this method are  
51 consistent with the various extraction protocols in the present case study.

52 **Conclusion:** This approach is a rapid and efficient tool to gain structural insight regarding lichen  
53 metabolites analyzed by high resolution mass spectrometry without relying on DDA by LC-  
54 MS/MS analysis. It may notably be of use during the development phase of novel MS-based

55 metabolomic approaches.

56

57 Keywords: Lichens, Metabolomics, Structural Classification, Extract Comparison

For Peer Review

## 58 INTRODUCTION

59 In recent years, significant advances in mass spectrometry have led to an increase in the  
60 measurement precision, sensitivity and frequency of acquisition of tandem mass spectra. These  
61 different parameters, combined with the extensive separation capabilities of liquid  
62 chromatography, have made LC-MS the gold standard in metabolomics.<sup>1,2</sup> It is indeed a preferred  
63 analytical technique because it provides an insight into the composition of an extract in a form that  
64 could then be used by the industry.<sup>3</sup> However, this also implies (i) that a time consuming extraction  
65 step must first be carried out and (ii) that a certain number of compounds will not be detected  
66 because they will be retained on the column during separation. The extraction step determines,  
67 through the choice of solvent, the fraction of the metabolome that can be studied.<sup>4</sup> To obtain a more  
68 comprehensive insight into the chemical composition of an organism, it becomes necessary to  
69 perform various extraction steps, which further increases the workload. Therefore, it is interesting  
70 to develop new tools for a comprehensive understanding of this composition by directly analyzing  
71 the biological material. Natural products (NPs), which are the first historical source of drugs and  
72 remain an important subject of study for the pharmaceutical industry today,<sup>5</sup> are of special concern.

73 With the establishment of LC-MS as a standard for metabolomics, many biocomputing  
74 tools have been developed and combined to facilitate the processing and interpretation of the  
75 massive amounts of data generated.<sup>6</sup> Most, if not all, of these tools have been created to process  
76 data containing tandem mass spectra, which are commonly acquired in a Data Dependent  
77 Acquisition (DDA).<sup>7</sup> It is possible to isolate features characterized by a dataset each comprising  
78 one  $m/z$  value, one retention time, and one tandem mass spectrum, which can be linked together  
79 by scores or shared patterns, notably on the GNPS platform.<sup>8-10</sup> Finally, these tandem mass spectra  
80 allow to propose identifications of a substantial number of molecules for each sample by comparing

81 them with experimental<sup>11,12</sup> or *in silico* databases.<sup>13</sup> This structural annotation can allow a  
82 hierarchical classification of the molecules composing an extract, in particular thanks to the  
83 ClassyFire tool.<sup>14</sup>

84 The use of tandem mass spectra, while being the strength of these tools, is also their  
85 weakness. For instance, when developing new mass spectrometry tools, it is not always possible to  
86 proceed *ab initio* to such an acquisition associating mass and tandem mass spectra. Data  
87 Independent Acquisition (DIA) is also possible and fragments all ions present in a wide range of  
88 *m/z* regardless of intensity. However, it then becomes difficult to interpret the resulting multiplexed  
89 spectra, and the *ad hoc* algorithms require substantial computation time, which is not suitable for  
90 a rapid screening.<sup>15</sup> As an alternative, the option was to develop a tool providing structural insight  
91 directly from the molecular formula (MF), which can readily be computed from high resolution  
92 MS data.<sup>16</sup> Indeed, an old tool from petroleomics – the Van Krevelen (VK) diagram<sup>17</sup> – was  
93 previously used to estimate the structural class associated with a MF through element ratios,  
94 notably for studying the composition of some alcohols.<sup>18,19</sup> To avoid the computational burden of  
95 MF determination, a local database search can be performed to determine element ratios for Van  
96 Krevelen analysis,<sup>20</sup> however in this approach the analysis ignores masses absent from the  
97 database, which can be problematic in a compound discovery context. Furthermore, a database  
98 search can sometimes generate mismatches because of cation adducts or the presence of ‘atypical’  
99 atoms – e.g. chlorinated metabolites, often described in lichens<sup>21</sup> – so, direct MF computation was  
100 chosen. For users wishing to try this kind of Van Krevelen-based approach, a Python package  
101 (PyKrev) has recently been published and offers a number of processing options, including a  
102 classification function – but not tailored for phytochemicals.<sup>22</sup>

103 Lichens, as organisms constituted of a fungus living in a symbiotic relationship with an alga



104 and/or a cyanobacterium,<sup>23</sup> possess an original chemistry comprising varied and unique  
105 compounds<sup>21</sup> and are thus well suited for the development and assessment of such a tool. They are  
106 also an interesting source of bioactive compounds,<sup>24,25</sup> and have been shown to be the ancestors of  
107 major fungal lineages (e.g. *Penicillium spp.*, *Aspergillus spp.*).<sup>26</sup>

108 This paper is the first part of a study aiming to establish a novel concept using Ambient  
109 Mass Spectrometry for NP research, focused on four lichen species: *Evernia prunastri* (L.) Ach.,  
110 *Lichina pygmaea* (Lightf.) C. Agardh., *Parmelia saxatilis* (L.) Ach. and *Roccella fuciformis* (L.)  
111 DC. The lichens were selected for their chemical diversity and an extensive list of the metabolites  
112 described in these species is provided as **Table S1**. In this paper a method capable of obtaining  
113 large structural insights into the composition of lichen extracts is presented without being reliant  
114 on tandem mass spectra.

## 115 **EXPERIMENTAL**

### 116 **General Procedures.**

117 All solvents and reagents used in this study were HPLC grade (Sigma-Aldrich) and were used  
118 without further purification. In order to avoid any contamination by plastic compounds, the use of  
119 glass containers was preferred and the use of colored plastics was strictly prohibited. Eppendorf  
120 micropipetting systems were used for metering the extract solutions. For experiments requiring the  
121 thermal desorption of extracts, samples were deposited on Marienfeld melting point capillaries.

### 122 **Lichen Material.**

123 This study was conducted on herbarium specimens. They were collected in: Cressensac, France (N  
124 45°0'19321" O 1°30'51.191"), August 2006 for *E. prunastri*; Roscoff, France (N 48°43'31.458" O  
125 3°58'8.651"), June 2016 for *L. pygmaea*; Girona, Spain (N 42°27' 53.399" O 1°47'23.816"), July

126 2017 for *P. saxatilis*; and Saint-Coulomb, France (N 48°41'29.299" O 1°56'45.855"), May 2003 for  
127 *R. fuciformis*. Voucher specimens are kept in the herbarium of the University of Rennes, France  
128 under the respective reference codes: JB/06/51, JB/16/204, JB/17/213 and JB/03/02.

### 129 **Extraction Protocol.**

130 The dried weights were chosen for each extraction so as to extract a sufficient quantity of material  
131 for all analyses (**Figure S1**). For each lichen, 2 extracts were obtained with methanol and acetone,  
132 and 2 extracts were from a successive extraction using cyclohexane, acetone then methanol (CAM)  
133 and using cyclohexane, acetone then water (CAW). CAM and CAW extracts were recombined *pro*  
134 *rata* to the extracted quantities. The selected quantities were 2 g for both the single-step 'Acetone'  
135 and the multiple-step 'CAM' extracts, 500 mg for the 'Methanol' extract and 4 g for the 'CAW'  
136 extract. All extractions were performed using an Accelerated Speed Extractor Büchi  
137 SpeedExtractor E-914 after dispersion of ground lichens in celite (1:1 ratio, m/m). A 2 g sand bed  
138 made of Büchi quartz sand of 0.3-0.9 mm granulometry was used to avoid clogging of the sintered  
139 part of the extraction cartridges. Each extraction step consisted of 3 maceration cycles of 10 min  
140 in 40 °C solvent, with a discharge time of 2 min and a degassing between every cycle.

### 141 **HR-MS Analyses.**

142 Direct infusion electrospray mass spectrometry (DI-ESI-MS) analyses were acquired on a Q-  
143 Exactive mass spectrometer (Thermo Scientific, Bremen, Germany) equipped with a Thermo  
144 HESI-II ion source. DI-ESI-MS analyses were performed on 15 µg/mL extract solutions using two  
145 ionization solvents, acetone and methanol. Acquisition time was set to 2 min (230 scans).  
146 Optimized parameters included a sheath gas set at 20 a.u., a transfer capillary voltage of 2800 V  
147 and an S-lens RF level of 50 a.u., the other parameters were set to default. The Thermo Scientific

148 Q-Exactive™ was calibrated with Pierce LTQ Velos ESI Positive and Negative Ion calibration  
149 solutions and lock-mass calibration was applied on palmitic acid during acquisition (deprotonated  
150 molecule  $[M-H]^-$   $m/z$  255.2330 and protonated molecule  $[M+H]^+$   $m/z$  257.2475). Experiments  
151 were performed in a single batch with analytical triplicates. The system was rinsed with the  
152 working solvent up to noise level between each sample. No loss of sensitivity was observed when  
153 rerunning samples as control.

154 Atmospheric Solids Analysis Probe Mass Spectrometry (ASAP-MS) analyses were  
155 acquired on a Maxis 4G spectrometer (Bruker Daltonics, Bremen, Germany) equipped with an  
156 APCI ion source and a direct insertion probe (DIP). The extracts were deposited on Marienfeld  
157 melting point capillaries as homogeneous solutions and dried with nitrogen before insertion in the  
158 APCI source. Preliminary tests having shown a major thermolability of some compounds, analyses  
159 were performed with a temperature ramp-up (50 °C every 15 sec) over 2 min. The temperature  
160 range covered was therefore 100-400 °C with 120 scans in the range  $m/z$  100-2000. Source  
161 parameters were optimized as following with a nebulization gas (nitrogen) pressure of 2.5 bars and  
162 a dry gas at 2.5 L.min<sup>-1</sup>. Voltages were set to: 500 V for the transfer capillary, -500 V for the end  
163 plate offset, with a 2000 V charging voltage and a 4000 V working voltage. An ion-cooler of 175  
164 V<sub>p-p</sub> was applied on the pusher. The Corona discharge was set to 500 nA for positive ionization and  
165 2000 nA for negative ionization. External calibration of the spectrometer was done with PEG 600  
166 in positive mode and PEG diacid 600 in negative mode. Lock-mass calibration was applied on  
167 palmitic acid by post-processing with R-script.<sup>27</sup> During the batch two replicates were performed  
168 to obtain a total of 8 acquisitions on the merged data for interpretation. Experiments were  
169 performed as a single batch with a thermal desorption to noise level between each run, no loss of  
170 sensitivity was observed on controls.

## 171 **Bioinformatic Data Processing.**

172 Thermo Scientific \*.RAW files were converted to \*.mzXML using readw 4.3.1 and Bruker \*.d  
173 folders were converted to \*.mzXML using Bruker CompassXport 3.0.6.9 from the command line.  
174 The obtained files were processed with R-script and Microsoft Visual Basic, following the Seven  
175 Golden Rules of MF determination,<sup>16</sup> every step was controlled manually while encoding the  
176 scripts, and exported in the form of \*.csv files.

177 Peak lists were extracted from the \*.mzXML files using R packages *xcms*,<sup>28–30</sup>  
178 *MSnbase*,<sup>31</sup> and *CAMERA*,<sup>32</sup> and automated instrument detection was performed with the  
179 *readMzXmlData*<sup>33</sup> package. The peak picking parameters were optimized for each instrument.  
180 Isotopic contributions calculation and clustering followed by deisotoping were performed with a  
181 3 ppm tolerance using element restrictions depending on *m/z* as described in the Seven Golden  
182 Rules. For this purpose the mass ranges were refined according to the compounds referenced in the  
183 online Dictionary of Natural Products (DNP) v26.2<sup>34</sup> (**Table 1**). After an in-depth study of the mass  
184 spectra, it appeared that isotopic contributions could not be seen for monoisotopic ions of less than  
185 0.01 % relative intensity, so this value was set as a threshold for the calculation of isotopic clusters.

186 The molecular formulae were then calculated by querying the online platform ChemCalc<sup>35</sup>  
187 with R-script, taking into account the calculated isotopic patterns, allowing an error of 4 carbons  
188 and 2 nitrogens.  $[M+Na]^+$  adducts were further accounted for, when generating the formulae for  
189 ESI-MS. When no isotope pattern was detected, **Table 1** features were considered. However, sulfur  
190 and chlorine were not considered unless detected. The absolute mass range error was set to 1.0  
191 milli-unified atomic mass (*mu*) for orbitrap data (DI-ESI-MS) and 1.7 *mu* for Q-TOF data (ASAP-  
192 MS). Based on the Double Bond Equivalent (DBE) the script filtered out the formulae  
193 corresponding to radical ions for ESI-MS. The results were coherent with manual MF

194 determination of known compounds and yielded most often a single formula, though the script was  
195 able to calculate the five best formulae.

196 The different elemental ratios were then calculated with Visual Basic in order to filter the  
197 formulae according to the Seven Golden Rules to remove the potentially incoherent ones. An  
198 adduct correction was applied according to the formulae and DBE with the following considered  
199 adducts: ASAP-MS:  $M^+\bullet$ ,  $[M+H]^+$ ,  $M^-\bullet$ ,  $[M-H]^-$  and ESI-MS:  $[M+H]^+$ ,  $[M+Na]^+$ ,  $[M-H]^-$ ,  
200  $[M+Cl]^-$ . The corresponding exact molecular masses were subsequently calculated.

201 Chemical classes associated with the Van Krevelen diagram were retrieved from the  
202 literature and double-checked against an in-house database of lichen compounds, the complete  
203 projet is available<sup>12</sup> (see **Technical validation**). The *Van Krevelen coordinates i.e. the H/C and*  
204 *O/C ratios* were then imported in R-script and merged in a single matrix that can be filtered by all  
205 experimental variables. A rectangle approximation of the chemical groups presented **Figure 1** was  
206 applied to tag the detected compounds, which were filtered for plotting in the form of a histogram  
207 (relative proportion within a method) stacked with a heatmap (absolute count of detected features).  
208 The plots were created with R package `plot3D`<sup>36</sup> and were made interactive with `plot3Drgl`.  
209 <sup>37</sup> Although the presented plots were constructed with the merged data of all lichens, the provided  
210 scripts allow for species filtering (e.g. **Figure S2**). The Van Krevelen coordinates were also used  
211 to plot the Van Krevelen diagrams using R-script.

212 The entire processed dataset is available at <https://osf.io/6pyuq/>  
213 (dataset\_solvents\_comp\_publi1.tsv) along with sample files for Q-Orbitrap and Q-TOF data in the  
214 \*.mzXML format. The scripts are deposited at  
215 <https://github.com/siollivier/directacquisitionproject>.

216 **Technical validation.**

217 **Database establishment.** An in-house database of lichen compounds was compiled from the  
218 literature. The total number of structures contained in this compiled database is about 2,000  
219 entries<sup>12</sup>. The database will be provided by the authors upon reasonable request.

220 The structural information was translated into InChIKey, a particularly popular format in  
221 chemoinformatics, because it is standardized and of fixed length, allowing easy data processing  
222 (example: mycosporine serinol, [VVTDHOIRNPCGTH-NSHDSACASA-N](#)).

223 A detailed classification of the lichen compounds was then obtained by submitting these  
224 InChIKeys to ClassyFire,<sup>14</sup> a tool to obtain a hierarchical classification of compounds, through the  
225 `classyfireR` package.<sup>38</sup> This structure-based bioinformatics approach allows for a  
226 comprehensive result through the establishment of a chemical taxonomy—in hierarchical order:  
227 chemical kingdom, superclass, class, subclass, and additional levels 5-9.

228 **Evaluation of the method against a state-of-the-art classification tool.** The accuracy of the  
229 classification established by **Figure 1** was examined in comparison to this database of lichen  
230 compounds. Within the database, compounds were selected according to different levels of their  
231 ClassyFire classification, which was compared to the classification obtained with Van Krevelen.

232 The same level of precision cannot be expected from a classification based only on the MF.  
233 However, depending on the structure of the compounds, it is possible to expect a result in one or  
234 more classes (more or less specific) of this diagram. Thus, some areas of the diagram (i.e.  
235 benzenoids or prenyl derivatives) can be considered as areas that extend more specific ones. During  
236 the evaluation of the protocol, the compounds were considered as correctly classified if their Van  
237 Krevelen annotation was in accordance with the expected matches between ClassyFire and Van

238 Krevelen that are shown in **Table 2**.

## 239 **RESULTS AND DISCUSSION**

### 240 **Determination and validation of the *Van Krevelen* chemical groups.**

241 First, a search was conducted in the literature to identify the positioning of the defined chemical  
242 groups on the Van Krevelen diagram.<sup>18,19,39–41</sup> These bibliographic results were then refined  
243 according to a database organized with ClassyFire and the final ranges are specified in **Figure 1**.  
244 The database used to validate these annotations mostly contains compounds of MW < 500 Da,  
245 including mainly, in terms of Classyfire classes, depsides and depsidones (> 20 %), as illustrated  
246 in **Figure 2**.

247 In order to guarantee the quality of the following results, it is first necessary to evaluate the  
248 accuracy of the classification on datasets not acquired in DDA. The proportion of correctly  
249 classified molecules in each class is an effective way to perform this validation.

250 However, two biases to this estimation should be considered: (i) the classification being  
251 based on H/C and O/C ratios, classes containing a large number of heteroatoms (N, S, *etc.*) such as  
252 amino acids are more likely to be incorrectly classified. Others have addressed this issue by  
253 including various elemental ratios, but this promising approach has yet to achieve the same level  
254 of specificity for classification;<sup>42</sup> (ii) some ClassyFire classifications include very structurally  
255 diverse derivatives within the same class, particularly for compounds with a lipid or carbohydrate  
256 function.

257 The proportions of molecules correctly classified in each of the classes considered are  
258 presented in **Figure 3**. As expected, amino acids and analogues, fatty acyls and carbohydrates have  
259 a lesser overlap between the two classifications, with respectively 56%, 59% and 68% of

260 compounds being correctly classified.

261 On the other hand, all the other structural classes considered have an overlap of more than  
262 70%. For 4 of the 10 structural types considered, the similarity between the two classifications is  
263 even higher than 90%. These are benzenoid derivatives, chromones, depsides and depsidones as  
264 well as prenol lipids, which include terpenes. These chemical families are among the most  
265 frequently reported in lichens, as illustrated in the recently published [tandem mass spectra database](#)  
266 of the most common lichen compounds.<sup>12</sup>

267 The encouraging results of the method provided by processing non-experimental data gave  
268 us the incentive to proceed to a practical evaluation. In the following case study, this method is  
269 used to obtain a rapid evaluation of the diversity of compounds detected by multiple ionization  
270 techniques, presenting a comparison of results from two DI-ESI-MS conditions vs ASAP-MS.

### 271 **Case Study: Evaluation of the Metabolic Diversity of Lichen Extracts.**

272 In order to cover a large panel of structures from lichen extracts, four lichen samples known to  
273 have very distinct metabolite profiles were chosen (the number of compound in the ClassyFire  
274 classes are given in **Table S2** for each lichen). Protocols used for each lichen species (two single-  
275 step and two three-step ones) were established to have a comparison with a simple extraction  
276 generally used in chemotaxonomic studies in lichenology and more exhaustive extractions (pooling  
277 successive extracts from polar to apolar ones) (**Figure S1**). These extracts were analyzed with DI-  
278 ESI-MS and ASAP-MS to compare the performance of these techniques to reveal lichen  
279 compounds and evaluate their distribution in structural families (**Figure S2**).

280 The DI-ESI-MS experiments were conducted with two ionization solvents, acetone and  
281 methanol since this aspect of the experiment heavily impacts the compound detection. Indeed, with



282 positive ion mode data, Principal Component Analysis (PCA) shows that the two ionization  
283 solvents form distinct groups on the PC2 component (**Figure S3**). PC1 clearly differentiates DI-  
284 ESI-MS from ASAP-MS analysis and post-treated data with an adduct correction resulted in a  
285 slightly better clustering of samples with regard to raw data. While the positive ion mode detects a  
286 huge number of compounds when compared to the negative ion mode, the Van Krevelen diagram  
287 provides a broad and clear overview of the chemical diversity of the diverse chemical families in  
288 lichens. The complementarity of these three methods is also revealed through the distinct zones  
289 mostly covered by these 3 MS conditions on the Van Krevelen diagram with all discriminant  
290 coordinates (**Figure S4**).

291 In other words, a single detection mode is not sufficient to cover the metabolic diversity of  
292 lichen compounds and each technique is more or less suitable for a class of compounds to be  
293 revealed. This assumption was verified using a bioinformatic quantification of the number of  
294 molecules ionized in various samples obtained with different extraction protocols: depending on  
295 their MF, a structural class was assigned to the compounds according to the cartography presented  
296 **Figure 1**.

297 With regard to the relative performance of the extraction protocols compared in this study  
298 the quantitative and the qualitative aspects are considered. Comparing extractions with a single  
299 solvent, yields are higher when a polar protic solvent (i.e. methanol) is used instead of a polar  
300 aprotic solvent (i.e. acetone). The Cyclohexane-Acetone-Methanol (CAM) and Cyclohexane-  
301 Acetone-Water (CAW) extractions gave yields in the same range compared to a single methanol  
302 extraction (**Figure S1**).

303 Considering the performance of the MS techniques used to reveal the composition of  
304 extracts, the results obtained with the two multi-step solvents were analyzed, which are considered

305 to be the more exhaustive. In **Figure 4** ASAP-MS allowed the detection of more compounds than  
306 ESI-MS in negative mode (average number of ions: 424/run for ASAP-MS vs. 116/run for ESI-  
307 MS in acetone), but fewer in positive mode (average: 482/run for ASAP-MS vs. 1434/run for ESI-  
308 MS in acetone). Furthermore, these results are coherent with what could be expected when  
309 considering the protic solvent selected, and the most appropriate extraction and ionization methods  
310 for each class of compounds are summarized in **Table 3**. In addition, the comparison with the in-  
311 house database<sup>12</sup> exhibits differences for Fatty acyls or Amino-acids, as examples, with higher  
312 frequencies in the experimental data than in the database.

313 *Regarding the extraction protocols.*

314 The successive extraction steps method appears to be better suited than the methanol extraction to  
315 gather terpenes and prenol derivatives (respectively 2.0 and 1.5 times more, in positive ASAP-  
316 MS), while 1.3 times more carbohydrates were extracted with water (negative ESI-MS in  
317 methanol).

318 A comparison with the single-step extractions (**Figure S5**) shows that an acetone extraction  
319 would appear as a good choice for studies targeting polyphenols & derivatives or condensed  
320 aromatic structures as well as terpenes. On the other hand, as methanol has the closest extraction  
321 profile to that of the successive solvent extraction and comparable yields, it seems to be a good  
322 compromise. However as expected with a protic solvent fewer terpenes are extracted with  
323 methanol.

324 *Regarding the MS methods.*

325 Compared to ESI-MS, ASAP-MS favors the ionization of terpenes (relative frequency within an  
326 ionization method up to r.f. 20 %) and unsaturated hydrocarbons (r.f. 10 %) in positive mode, and

327 benzenoids (r.f. 35 %) in negative mode. However, ASAP-MS appears less favorable for  
328 polyphenols & derivatives but, interestingly, quite favorable for condensed aromatic compounds  
329 (up to r.f. 20% in some conditions, see **Figure 4d**).

330 This piece of data also shows differences between the two DI-ESI-MS ionization solvents  
331 that may be of use for targeted studies: in negative mode, carbohydrates and fatty acyls' ionization  
332 is superior in methanol (15 vs. 10 % and 20 vs. 10 % r.f. respectively) whereas that of polyphenols  
333 & derivatives and condensed aromatic compounds is better in acetone (40 vs. 25 % and 15 vs. 10  
334 % respectively). Differences are less marked in positive ionization, except for the condensed  
335 aromatic structures that are better ionized in methanol (> 15 vs. 10 %). Such compounds are  
336 common occurrences in lichens (e.g. depsides for polyphenols & derivatives and xanthones for  
337 condensed aromatic compounds).<sup>21,43</sup>

338 Furthermore, in positive mode there appears to be a similar ionization profile for benzenoids  
339 in all ionization methods (**Figure 4, Figure S5**), questioning the complementarity of the methods  
340 in this aspect. A referral to the Van Krevelen diagram however reveals that the ionized compounds  
341 have some discriminant elemental ratios with each method (Figure 5). VK discriminant coordinates  
342 are clearly found in low H/C ratio for ASAP-MS and in high H/C ratio for DI-ESI-MS in Methanol,  
343 whereas DI-ESI-MS in Acetone as few discriminant coordinates, showing a complementarity  
344 within the ionization modes. A referral to the Van Krevelen diagram however reveals that some  
345 elemental ratios are common to 2 or 3 ionization modes (**Figure 6**). Moreover, an overall higher  
346 number of occurrences in acetone extract is clearly observed, with a specificity in high H/C ratio—  
347 once again showing the usefulness of the Van Krevelen diagram for natural products metabolomics.  
348 To illustrate the potential of the use of Ambient MS for validation or new identification of Natural  
349 Product, in comparison with the in-house database, classical expected compounds but also new

350 coordinates are found in DI-ESI-MS and ASAP-MS as proofs of principle (**Figures S6-S10**).

351 Overall, the interest of the Van Krevelen diagram was demonstrated for the structural  
352 classification of metabolites (without the need to acquire tandem mass spectra data), and its  
353 applicability was shown with real-life cases by studying the chemical composition of lichens.

## 354 REFERENCES

- 355 1. Wolfender J-L, Marti G, Thomas A, Bertrand S. Current approaches and challenges for the  
356 metabolite profiling of complex natural extracts. *J Chromatogr A*. 2015;1382:136-164.  
357 doi:10.1016/j.chroma.2014.10.091
- 358 2. Wolfender JL, Nuzillard JM, Van J der H, Renault JH, Bertrand S. Accelerating metabolite  
359 identification in natural product research: toward an ideal combination of LC-HRMS/MS  
360 and NMR profiling, in silico databases and chemometrics. *Anal Chem*. Published online  
361 November 2018. doi:10.1021/acs.analchem.8b05112
- 362 3. Sarker SD, Latif Z, Gray AI. Natural Product Isolation. In: Sarker SD, Latif Z, Gray AI, eds.  
363 *Natural Products Isolation*. Methods in Biotechnology. Humana Press; 2005:1-25.  
364 doi:10.1385/1-59259-955-9:1
- 365 4. Choi YH, Verpoorte R. Metabolomics: What You See is What You Extract. *Phytochem*  
366 *Anal*. 2014;25(4):289-290. doi:10.1002/pca.2513
- 367 5. Basmadjian C, Zhao Q, Bentouhami E, et al. Cancer wars: natural products strike back.  
368 *Front Chem*. 2014;2. doi:10.3389/fchem.2014.00020
- 369 6. Ernst M, Kang KB, Caraballo-Rodríguez AM, et al. MolNetEnhancer: Enhanced Molecular  
370 Networks by Integrating Metabolome Mining and Annotation Tools. *Metabolites*.  
371 2019;9(7):144. doi:10.3390/metabo9070144
- 372 7. Allard P-M, Genta-Jouve G, Wolfender J-L. Deep metabolome annotation in natural  
373 products research: towards a virtuous cycle in metabolite identification. *Curr Opin Chem*  
374 *Biol*. 2017;36:40-49. doi:10.1016/j.cbpa.2016.12.022
- 375 8. Wang M, Carver JJ, Phelan VV, et al. Sharing and community curation of mass  
376 spectrometry data with Global Natural Products Social Molecular Networking. *Nat*  
377 *Biotechnol*. 2016;34:828-837. doi:10.1038/nbt.3597
- 378 9. Wandy J, Zhu Y, van der Hooft JJJ, Daly R, Barrett MP, Rogers S. Ms2lda.org: web-based  
379 topic modelling for substructure discovery in mass spectrometry. *Bioinformatics*.  
380 2018;34(2):317-318. doi:10.1093/bioinformatics/btx582
- 381 10. Hooft JJJ van der, Wandy J, Barrett MP, Burgess KEV, Rogers S. Topic modeling for  
382 untargeted substructure exploration in metabolomics. *Proc Natl Acad Sci*.

- 383 2016;113(48):13738-13743. doi:10.1073/pnas.1608041113
- 384 11. Ramos AEF, Le Pogam P, Alcover CF, et al. Collected mass spectrometry data on  
385 monoterpene indole alkaloids from natural product chemistry research. *Sci Data*.  
386 2019;6(1):1-6. doi:10.1038/s41597-019-0028-3
- 387 12. Olivier-Jimenez D, Chollet-Krugler M, Rondeau D, et al. A database of high-resolution  
388 MS/MS spectra for lichen metabolites. *Sci Data*. 2019;6(1):1-11. doi:10.1038/s41597-019-  
389 0305-1
- 390 13. Allard P-M, Péresse T, Bisson J, et al. Integration of Molecular Networking and In-Silico  
391 MS/MS Fragmentation for Natural Products Dereplication. *Anal Chem*. 2016;88:3317-3323.  
392 doi:10.1021/acs.analchem.5b04804
- 393 14. Djoumbou Feunang Y, Eisner R, Knox C, et al. ClassyFire: automated chemical  
394 classification with a comprehensive, computable taxonomy. *J Cheminformatics*.  
395 2016;8(1):61. doi:10.1186/s13321-016-0174-y
- 396 15. Bilbao A, Varesio E, Luban J, et al. Processing strategies and software solutions for data-  
397 independent acquisition in mass spectrometry. *Proteomics*. 2015;15(5-6):964-980.  
398 doi:10.1002/pmic.201400323
- 399 16. Kind T, Fiehn O. Seven Golden Rules for heuristic filtering of molecular formulas obtained  
400 by accurate mass spectrometry. *BMC Bioinformatics*. 2007;8:105. doi:10.1186/1471-2105-  
401 8-105
- 402 17. Van Krevelen DW. Graphical-statistical method for the study of structure and reaction  
403 processes of coal. *Fuel*. 1950;29:269-284.
- 404 18. Roullier-Gall C, Witting M, Gougeon RD, Schmitt-Kopplin P. High precision mass  
405 measurements for wine metabolomics. *Front Chem*. 2014;2. doi:10.3389/fchem.2014.00102
- 406 19. Roullier-Gall C, Lucio M, Noret L, Schmitt-Kopplin P, Gougeon R. How Subtle Is the  
407 “Terroir” Effect? Chemistry-Related Signatures of Two “Climats de Bourgogne.” *PLoS One*.  
408 2014;9:e97615. doi:10.1371/journal.pone.0097615
- 409 20. Brockman SA, Roden EV, Hegeman AD. Van Krevelen diagram visualization of high  
410 resolution-mass spectrometry metabolomics data with OpenVanKrevelen. *Metabolomics*.  
411 2018;14(4):48. doi:10.1007/s11306-018-1343-y
- 412 21. Huneck S, Yoshimura I. *Identification of Lichen Substances*. Springer Berlin Heidelberg;  
413 1996. Accessed February 17, 2019.  
414 <http://public.eblib.com/choice/publicfullrecord.aspx?p=3097560>
- 415 22. Kitson E, Kew W, Ding W, Bell NGA. PyKrev: A Python Library for the Analysis of  
416 Complex Mixture FT-MS Data. *J Am Soc Mass Spectrom*. 2021;32(5):1263-1267.  
417 doi:10.1021/jasms.1c00064

- 418 23. Honegger R. The Lichen Symbiosis—What is so Spectacular about it?\*. *The Lichenologist*.  
419 1998;30(3):193-212. doi:10.1006/lich.1998.0140
- 420 24. Boustie J, Grube M. Lichens—a promising source of bioactive secondary metabolites. *Plant*  
421 *Genet Resour*. 2005;3(2):273-287. doi:10.1079/PGR200572
- 422 25. Ranković B, Kosanić M. Lichens as a Potential Source of Bioactive Secondary Metabolites.  
423 In: Ranković B, ed. *Lichen Secondary Metabolites: Bioactive Properties and*  
424 *Pharmaceutical Potential*. Springer International Publishing; 2015:1-26. doi:10.1007/978-3-  
425 319-13374-4\_1
- 426 26. Lutzoni F, Pagel M, Reeb V. Major fungal lineages are derived from lichen symbiotic  
427 ancestors. *Nature*. 2001;411(6840):937-940. doi:10.1038/35082053
- 428 27. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for  
429 Statistical Computing, Vienna, Austria; 2018. <https://www.R-project.org/>
- 430 28. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. XCMS: processing mass  
431 spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and  
432 identification. *Anal Chem*. 2006;78(3):779-787. doi:10.1021/ac051437y
- 433 29. Tautenhahn R, Böttcher C, Neumann S. Highly sensitive feature detection for high  
434 resolution LC/MS. *BMC Bioinformatics*. 2008;9(1):504. doi:10.1186/1471-2105-9-504
- 435 30. Benton HP, Want EJ, Ebbels TMD. Correction of mass calibration gaps in liquid  
436 chromatography–mass spectrometry metabolomics data. *Bioinformatics*. 2010;26(19):2488-  
437 2489. doi:10.1093/bioinformatics/btq441
- 438 31. Gatto L, Lilley KS. MSnbase—an R/Bioconductor package for isobaric tagged mass  
439 spectrometry data visualization, processing and quantitation. *Bioinformatics*.  
440 2012;28(2):288-289. doi:10.1093/bioinformatics/btr645
- 441 32. Kuhl C, Tautenhahn R, Böttcher C, Larson TR, Neumann S. CAMERA: an integrated  
442 strategy for compound spectra extraction and annotation of liquid chromatography/mass  
443 spectrometry data sets. *Anal Chem*. 2012;84(1):283-289. doi:10.1021/ac202450g
- 444 33. Gibb S. *ReadMzXmlData: Reads Mass Spectrometry Data in MzXML Format.*; 2015.  
445 Accessed February 21, 2019. <https://CRAN.R-project.org/package=readMzXmlData>
- 446 34. Dictionary of Natural Products 26.2. Published April 13, 2018. Accessed April 13, 2018.  
447 <http://dnp.chemnetbase.com/faces/chemical/ChemicalSearch.xhtml>
- 448 35. Patiny L, Borel A. ChemCalc: A Building Block for Tomorrow's Chemical Infrastructure. *J*  
449 *Chem Inf Model*. 2013;53(5):1223-1228. doi:10.1021/ci300563h
- 450 36. Soetaert K. *Plot3D: Plotting Multi-Dimensional Data.*; 2017. Accessed February 21, 2019.  
451 <https://CRAN.R-project.org/package=plot3D>

- 452 37. Soetaert K. *Plot3Drgl: Plotting Multi-Dimensional Data - Using "Rgl."*; 2016. Accessed  
453 February 21, 2019. <https://CRAN.R-project.org/package=plot3Drgl>
- 454 38. Wilson T, Finch J. *ClassyfireR: R Interface to the ClassyFire RESTful API.*; 2019. Accessed  
455 August 16, 2019. <https://CRAN.R-project.org/package=classyfireR>
- 456 39. D'Andrilli J, Cooper W, Foreman C, G. Marshall A. An ultrahigh-resolution mass  
457 spectrometry index to estimate natural organic matter lability. *Rapid Commun Mass*  
458 *Spectrom.* 2015;29:2385-2401. doi:10.1002/rcm.7400
- 459 40. Kuhnert N, Dairpoosh F, Yassin G, Golon A, Jaiswal R. *What Is under the Hump? Mass*  
460 *Spectrometry Based Analysis of Complex Mixtures in Processed Food – Lessons from the*  
461 *Characterisation of Black Tea Thearubigins, Coffee Melanoidines and Caramel.* Vol 4.;  
462 2013. doi:10.1039/c3fo30385c
- 463 41. Lu Y, Li X, Mesfioui R, et al. Use of ESI-FTICR-MS to Characterize Dissolved Organic  
464 Matter in Headwater Streams Draining Forest-Dominated and Pasture-Dominated  
465 Watersheds. *PLoS One.* 2015;10:e0145639. doi:10.1371/journal.pone.0145639
- 466 42. Rivas-Ubach A, Liu Y, Bianchi TS, Tolić N, Jansson C, Paša-Tolić L. Moving beyond the  
467 van Krevelen Diagram: A New Stoichiometric Approach for Compound Classification in  
468 Organisms. *Anal Chem.* 2018;90(10):6152-6160. doi:10.1021/acs.analchem.8b00529
- 469 43. Le Pogam P, Boustie J. Xanthonenes of Lichen Source: A 2016 Update. *Molecules.*  
470 2016;21(3):294. doi:10.3390/molecules21030294

471

472 **TABLES**

473 **Table 1.** Element restriction ranges established from the online Dictionary of Natural Products  
474 v26.2.

Mass range [Da]	C max	H max	N max	O max	S max	Cl max
< 200	15	30	8	7	6	4
< 400	30	58	10	14	12	7
< 600	42	86	13	21	12	8
< 800	56	108	16	25	20	10
< 1000	66	126	25	37	20	11
< 1500	100	182	26	44	20	11

475



476 **Table 2.** Matches between ClassyFire and the Van Krevelen diagram for the most common lichen  
 477 metabolites. The matches were determined to fit properly with the usual classification of lichen  
 478 compounds by Huneck and Yoshimura.<sup>21</sup>

ClassyFire Classification (level)	Expected Area(s) in the Van Krevelen Diagram
Amino Acids and analogues (subclass)	Amino acids; Nucleic acids
Anthraquinones (subclass)	Condensed aromatic compounds
Benzene derivatives (class)	<i>Polyphenols and derivatives</i> ; Benzenoids
Carbohydrates and conjugates (subclass)	Carbohydrates
Chromones (level 5)	<i>Polyphenols and derivatives</i> ; Benzenoids
Depsidones and depsidones (class)	<i>Polyphenols and derivatives</i> ; Benzenoids
Dibenzofurans (subclass)	Condensed aromatic compounds; Benzenoids
Fatty acyls (class)	Fatty acyls; Prenol derivatives
Prenol lipids (class)	Terpenes; Prenol derivatives; Unsaturated hydrocarbons
Xanthonones (level 7)	Condensed aromatic compounds

479

480 **Table 3.** An overview of the most efficient parameters in this study for the extraction and MS  
 481 analysis of lichen metabolites, with regards to the number of compounds ionized and the amount  
 482 of other metabolites that may hinder the interpretation for each structural class.

Class	Extraction	Ionization	MS Polarity
<i>Carbohydrates</i>	CAW	ESI in methanol	NEG
<i>Amino Acids</i>	Methanol	ESI in acetone	POS
<i>Nucleic Acids</i>	CAW	ESI in acetone	NEG
<i>Prenol Derivatives</i>	CAM	ASAP (APCI)	POS
<i>Fatty Acyls</i>	Methanol	ESI in methanol	NEG
<i>Unsaturated Hydrocarbons</i>	CAM	ASAP (APCI)	POS



<i>Terpenes</i>	CAM or Acetone	ASAP (APCI)	POS
<i>Benzenoids</i>	Acetone	ESI in acetone	POS
<i>Polyphenols and derivatives</i>	Acetone	ESI in acetone	NEG
<i>Condensed Aromatic Compounds</i>	Acetone	ESI in methanol	POS

483

484 **FIGURE LEGENDS**485 **Figure 1.** Identification of the chemical groups on the Van Krevelen diagram.

486 **Figure 2.** (a) The distribution of the molecular weights of the compounds contained in the  
 487 validation database (LDB-Lit, solid line) shows a good overlap with the experimental data (dotted  
 488 line); (b) The Classyfire classes of the compounds in the LDB-Lit illustrate the diversity of the  
 489 molecules used for technical validation.

490 **Figure 3.** Proportions of molecules in the database accurately classified by the Van Krevelen  
 491 diagram (in accordance with **Table 2**).

492 **Figure 4.** Evaluation of the chemical diversity for the merged data of the three-step extractions:  
 493 ‘Cyclohexane-Acetone-Methanol’ in positive mode (a) and negative mode (b), and ‘Cyclohexane-  
 494 Acetone-Water’ in positive mode (c) and negative mode (d). Both the relative frequency for each  
 495 technique (histogram, vertical axis) and the absolute number of detected ions (heatmap) are  
 496 presented. The relative frequency represents the proportion of a chemical class within a given  
 497 ionization method. The database content is proposed in comparison to show the existing potential  
 498 in new identifications.

499 **Figure 5.** Extracted Van Krevelen diagram for benzenoids i.e. in the (O/C 0.2-0.4; H/C 0.8-1.8)  
 500 range, in all three ionization methods for the acetone (a) and Cyclohexane-Acetone-Methanol  
 501 extracts (b). The size of the circle, represented by a coefficient of expansion (cex), is proportional  
 502 to the number of occurrences of the (O/C, H/C) coordinates without correction factor (100 gives a  
 503 cex of 100). The colors (green, blue and red, alpha of 0.4) indicate respectively discriminant VK  
 504 coordinates for DI-ESI-MS in Methanol, ASAP-MS and DI-ESI-MS in Acetone ionization mode.

505 **Figure 6.** Extracted Van Krevelen diagram for benzenoids i.e. in the (O/C 0.2-0.4; H/C 0.8-1.8)  
 506 range, in all three ionization methods for the acetone (a) and Cyclohexane-Acetone-Methanol  
 507 extracts (b). The size of the circle is proportional to the number of occurrences of the (O/C, H/C)  
 508 coordinates with a correction factor of 0.1 (100 gives a cex of 10). Non-discriminant VK  
 509 coordinates for a single ionization mode are plotted in violet.

510

511 **Acknowledgements**

512 The authors thank Isabelle Rouaud (ISCR, CorInt, Rennes) for her technical assistance regarding  
513 the extraction protocols.

514 **Author Notes**

515 SO, PJ, FLD, JB and NLY designed the experiments. SO and FLD selected the lichen material,  
516 which was identified by JB. DO optimized the ASE protocol; SO, PJ and FL carried out the  
517 experiments. SO and NLY encoded the R-scripts and Visual Basic macro used for data processing.  
518 DO compiled the database for validation, DO and SO curated the database. All authors contributed  
519 to the redaction of the manuscript. None of the authors declare any competing financial interest.

520

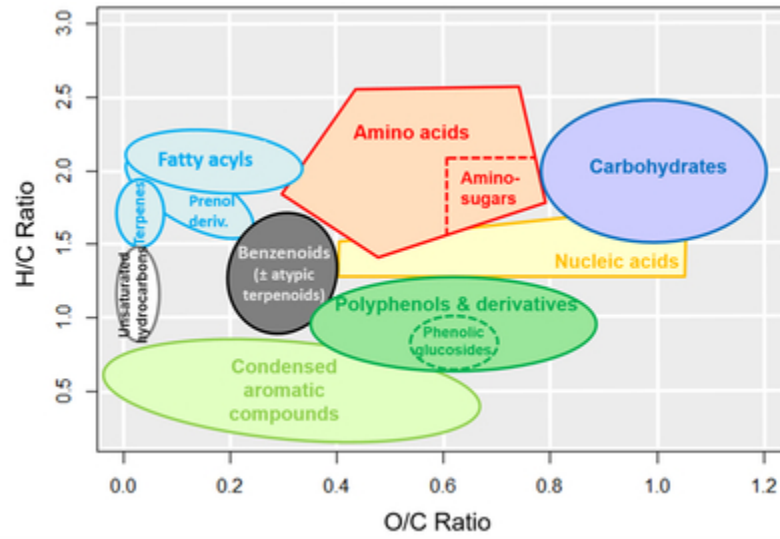


Figure 1. Identification of the chemical groups on the Van Krevelen diagram.

35x23mm (300 x 300 DPI)

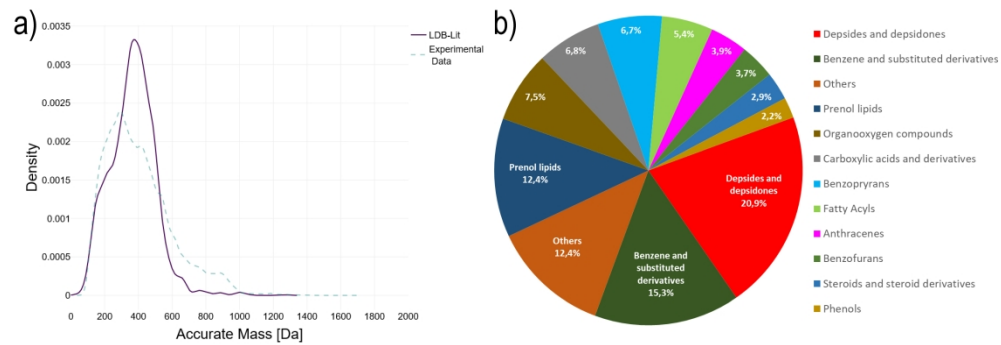


Figure 2. (a) The distribution of the molecular weights of the compounds contained in the validation database (LDB-Lit, solid line) shows a good overlap with the experimental data (dotted line); (b) The Classyfire classes of the compounds in the LDB-Lit illustrate the diversity of the molecules used for technical validation.

950x328mm (200 x 200 DPI)

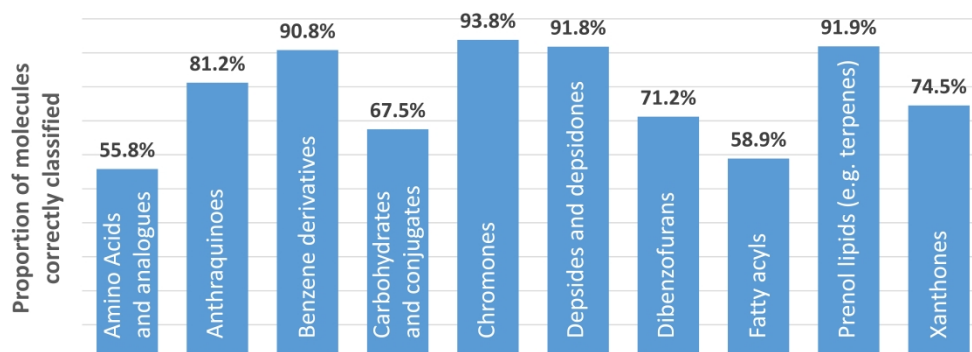


Figure 3. Proportions of molecules in the database accurately classified by the Van Krevelen diagram (in accordance with Table 2).

3102x1281mm (72 x 72 DPI)

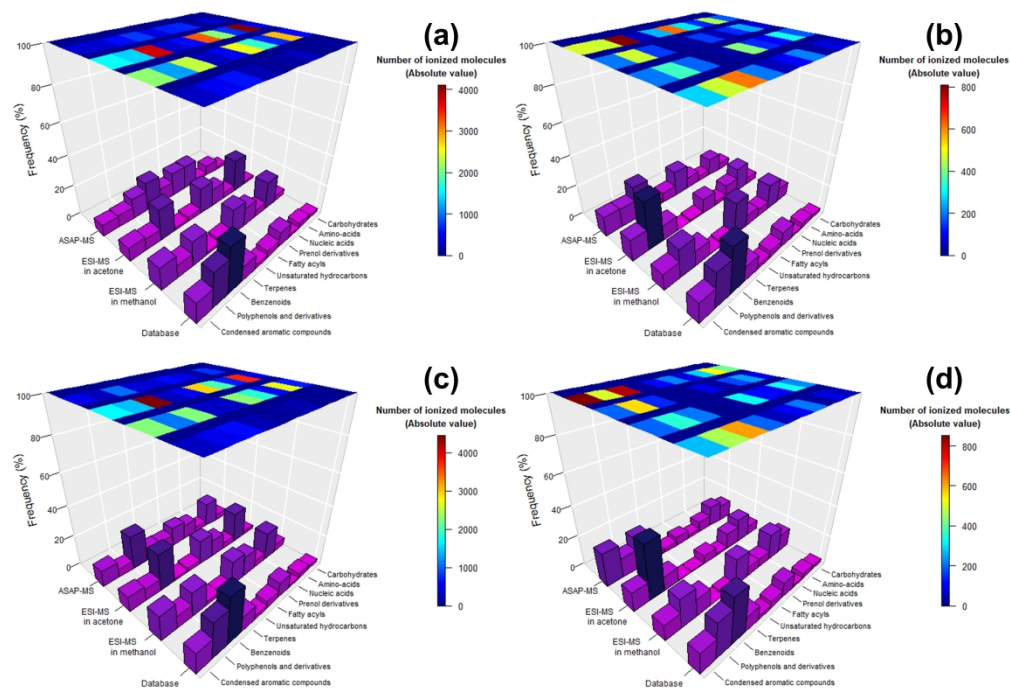


Figure 4. Evaluation of the chemical diversity for the merged data of the three-step extractions: 'Cyclohexane-Acetone-Methanol' in positive mode (a) and negative mode (b), and 'Cyclohexane-Acetone-Water' in positive mode (c) and negative mode (d). Both the relative frequency for each technique (histogram, vertical axis) and the absolute number of detected ions (heatmap) are presented. The relative frequency represents the proportion of a chemical class within a given ionization method. The database content is proposed in comparison to show the existing potential in new identifications.

1420x969mm (96 x 96 DPI)

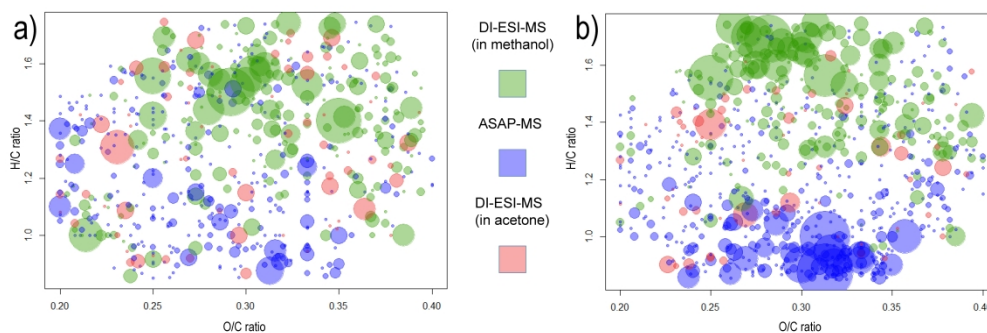


Figure 5. Extracted Van Krevelen diagram for benzenoids i.e. in the (O/C 0.2-0.4; H/C 0.8-1.8) range, in all three ionization methods for the acetone (a) and Cyclohexane-Acetone-Methanol extracts (b). The size of the circle, represented by a coefficient of expansion (cex), is proportional to the number of occurrences of the (O/C, H/C) coordinates without correction factor (100 gives a cex of 100). The colors (green, blue and red, alpha of 0.4) indicate respectively discriminant VK coordinates for DI-ESI-MS in Methanol, ASAP-MS and DI-ESI-MS in Acetone ionization mode.

950x328mm (200 x 200 DPI)

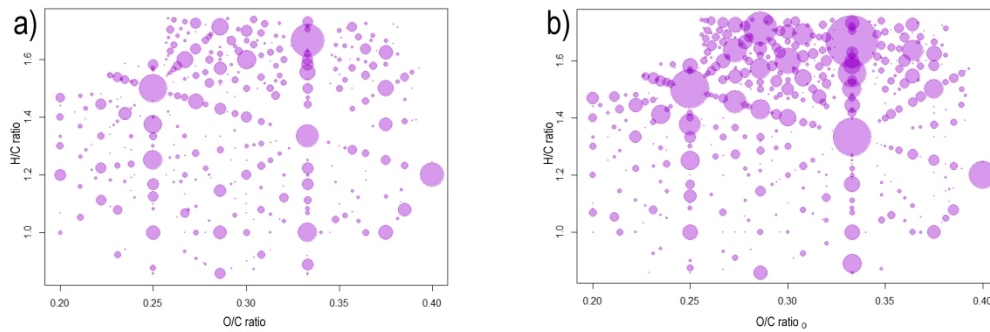


Figure 6. Extracted Van Krevelen diagram for benzenoids i.e. in the (O/C 0.2-0.4; H/C 0.8-1.8) range, in all three ionization methods for the acetone (a) and Cyclohexane-Acetone-Methanol extracts (b). The size of the circle is proportional to the number of occurrences of the (O/C, H/C) coordinates with a correction factor of 0.1 (100 gives a cex of 10). Non-discriminant VK coordinates for a single ionization mode are plotted in violet.

950x328mm (200 x 200 DPI)