



**HAL**  
open science

## Contribution of NAD 2D-NMR in liquid crystals to the determination of hydrogen isotope profile of methyl groups in miliacin

Philippe Berdagué, Philippe Lesot, Jérémy Jacob, Valery T. Terwilliger,  
Claude Le Milbeau

### ► To cite this version:

Philippe Berdagué, Philippe Lesot, Jérémy Jacob, Valery T. Terwilliger, Claude Le Milbeau. Contribution of NAD 2D-NMR in liquid crystals to the determination of hydrogen isotope profile of methyl groups in miliacin. *Geochimica et Cosmochimica Acta*, 2016, 173, pp.337-351. 10.1016/j.gca.2015.10.004 . insu-01225227

**HAL Id: insu-01225227**

**<https://insu.hal.science/insu-01225227v1>**

Submitted on 5 Nov 2015

**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.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

## Accepted Manuscript

Contribution of NAD 2D-NMR in liquid crystals to the determination of hydrogen isotope profile of methyl groups in miliacin

Philippe Berdagué, Philippe Lesot, Jérémy Jacob, Valery J. Terwilliger, Claude Le Milbeau

PII: S0016-7037(15)00586-4  
DOI: <http://dx.doi.org/10.1016/j.gca.2015.10.004>  
Reference: GCA 9458

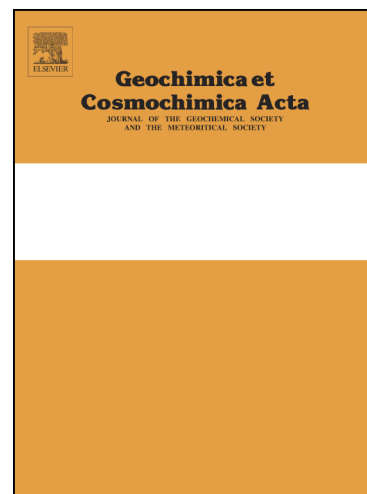
To appear in: *Geochimica et Cosmochimica Acta*

Received Date: 20 February 2015

Accepted Date: 6 October 2015

Please cite this article as: Berdagué, P., Lesot, P., Jacob, J., Terwilliger, V.J., Milbeau, C.L., Contribution of NAD 2D-NMR in liquid crystals to the determination of hydrogen isotope profile of methyl groups in miliacin, *Geochimica et Cosmochimica Acta* (2015), doi: <http://dx.doi.org/10.1016/j.gca.2015.10.004>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



# Contribution of NAD 2D-NMR in liquid crystals to the determination of hydrogen isotope profile of methyl groups in miliacin

Philippe Berdagué<sup>a,b</sup>, Philippe Lesot<sup>a,b,\*</sup>, Jérémy Jacob<sup>c,d,e</sup>, Valéry J. Terwilliger<sup>c,d,e,f,g</sup>

and Claude Le Milbeau<sup>c,d,e</sup>

<sup>a</sup>RMN en Milieu Orienté, ICMMO, UMR CNRS 8182, Université de Paris-Sud, Université Paris-Saclay, Bât. 410, 91405 Orsay cedex, France.

<sup>b</sup>CNRS/INC, ICMMO, UMR 8182, 91405 Orsay cedex, France.

<sup>c</sup>Université d'Orléans, ISTO, UMR 7327, 45071, Orléans, France.

<sup>d</sup>CNRS/INSU, ISTO, UMR 7327, 45071 Orléans, France.

<sup>e</sup>BRGM, ISTO, UMR 7327, BP 36009, 45060 Orléans, France.

<sup>f</sup>LE STUDIUM<sup>®</sup>, Loire Valley Institute for Advanced Studies, France.

<sup>g</sup>Department of Geography, University of Kansas, Lawrence, KS 66045, USA.

## Abstract

The hydrogen isotopic composition ( $\delta D$  or (D/H) value) of molecular biomarkers preserved in sedimentary archives is increasingly used to provide clues about the evolution of past climatic conditions. The rationale is that intact biomarkers retain isotopic information related to the climatic conditions that prevailed at the time of their synthesis. Some of these biomarkers may be degraded during diagenesis, however. The extent to which these degradations alter the original  $\delta D$  value of the source biomarker is presently debated and the capacity to resolve the debate by determinations of compound-specific  $\delta D$  values alone is limited. The “bulk” or “global”  $\delta D$  value of any molecule is in fact a composite of  $\delta D$  values at each site within this molecule ( $(\delta D_i$  or (D/H)<sub>*i*</sub> with *i*= number of hydrogen/deuterium atoms in the considered molecule). Determination of this site-specific  $\delta D_i$  value in biomarkers could not only yield outstanding paleoenvironmental information but also help forecast the impacts of diagenesis and define essential steps in biosynthetic pathways.

This task is analytically challenging. Here, we examined the capabilities of natural abundance deuterium 2D-NMR (NAD 2D-NMR) using homopolypeptide liquid crystals as an NMR solvent to: i) analyze the NAD spectra of biomarkers; ii) determine the site-specific distribution of hydrogen in the nine methyl groups ( $\delta D_{Mei}$  with *i*= 23 to 31) of miliacin, a pentacyclic triterpene of the amyrin family and key biomarker for broomcorn millet in sedimentary archives. Relative (D/H)<sub>Mei</sub> values were established by anisotropic NAD 2D-NMR. Then absolute  $\delta D_{Mei}$  values were obtained by determining  $\delta D_{Mei}$  value of the methoxy group of miliacin using two independent approaches: isotropic NAD NMR (SNIF-NMR<sup>TM</sup>) and GC-irMS. The resulting isotope profile for miliacin shows, for the first time, large variations in  $\delta D_{Mei}$  values that can directly be explained by biosynthetic processes. This approach has also the potential to permit predicting fractionations during pentacyclic triterpene diagenesis.

**Corresponding author:** Dr. Philippe Lesot

Address: RMN en Milieu Orienté, ICMMO, UMR CNRS 8182, Université de Paris-Sud, 91405

Orsay cedex, France, Tel.: +33 (0) 1 69 15 47 59, Fax: +33 (0) 1 69 15 81 05

E-mail address: [philippe.lesot@u-psud.fr](mailto:philippe.lesot@u-psud.fr) (Philippe Lesot)

**Keywords:** Deuterium NMR, Isotope profile, Site-specific, Natural abundance, Miliacin, Pentacyclic triterpene, Liquid crystal, Biosynthesis, Diagenesis

**Type of article:** Research article

**Supplementary information:** Yes

50 **1. INTRODUCTION**

51  
52 Higher plant pentacyclic triterpenes are natural products that bear an oxygenated group at the C-3  
53 position (**Fig. 1**). The large diversity of structures, isomerisms and functional groups resulting from  
54 specific enzymatic processes have led to the use of compounds in this family for establishing precise  
55 chemotaxonomic relationships (Das and Mahato, 1983; Mahato et al., 1992; Mahato and Sen, 1997).  
56 Although  $\beta$ - and  $\alpha$ -amyrins are very common structures in angiosperms, several classes of  
57 compounds are more or less specific to given taxa. For example, pentacyclic triterpenes with an  
58 acetate function at C-3 are largely restricted to Asteraceae species (Lavrioux et al., 2011) whereas  
59 those bearing a methoxy group at the C-3 position are mostly synthesized by Poaceae (Ohmoto et al.,  
60 1970; Jacob et al., 2005). After the death of their source organisms, these compounds may be  
61 archived in geological deposits such as soils or sediments. If diagenesis does not strongly modify  
62 their structure, these compounds can be analyzed for presence/absence and concentrations in  
63 sedimentary records in order to reconstruct past vegetation dynamics (e.g., Cranwell et al., 1984;  
64 Killops et al., 1995; van Aarsen et al., 2000; Regnery et al., 2013).

65  $3\beta$ -methoxyolean-18-ene, usually denoted miliacin or germanicol methyl ether (**Fig. 1**), is such a  
66 specific molecular biomarker. It is a pentacyclic triterpene bearing a methoxy group at position C-3  
67 that is predominantly produced by Poaceae species. Furthermore, this compound is most abundant in  
68 broomcorn millet (*Panicum miliaceum* L) (Itô et al., 1934; Ohmoto et al., 1970; Bossard, 2013).  
69 Miliacin concentrations in lake sediments and soils have thus been used to reconstruct the history of  
70 millet cultivation (Jacob et al., 2008, 2008b, 2009; Simonneau et al., 2013; Motuzaitė-Matuceviciute  
71 et al., 2013).

72 In addition to these crucial information on past vegetation, the hydrogen isotopic compositions  
73 ( $\delta D$  or D/H) of lipid compounds, such as miliacin, in geological deposits are providing superior  
74 insights about paleoclimate (Sachse et al., 2012; Terwilliger et al., 2013; Terwilliger and Jacob,  
75 2013). This is because meteoric water is the original source of hydrogen atoms in biochemicals  
76 produced by autotrophs and the  $\delta D$  values of meteoric water are strongly influenced by climate  
77 (Dansgaard et al., 1964; Gat et al., 1996). Further studies are needed in order to fully exploit the  
78 paleoclimate potential of  $\delta D$  values of lipid compounds for providing clues about past climates, with  
79 distinct levels of understanding on the processes occurring at different steps (Sachse et al., 2012).  
80 The impacts of environmental factors such as soil evaporation and leaf transpiration are relatively  
81 well identified. Leaf transpiration can alter the  $\delta D$  value of meteoric water along its path to the sites  
82 of photosynthesis where its hydrogen atoms from leaf water will be used to produce organic  
83 compounds via biosynthetic pathways (Sessions et al., 1999; Chikaraishi and Narakoa, 2003; Hou et  
84 al., 2007b; McInerney et al., 2011). Additionally, it has been demonstrated that plant type can cause  
85 differences in  $\delta D$  values of the most common lipid compounds such as *n*-alkyl lipids (Sachse et al.,

86 2006; Liu et al., 2006; Smith and Freeman, 2006; Hou et al., 2007; Feakins and Sessions, 2010;  
87 Wang et al., 2013). Because it would not be affected by variations caused by differences in plant-  
88 source taxa, the  $\delta D$  value of more specific compounds such as sedimentary miliacin is expected to  
89 solve this issue and could provide even higher resolution information about climatic conditions at the  
90 time of their synthesis (Bossard et al., 2011; Bossard, 2013; Terwilliger et al., 2013; Schwab et al.,  
91 2015). A better understanding of the biosynthetic pathway that designs the  $\delta D$  values of a given  
92 compound is also necessary since distinct NADPH sources and enzymatic processes can lead to  
93 highly variable fractionations (i.e. Sauer et al., 2001; Chikaraishi et al., 2004, 2009). Although the  
94 biosynthetic pathway of pentacyclic triterpenes such as miliacin is rather well known (Xu et al.,  
95 2004), associated H isotopic fractionations are less understood.

96 Finally, it is crucial to determine whether diagenetic derivatives retain the original paleoclimate  
97 information acquired by biochemicals during biosynthesis. For instance, pentacyclic triterpenes can  
98 undergo several transformations during early diagenesis such as the loss of functional groups, the loss  
99 of rings, double bond migration, aromatization, or epimerization of methyl groups (Corbet, 1980;  
100 Trendel, 1985; Lohmann, 1988; Jacob et al., 2007). Only isotopic exchanges with water have been  
101 addressed on polycyclic compounds (Wang et al., 2013) but the impact of diagenetic transformations  
102 on  $\delta D$  values of diagenetic by-products compared to their precursors is yet unknown.

103  
104 Our hypothesis is that a site-specific (or intramolecular) heterogeneous distribution of hydrogen  
105 isotopes in a biochemical precursor would explain any isotopic difference with its derivatives caused  
106 by the preferential loss of D-enriched or D-depleted groups during diagenesis. Identifying such  
107 isotopic difference between biochemicals and their diagenetic derivatives will allow better climate  
108 reconstructions from  $\delta D$  values of diagenetic derivatives. In addition, obtaining the site-specific  
109 distribution of hydrogen isotopes could provide useful information on biosynthetic pathways. As  
110 emphasized by Eglinton and Eglinton (2008), "*intramolecular compound-specific isotope*  
111 *measurements (i.e., analyses of the distribution of isotopes within a molecular structure) are*  
112 *analytically challenging and presently rarely attempted, but may ultimately prove of value*" and so  
113 addressing this hypothesis requires the development of new tools.

114 The two main objectives of our study were to test the hypothesis that hydrogen isotopes are not  
115 randomly distributed in molecular biomarkers and to identify the means of analyzing the site-specific  
116 distribution of hydrogen isotopes necessary for this test. We use miliacin as a representative of  
117 pentacyclic triterpenes. Specifically, we determined the (D/H) ratios ( $\delta D$ ) of hydrogen located in the  
118 methyl groups of miliacin by Natural Abundance Deuterium (NAD), two-dimensional Nuclear  
119 Magnetic Resonance (2D-NMR) spectroscopy using aligning organic media (anisotropic systems) as  
120 NMR solvents. This advanced technique modifies the well-known SNIF-NMR<sup>TM</sup> method pioneered

121 by Martin and co-workers in the 80's (Martin et al., 1980; 1986) by dissolving the analyte in a liquid  
 122 crystal (LC) instead of a liquid solvent (isotropic system). The replacement of isotropic to oriented  
 123 solvent is analytically highly valuable because we have access to  $^2\text{H}$  residual quadrupolar couplings  
 124 in solutes, leading to the presence on  $^2\text{H}$  NMR spectra of quadrupolar doublets (QD) for each  
 125 inequivalent deuteron instead of singlets. The range of splittings of each QD leads in turn to  
 126 distribute the NAD signals on larger spectral windows, thus reducing the peak overlaps.  
 127 Concomitantly the use of 2D-NMR experiments simplifies the analysis of entangled, overcrowded  
 128 anisotropic NAD 1D-NMR spectra. As we will see, this approach makes it possible to distinguish  
 129 several NAD signals from H in methyl groups of miliacin that resonated at the same (or very close)  
 130 frequencies in isotropic solvents. This unusual strategy has been recently pioneered by Lesot and  
 131 coworkers for analyzing the site-specific isotope ratios,  $(\text{D}/\text{H})_i$ , in fatty acids and then triglycerides  
 132 (Lesot et al., 2008, 2009, 2011, 2012; Serhan et al., 2010). We report here its first application for  
 133 investigating biomarkers of interest such as miliacin.

134

## 135 2. ANISOTROPIC $^2\text{H}$ NMR SPECTROSCOPY

136

137 When a molecule in the spectrometer's magnetic field does not tumble isotropically, all  
 138 orientation-dependent NMR interactions become visible on spectra. Thus  $^2\text{H}$  NMR spectra of solutes  
 139 in a mesophase contains new informative NMR interactions such as the chemical shift anisotropy ( $^2\text{H}$   
 140 CSA); ii) the dipolar interactions (D); iii) and the quadrupolar interaction ( $I \geq 1$ ) (Sarfati et al., 2000;  
 141 Lesot et al., 2015) (isotopically normal or enriched).

142 Assuming that  $^2\text{H}$  CSA is generally negligible while the  $^2\text{H}$ - $^1\text{H}$  couplings can be removed by  
 143  $^1\text{H}$  decoupling, the  $^2\text{H}$  signal of aligned monodeuterated molecule is a simple QD (two transitions)  
 144 centered on the anisotropic  $^2\text{H}$  chemical shift,  $\delta(^2\text{H})^{\text{aniso}}$ , (generally  $\delta(^2\text{H})^{\text{aniso}} \approx \delta(^2\text{H})^{\text{iso}}$ ) (Lesot et al.,  
 145 2009). The separation (in Hz) between components of doublet is referred to as  $^2\text{H}$  quadrupolar  
 146 splitting (or the residual quadrupolar coupling) and will be noted  $\Delta\nu_{\text{Q}_i}(^2\text{H})$  therein (**Fig. S-4**,  
 147 **Appendix A.II**). For a C-D internuclear directions oriented in a weakly-aligning LC, the electric  
 148 field gradient ( $V_{\text{C-D}}$ ) is usually assumed to be axially symmetric along the C-D vector and the value  
 149 of  $\Delta\nu_{\text{Q}_i}$  relative to the  $\mathbf{B}_0$  can be expressed as (**Eq. 1**; Sarfati et al., 2000; Lesot et al., 2009):

150

$$151 \Delta\nu_{\text{Q}} = \frac{3}{2} C_{\text{Q}}^{\text{D}_i} \times S_{\text{C-D}_i} = \frac{3}{2} \left( \frac{eQ_{\text{D}_i} V_{\text{C-D}_i}}{h} \right) \times \left( \frac{1}{2} \langle 3 \cos^2(\theta_{\text{C-D}_i}) - 1 \rangle \right) \quad (1)$$

152

153 where  $C_{\text{Q}}^{\text{D}_i}$  is the quadrupolar coupling constant at deuterium atom  $i$ ,  $S_{\text{C-D}_i}$  is the order parameter of  
 154 the C-D <sub>$i$</sub>  vector and  $\theta_{\text{C-D}_i}$  is the angle between the  $\mathbf{B}_0$  magnetic field axis and the C-D <sub>$i$</sub>  bond axis.

155 The  $\langle \dots \rangle$  denotes an ensemble average. As the trigonometric function in brackets varies between 1

156 (when  $\theta_{C-D_i} // \mathbf{B}_0$ ) and -0.5 (when  $\theta_{C-D_i} \perp \mathbf{B}_0$ ),  $\Delta\nu_{Q_i}$  can be positive or negative. When  $\theta_{C-D_i}$  is  
157 equal to  $54.7^\circ$  (the so-called ‘magic angle’),  $S_{C-D_i}$  becomes null, and so the doublet collapses to a  
158 single line.

159 As the second stable isotope of hydrogen, deuterium atoms are naturally in all organic  
160 compounds, and hence they can be used as nuclear probe to study the site-specific natural isotope  
161 distribution of molecules. Even with low natural abundance of deuterons (around  $1.55 \cdot 10^{-2} \%$  V-  
162 SMOW value; De Laeter et al., 2003), routine NMR spectrometers can detect quite easily (with  
163 reasonable experimental times) their NMR signals both in isotropic and anisotropic solvents,  
164 successful results depending also on the concentration of the analyte and its molecular weight (Lesot  
165 et al., 1998, 2009).

166 Proton-decoupled NAD NMR spectra of solutes dissolved in homogeneous mesophases  
167 consist of a sum of resolved independent quadrupolar doublets instead of single lines as in liquids,  
168 each of them corresponding to the various inequivalent  $^2\text{H}$  sites in the molecule. For middle to large-  
169 sized molecules, the distribution of positions and the splitting of  $^2\text{H}$  doublets over the spectral  
170 window can generate complex spectra (with possible overlaps of resonances), where the assignment  
171 of QD can be ambiguous. Various two and three-dimensional (2D/3D) NMR sequences belonging to  
172 NMR experiment class entitled “QUadropole Ordered SpectroscopY” (QUOSY) have been  
173 specifically developed to achieve clearer analyses of overcrowded anisotropic NAD 1D spectra  
174 (Merlet et al, 1999; Lesot and Lafon, 2008b). All of them allow the spectral information ( $\delta$ ,  $\Delta\nu_Q$ ) to  
175 be distributed on several (two or three) spectral dimensions ( $F_1$ ,  $F_2$  and  $F_3$ ), thus facilitating the  
176 assignment of  $^2\text{H}$  quadrupolar doublets.

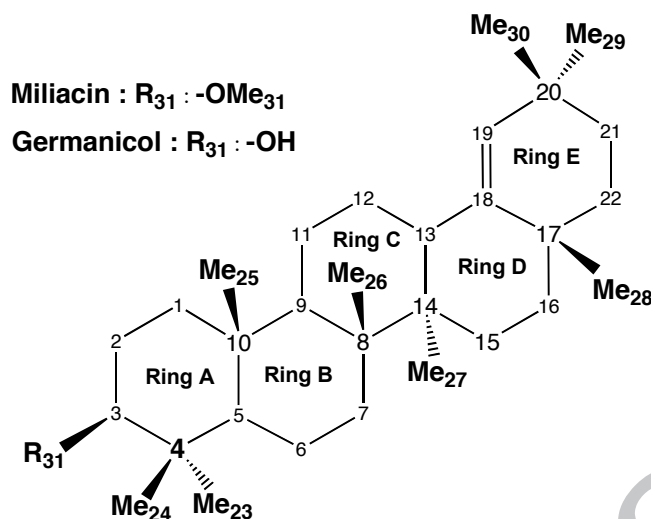
177

### 178 3. ANALYTES, MATERIALS AND METHOD

179

#### 180 3.1. Structure of miliacin and germanicol

181 Like other pentacyclic triterpenes, miliacin is formed from mevalonic acid *via* the mevalonate  
182 pathway through the cyclisation of squalene epoxide by an oxidosqualene cyclase (Xu et al., 2004;  
183 **see Section 6**). This scheme leads to the formation of germanicol, a triterpene alcohol constituted of  
184 five six-membered rings (**Fig. 1**) that bear eight methyl groups (Me23 to Me30). The final conversion  
185 of the alcohol function into a methyl ether function, which principally takes place in plants of the  
186 family Poaceae (Jacob et al., 2005), transforms germanicol into miliacin, by the addition of a methyl  
187 group (Me31) linked to the oxygen on the C-3 position. In miliacin, the nine methyl groups are  
188 numbered from Me23 to Me31 (**Fig. 1**). For simplicity, the five 6-membered rings of miliacin have  
189 been labeled A-E and contain two or three methyl groups each (ring A: Me23, Me24 and Me 25; ring  
190 C: Me26 and Me27; ring E: Me28, Me29 and Me30).



**Figure 1:** Structure of miliacin where  $R_{31} = -OMe$  and germanicol where  $R_{31} = -OH$ , together with the atomic numbering used and notation of rings (A to E) and methyl groups (23 to 31).

### 3.2. Miliacin and germanicol extraction and purification

Miliacin and germanicol were recovered from millet oil (provided by HITEX, Vannes, France) of > 98% purity as determined by Gas Chromatography/Mass Spectrometry (GC/MS). 6 g of miliacin were directly recovered from 750 ml of oil by repeated precipitation in cold acetone (Itô, 1934). From these 6 g of analyte, only about 100 mg of miliacin was necessary to prepare NMR samples used for NAD NMR analyses described in this work. Germanicol, which is far less abundant than miliacin in the oil, was recovered from a longer process of separating the oil into neutral, acidic and polar fractions by Solid Phase Extraction using aminopropyl bonded silica (Jacob et al., 2005). The neutral fraction was recovered by elution with  $CH_2Cl_2$ :isopropanol (9:1) and further separated on activated silica (120 °C during 24 h) deactivated with 5% weight water. After elution of aliphatic hydrocarbons, aromatics, esters and ether, and ketones, the alcohol fraction that contains germanicol was eluted with 2 ml of hexane:ethyl acetate (85:15) and then 2 ml of hexane:ethyl acetate (80:20) as described by Jacob et al. (2008b). To eliminate the large amounts of sitosterol that coelutes with germanicol, we then separated the alcohol fraction on activated silica by splitting the alcohol fraction into sub-fractions of 500  $\mu$ l of hexane:ethyl acetate (85:15). After GC/MS analysis of an acetylated aliquot of these sub-fractions, we combined fractions containing germanicol.

### 3.3. Miliacin and germanicol $\delta D$ values by GC-irMS

Germanicol was acetylated into germanicol acetate with acetic anhydride and pyridine (1:1) during 1h at 80°C prior to its GC-irMS analysis. Acetic anhydride with an offline-determined  $\delta D$  value of  $-133 \pm 2$  ‰ was acquired from A. Schimmelmann (Indiana University, USA). After drying under nitrogen, the fraction was dissolved in isooctane prior to analysis by Gas Chromatography-isotope ratio Mass Spectrometry (GC-irMS) in order to determine its  $\delta D$  value. Miliacin from the 6 g



219 in-house stock was also dissolved in isooctane prior to GC-irMS analysis.

220  $\delta D$  values were determined on a TraceGC chromatograph equipped with a Triplus  
 221 Autosampler coupled to a DeltaV Advantage isotope ratio mass spectrometer through a GC-Isolink  
 222 interface and a Conflo IV dilution system (all from Thermo Scientific, Bremen, Germany). The gas  
 223 chromatograph was fitted with a TG5-MS capillary column (60 m, 0.25 mm i.d., 0.25  $\mu\text{m}$  film  
 224 thickness; ThermoScientific). The temperature of the column was held at 80°C for 1 min, and then  
 225 increased from 80 to 300°C at 20°C.min<sup>-1</sup>, with a final isothermal hold at 300°C for 30 min. Precision  
 226 and accuracy were determined by using a mixture of 15 *n*-alkanes (*n*-C<sub>16</sub> to *n*-C<sub>30</sub>) with  $\delta D$  values  
 227 ranging from -9 to -254 ‰ determined off-line (A5 mixture, A. Schimmelmann, Indiana University,  
 228 USA) and allowed us to normalize the  $\delta D$  values of miliacin ( $\delta D_{\text{miliacin}}$ ) and of germanicol ( $\delta D_{\text{germanicol}}$ )  
 229 to the V-SMOW isotopic scale. The measured  $\delta D$  values for the standard *n*-alkanes are in good  
 230 agreement with those measured off-line ( $r^2 = 0.987$ ). The overall precision for the *n*-alkanes standards  
 231 was  $\leq 4\text{‰}$ . Samples were injected at least three times and were accurate to  $\pm 5\text{‰}$  for miliacin and  
 232 germanicol. The  $\delta D$  value of germanicol acetate was mathematically corrected from the contribution  
 233 of acetic anhydride to obtain the  $\delta D_{\text{germanicol}}$  value. Our results are shown as both (D/H) ratios in ppm  
 234 and  $\delta D$  values in ‰, calculated using a (D/H) value of V-SMOW of  $155.76 \pm 0.05$  ppm, and  
 235 converting from ppm values to ‰ values using Equation 2:

$$236 \quad \delta D = \left( \frac{(D/H)_i}{(D/H)_{\text{V-SMOW}}} - 1 \right) \times 1000 \quad (2)$$

### 237 3.4. Isotropic <sup>1</sup>H and <sup>13</sup>C NMR of miliacin

239 The full assignment of the proton and carbon-13 signals recorded in liquid solvents (CDCl<sub>3</sub>)  
 240 was achieved using homonuclear <sup>1</sup>H-<sup>1</sup>H 2D COSY, heteronuclear <sup>13</sup>C-<sup>1</sup>H HMBC and HMQC 2D  
 241 experiments combined with the analysis of <sup>13</sup>C-<sup>13</sup>C ADEQUATE 2D spectrum. Several isotropic  
 242 NMR spectra of miliacin are given in **Figs. S-1, S-2 and S-3** in **Appendix A.II**. All isotropic NMR  
 243 data were recorded on a Bruker 9.4 T (400 MHz) Advance I spectrometer equipped with 5 mm o.d.  
 244 BBI probe.

### 245 3.5. Preparation of polypeptide oriented samples

246 The ability to solubilize a given analyte but also to form a polypeptidic homogeneous liquid-  
 247 crystalline phase is an important factor that depends on what organic co-solvent is chosen. We were  
 248 best able to solubilize miliacin using 30 mg of miliacin, 80 mg of poly- $\gamma$ -benzyl-*L*-glutamate (PBLG)  
 249 with a degree of polymerization of 743 (MW (viscosity) = 162 900 g/mol) and 810 mg of  
 250 chloroform, corresponding to 8.7% w/w (PBLG) and 3.2 ‰ w/w (miliacin). Over this latter mass  
 251 ratio, miliacin is not totally dissolved in the mesophase, and solid particles are visible, even at higher  
 252 temperatures. Other polar and apolar organic solvents that we experimented and that proved less

253 satisfactory include  $\text{CH}_2\text{Cl}_2$ , tetrachloroethane, THF, dioxane, pyridine. For all of them, the solubility  
254 of miliacin was found to be extremely low. Note that even DMSO (incompatible with PBLG) is not  
255 suitable.

256 As the polypeptide is chiral, oriented solutions of PBLG is able to discriminate between NMR  
257 signals of enantiomers of wide range of organic molecules but also enantiotopic directions in  
258 prochiral compounds (Lesot et al., 2000, 2009). Nevertheless, in this work, PBLG was selected not  
259 for its enantiodiscriminating properties, which are useless for miliacin (enantiopure biocompound), but  
260 for its ability to homogeneously orient a guest molecule with a rather low degree of alignment (Lesot  
261 et al., 2000).

262 Due to the high molecular weight ( $\text{MW} = 440.7 \text{ g/mol}$ ) and the small amount of miliacin that  
263 can be solubilized in the PBLG/ $\text{CHCl}_3$  mesophase (only 30 mg), the molar amount in  
264 monodeuterated isotopomers is equal to  $3.17 \times 10^{-8} \text{ mol}$  for methyl groups and  $1.06 \times 10^{-8} \text{ mol}$  for  
265 methines, if  $(\text{D/H}) = 1.55 \times 10^{-2}\%$ . Assuming that density of the mixture is equal to 1, the  
266 concentration of isotopomers in the sample at methyl and methine sites is evaluated at about 3.4  
267  $\mu\text{mol/l}$  and 1.14  $\mu\text{mol/l}$ , respectively. This isotopomeric concentration is weak (30 fold less)  
268 compared to the value of chloroform. However recording anisotropic NAD spectra of solute with  
269 rather short recycling delays ( $T_R = 1 \text{ sec}$ ) allows to significantly reduce the intensity of  $^2\text{H}$  signal of  
270  $\text{CDCl}_3$  (and avoid a bad digitization of solute signals) whereas quantitative conditions ( $T_R = 5T_1(^2\text{H})$ )  
271 were maintained to record the  $^2\text{H}$  signals of miliacin (Lesot et al., 2008).

### 272 273 **3.6. Anisotropic NAD 2D-NMR in LC**

274 All anisotropic NAD 2D-NMR spectra were recorded with a 14.1 T Bruker Advance II  
275 spectrometer (92.1 MHz for deuterium nuclei) equipped with a 5 mm o.d. selective  $^2\text{H}$  cryogenic  
276 probe. NAD  $Q$ -resolved Fz 2D experiments combined with proton Waltz-16 decoupling sequence  
277 was chosen to analyze of NAD NMR spectra of miliacin dissolved in PBLG/ $\text{CHCl}_3$  (Lafon et al.,  
278 2009). In this experiment, the  $^2\text{H}$  chemical shifts are refocused in the  $F_1$  dimension. After a double  
279 FT, the 2D map can be tilted (see the principle on **Figure S-5** in **Appendix A.II**), and then  
280 symmetrized, thereby spreading the  $^2\text{H}$  chemical shifts,  $\delta(^2\text{H})$ , and quadrupolar splittings,  $\Delta\nu_Q(^2\text{H})$   
281 information in the  $F_2$  (horizontal axis) and  $F_1$  (vertical axis) dimensions, respectively. Thus a series  
282 of  $^2\text{H}$  singlets are displayed in  $F_2$  whilst a series of independent  $^2\text{H}$  doublets are observed in  $F_1$ .

283 The anisotropic  $Q$ -resolved Fz 2D spectra were recorded using a 2D matrix of 2652 ( $t_1$ )  $\times$  300  
284 ( $t_2$ ) data points, then zero-filled to 4096 ( $t_1$ )  $\times$  1024 ( $t_2$ ) data points *prior* the double FT. 256 scans  
285 were added at each  $t_1$  increment. A lorentzian filtering was applied in both dimensions ( $\text{LB} = 1.5 \text{ Hz}$ ).  
286 The recycling delay ( $T_R$ ) of anisotropic NAD 2D experiments was set to 1 sec, leading to a total  
287 acquisition time of 24 hours. Such a delay is large enough to record the anisotropic NAD spectrum

288 of miliacin with quantitative conditions and reliable evaluation of peak integrals (on 1D spectra) or  
 289 2D volume on 2D spectra of miliacin is then possible (Lesot et al., 2008). “Integral of two  
 290 components of NAD doublets measured on 1D traces extracted to 2D maps (see below and sup info)  
 291 were determined by peak deconvolution (Top-spin 3.0 software)“. To improve the accuracy of the  
 292 measure of (D/H) ratios and evaluate the standard deviation, six identical NAD 2D-NMR  
 293 experiments on the same NMR sample of miliacin were recorded and analysed.

### 295 3.7. SNIF-NMR™ protocol and experimental details

296 According to the SNIF-NMR protocol, isotropic NAD 1D-NMR spectra of analytes are  
 297 recorded under quantitative conditions ( $T_R = 5T_1$ ) with internal isotopic reference whose D/H ratio is  
 298 calibrated. If the mass of compounds (analyte and reference) is accurately weighted, the site-specific  
 299 (D/H)<sub>i</sub> ratios (expressed in ppm) can be calculated from evaluation of peaks surface on NMR spectra  
 300 as follows (Eq. 3; Lesot et al., 2011; Martin et al., 1986):

$$(D/H)_i^{anal} = \left( \frac{A_i^{anal}}{A_i^{ref}} \right) \times \left( \frac{P_i^{ref} \times m^{ref} \times M^{anal}}{P_i^{anal} \times m^{anal} \times M^{ref}} \right) \times (D/H)_i^{ref} \quad (3)$$

301 In Eq. 3,  $A_i^{anal}$  and  $A_i^{ref}$  are the areas of the signals at site *i* for the analyte and the reference,  $P_i^{anal}$  and  
 302  $P_i^{ref}$  are their stoichiometric numbers of hydrogen at site *i*,  $M^{anal}$ ,  $M^{ref}$ ,  $m^{anal}$  and  $m^{ref}$  are the molecular  
 303 weights and masses of the analyte and the reference, respectively. Finally,  $(D/H)_i^{ref}$  is the isotope  
 304 ratio of reference (at site *i*).

305 From the experimental conditions viewpoint, the SNIF-NMR 1D spectrum of miliacin was  
 306 recorded in chloroform in the presence of isotopically calibrated tetramethyl urea (TMU) as internal  
 307 chemical reference ((D/H)<sub>TMU</sub> = 84.4 ± 2 ppm, Lot 0116XAC from Eurofins). The temperature of  
 308 305 K used provided a good compromise between a reasonable experimental time for the NAD 1D  
 309 experiments (< 15 h) and a full solubilization of miliacin. At 305 K, the  $T_1(^2H)$  of TMU was  
 310 measured at 1.17 sec (average of three NAD inversion-recovery 1D-NMR experiments), leading to a  
 311 recycling delay,  $T_R$ , of 6 sec (with 3.5 sec of acquisition time) and in turns to a total acquisition time  
 312 of 15 hours for 9000 scans added. Six experiments were recorded with 6k data points (zerofilled to  
 313 16K), filtered by an exponential function (1 Hz) prior FT and peak integral for Me31 and TMU  
 314 signal were quantified by peak deconvolution after a baseline correction.

315 The mixture was prepared in a small flask tube by adding 39.9 mg of miliacin and 18.6 mg of  
 316 TMU (MW = 116.16 g.mol<sup>-1</sup>) and 1.640 g of dry CHCl<sub>3</sub>. The flask was then actively mixed (by  
 317 strongly shaking) prior to a part of mixture (≈ 700 mg; this amount is optimized for the length of the  
 318 coil in cryogenic probe) was transferred into the 5-mm NMR tube. The tube has been fire-sealed to  
 319 avoid long-term solvent evaporation. Due to the low solubility of miliacin in CHCl<sub>3</sub>, the sample was  
 320

321 prepared with larger amounts of components in order to reduce the error on the weightings of small  
 322 masses. Also, the mass of TMU used (twelve deuterium equivalent sites) was calculated to reach a  
 323 ratio in peak intensity of TMU and Me31 (three deuterium equivalent sites) equal to about 4, thus  
 324 optimizing the S/N ratio of TMU.

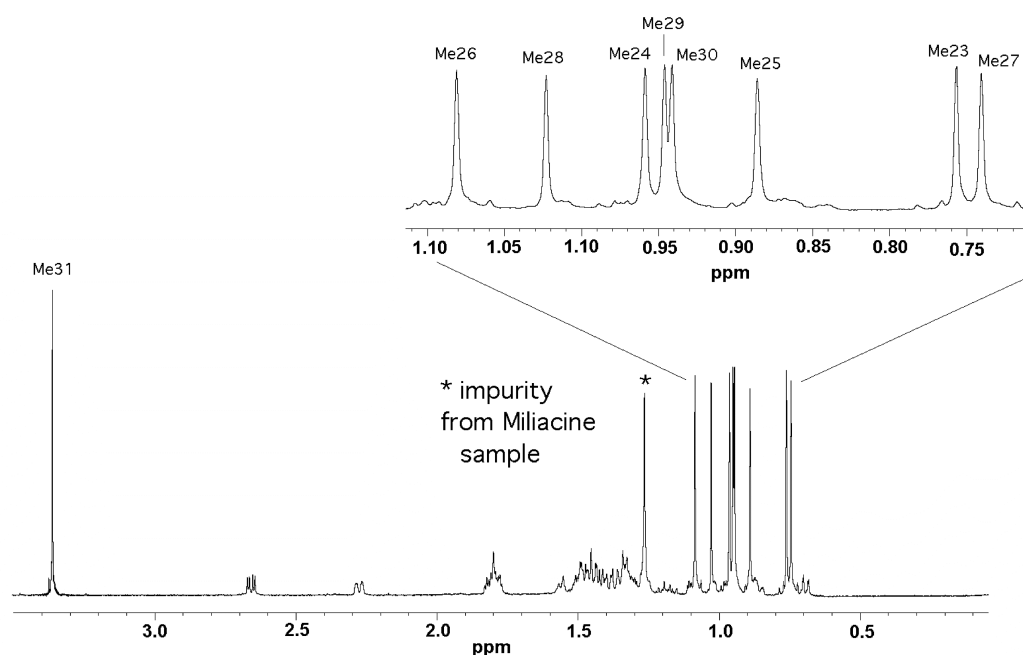
325

## 326 4. RESULTS

327

### 328 4.1. Spectral assignment of isotropic $^1\text{H}$ and $^{13}\text{C}$ resonances

329 In order to avoid any misinterpretation of the isotopic profile of methyl groups on miliacin,  
 330 we re-examined the assignment of  $^1\text{H}$  signals in combination with  $^{13}\text{C}$  signals before analyzing NAD  
 331 2D spectra in oriented solvent and then evaluating the site-specific isotope distribution. **Fig. 2** gives  
 332 the 600 MHz  $^1\text{H}$  spectrum of miliacin dissolved in chloroform at 295 K. Due to the presence of  
 333 several stereogenic centers (asymmetric carbons), the geminal protons in  $\text{CH}_2$  groups and the *gem*-  
 334 dimethyl groups are diastereotopic, and hence yield distinct  $^1\text{H}$  (and also  $^2\text{H}$ ) signals. The proton-  
 335 decoupled carbon-13 1D spectrum ( $^{13}\text{C}\{-^1\text{H}\}$ ) as well as  $^{13}\text{C}/^1\text{H}$  heteronuclear HSQC and HMBC 2D  
 336 spectra of miliacin dissolved in chloroform at 295 K and recorded at 9.4 T (100 MHz) (**Fig. S-1**)  
 337 (**Figs. S-2 and S-3**) are presented in **Appendix A.II**. The combined analysis of all spectral data  
 338 allows an unambiguous assignment of all  $^{13}\text{C}$  and  $^1\text{H}$  resonances. The full assignment of  $^{13}\text{C}$  and  $^1\text{H}$   
 339 signals is listed in **Tables S-1, S-2 and S-3** in **Appendix A.I**. Note here that our assignment agrees  
 340 with Smetanina et al. (2001), but the two ambiguities pointed out by these authors have been  
 341 eliminated (**Fig. S-5** in **Appendix A.II**).



342

343 **Figure 2:** 600.1 MHz proton 1D-NMR spectrum of miliacin in  $\text{CDCl}_3$  (295 K). The assignment of  
 344 resonances of methyl groups is given in **Table S-1, S-2 and S-3** in **Appendix A.I** of **Supp.**  
 345 **Info.**). The peak marked by an asterisk ( $\delta = 1.24$  ppm) is an unidentified impurity.

346

347 **4.2. Analysis of anisotropic NAD signals of methyl groups**

348 Miliacin is a very illustrative example of the analytical performance of anisotropic NAD 2D-  
 349 NMR to determine the intramolecular site-specific isotope distribution. Indeed, on the isotropic NAD  
 350 1D-NMR spectrum of miliacin, the resonances of methyl groups, Me<sub>24</sub>, Me<sub>29</sub> and Me<sub>30</sub> as well as  
 351 Me<sub>23</sub> and Me<sub>27</sub> are not spectrally resolved (even at 92.1 MHz) (see **Fig. S-4** in **Appendix A.I of**  
 352 **Supp. Info.**). As a consequence, only two broad <sup>2</sup>H resonances are observed, thus preventing a  
 353 specific determination of their respective peak areas, and consequently the associated (D/H) ratios at  
 354 each individual site.

355 **Fig. 3a** shows the shielded region of methyl groups (Me<sub>23</sub>- Me<sub>30</sub>) of the tilted *Q*-resolved Fz  
 356 2D map of miliacin recorded in PBLG/CHCl<sub>3</sub> at 295 K (see **Fig. S-7, Appendix A.II of Supp. Info.**  
 357 for the full map). As expected, all methyl signals are now separated on the basis of their  $\delta(^2\text{H})$   
 358 and  $\Delta\nu_Q(^2\text{H})$ . The sixteen <sup>2</sup>H resonances (8 quadrupolar doublets) detected on the 2D map correspond  
 359 to eight inequivalent methyl sites, Me<sub>23</sub>-Me<sub>30</sub>; each of them giving one integrated signal of three <sup>2</sup>H  
 360 equivalent sites. As expected, almost all QDs of methylene and methine ethylene groups of miliacin  
 361 do not clearly emerge from the spectral noise (baseline) and cannot be used reliably for quantitative  
 362 purposes. In contrast, the doublets of all methyl groups are detected on 2D map with S/N ratios  
 363 varying between 8 (Me<sub>30</sub>) and 25 (Me<sub>26</sub>) whereas a value of 35 is found for the methyl of methoxy  
 364 group (Me<sub>31</sub>) (**Table 1**). Due to the small amount of miliacin solubilized in the mesophase and its  
 365 high MW, the S/N ratios are unusually low. However a reliable (acceptable) evaluation of peak areas  
 366 leading to reliable (D/H) ratios is still possible.

367

368 **Table 1:** Spectral data associated to anisotropic NAD signals of methyl groups  
 369 of miliacin at 295 K.

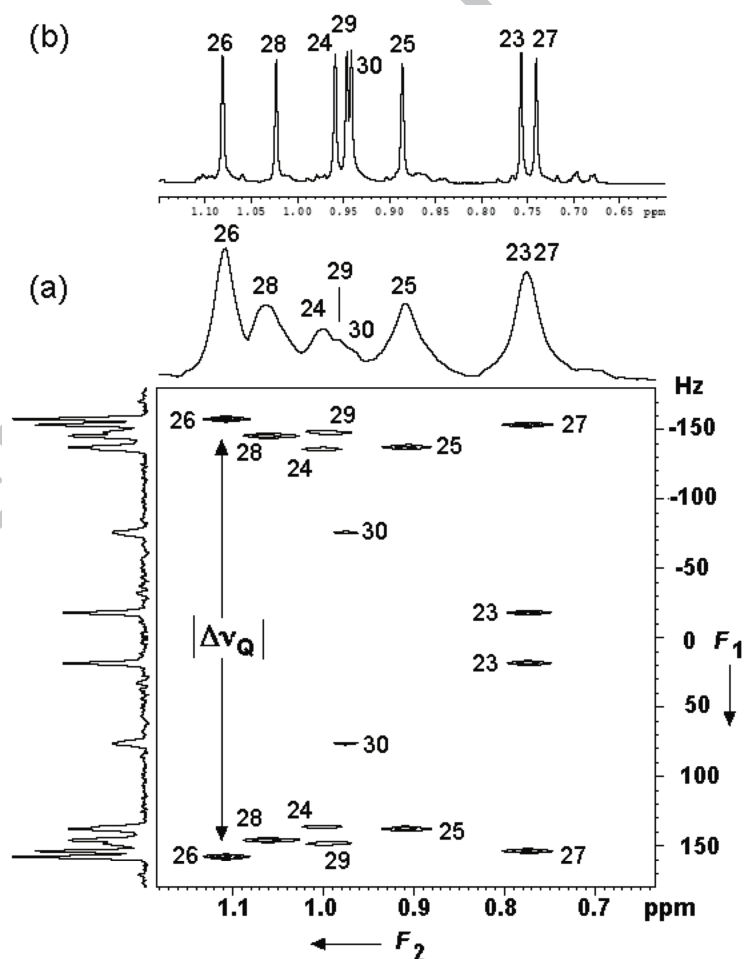
| Methyl group      | Ring | $\delta(^2\text{H})^a$<br>(ppm) | $\Delta_{1/2}$<br>(Hz) | S/N                 | $ \Delta\nu_Q $<br>(Hz) |
|-------------------|------|---------------------------------|------------------------|---------------------|-------------------------|
| Me <sub>27</sub>  | C    | 0.771 (0.746)                   | 2.4                    | 24                  | 309                     |
| Me <sub>23</sub>  | A    | 0.774 (0.763)                   | 2.8                    | 22                  | 36                      |
| Me <sub>25</sub>  | A    | 0.905 (0.891)                   | 3.3                    | 12                  | 275                     |
| Me <sub>30</sub>  | E    | 0.978 (0.948)                   | 3.6                    | 8                   | 272                     |
| Me <sub>29</sub>  | E    | 0.999 (0.952)                   | 5.6                    | 9                   | 152                     |
| Me <sub>24</sub>  | A    | 0.982 (0.965)                   | 3.3                    | 9                   | 296                     |
| Me <sub>28</sub>  | E    | 1.057 (1.029)                   | 3.3                    | 17                  | 292                     |
| Me <sub>26</sub>  | C    | 1.105 (1.086)                   | 2.3                    | 25                  | 314                     |
| Me <sub>31</sub>  | A    | 3.399 (3.368)                   | 2.1                    | 35                  | 193                     |
| CHCl <sub>3</sub> | —    | 7.270 (7.270)                   | 1.8                    | 3000 <sup>(b)</sup> | 273                     |

370 <sup>(a)</sup>  $\delta(^2\text{H})$  measured on the  $F_2$  projection of anisotropic NAD 2D spectrum ( $\delta_{\text{chloroform}} = 7.27$  ppm).

371 In parenthesis are reported the  $\delta(^1\text{H})$  measured on the isotropic <sup>1</sup>H 1D-NMR spectrum ( $\delta_{\text{chloroform}} = 7.27$  ppm).

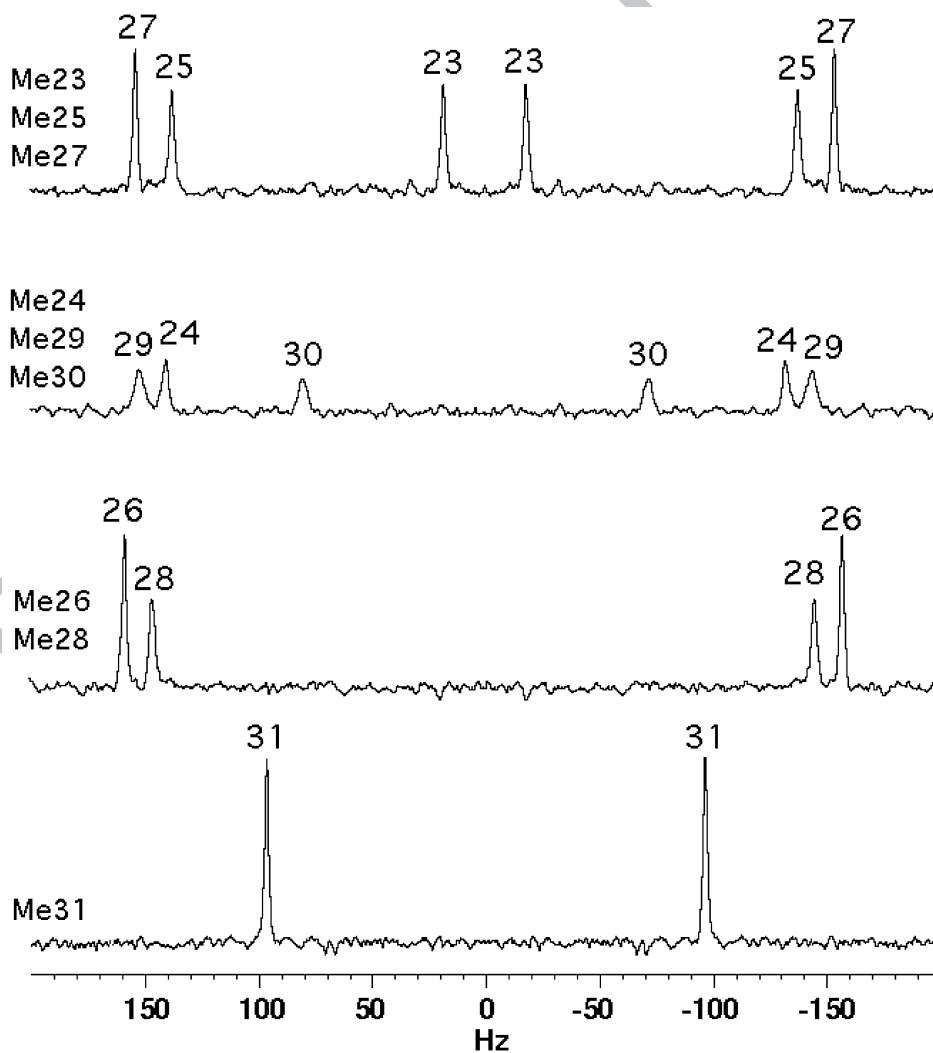
372 <sup>(b)</sup> The large difference between the S/N ratios of Me groups and CHCl<sub>3</sub> originates from the  
 373 difference of deuterated isotopomer concentrations and the linewidths of <sup>2</sup>H signals.

374 On the tilted 2D map, it becomes possible to extract a single or several columns (projection of  
 375 columns) associated to a single or several  $^2\text{H}$  site(s) (Lesot et al., 2008). Four NAD 1D sub-spectra  
 376 (containing the signal associated to one to four  $^2\text{H}$  sites) were selected (**Fig. 4**) in order to separate  
 377 reliably all quadrupolar doublets (regions selected are shown in **Fig. S-8** in **Appendix A.II. of Supp.**  
 378 **Info.**) **Table 1** lists the  $^2\text{H}$  quadrupolar splittings in Hz (distance in Hz between two components of  
 379 each  $^2\text{H}$  doublet). As seen, the  $\Delta\nu_Q$ 's (in absolute value) varies from 36 to 314 Hz, but the majority of  
 380 doublets (6 / 9) range from 272 to 314 Hz. As expected by using 2D NMR approach, the NAD  
 381 signals associated to each methyl group of miliacin are spectrally distributed (on the 2D map) on the  
 382 basis of their  $^2\text{H}$  chemical shifts and quadrupolar splittings. Hence it becomes possible to evaluate the  
 383 areas of each doublet by peak deconvolution. For the same molecule, these areas are proportional to  
 384 the number of monodeuterated isotopomers contributing to the NAD signal (three for a methyl  
 385 group) and the  $(\text{D}/\text{H})_i^{\text{Mc}}$  ratio at methyl site  $i$  (results, discussion in **Section 4.4**).  
 386

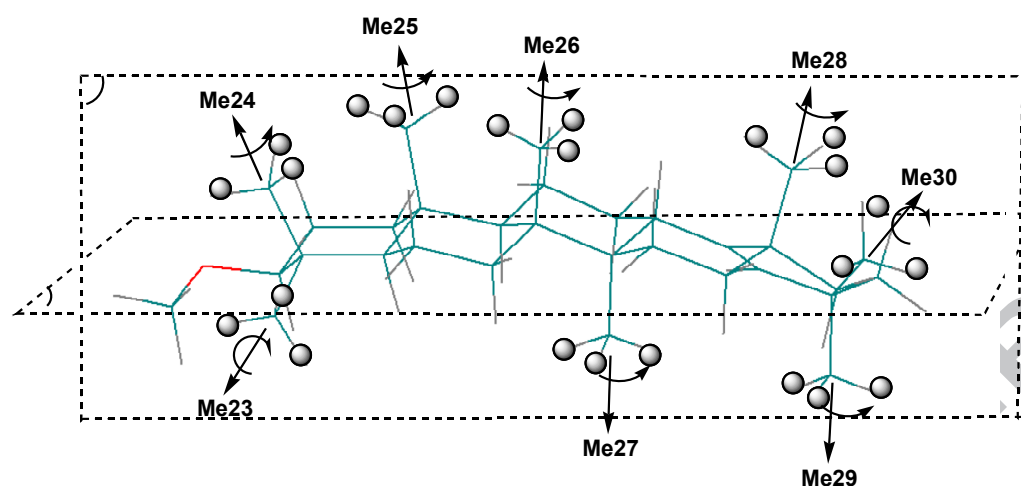


387  
 388 **Figure 3:** (a) Methyl region (Me31 not shown) of the 92.1 MHz NAD tilted  $Q$ -resolved  $F_z$  2D map  
 389 of miliacin dissolved in PBLG/ $\text{CHCl}_3$  and recorded at 295 K. (b) 600 MHz proton 1D  
 390 spectrum of miliacin in  $\text{CHCl}_3$  plot at the horizontal same scale (in ppm). A lorentzian filtering  
 391 (LB = 1.5 Hz) has been applied in both spectral dimensions. On the  $F_1$  and  $F_2$  axes are shown  
 392 the spectral 2D projections. Note the significant variations of peak intensity of  $^2\text{H}$  doublets in  
 393  $F_1$  dimension illustrating the enrichment/depletion in deuterons.

394  
 395 As far as possible, the assignment of  $^2\text{H}$  resonances in  $F_2$  dimension on the  $Q$ -resolved  $F_z$  map  
 396 of miliacin can be derived from the position of  $^1\text{H}$  lines of methyl groups observed the isotropic  $^1\text{H}$   
 397 spectrum (see **Fig. 3**). However when ambiguities exist (as for sites 23 and 27), further arguments  
 398 based on the orientation of the C- $\text{CH}_3$  axis (see **Fig. 5**) can help to assign the quadrupolar doublets on  
 399 the basis of their magnitude. Indeed as the quadrupolar splittings are proportional to the order  
 400 parameter,  $S_{\text{C-D}}$ , of the C-D bond (see **Eq. 1**), they can be related to the relative positions of  
 401 associated C-D vectors in a (rigid) molecular structure (Aroulanda et al., 2001; Lesot et al., 2015).  
 402 We can therefore correlate the magnitude of  $^2\text{H}$  doublets to geometrical data, thus providing  
 403 complementary information for their assignment. This approach was typically useful to assign the  
 404 quadrupolar doublets associated to Me23 and Me27, for instance. Detailed explanations and further  
 405 discussions on the assignment of doublets are given in **Appendix A.III** of **Supp. Info**.  
 406



407  
 408 **Figure 4:** Series of four NAD 1D sub-spectra extracted from the tilted  $Q$ -resolved  $F_z$  2D map  
 409 showing the nine methyl groups of miliacin along with their assignments (see also **Fig. S-9**).  
 410 Each sub-spectrum is plotted with the same amplitude of noise. Note the variation of peak  
 411 intensity of  $^2\text{H}$  doublets. The variations observed reveal the differences of (D/H) for each site.  
 412



**Figure 5:** 3D view of the minimal energy conformation of miliacin calculated with Hyperchem 3.0 software. A semi-empirical quantum chemistry methods based on the Hartree–Fock formalism (Austin Model 1 or AM1) has been used for modeling the molecule. The minimal energy was found at -8177.99 kJ/mol. Except for ring E distorted by the double bond, all other cycles have adopted a chair conformation. Note the difference of orientation axes of C-CH<sub>3</sub> vectors for Me23 and Me30 (close to the median horizontal molecular plan) compared to Me24 to Me29 (see also **Fig. S-9** in **Appendix A.II** of **Supp. Info.**).

#### 4.3. Absolute (D/H) and $\delta D$ values of methyl groups

In contrast to their <sup>1</sup>H signals (**Fig. 2**), significant variations in the intensity of <sup>2</sup>H doublets for the nine methyl groups are observed on the 2D map or the associated  $F_1$  and  $F_2$  projections (**Figs. 3** and **4**). As NAD 2D-NMR spectra have been recorded with quantitative experimental conditions, the peak areas on projections (or peak volume on the 2D map) are in principle proportional to the number of isotopomers contributing to the NAD signal (three for a methyl group) and the isotope ratio (D/H) at each methyl site.

##### 4.3.1. Evaluation of relative <sup>2</sup>H peak surface by anisotropic NAD 2D-NMR

When the direct quantification of peak surfaces on anisotropic 2D spectra (using an isotopically calibrated, internal chemical reference) is experimentally difficult as in case of miliacin (problem with mesophase, interference between the <sup>2</sup>H solute and reference signals), other strategies must be proposed. Thus the use of an electronic reference signal (ERETIC method) is a possible alternative but required special equipment not implemented in the majority of spectrometers (Akoka et al., 1999). For the study of miliacin, we have adopted a two-step protocol for quantification of relative D concentrations in methyl groups (or any other sites) consisting into: i) evaluating the relative variation of <sup>2</sup>H surfaces of each doublet using the NAD signal of methyl group of methoxy (Me31) as internal reference value; ii) determining the (D/H) and  $\delta D$  values of Me31 using either GC-irMS (see **Section 4.3.2**) or isotropic quantitative NAD NMR in presence of TMU (see **Section 4.3.3**). Note that the choice of the Me31 as internal reference is advantageous for both methods. From GC-irMS viewpoint, the (D/H)<sub>Me31</sub> ratio can be easily quantified by comparing the global  $\delta D$  of miliacin and the one of germanicol. From an NMR viewpoint, the isotropic <sup>2</sup>H signal of Me31 is



443 isolated from other signals of miliacin, reducing errors on the measure of peak surface by  
 444 deconvolution, and does not interfere with the signal of TMU.

445 The measurement of relative surface of peaks can be either directly performed on the 2D map  
 446 or using 1D projections of selected slices extracted from the map (Massou et al., 2007; Lesot et al.,  
 447 2008). Adopting this second approach, the evaluation of relative surface of doublet of methyl groups  
 448 was performed as following: i) selection of 10 columns centered on the NAD signal of each  
 449 inequivalent  $^2\text{H}$  site and calculation of their respective  $F_1$  projection, ii) addition of NAD signal of  
 450 Me31 (10 columns) to each NAD 1D sub-spectrum; iii) determination (by signal deconvolution) of  
 451 the peak surface for each methyl group, iv) calculation of the average value obtained for the six  
 452 different NAD 2D spectra recorded with the same experimental conditions (acquisition and  
 453 processing), assuming a reference value for the Me31 signal. In order to determine the reliability of  
 454 this approach, we have compared the peak area of Me31 for the six experiments. The variation  
 455 between each experiment is below 2%. In column 3 of **Table 2** are listed the averaged surfaces  
 456 measured for the six NMR experiments, when the value for the Me31 is arbitrary set at 100. Values  
 457 for each experiment are given in **Table S-4** and plotted in **Fig. S-10** whereas **Fig. S-11** in **Appendix**  
 458 **A.II of Supp. Info.** displays the associated average values together with the SD.

459

#### 460 4.3.2. $\delta D_{\text{Me31}}$ determined by GC-irMS and SNIF-NMR<sup>TM</sup>

461  $\delta D_{\text{Me31}}$  was first estimated by comparing the GC-irMS  $\delta D$  values of miliacin ( $\delta D_{\text{miliacin}}$ ) and  
 462 of germanicol ( $\delta D_{\text{germanicol}}$ ) extracted from millet oil according to the following rationale and  
 463 equations. We assumed no other hydrogen isotopic difference between miliacin and germanicol (see  
 464 **Fig. 1**) than the addition of three hydrogen atoms from the Me31.

465 Experimentally, we obtained  $\delta D_{\text{miliacin}} = -100 \pm 5 \text{ ‰}$  and  $\delta D_{\text{germanicolacetate}} = -110 \pm 5 \text{ ‰}$ , which  
 466 gives  $\delta D_{\text{germanicol}} = -108.6 \pm 5 \text{ ‰}$ , after mathematically removing the contribution from acetic  
 467 anhydride ( $\delta D_{\text{acanh}} = -133 \pm 2 \text{ ‰}$ ) according to **Eq. 4**:

468

$$469 \quad \delta D_{\text{germanicol}} = \left( \frac{52 \times \delta D_{\text{germanicolacetate}} - 3 \times \delta D_{\text{acanh}}}{49} \right). \quad (4)$$

470 The error on  $\delta D_{\text{germanicol}}$  was classically calculated as  $\sqrt{(52/49)^2 \times (\delta D_{\text{germanicolacetate}})^2 + (3/49)^2 \times (\delta D_{\text{acanh}})^2}$ , thus  
 471 taking into account the multiplicative factors of **Eq. 4** and assuming that the correlative term is null  
 472 (independent variables). Note here that this  $\delta D_{\text{germanicol}}$  value does not take into account the H atom of  
 473 the alcohol function that is highly exchangeable and was lost during acetylation. Hence, when  
 474 calculating  $\delta D_{\text{Me31}}$  in mass balance equations **Eq. 4** and **Eq. 5**, we consider 49 H atoms in germanicol  
 475 (instead of 50) and 52 H atoms in miliacin.

476

$$\delta D_{Me31} = \left( \frac{52 \times \delta D_{miliacin} - 49 \times \delta D_{germanicol}}{3} \right). \quad (5)$$

477  
478  
479 The calculation in **Eq. 5** gives  $\delta D_{Me31}$  equal to  $+41 \text{ ‰} \pm 119 \text{ ‰}$ . Propagation of uncertainties on the  
480  $\delta D_{Me31}$  was calculated as for **Eq. 4** (see **Table 2**, footnote i). Using **Eq. 2**, this corresponds to a  
481  $(D/H)_{Me31}^{GC-irMS}$  value of  $162 \pm 19$  ppm. Details for the conversion from  $\delta D$  to  $(D/H)$  are given in footnote g  
482 of **Table 2**.

483 A possible experimental alternative to GC-irMS consists to determine  $\delta D_{Me31}$  value applying  
484 the so-called SNIF-NMR<sup>TM</sup> protocol using NAD NMR in isotropic solvent in presence of calibrated  
485 TMU (see **Section 3.7**). The analysis of experiments recorded indicates that standard deviation on the  
486 area ratio ( $R = A_{Me31}/A_{TMU}$ ) was found to be equal to 0.28 (e.g.  $R = 25.2 \pm 0.6$ ), thus indicating a very  
487 good repeatability of isotropic NAD 1D NMR experiments (see Table S-5). Knowing the  $(D/H)_{TMU}$   
488 and the masses of TMU and miliacin (purity of 98%) used, the  $(D/H)_{Me31}$  was obtained using **Eq. 3**;  
489 the average value was found to be equal to 154 ppm. The uncertainty value on  $(D/H)_{Me31}$  was  
490 classically calculated as  $\Delta X = X \times \sqrt{\sum_i (\Delta g_i / g_i)^2}$  where  $g_i$  are the key variables of **Eq. 3** (see **Table 2**  
491 and **S-5**) and found to be equal to  $\pm 10$  ppm. Expressed in  $\delta D_{Me31}$  (from the **Eq. 2**), the value is equal  
492 to  $-11 \pm 64 \text{ ‰}$  (see **Table 2**, footnote e).

493 Conclusively, it is worthwhile noting that, although they were determined by two very distinct  
494 analytical methods, the  $(D/H)_{Me31} / (\delta D_{Me31})$  values are consistent within the uncertainty range ( $162 \pm$   
495  $19$  ppm *versus*  $154 \pm 10$  ppm ( $+41 \pm 119 \text{ ‰}$  *versus*  $-11 \pm 64 \text{ ‰}$ ). The difference on the nominal  
496  $(D/H)_{Me31}$  value (relative discrepancy of 5%) could be explained by the lower sensitivity of NAD  
497 NMR spectroscopy (compared to GC-irMS) in particular when samples cannot be enough  
498 concentrated in analyte.

#### 499 4.3.3. Absolute $(D/H)$ and $\delta D$ values of methyl groups 23 to 30

500 From  $(D/H)_{Me31}$  value determined by the SNIF-NMR<sup>TM</sup> protocol or GC-irMS, it becomes  
501 possible to evaluate the  $(D/H)$  value for each methyl group  $(D/H)_{Mei}$  and corresponding  $\delta D_{Mei}$  with  $i =$   
502 23 to 30) in miliacin, using relative  $(D/H)$  values previously determined using the anisotropic NAD  
503 NMR experiment (**Section 4.3.1**). All data and details of the calculation are given in **Table 2** and  
504 footnotes. A graphical comparison of the  $(D/H)_{Mei}$  and  $\delta D_{Mei}$  variations *versus* the methyl groups is  
505 proposed in **Figs 6a** and **6b**. Large variations of  $(D/H)_{Mei}$  values (up to a factor  $\approx 2$ ) exist from one  
506 site to another one. Except for Me31, the  $(D/H)_{Mei}$  values are significantly below the V-SMOW value.  
507 This deuterium depletion in lipids produced by autotrophs is expected from biosynthetic fractionation  
508 (Sessions et al., 1999). Besides for Me31, Me26 and Me27 appear to be the less depleted methyl  
509

511 **Table 2:** Values of  $(D/H)_{Me_i}$  (in ppm) and  $\delta D_{Me_i}$  (in ‰) of the methyl groups of miliacin.

| Ring | Methyl <sup>a</sup><br>group | Peak area <sup>b</sup><br>no unit | Anisotropic NMR / SNIF-NMR <sup>TM</sup> |                                     | Anisotropic NMR / GC-irMS          |                                     |
|------|------------------------------|-----------------------------------|--|-------------------------------------|------------------------------------|-------------------------------------|
|      |                              |                                   | $(D/H)_{Me_i}$ <sup>c</sup><br>ppm       | $\delta D_{Me_i}$ <sup>e</sup><br>‰ | $(D/H)_{Me_i}$ <sup>f</sup><br>ppm | $\delta D_{Me_i}$ <sup>h</sup><br>‰ |
| A    | Me <sub>31</sub>             | 100 (0.0)                         | 154 ± 10 <sup>d</sup>                    | -11 ± 64                            | 162 ± 19 <sup>g</sup>              | +41 ± 119 <sup>i</sup>              |
| A    | Me <sub>23</sub>             | 74.1 ± 13.8 (6.9)                 | 114 ± 23                                 | -268 ± 148                          | 120 ± 25                           | -229 ± 162                          |
|      | Me <sub>24</sub>             | 54.6 ± 17.6 (8.8)                 | 84 ± 28                                  | -461 ± 180                          | 88 ± 30                            | -432 ± 193                          |
|      | Me <sub>25</sub>             | 73.5 ± 14.8 (7.4)                 | 113 ± 24                                 | -274 ± 154                          | 119 ± 28                           | -235 ± 177                          |
| C    | Me <sub>26</sub>             | 90.3 ± 10.2 (5.1)                 | 139 ± 18                                 | -107 ± 122                          | 146 ± 23                           | -60 ± 151                           |
|      | Me <sub>27</sub>             | 82.4 ± 9.6 (4.8)                  | 127 ± 17                                 | -186 ± 116                          | 133 ± 22                           | -143 ± 139                          |
| E    | Me <sub>28</sub>             | 76.8 ± 14.2 (7.1)                 | 118 ± 23                                 | -242 ± 154                          | 124 ± 27                           | -201 ± 173                          |
|      | Me <sub>29</sub>             | 66.8 ± 12.0 (6.0)                 | 103 ± 20                                 | -340 ± 128                          | 108 ± 23                           | -305 ± 148                          |
|      | Me <sub>30</sub>             | 41.5 ± 10.4 (5.4)                 | 64 ± 17                                  | -590 ± 109                          | 67 ± 18                            | -568 ± 116                          |

512 a) The methyl group Me31 is used as internal reference.

513 b) Average value of peak areas of NAD signals of methyl groups measured on anisotropic NAD 2D spectra (six NMR  
514 experiments) when the methoxy group is arbitrary calibrated at 100. In parenthesis is given the value of the simple  
515 standard deviation ( $\sigma_n$ ) on peak areas calculated as:  $\sigma_n = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$  where n = 6. The uncertainty, noted  $\Delta X$ , on value X (such

516 as  $X = x \pm \Delta X$ ) is calculated (in ppm) as  $\Delta X = k \times \sigma_n$  where k = 2, namely leading to a confidence interval of 95% [ $x -$   
517  $1.96\sigma$ ;  $x + 1.96\sigma$ ].

519 c) The  $(D/H)_{Me23-Me30}$  values in ppm (methyl groups 23 to 30) are calculated as  $(D/H)_{Me23-Me30} = (\%A_{Me23-Me30}^{Aniso\ NMR}) \times (D/H)_{Me31}^{SNIF-NMR}$ ,  
520 where  $(D/H)_{Me31}^{NAD\ NMR}$  is determined by isotropic NAD 1D-NMR (i.e.  $(D/H)_{Me31}^{SNIF-NMR} = 154 \pm 10$  ppm, see Table S-4). The  
521 uncertainty on  $(D/H)_{Me23-Me30}$  values is evaluated (in ppm) as:  $\Delta(D/H)_i = (D/H)_i \times \sqrt{\left(\frac{\Delta(Area)_i}{Area_i}\right)^2 + \left(\frac{\Delta(D/H)_{Me31}^{SNIF-NMR}}{(D/H)_{Me31}^{SNIF-NMR}}\right)^2}$ .

522 d) The  $(D/H)_{Me31}^{SNIF-NMR}$  value in ‰ (methyl group 31) is experimentally determined by SNIF-NMR<sup>TM</sup> protocol in isotropic  
523 solvent. See Table S-5 for the calculation of uncertainty on  $(D/H)_{Me31}^{NAD\ NMR}$ .

524 e) The  $\delta D_{Me23-Me31}$  values in ‰ are calculated using Eq. 2. Considering that the  $(D/H)_{V-SMOW}$  value is a constant  
525 (international standard value) and not a variable, the uncertainty on  $\delta D_{Me23-Me31}$  values is simply evaluated as  
526  $\Delta\delta_{Me_i} = \sqrt{\left(1000 \times \frac{\Delta(D/H)_{Me_i}}{(D/H)_{V-SMOW}}\right)^2}$ , assuming that correlation term is null.

527 f) The  $(D/H)_{Me23-Me30}$  values in ppm are calculated as  $(D/H)_{Me23-Me30} = (\%A_{Me23-Me30}^{Aniso\ NMR}) \times (D/H)_{Me31}^{GC-irMS}$ , namely from the relative peak  
528 surfaces (%) measured on anisotropic NMR spectra (see Table S-4) and using the  $(D/H)_{Me31}^{GC-irMS}$  value (ppm) as reference.  
529 The uncertainty on  $(D/H)_{Me23-Me30}$  is evaluated (in ppm) as:  $\Delta(D/H)_i = (D/H)_i \times \sqrt{\left(\frac{\Delta(Area)_i}{Area_i}\right)^2 + \left(\frac{\Delta(D/H)_{Me31}^{GC-irMS}}{(D/H)_{Me31}^{GC-irMS}}\right)^2}$ .

530 g) The  $(D/H)_{Me31}^{GC-irMS}$  value in ppm is calculated as  $(D/H)_{Me31} = \frac{\delta D_{Me31}^{GC-irMS} \times (D/H)_{V-SMOW}}{1000} + (D/H)_{V-SMOW}$ . From the same reason evoked in  
531 footnote e, the uncertainty on  $(D/H)_{Me31}^{GC-irMS}$  is evaluated as  $\Delta(D/H)_{Me_i} = \sqrt{\left(\frac{(D/H)_{V-SMOW} \times \Delta\delta_{Me_i}}{1000}\right)^2}$ , leading to a value of  $\pm 19$  ppm.

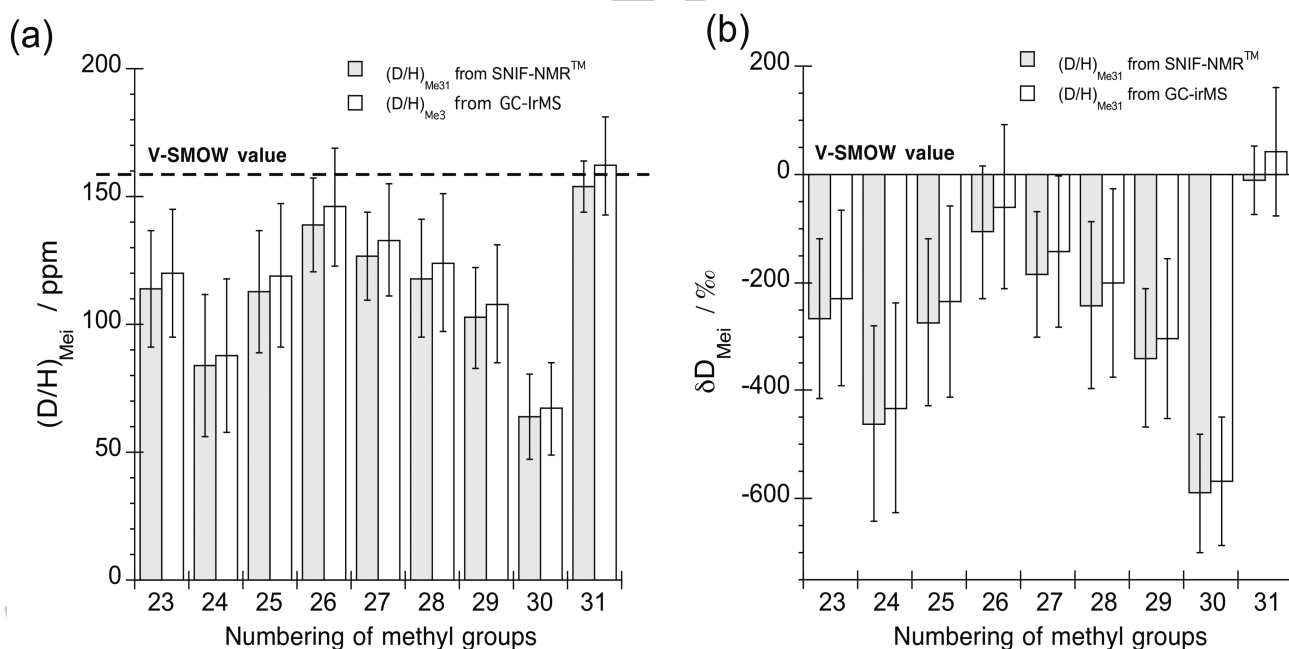
532 h) The  $\delta D_{\text{Me23-Me30}}$  values in ‰ are calculated as  $\delta D_{\text{Me23-Me30}} = (\%A_{\text{Me23-30}}^{\text{Aniso NMR}}) \times \frac{(D/H)_{\text{Me23-Me30}}^{\text{Aniso NMR}}}{(D/H)_{\text{V-SMOW}}}$ . Note that these values cannot  
 533 directly be obtained from  $\%A_{\text{Me23-30}}^{\text{Aniso NMR}}$  and  $\delta D_{\text{Me31}}^{\text{GC-irMS}}$  values due to the non-linearity of **Eq. 2**. For the same reason evoked  
 534 in footnote e, the uncertainty on  $\delta D_{\text{Mei}}$  is simply evaluated as  $\Delta \delta_{\text{Mei}} = 1000 \times \frac{\Delta(D/H)_{\text{Mei}}}{(D/H)_{\text{V-SMOW}}}$ .

535 i) The  $\delta D_{\text{Me31}}^{\text{GC-irMS}}$  value in ‰ was experimentally determined by GC-irMS. The calculation of uncertainty on  $\delta D_{\text{Me31}}^{\text{GC-irMS}}$  is  
 536 evaluated as  $\Delta \delta D_{\text{Me31}}^{\text{GC-irMS}} = \sqrt{(52/3)^2 \times (\delta D_{\text{germanicolacetate}})^2 + (49/3)^2 \times (\delta D_{\text{acanh}})^2}$  and gives  $\pm 119$  ‰ (see main text).

537  
 538 groups ( $\delta D_{\text{Mei}} \sim -100$  ‰), followed by Me23, Me25, Me28 and Me29 ( $\delta D_{\text{Mei}} \sim -250$  ‰), whereas  
 539 Me30 and Me24 are strongly depleted ( $\delta D_{\text{Mei}} \sim -500$  ‰).

540 More interesting is actually the global variation of  $(D/H)_{\text{Me23-Me30}}$  values (the exogenous methyl  
 541 31 group is disregarded), that can be modeled as a quasi symmetrical, bell curve, the biosynthetic  
 542 implication of which is discussed below. Note also that these conclusions may still hold based on the  
 543 relative variation of peak areas measured on anisotropic NAD 2D map of miliacin (see **Table S-4** and  
 544 **Figure S-12**.)”

545



546

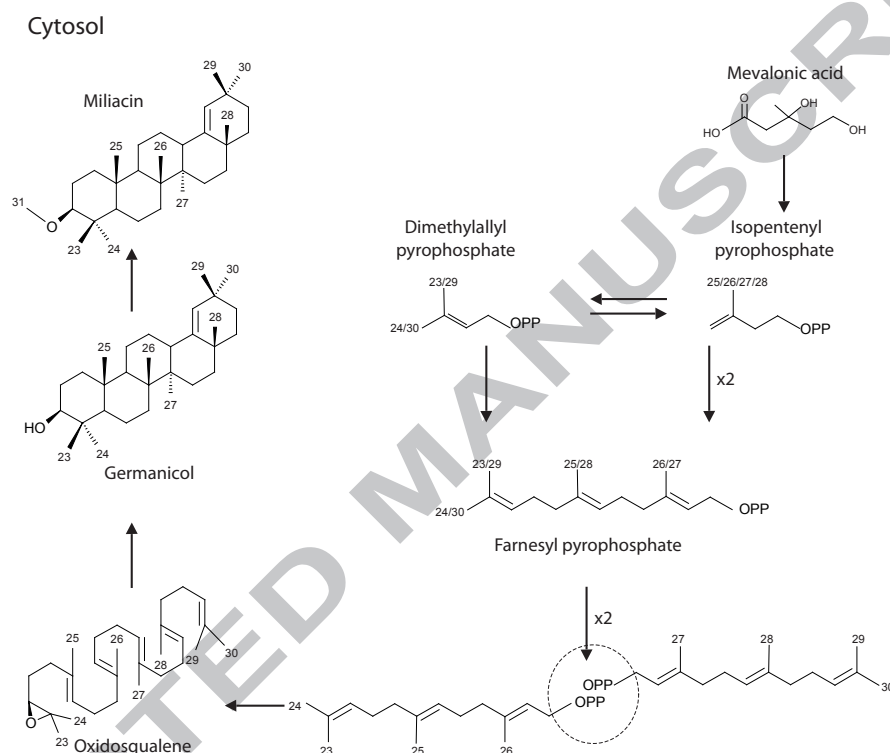
547 **Figure 6:** Variation of (a)  $(D/H)$  ratios in ppm and (b)  $\delta D$  in ‰ for each methyl group of miliacin  
 548 when  $(D/H)_{\text{Me31}}$  is evaluated by the SNIF-NMR<sup>TM</sup> ( $154 \pm 10$  ppm or  $-11 \pm 64$  ‰) and by GC-  
 549 irMS ( $162 \pm 19$  ppm or  $41 \pm 119$  ‰). Excluding the exogenous methyl group 31, note the  
 550 (inverted) quasi-symmetrical bell shape of both plots.  
 551

## 542 5. DISCUSSION

### 543 5.1. Implications for biosynthesis

544 The heterogeneous distribution of  $\delta D_{\text{Mei}}$  values in miliacin can be explained by the different  
 545 steps involved in the biosynthesis of this molecule, by tracking back the origin of methyl groups (**Fig.**

556 7). As for any C-3 oxygenated pentacyclic triterpenes in other vascular plants, miliacin is produced in  
 557 the cytosol from mevalonic acid, through the mevalonate pathway (i.e. Sachse et al., 2012).  
 558 Mevalonic acid is converted into isopentenyl pyrophosphate that can be isomerized into dimethylallyl  
 559 pyrophosphate. Then, farnesyl pyrophosphate is synthesized by the addition of two isopentenyl  
 560 pyrophosphate units and a dimethylallyl pyrophosphate unit. Oxidosqualene is synthesized by the  
 561 tail-to-tail combination of two farnesyl pyrophosphate units and is then cyclized by an oxidosqualene  
 562 cyclase.



563  
 564 **Figure 7:** Synthetic pathway of miliacin in the cytosol. Modified from Sachse et al. (2012).  
 565

566 Hence, from a biosynthetic point of view, Me24-Me30, Me23-Me29, Me25-Me28 and Me26-Me27  
 567 are isotopically equivalent because they derive from the same methyl groups in the farnesyl  
 568 pyrophosphate unit. This biosynthetic equivalence explains reasonably the symmetry and the “bell-  
 569 curve” variation of  $(D/H)_{Mei} / (\delta D_{Mei})$  values of Me23 to Me30 as observed in **Fig. 6**. Farnesyl  
 570 pyrophosphate is produced by head-to-tail combination of two isopentenyl pyrophosphate units and a  
 571 dimethylallyl unit. Thus, Me25, Me26, Me27 and Me28 in miliacin originate from the same methyl  
 572 groups in isopentenyl pyrophosphate, whereas Me23, Me29, Me24 and Me30 originate from the  
 573 dimethylallyl pyrophosphate. The dimethylallyl pyrophosphate itself results from the isomerization  
 574 of isopentenyl pyrophosphate by an isopentenyl diphosphate isomerase (Sachse et al., 2012).

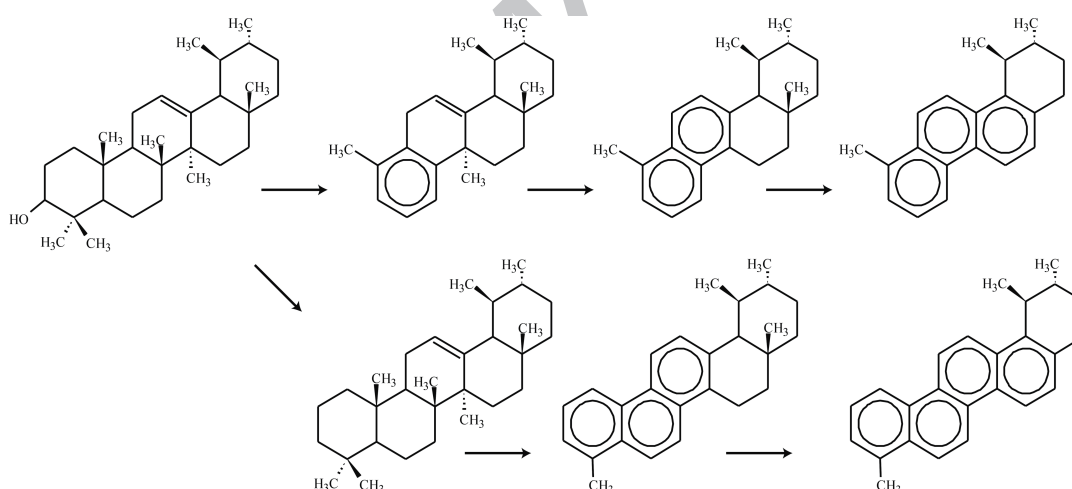
575 During this process, the formation of a carbocation affects hydrogen atoms of Me24/Me30.  
 576 This could explain their D depletion when compared to other methyl groups. As already noted, Me31  
 577 is highly enriched when compared to other methyl groups. We have currently no information on the

578 origin of this methyl group that would explain such a relative D enrichment. These results for  
 579 miliacin can be directly extrapolated to any compound with an oleanane, and potentially an ursane or  
 580 lupane structure. For compounds that originate from oleanane after successive migrations of methyl  
 581 groups (friedo rearrangements leading to structures such as taraxerane, glutinane, multiflorane and  
 582 friedelane; i.e Xu et al., 2004), additional fractionations of hydrogen can be expected.

## 583 5.2. Implications for diagenesis

584 Diagenesis of vascular plant C-3 oxygenated pentacyclic triterpenes can potentially consist of  
 585 several structural modifications that comprise loss of functional groups, ring-A degradation,  
 586 epimerisations, double bond migrations, loss or migration of methyl groups (Rullkötter et al., 1994)  
 587 and aromatization (i.e. **Fig. 8** and Wolff et al., 1989). No discernible difference in  $\delta^{13}\text{C}$  values has  
 588 been found during aromatization between biochemical precursors and diagenetic byproducts  
 589 (Freeman et al., 1994). Preliminary results in organic archaeological layers showed significant D  
 590 depletion of diagenetic derivatives compared to their potential biochemical precursors whereas no  
 591 carbon isotopic shift was recorded (Jacob et al., 2011).

592



593  
 594

595 **Figure 8:** Potential diagenetic pathways of pentacyclic triterpenes (here starting from  $\alpha$ -amyrin)  
 596 illustrating the loss of hydrogen atoms related to methyl groups, methylene/methine.  
 597 Modified from Wolff et al. (1989).

598

599 Our results have shown significant variations in  $\delta\text{D}_{\text{Mei}}$  values of methyl groups in miliacin  
 600 with a range  $> 500$  ‰. Thus, removing of methyl groups during diagenesis could significantly enrich  
 601 diagenetic by-products compared to their biochemical precursor. But, as depicted in **Fig. 8**, the loss  
 602 of methyl groups during diagenesis is always associated with the loss of hydrogen atoms in  
 603 methylene and methine groups associated to rings. Thus, in order to fully understand any isotopic  
 604 shift during diagenesis, it is necessary to take into account  $\delta\text{D}$  values of methylene and methine  
 605 ( $\delta\text{D}_{\text{meth}}$ ). Miliacin is composed by 52 hydrogen atoms of which 25 are in methine or methylene  
 606 groups and  $3 \times 9 = 27$  compose methyl groups. The average  $\delta\text{D}_{\text{meth}}$  value can thus be accessed by mass

607 balance according to equation **Eq. 5**.

$$608 \quad \delta D_{\text{Meth}} = \left( \frac{52 \times \delta D_{\text{miliacin}} - 3 \times \sum \delta D_{\text{Mei}}}{25} \right). \quad (6)$$

609  
610 Using  $\delta D_{\text{Mei}}$  values of methyl groups from anisotropic NMR / SNIF-NMR<sup>TM</sup> combined  
611 approach (**Table 2, column 5**) and the  $\delta D_{\text{miliacin}} = -100 \text{ ‰}$  in Eq. 6, we obtain an average  $\delta D_{\text{meth}} =$   
612  $89 \text{ ‰}$ . Using  $\delta D_{\text{Mei}}$  values of methyl groups from GC-irMS, we obtain  $\delta D_{\text{meth}} = 48 \text{ ‰}$ . From these  
613 results it appears that methylene and methine are, on average, largely enriched in D compared to  
614 methyl groups. One could thus expect that, during diagenesis, any D-enrichment of the diagenetic  
615 byproduct compared to its precursor due to the loss of depleted methyl groups could be  
616 counterbalanced by D-depletion due to the associated loss of D-enriched methylene and methine.  
617 Nevertheless, in the absence of information on the heterogeneous distribution of D in methylene and  
618 methine, it is yet impossible to fully forecast the impacts of pentacyclic triterpenes degradation on the  
619 hydrogen isotopic composition of their by-products. In addition, only removal of hydrogen atoms is  
620 here considered but fractionation processes can be expected from the preferential biotic or abiotic  
621 degradation of biochemical precursors.

## 622 623 **6. CONCLUSION**

624  
625 Obtaining the site-specific H isotope distribution in biomarkers by a reliable method is crucial  
626 but has been analytically challenging. GC-irMS allows  $\delta D$  values determination at a molecular scale  
627 but does not inform on any preferential position of H isotopes on molecular sites. In contrast,  
628 anisotropic NAD 2D-NMR spectroscopy provides more detailed information on isotope location  
629 within molecules. This study explains how and why NAD 2D-NMR using weakly aligning LC is a  
630 fruitful analytical tool able to overcome difficulties associated to isotropic NAD 1D-NMR (SNIF-  
631 NMR), with the final aim of studying the site-specific distribution of hydrogen isotopes. By coupling  
632 anisotropic and isotropic NAD NMR techniques or anisotropic NAD NMR and GC-irMS  
633 measurements, we were, for the first time, able not only to establish the relative distribution of  
634 hydrogen isotopes in methyl groups of miliacin, but also to afford quantitative (D/H) and  $\delta D$  values  
635 for each methyl group.

636  
637 The resulting distribution of (D/H) values of methyl groups in miliacin provides detailed  
638 information that is consistent with its biosynthetic pathways; i.e. miliacin results from the addition of  
639 a methyl group on germanicol that is synthesized through the cyclisation of oxidosqualene. This later  
640 is produced by the condensation of two farnesyl pyrophosphate units, themselves built from two  
641 isopentenyl pyrophosphate units and a dimethylallyl pyrophosphate unit. The heterogeneous  
642 distribution of site-specific  $\delta D$  values revealed in this study could have major consequences on the

642  $\delta D$  values of diagenetic derivatives of pentacyclic triterpenes since removal of methyl groups is often  
643 observed in diagenetic pathways. Although this work paves the way to exciting prospects in the  
644 frame of a better understanding of hydrogen isotopic changes during pentacyclic triterpene diagenesis,  
645 thus improving any climate reconstruction based on these derivatives, the next challenge will consist  
646 of experimentally evaluating the  $\delta D$  values of all hydrogen sites in biomarkers in order to have a  
647 complete picture of their heterogeneous distribution. By achieving this task, we expect being able to  
648 forecast  $\delta D$  values of diagenetic derivatives from their precursors or to calculate the  $\delta D$  values of  
649 biochemical precursors through  $\delta D$  values of their diagenetic derivatives.

650 Various analytical difficulties (assignment of the anisotropic NAD signals, low S/N ratios, ...)   
651 met and discussed in this study are both related to the structure and the low solubility of miliacin.  
652 They are therefore highly specific to this structurally-complex geological biomarker, and will not  
653 necessarily come up gain when other (simpler) organosoluble biomarkers will be studied. As seen  
654 here, possible protocols/strategies/solutions were proposed to overcome them, thus showing that  
655 NAD 2D-NMR in LC remains an adaptable method able to face the majority of problems associated  
656 to given analyte. In all cases, the possibility to significantly increase the concentration of deuterio-  
657 isotopomers in NMR samples and/or record the NAD experiments by using very high-field NMR  
658 spectrometers ( $B_0 > 14$  T) still equipped by a  $^2H$  cryoprobe are the most direct way to overcome the  
659 sensitivity problem, rendering the accuracy of measurements higher.

660

## 661 **ACKNOWLEDGMENTS**

662

663 French authors thank CNRS and University of Paris-Sud for their continuous financial  
664 support. P.L. thanks also Dr. Isabelle Billault for the gift of calibrated TMU and J.-P. Baltaze for his  
665 help. J.J. is grateful to R. Boscardin for her technical support and I. Lamour (HITEX, Vannes,  
666 France) for providing millet oil. This work is also a part of the ANR Project PalHydroMil project,  
667 supported by Agence Nationale de la Recherche grant (ANR JCJC, 2011-2013). V.T. was supported  
668 by PalHydroMil and Le STUDIUM Loire Valley Institute for Advanced Studies for her participation  
669 in this study. The authors wish to thank Pr. John Hayes and the two anonymous reviewers for their  
670 constructive comments on an earlier version of the manuscript.

671

## 672 **REFERENCES**

673

674 van Aarssen B.G.K., Alexander R. and Kagi R.I. (2000) Higher plant biomarkers reflect  
675 palaeovegetation changes during Jurassic times. *Geochim. et Cosmochim. Acta* **64**, 1417-  
676 1424.

677 Akoka. S., Barantin L., Trierweiler M. (1999) Concentration measurement by proton NMR using the



- 678 ERETIC method. *Anal. Chem.* **13**, 2554-2557.
- 679 Aroulanda C., Lesot P., Merlet D., and Courtieu J. (2001) Structural ambiguities in bridged ring  
680 systems resolved using natural abundance deuterium NMR in chiral liquid crystals. *J. Phys.*  
681 *Chem. A.*, **107**, 10911-10918.
- 682 Bossard N., Jacob J., Le Milbeau C., Lallier-Vergès E., Terwilliger V.J. and Boscardin R. (2011)  
683 Variation in  $\delta D$  values of a single, species-specific molecular biomarker: a study of miliacin  
684 throughout a field of broomcorn millet (*Panicum miliaceum* L.). *Rapid Com. Mass Spec.* **25**,  
685 1-9.
- 686 Bossard, N. (2013). Pertinence et calibration d'un nouveau marqueur paléohydrologique : Le rapport  
687 isotopique de l'hydrogène mesuré sur la miliacine. Ph.D. thesis, Université d'Orléans,  
688 Orléans, France.
- 689 Bossard N., Jacob, J. LeMilbeau C., Sauze J., Terwilliger V.J., Poissonnier B., and Vergès E. (2013).  
690 Distribution of miliacin (olean-18-en-3 $\beta$ -ol methyl ether) in broomcorn millet (*Panicum*  
691 *miliaceum*), and other reputed potential sources. Consequences on the use of sedimentary  
692 miliacin as a tracer of millet. *Org. Geochem.* **63**, 48-55.
- 693 Chikaraishi Y. and Narakoa H. (2003) Compound-specific  $\delta D$ - $\delta^{13}C$  analyses of n-alkanes extracted  
694 from terrestrial and aquatic plants. *Phytochem.* **63**, 361-371.
- 695 Chikaraishi, Y., Naraoka H. and Poulson S.R. (2004) Hydrogen and carbon isotopic fractionations of  
696 lipid biosynthesis among terre (C3, C4 and CAM) and aquatic plants. *Phytochem.* **65**, 1369-  
697 1381.
- 698 Chikaraishi Y., Tanaka R., Tanaka A. and Ohkouchi N. (2009) Fractionation of hydrogen isotopes  
699 during phytol biosynthesis. *Org. Geochem.* **40**, 569-573.
- 700 Corbet B. (1980) Origine et transformation des triterpènes dans les sédiments récents. Ph.D. thesis,  
701 Université Louis Pasteur, Strasbourg, France, 106 p.
- 702 Cranwell P.A. (1984) Organic geochemistry of lacustrine sediments: triterpenoids of higher plant  
703 origin reflecting post-glacial vegetational succession. In *Lakes Sediments and Environmental*  
704 *History* (Eds. E.Y., Haworth and J.W.G. Lund). University Press, Leicester, pp. 69-92.
- 705 Das M. and Mahato S. (1983) Triterpenoids. *Phytochem.* **22**, 1071-1085.
- 706 Dansgaard W. (1964) Stable isotopes in precipitation. *Tellus* **16**, 436-468.
- 707 de Laeter J.R., Böhlke J.K., De Bièvre P., Hidaka H., Pieser H.S., Rosman K.J.R. and Taylo, P.D.P.,  
708 (2003) Atomic weights of the elements. *Pure Appl. Chem.* **75**, (6), 683-800.
- 709 Eglinton I.E. and Eglinton G. (2008) Molecular proxies for paleoclimatology. *Earth and Planet. Sci.*  
710 *Let.* **275**, 1-16.
- 711 Feakins S.J. and Sessions A.L. (2010) Controls on the D/H ratios of plant leaf waxes in an arid  
712 ecosystem. *Geochim. Cosmochim. Acta* **74**, 2128-2141.

- 713 Freeman K. H., Boreham C.J., Summons R.E. and Hayes J. M. (1994) The effect of aromatization on  
714 the isotopic compositions of hydrocarbons during early diagenesis. *Org. Geochem.* **21**, 1037-  
715 1049.
- 716 Gleixner G. and Mügler I. (2007) Compound-specific hydrogen isotope ratios of biomarkers: tracing  
717 climatic changes in the past. In *Stable Isotopes as Tracers for Ecological Change* (Eds. T.  
718 Dawson and R. Siegwolf). Terrestrial Ecology Series, Academic Press, Amsterdam, pp. 249-  
719 266.
- 720 Hou J., D'Andrea W.J., McDonald D. and Huang Y. (2007) Hydrogen isotopic variability in leaf  
721 waxes among terrestrial and aquatic plants around Blood Pond, Massachusetts (USA). *Org.*  
722 *Geochem.* **38**, 977-984.
- 723 Hyperchem™, Professional 6.1, Hypercube, Inc, 1115 NN 4th street, Gainesville, Florida, 32601,  
724 USA.
- 725 Itô H. (1934) The chemical investigation of some gramineae oils. *J. Fac. of Agri., Hokkaido Imperial*  
726 *University*, **XXXVII**, 1-40.
- 727 Jacob J., Disnar J.-R., Boussafir M., Sifeddine A., Albuquerque A.L.S. and Turcq B. (2005)  
728 Pentacyclic triterpene methyl ethers in recent lacustrine sediments (Lagoa do Caco, Brazil).  
729 *Org. Geochem.* **36**, 449-458.
- 730 Jacob J., Disnar J.R., Boussafir M., Albuquerque A.L.S., Sifeddine A. and Turcq B. (2007)  
731 Contrasted distributions of triterpene derivatives in the sediments of Lake Caçó reflect  
732 paleoenvironmental changes during the last 20,000 years in NE Brazil. *Org. Geochem.* **38**,  
733 180-197.
- 734 Jacob J., Disnar J.-R., Arnaud F., Chapron E., Debret M., Lallier-Vergès E., Desmet M. and Revel-  
735 Rolland M. (2008a) Millet cultivation history in the French Alps as evidenced by a  
736 sedimentary molecule. *J. Archaeol. Sci.* **35**, 814-820.
- 737 Jacob J., Disnar J.-R. and Bardoux G. (2008b) Carbon isotope evidence for sedimentary miliacin as a  
738 tracer of *Panicum miliaceum* (broomcorn millet) in the sediments of Lake le Bourget (French  
739 Alps). *Org. Geochem.* **39**, 1077-1080.
- 740 Jacob J., Disnar J.-R., Arnaud F., Gauthier E., Billaud Y., Chapron E. and Bardoux G. (2009)  
741 Impacts of new agricultural practices on soil erosion during the Bronze Age in French  
742 Prealps. *The Holocene* **19**, 241-249.
- 743 Jacob J., LeMilbeau C., Bossard N., Disnar J.R. and Billaud Y. (2011) Combined  $\delta^{13}\text{C}$  -  $\delta\text{D}$  analysis  
744 of pentacyclic triterpenes and their derivatives. Studium Conference: "Hydrogen Isotopes as  
745 Environmental Recorders", 15-16 September 2011, Orléans, France.
- 746 Killops S.D., Raine J.I., Woolhouse A.D. and Weston R.J. (1995) Chemostratigraphic evidence of  
747 higher plant evolution in the Taranaki Basin, New Zealand. *Org. Geochem.* **23**, 429-445.

- 748 Lafon O., Lesot P., Merlet D. and Courtieu J. (2004) Modified z-gradient filtering as a new mean to  
749 obtain phased deuterium autocorrelation 2D-NMR spectra in oriented solvents. *J. Magn.*  
750 *Reson.* **171**, 135-142.
- 751 Lavrieux M., Jacob J., Le Milbeau C., Zocatelli R., Masuda K., Bréheret J. G. and Disnar J. R. (2011)  
752 Occurrence of triterpenyl acetates in soil and their potential as chemotaxonomical markers of  
753 Asteraceae. *Org. Geochem.* **42**, 1315-1323.
- 754 Lesot P., Merlet D., Lowenstein A. and Courtieu J. (1998) Enantiomeric Visualisation using Proton-  
755 Decoupled Natural Abundance Deuterium NMR in Poly( $\gamma$ -benzyl-L-glutamate) Liquid  
756 Crystalline Solutions. *Tetrahedron: Asymmetry* **9**, 1871-1881.
- 757 Lesot P., Baillif V. and Billault I. (2008) Combined analysis of four C-18 unsaturated fatty acids  
758 using Natural Abundance Deuterium 2D-NMR spectroscopy in chiral oriented solvents. *Anal.*  
759 *Chem.* **80**, 2963-2972.
- 760 Lesot P. and Lafon O. (2008) Enantiomeric Analysis using Natural Abundance Deuterium 3D-NMR  
761 Spectroscopy in Polypeptide Chiral Oriented Media. *Chem. Phys. Lett.* **458**, 219-222.
- 762 Lesot P. and Courtieu J. (2009) Natural abundance deuterium NMR spectroscopy: Developments and  
763 analytical applications in liquids, liquid crystals and solid phases. *Prog. Nucl. Magn. Reson.*  
764 *Spectrosc.* **55**, 128-159.
- 765 Lesot, P., Serhan, Z. and Billault, I. (2011) Recent advances in the analysis of the site-specific  
766 isotopic fractionation of metabolites such as fatty acids using anisotropic natural abundance  
767  $^2\text{H}$  NMR spectroscopy: Application on conjugated linolenic methyl esters. *Anal. Bioanal.*  
768 *Chem.* **399**, 1187-1200.
- 769 Lesot P., Serhan Z. Aroulanda C. and Billault I. (2012) Analytical contribution of NAD 2D-NMR  
770 spectroscopy in polypeptide mesophases to the investigation of triglycerides *Magn. Reson.*  
771 *in Chem.* **50**, S2-S11.
- 772 Lesot P., Aroulanda C., Zimmermann, H. and Luz Z. (2015) Enantiotopic Discrimination in the NMR  
773 Spectrum of Prochiral Solutes in Chiral Liquid Crystals, *Chem. Soc. Rev.* **44**, 2330.
- 774 Massou S., Nicolas C., Letisse F., Portais J.-C., Nmr-Based Fluoxomics: Quantitative 2D-NMR  
775 methods for isotopomers Analysis. *Phytochemistry*, 2007, **68**, 2330-2340.
- 776 Liu W., Yang H. and Li L. (2006) Hydrogen isotopic composition of *n*-alkanes from terrestrial plants  
777 correlate with their ecological life form. *Oecologia* **150**, 330-338.
- 778 Lohmann F. (1988) Aromatisations microbiennes de triterpènes végétaux. Ph.D. thesis, Université  
779 Louis Pasteur, Strasbourg, France.
- 780 Mahato S.B., Nandy A.K. and Roy G. (1992) Triterpenoids, *Phytochem.* **31**, 2199-2249.
- 781 Mahato S.B. and Sen S. (1997) Advances in triterpenoid research, 1990-1994. *Phytochem.* **44**, 1185-  
782 1236.

- 783 Martin G.J. and Martin M.L., (1981) Deuterium labelling at the natural abundance level as studied by  
784 high field quantitative  $^2\text{H}$  NMR, *Tet. Lett.* **22**, 3525-3528.
- 785 Martin G.J., Zhang B.-L., Naulet N. and Martin M.L., (1986) Deuterium transfer in the bioconversion  
786 of glucose to ethanol studied by specific labeling at the natural abundance level. *J. Am. Chem.*  
787 *Soc.* **108**, 5116-5122.
- 788 McInerney F.A., Helliker B.R. and Freeman K.H. (2011) Hydrogen isotope ratios of leaf wax n-  
789 alkanes in grasses are insensitive to transpiration. *Geochim. Cosmochim. Acta* **75**, 541-554.
- 790 Merlet D., Ancian B., Courtieu D. and Lesot P. (1999) Two-dimensional Deuterium NMR  
791 spectroscopy of chiral molecules oriented in a polypeptide liquid crystal: application for the  
792 enantiomeric analysis through natural abundance deuterium NMR. *J. Am. Chem. Soc.* **121**,  
793 5249-5258.
- 794 Motuzaitė-Matuzevičiūtė G., Jacob J., Telizhenko S., Jones M.K. (2013) Miliacin in paleosoils from  
795 early Iron Age in Ukraine reveals paleofields of broomcorn millet. *Archaeological and*  
796 *Anthropological Sciences*, 1-8.
- 797 Ohmoto T., Ikuse M. and Natori S. (1970) Triterpenoids of the Gramineae. *Phytochem.* **9**, 2137-  
798 2148.
- 799 Regnery J., Püttmann W., Koutsodendris A., Mulch A. and Pross J. (2013) Comparison of the  
800 paleoclimatic significance of higher land plant biomarker concentrations and pollen data: A  
801 case study of lake sediments from the Holsteinian interglacial. *Org. Geochem.* **61**, 73-84.
- 802 Rullkötter J., Peakman T.M. and ten Haven, H.L. (1994) Early diagenesis of terrigenous triterpenoids  
803 and its implications for petroleum geochemistry. *Org. Geochem.* **21**, 215-233.
- 804 Sachse D., Radke J. and Gleixner G. (2006)  $\delta\text{D}$  values of individual n-alkanes from terrestrial plants  
805 along a climatic gradient – Implications for the sedimentary biomarker record. *Org. Geochem.*  
806 **37**, 469-483.
- 807 Sarfati M., Lesot P., Merlet D. and Courtieu J. (2000) Theoretical and experimental aspects of  
808 enantiomeric differentiation using natural abundance multinuclear NMR spectroscopy in  
809 polypeptide liquid crystals. *Chem. Commun.* **21**, 2069-2081.
- 810 Sauer P.E., Eglinton T.I., Hayes J.M., Schiommelmann A., Sessions A.L. (2001) Compound-specific  
811 D/H ratios of lipid biomarkers from sediments as a proxy for environmental and climatic  
812 conditions. *Geochim. Cosmochim. Acta* **65**, 213-222.
- 813 Schwab V.F., Garcin Y., Sachse S., Todou G., Séné O., Onana J.-M., Houndong G. and Gleixner G.  
814 (2015) Effect of aridity on  $\delta^{13}\text{C}$  and  $\delta\text{D}$  values of C3 plant- and C4 graminoid-derived leaf  
815 wax lipids from soils along an environmental gradient in Cameroon (Western Central Africa)  
816 *Org. Geochem.* **78**, 99-109.

- 817 Serhan Z., Martel L., Billault I. and Lesot P. (2010) Complete determination of site specific bio-  
818 enantiomeric excesses in linoleic acid using natural abundance deuterium 2D-NMR in  
819 polypeptide mesophase. *Chem. Commun.* **46**, 6599-6601.
- 820 Sessions A.L., Burgoyne T.W., Schimmelmann A. and Hayes J.M. (1999) Fractionation of hydrogen  
821 isotopes in lipid biosynthesis. *Org. Geochem.* **30**, 1193-1200.
- 822 Smetanina O.F., Kuznetzova T.A., Denisenko V.A., Pivkin M.V., Khudyakova Y.V., Gerasimenko  
823 A.V., Popov D.Y., Il'in S.G. and Elyakov G.B. (2001) 3 $\beta$ -Methoxyolean-18-ene (miliacin)  
824 from the marine fungus *Chaetomium olivaceum*. *Russ. Chem. Bull.* **50**, 2463-2465.
- 825 Smith F.A. and Freeman K.H. (2006) Influence of physiology and climate  $\delta$ D of leaf wax n-alkanes  
826 from C3 and C4 grasses. *Geochim. Cosmochim. Acta* **70**, 1172-1187.
- 827 Terwilliger V.J., Eshetu Z., Disnar J.R., Jacob J., Adderley W.P., Huang Y., Alexandre M. and Fogel  
828 M. (2013) Environmental changes and the rise and fall of civilizations in the northern Horn of  
829 Africa: an approach combining  $\delta$ D analyses of land-plant derived fatty acids with multiple  
830 proxies in soil. *Geochim. et Cosmochim. Acta* **111**, 140-161.
- 831 Terwilliger V.J. and Jacob J. (2013) Introduction: Hydrogen isotopes as environmental recorders.  
832 *Geochim. et Cosmochim. Acta* **111**, 1-4.
- 833 Trendel J. (1985) Dégradation des triterpènes dans les sédiments. Aspects photochimiques et  
834 microbiologiques. Ph.D. thesis, Université Louis Pasteur, Strasbourg, France, 126 p.
- 835 Wang Y., Sessions A.L., Nielsen R.J. and Goddard W.A. (2013) Equilibrium  $^2\text{H}/^1\text{H}$  fractionation in  
836 organic molecules: III. Cyclic ketones and hydrocarbons. *Geochim. et Cosmochim. Acta* **107**,  
837 82-95.
- 838 Wang Y.V., Larsen T., Leduc G., Andersen N., Blanz T. and Schneider R.R. (2013) What does leaf  
839 wax  $\delta$ D from a mixed C3/C4 vegetation region tell us? *Geochim. et Cosmochim. Acta* **111**,  
840 128-139.
- 841 Wolff G.A., Trendel J.M. and Albrecht P. (1989) Novel Monoaromatic triterpenoids hydrocarbons  
842 occurring in sediments. *Tetrahedron* **45**, 6721-6728.
- 843 Xu R., Fazio G.C. and Matsuda S.P.T. (2004) On the origins of triterpenoid skeletal diversity.  
844 *Phytochem.* **65**, 261-291.
- 845

846 **FIGURE CAPTIONS**

847

848 **Figure 1:** Structure of miliacin where  $R_{31} = -\text{OMe}$  and gemanicol where  $R_{31} = -\text{OH}$ , together with the  
 849 atomic numbering used and notation of rings (A to E) and methyl groups (23 to 31).

850

851 **Figure 2:** 600.1 MHz proton 1D-NMR spectrum of miliacin in  $\text{CDCl}_3$  (295 K). The assignment of  
 852 resonances of methyl groups is given in **Table S-1, S-2** and **S-3** in **Appendix A.I of Supp.**  
 853 **Info.**). The peak marked by an asterisk ( $\delta = 1.24$  ppm) is an unidentified impurity.

854

855 **Figure 3:** (a) Methyl region (Me31 not shown) of the 92.1 MHz NAD tilted  $Q$ -resolved Fz 2D map  
 856 of miliacin dissolved in PBLG/ $\text{CHCl}_3$  and recorded at 295 K. (b) 600 MHz proton 1D  
 857 spectrum of miliacin in  $\text{CHCl}_3$  plot at the horizontal same scale (in ppm). A lorentzian filtering  
 858 ( $\text{LB} = 1.5$  Hz) has been applied in both spectral dimensions. On the  $F_1$  and  $F_2$  axes are shown  
 859 the spectral 2D projections. Note the significant variations of peak intensity of  $^2\text{H}$  doublets in  
 860  $F_1$  dimension illustrating the enrichment/depletion in deuterons.

861

862 **Figure 4:** Series of four NAD 1D sub-spectra extracted from the tilted  $Q$ -resolved Fz 2D map  
 863 showing the nine methyl groups of miliacin along with their assignments (see also **Fig. S-9**).  
 864 Each sub-spectrum is plotted with the same amplitude of noise. Note the variation of peak  
 865 intensity of  $^2\text{H}$  doublets. The variations observed reveal the differences of (D/H) for each site.

866

867 **Figure 5:** 3D view of the minimal energy conformation of miliacin calculated with Hyperchem 3.0  
 868 software. A semi-empirical quantum chemistry methods based on the Hartree-Fock  
 869 formalism (Austin Model 1 or AM1) has been used for modeling the molecule. The minimal  
 870 energy was found at  $-8177.99$  kJ/mol. Except for ring E distorted by the double bond, all other  
 871 cycles have adopted a chair conformation. Note the difference of orientation axes of C- $\text{CH}_3$   
 872 vectors for Me23 and Me30 (close to the median horizontal molecular plan) compared to  
 873 Me24 to Me29 (see also **Fig. S-9** in **Appendix A.II of Supp. Info.**).

874

875 **Figure 6:** Variation of (a) (D/H) ratios in ppm and (b)  $\delta\text{D}$  in ‰ for each methyl group of miliacin  
 876 when  $(\text{D}/\text{H})_{\text{Me31}}$  is evaluated by the two-step NAD NMR approach. Excluding the exogenous  
 877 methyl group 31, note the (inverted) quasi-symmetrical bell shape of both plots.

878

879 **Figure 7:** Synthetic pathway leading to miliacin. Modified from Sachse et al. (2012).

880

881 **Figure 8:** Potential diagenetic pathways of pentacyclic triterpenes (here starting from  $\alpha$ -amyrin)  
 882 illustrating the loss of hydrogen atoms related to methyl groups, methylene/methine. Modified  
 883 from Wolff et al. (1989).

884