1   Mono- and dihydroxyl Glycerol Dibiphytanyl Glycerol Tetraethers in

2   marine sediments: identification of both core and intact polar lipid forms

3

4   Xiao-Lei Liu[a][*], Julius S. Lipp[a], Jeffrey H. Simpson[b], Yu-Shih Lin[a], Roger E. Summons[c],

5   Kai-Uwe Hinrichs[a]

6

7   [a] *Organic Geochemistry Group, MARUM Center for Marine Environmental Sciences & Dept. of Geosciences,*

8   *University of Bremen, 28359 Bremen, Germany*

9   [b] *Department of Chemistry, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139-*

10 *4307, USA*

11 [c] *Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, 77 Massachusetts*

12 *Avenue, Cambridge, MA 02139-4307, USA*

13 ** Corresponding author. E-mail address: xliu@uni-bremen.de (Xiao-Lei Liu).*

14

15 **ABSTRACT**

16   Hydroxylated glycerol dialkyl glycerol tetraethers (hydroxy-GDGTs) were

17 detected in marine sediments of diverse depositional regimes and ages. Mass

18 spectrometric evidence, complemented by information gleaned from two-dimensional

19 (2D) ^{1}H-^{13}C nuclear magnetic resonance (NMR) spectroscopy on minute quantities of

20 target analyte isolated from marine sediment, allowed us to identify one major compound

21 as a monohydroxy-GDGT with acyclic biphytanyl moieties (OH-GDGT-0). NMR

22 spectroscopic and mass spectrometric data indicate the presence of a tertiary hydroxyl

23 group suggesting the compounds are the tetraether analogues of the widespread

24 hydroxylated archaeol derivatives that have received great attention in geochemical

25   studies of the last two decades. Three other related compounds were assigned as acyclic

26   dihydroxy-GDGT (2OH-GDGT-0) and monohydroxy-GDGT with one (OH-GDGT-1)

27   and two cyclopentane rings (OH-GDGT-2). Based on the identification of hydroxy-

28   GDGT core lipids, a group of previously reported unknown intact polar lipids (IPLs),

29   including the ubiquitously distributed H341-GDGT (Lipp and Hinrichs, 2009, Structural

30   diversity and fate of intact polar lipids in marine sediments. *Geochim. Cosmochim. Acta*

31   73, 6816–6833), and its analogues were tentatively identified as glycosidic hydroxy-

32   GDGTs. In addition to marine sediments, we also detected hydroxy-GDGTs in a culture

33   of *Methanothermococcus thermolithotrophicus*. Given the previous finding of the

34   putative polar precursor H341-GDGT in the planktonic marine crenarchaeon

35   *Nitrosopumilus maritimus*, these compounds are synthesized by representatives of both

36   cren- and euryarchaeota. The ubiquitous distribution and apparent substantial abundance

37   of hydroxy-GDGT core lipids in marine sediments (up to 8% of total isoprenoid core

38   GDGTs) point to their potential as proxies.

39

40

41   **1. Introduction**

42

43   As one of the three domains of life, Archaea differ from other microorganisms by

44   not only genetic affiliation but also through their distinctive isoprenoidal glycerol ether

45   lipids. Since the first report of glycerol diether lipids in a halophilic archaeon,

46   *Halobacterium salinarum* (Kates et al., 1963), various ether lipids have been discovered

47   in different cultivated species of extremophiles (Langworthy, 1977), methanogens

48 (Tornabene and Langworthy, 1979; Koga et al., 1993) and also in the uncultivated

49 mesophilic marine crenarchaeota (DeLong et al., 1998). Among the diverse structures

50 now identified in the lipids of archaea, including glycerol dibiphytanyl glycerol

51 tetraethers (GDGTs, see molecular structures in Fig. 1) with different numbers of

52 cyclopentane rings (e.g., De Rosa et al., 1980), multi-unsaturated diether lipids (Gonthier

53 et al., 2001) and the macrocyclic diether (Comita and Gagosian, 1983), hydroxylated

54 diethers are recognized as distinctive compound class. Ferrante et al. (1988a) first

55 reported the identification of hydroxyarchaeol isolated from *Methanothrix concilii*.

56 Nuclear magnetic resonance (NMR) analysis indicated a tertiary hydroxyl group on C-3

57 of the phytanyl chain at the *sn*-3 position of glycerol (see molecular structures in Fig. 1).

58 In the same species, the intact polar lipid (IPL) of hydroxyarchaeol, a diglycosidic

59 hydroxyarchaeol, was also identified (Ferrante et al., 1988b). According to the

60 biosynthetic pathway of GDGT in cultivated archaea, the biphytane skeleton was

61 proposed to be formed via head-to-head condensation of two archaeol units (Nemoto et

62 al., 2003). Therefore, given the presence of hydroxyarchaeol, synthesis of hydroxy-

63 GDGT appears plausible but has not yet been conclusively demonstrated. Summons et al.

64 (2002) detected hydroxylated biphytanes in *Methanothermococcus thermolithotrophicus*

65 after ether cleavage of intact lipids, and Lipp and Hinrichs (2009) observed one unknown

66 IPL producing a very weak MS^2 signal of putative hydroxy-GDGT, but to date the

67 identification of hydroxy-GDGT remains to be confirmed.

68 Based on the knowledge gained from lipids in microbial cultures, archaeal ether

69 lipids have also been intensively documented in the sedimentary record. For instance, the

70 existence of methane-consuming archaea in marine sediment was revealed by the

71 detection of ^{13}C depleted archaeol and hydroxyarchaeol (Hinrichs et al., 1999). The fossil

72 archaeal lipids in Mid-Cretaceous black shale were reported to indicate the massive

73 expansion of planktonic archaea during ocean anoxic events (Kuypers et al., 2001). On

74 the other hand, the presence of archaeal IPLs was proposed to reflect the *in-situ* activity

75 of live archaea in specific environments, such as the anoxic water column of the Black

76 Sea (Schubotz et al., 2009), marine methane seep sites (Rossel et al., 2008; 2011) and

77 subseafloor sediments (Biddle et al., 2006), although their applicability to low-activity

78 subseafloor sediment remains controversial (cf. Lipp and Hinrichs, 2009; Schouten et al.,

79 2010) and will require further validation. In several recent IPL studies, one unknown

80 compound, termed H341-GDGT by Lipp and Hinrichs (2009) according to its

81 fragmentation behavior during mass spectrometric analysis, was designated as a major

82 putative archaeal IPL in a wide range of depositional settings (e.g., Sturt et al., 2004;

83 Lipp and Hinrichs, 2009; Schubotz et al., 2009) and thaumarchaeotal cultures (Schouten

84 et al., 2008; Pitcher et al., 2011). Knowing the structure and biological source of this

85 widely occurring, but uncharacterized compound H341-GDGT will improve

86 understanding of both the composition of archaeal communities and their activity in these

87 environments.

88   With the development of liquid chromatography mass spectrometry (LC-MS),

89 identification and quantification of large and polar archaeal lipids has been facilitated,

90 such as the analysis of GDGT core lipids with HPLC-atmospheric pressure chemical

91 ionization (APCI)-MS (Hopmans et al., 2000) and the analysis of archaeal IPLs with

92 HPLC-electrospray ionization (ESI)-MS (Sturt et al., 2004). In this study, we describe the

93   identification of novel hydroxylated GDGT core lipids and a corresponding series of

94   glycosidic hydroxy-GDGTs.

95

96   **2. Materials and methods**

97

98   *2.1. Materials and sample preparation*

99

100   Twelve globally distributed marine sediments from various geological settings

101   were included in this study (Table 1). Detailed descriptions pertaining to these samples

102   were published previously in Liu et al. (2011). For isolating target compounds for NMR

103   analysis, we extracted more than 1 kg of composite dry sediment samples from Aarhus

104   Bay (position: 56°07.06'N, 10°20.85'E, 15 m water depth and 6-7 m sediment depth

105   below surface) and the Baltic Sea (position: 58°56.04'N, 17°43.81'E, 52 m of water depth,

106   surface sediment). All samples were freeze-dried and extracted using the modified Bligh

107   and Dyer protocol as described by Sturt et al. (2004). The total lipid extracts (TLE) of the

108   twelve marine sediments were first separated into core lipid and IPL fractions with

109   preparative LC before analysis in order to increase sensitivity for the detection of minor

110   compounds. Fractions were collected with a fraction collector (GILSON, FC 204) using a

111   semi-preparative LiChrospher Si60 column (250 x 10 mm, 5 μm, Alltech, Germany) and

112   the mobile phase gradient and instrument settings previously described by Biddle et al.

113   (2006).

114   In addition, one archaeal culture, *Methanothermococcus thermolithotrophicus*

115   strain (DSM 2095), grown at 85°C in enamel-protected fermentors with stirring (400 rpm)

116   and continuous gassing (H_2/CO_2, 80:20) was provided by M. Baumgartner and K. Stetter

117   (University of Regensburg, Germany). Using an ultrasonic probe (HD 2200, Bandelin

118   electronic GmbH & Co. KG, Germany), 0.5 g freeze-dried biomass of

119   *Methanothermococcus thermolithotrophicus* was extracted (4 x) with dichloromethane

120   (DCM):MeOH (20 ml;1:1, v/v) for 15 min. One aliquot (1%) of the TLE of

121   *Methanothermococcus thermolithotrophicus* was evaporated to dryness in a 2-mL vial

122   and 1 mL of 6 M HCl/MeOH/DCM (1:9:1, v/v/v) was added before reaction at 70°C for

123   3 h to hydrolyze the glycosidic IPLs into core lipids.

124

125   *2.2. HPLC-APCI-MS analysis of hydroxy-GDGT core lipids*

126

127   One aliquot of the core lipid fractions of each sample was dissolved in 100 μL *n*-

128   hexane/propan-2-ol (99:1, v/v) for HPLC-MS analysis. Separation of compounds was

129   performed on a Prevail Cyano column (2.1×150 mm, 3 μm; Alltech, Grace) maintained at

130   35°C with the following gradient program: flow rate of 0.25 mL min^{-1}, the gradient of the

131   mobile phase was first held at 100% of eluent A (*n*-hexane/propan-2-ol, 99:1, v/v) for

132   5 min, followed by a linear gradient to 90% of A and 10% B (n-hexane/propan-2-ol,

133   90:10, v/v) in 20 min, followed by a linear gradient to 100% B at 35 min, after holding

134   100% B for 5 min the column was re-equilibrated with 100% A at a flow rate of 0.6 mL

135   min^{-1} for 5 min before the next injection. Detection was achieved with an Agilent 6130

136   MSD single quadrupole mass spectrometer, coupled to an Agilent 1200 series HPLC via

137   multimode ion source set in APCI mode. APCI settings were nebulizer pressure 60 psi,

138   vaporizer temperature 250°C, drying gas (N_2) flow 6 L min^{-1} and drying gas temperature

139 200°C, capillary voltage 2 kV, and corona current 5 μA. The detector was set for

140 selective ion monitoring (SIM) of [M+H]$^+$ ions (m/z 1302, 1300, 1298, 1296, 1292, 1246,

141 1244, 1242, 1240 and 1236; fragmentor voltage 190 V).

142   In order to obtain the MS^2 spectra with detailed fragmentation features, selected

143 samples with higher abundance of target compounds were also analyzed with an Agilent

144 1200 series HPLC system coupled to an Agilent 6520 quadrupole time-of-flight (qTOF)

145 mass spectrometer through an APCI interface. The APCI drying gas temperature was set

146 at 350°C with a gas (N$_2$) flow of 4 L min^{-1}. The qTOF parameters were set to: capillary

147 voltage 1 kV, corona current 5 μA, fragmentor voltage 150 V; skimmer 65 V and

148 octopole 750 V in auto MS/MS scanning mode with MS^1 range of m/z 500-2000 and

149 MS^2 mass range of m/z 100-2000. To achieve the ideal fragmentation of targeted

150 compounds, the mass-dependent collision energy for each precursor was automatically

151 set by the Agilent MassHunter control software (version B.03.01) according to the

152 following equation: [(m/z) / 100] x slope + intercept, where the slope was 3.0 eV and the

153 intercept was -0.5 eV.

154

155 *2.3. Isolation of OH-GDGT-0*

156

157   The TLE of the combined sediment sample from Aarhus Bay and Baltic Sea was

158 dissolved with hexane/propan-2-ol (99:1, v/v). For each run 200 μL of sample,

159 approximately 0.5% of TLE, was loaded onto a PerfectSil 100 CN-3 preparative LC

160 column (250 x 10 mm, 5 μm particle size, MZ Analysentechnik, Germany) equipped

161 with a guard column of the same packing material. With the Agilent 1200 series HPLC

162   system compounds were separated by applying the following solvent gradient at an eluent

163   flow rate of 3 mL min^{-1}: first 100% of eluent A (n-hexane/propan-2-ol, 99:1 [v/v]) for

164   3 min, then to 90% of A and 10% B (*n*-hexane/propan-2-ol, 90:10 [v/v]) with a linear

165   gradient in 12 min, then to 100% B at 20 min and hold 100% B for 10 min, before the

166   column was equilibrated with 100% A for 15 min for the next injection. OH-GDGT-0

167   was collected with a fraction collector (Waters Fraction Collector III) in the time interval

168   from 22 to 23 min.

169

170   *2.4. NMR analysis of OH-GDGT-0*

171

172   After two cycles of purification with preparative HPLC, around 0.6 mg of

173   putative OH-GDGT-0 was isolated from more than 1 kg (dry mass) of composite

174   sediment sample from Aarhus Bay and Baltic Sea. For NMR analysis, purified OH-

175   GDGT-0 was dissolved with 40 µL deuterated benzene (benzene-D$_6$) and transferred into

176   a 40-µL Zirconia nanoprobe sample tube for 2D ^{1}H-^{13}C-NMR analysis. All spectra were

177   obtained using a Varian Unity-INOVA 500 MHz NMR spectrometer with a Varian

178   Nanoprobe which spins the 40-µL sample tube at ca. 2 kHz at the magic-angle (54.7°)

179   relative to the applied field axis. The 90° ^{1}H pulse was 10.2 µs and, for the indirect

180   detection experiments, the ^{13}C 90° pulse was 14.3 µs.

181   The 1-D 500 MHz ^{1}H NMR spectrum was obtained by collecting 64 transients.

182   Each transient was collected for 3 s, with an additional 2 s relaxation delay prior to each

183   acquisition. The ^{1}H spectral window spanned 4.3 kHz (8.6 ppm) and contained 12881

184   complex (25762 actual) points. No apodization was applied. The time domain data was

185   zero-filled to obtain a total size of 128k complex (256k actual) data points prior to

186   Fourier transformation. Phase correction and a small amount of baseline correction were

187   applied. The total experiment time was 5 min and 20 s.

188   The 2-D 500 MHz ^{1}H-^{1}H gradient-selected correlation spectroscopy (gCOSY)

189   NMR spectrum was obtained using the same spectral window as was used for the

190   collection of the 1-D ^{1}H NMR spectrum. The total number of points in the free induction

191   decay (FID) was 1024 complex (2048 actual) point with 256 t_1 time increments, and so

192   the dwell time dictated by the sweep width of 4.3 kHz gave an acquisition time for the

193   FID of 238 ms. An additional relaxation delay of 1 s made the total time per scan 1.29 s.

194   For each FID, eight scans were collected. 256 FIDs were collected in the t1 time domain,

195   and linear prediction was used to double the t_1 time domain data. A squared sine bell

196   function was applied to both the t_1 and t_2 time domain data dimensions in the 2-D data

197   matrix. The period of the t_2 squared sine bell was 119 ms (zero to maximum amplitude),

198   and in the t_1 dimension, the squared sine bell had a period of 60 ms. Following Fourier

199   transformation of both time domains into their respective frequency domains (1k x 1k

200   actual resolution), the data rows in the f_1 frequency domain were baseline corrected and

201   matrix symetrization was performed.

202   The 2-D 500 MHz ^{1}H-^{13}C heteronuclear single quantum correlation (HSQC)

203   NMR spectrum was obtained using the same ^{1}H spectral window parameters, except that

204   the digitization time of the FID was restricted to 238.5 ms and the relaxation delay was

205   1 s, making the total time required for a single pass through the pulse sequence 1.29 s.

206   For each t_1 evolution time, 64 transients were averaged together. 200 complex points

207   were collected in the t_1 time domain (the ^{13}C dimension), but linear prediction was used

208   to extend the data set to 512 complex points (312 complex points were added to the 200

209   that were collected). The ^{13}C spectral window spanned 10 kHz (80 ppm) and was

210   centered at 40 ppm. The t_2 time domain data was apodized with a Gaussian function of

211   0.110 s, and the t_1 time domain data was apodized with a Gaussian function of 0.024 s.

212   Baseline correction parallel to the f_2 frequency axis was performed on the half-Fourier-

213   transformed interferogram prior to Fourier transformation of the t_1 time domain into the

214   f_1 frequency domain. The final data matrix was 1k x 1k complex (2k x 2k actual).

215   The 2-D 500 MHz ^1H-^{13}C gradient-selected heteronuclear multiple bond

216   correlation (gHMBC) NMR spectrum was obtained with parameters similar to those used

217   in obtaining the 2-D HSQC NMR spectrum, except that 400 FIDs with unique t_1

218   evolution times were collected and no linear prediction was used. Each FID was the

219   result of 128 scans. The t_2 time domain data was apodized with a sine bell with a period

220   of 0.119 s (minimum to maximum), and the t_1 time domain data was apodized with a sine

221   bell function with a period of 0.020 s. The final size of the data matrix was 1k x 1k.

222

223   *2.5. HPLC-ESI-MS analysis of glycosidic hydroxy-GDGTs*

224

225   For each sample one aliquot of IPL fraction was dissolved in 100 μL of MeOH

226   and DCM (1:5, v/v), and analysis was performed on an Agilent 1200 series HPLC system

227   coupled to a ThermoFinnigan LCQ Deca XP Plus ion trap mass spectrometer via ESI

228   interface. 10 μL of each sample was injected onto a LiChrosphere Diol-100 column (150

229   x 2.1 mm, 5 μm particle size; Alltech, Germany) equipped with a guard column of the

230   same packing material.

231   Separation of lipids was achieved with a solvent gradient modified from Sturt et

232   al. (2004). A flow rate of 0.2 mL min^{-1} was used with the following gradient program:

233   from 100% A (79:20:0.12:0.04 [v:v:v:v] of hexane/propan-2-ol/formic acid/14.8 M NH$_{3,}$

234    $_{aq}$) to 35% A and 65% B (88:10:0.12:0.04 [v:v:v:v] of propan-2-ol/water/formic acid/14.8

235   M NH$_{3, aq}$) over 45 min, then 100% A for 15 min to re-equilibrate the column for the next

236   analysis. MS analysis was performed in data-dependent mode with fragmentation of the

237   base peak ion in MS^2 under conditions described by Sturt et al. (2004). Intact polar

238   GDGTs with different polar head groups were identified according to the characteristic

239   fragmentation in MS^2 spectra (Sturt et al., 2004; Lipp and Hinrichs, 2009). In addition,

240   different levels of collision energy (normalized collision energy 25% and 100%) for the

241   ion trap setting were applied to optimize detection of labile ions.

242

243   *2.6. Degradation test on OH-GDGT-0*

244

245   An aliquot of purified OH-GDGT-0 was transferred into a 2mL-vial, sealed and

246   subjected to a degradation test under acidic conditions. After drying with a flow of N$_2$,

247   1 mL of 6 M HCl/MeOH/DCM (1:9:1, v/v/v) was added before reaction at 70°C for 3 h.

248   The reaction mixture was blown to dryness with N$_2$ and re-dissolved in 100 µL of *n*-

249   hexane/propan-2-ol (99:1, v/v) for analysis.

250

251   *2.7. GC-MS on ether cleavage released hydrocarbons of OH-GDGT-0*

252

11

253   Another aliquot of purified OH-GDGT-0 was transferred into a 2mL-vial and

254   dried with N_2, then added 0.5 ml BBr_3 (1 M in DCM; Aldrich) to react at 60°C for 2 h.

255   After evaporating the solvent and residual BBr_3 with N_2, the bromides were reduced to

256   hydrocarbons by adding 0.5 ml superhydride ($LiEt_3BH$, in THF, Aldrich) and maintained

257   at 60°C for 2 h. A few drops of water were then added to quench the reaction. Released

258   hydrocarbons were extracted (3 x) with *n*-hexane and combined. For analyzing potential

259   hydroxylated biphytanes, the ether cleavage/reduction products were transferred into a 2-

260   ml vial, dried with a flow of N_2 and mixed with 100 µl BSTFA (N,O-

261   bis(trimethylsilyl)trifluoroacetamide) and 100 µl pyridine at 70°C for 1 h. GC-MS was

262   performed with an Agilent 5973 inert MSD system equipped with a Restek Rxi-5ms

263   column (30 m x 250 µm x 0.25 µm). Separation was achieved using an oven temperature

264   program of 60°C (1 min) to 150°C at 10°C min^{-1} and then to 310°C (held 20 min) at 4°C

265   min^{-1}.

266

267   **3. Results and discussion**

268

269   *3.1. Detection of hydroxy-GDGTs with HPLC-APCI-MS*

270

271   By increasing both the flow rate and the polar gradient of the mobile phase used

272   by the established GDGT core lipid analysis (Hopmans et al., 2000), a series of

273   unknowns (compound a, b, c and d), with nearly identical patterns of molecular and

274   fragment ions but higher retention times relative to the known archaeal GDGTs, was

275   generally observed during normal phase HPLC-APCI-MS analysis of marine sediments

276 (Fig. 2). For example, in Fig. 2 the extracted ion chromatogram (EIC) of m/z 1298

277 revealed three compounds, GDGT-2 and two unknowns designated compound b and

278 compound d. By selecting m/z 1298 as the precursor ion these three compounds

279 generated a similar pattern of MS^2 fragments (Fig. 3a) suggesting similar molecular

280 structures. One small difference is that GDGT-2 afforded a major fragment ion of

281 m/z 741, which represents two glycerol units with one monocyclic biphytane, while

282 compounds b and d both yielded the fragment of m/z 743 resulting from the loss of one

283 biphytane moiety, and thus representing a fragment comprising two glycerol units with

284 one acyclic biphytane. We also noticed that even though compounds b and d gave the

285 major MS^1 signal of m/z 1298, there were also minor MS^1 signals of m/z 1316 for

286 compound b and m/z 1316 and 1334 for compound d (Fig. 3b). The selected precursor

287 ion m/z 1316 of compound b gave m/z 1298 as its daughter ion MS^2 while a precursor

288 ion m/z 1334 of compound d then generated the daughter ions of m/z 1316 and 1298. The

289 observed mass differences of 18 and 36 Da logically represent loss of one and two

290 hydroxyl groups as H_2O moieties, respectively.

291   Based on the above observation, these two series of unknown compounds a to d

292 were tentatively identified as OH-GDGT-0, -1, -2, where numbers refer to the numbers of

293 cyclopentane rings, and 2OH-GDGT-0. Due to the lack of NMR data on OH-GDGT-1, -2

294 and 2OH-GDGT-0, the exact structure of these compounds cannot be determined

295 unambiguously. However, according to the fragment ion of m/z 743 in the MS^2 spectrum

296 of 2OH-GDGT-0 (Fig. 3a), representing two glycerol units that are ether-bound to one

297 acyclic biphytane, the two hydroxyl groups of 2OH-GDGT-0 are tentatively assigned to a

298 single biphytane moiety. Isomers with different positions of the hydroxyl group(s) may

13

299   also exist but cannot be distinguished with the analytical protocols applied in this study.

300   Compound a, that is OH-GDGT-0, detected around 27 minutes under our

301   chromatographic conditions, is usually the most abundant component among these novel

302   hydroxy-GDGTs, and was therefore isolated for further structural verification with NMR.

303

304   *3.2. Identification of OH-GDGT-0 by NMR*

305

306   Given that the carbon skeleton of OH-GDGT-0 has been previously analyzed with

307   NMR for initial determination of acyclic isoprenoidal GDGT (cf. Langworthy, 1977;

308   Heathcock, et al., 1988; Sinninghe Damsté, et al., 2002), we here focus our discussion of

309   NMR data on locating the extra hydroxyl group, assuming that the carbon skeleton is as

310   previously reported. The extremely small amount of this material required the use of an

311   Agilent inverse z-gradient nanoprobe with a 40 µL sample volume. Only [1]H-detected

312   experiments were possible given the small amount of material, and so ^{13}C shifts of

313   protonated carbon sites were determined (indirectly) using the 2D HSQC spectrum. The

314   2D HSQC spectrum also revealed (just as an APT or DEPT ^{13}C spectrum might) whether

315   a given protonated carbon site had an even or odd number of attached [1]H's. The ^{13}C shifts

316   of non-protonated carbon sites were determined (again indirectly) using the 2D gHMBC

317   NMR spectrum.

318   The extra hydroxyl group is attached to one of the isopranyl chains at a position

319   that was formerly a methine. The evidence for this is two-fold: (1) the ^{1}H NMR

320   resonance arising from the methyl group near the hydroxyl group is a singlet, thus

321   indicating an isolated spin system (these methyl [1]H's do not J-couple to any other [1]H's 3

322 bonds away), and (2) the ^{1}H resonance from the methyl group couples (in the 2D

323 gHMBC NMR spectrum) to a carbon at 71.7 ppm which cannot be the carbon in the

324 propyl end group. There is no plausible chemical-shift-compatible arrangement by which

325 one can position a methyl group on a non-protonated carbon near the methylene group

326 that is part of the propyl end group whose ^{13}C shift is 71.7 ppm (C3' in Table 2 and

327 carbon labels in the numbered molecular skeleton in Fig. 1). No protonated carbon

328 resonances besides one from the propyl end group (C3') appear in the 2D HSQC NMR

329 spectrum with this chemical shift, thus indicating that this carbon shift arises from a non-

330 protonated carbon site. The observed ^{1}H and ^{13}C chemical shifts of the isolated

331 (uncoupled) methyl group are consistent with the proposed structure, which shows this

332 methyl at position B17.

333   Having located the hydroxyl group on a non-protonated carbon site for one of the

334 isopranyl units, we see that the singlet methyl ^{1}H resonance (near the hydroxyl group)

335 correlates with two ^{13}C resonances, one at 43.3 ppm and the other at 40.3 ppm. If the site

336 of the extra hydroxyl group were well removed from the end of the chain (i.e., on the

337 second or third methylated C-atom in the biphytanyl chain), the two adjacent carbon

338 resonances for the methylene groups (CH$_2$'s) adjacent to the non-protonated (OH-bearing)

339 carbon would be nearly the same. This cannot explain the observed difference of more

340 than 3 ppm, unless the hydroxyl group is located at C-3 in one of the biphytanyl chains.

341 And the 40.3 and 43.3 ppm chemical shifts are comparable to the reported values of C-2

342 and C-4 (39.78 and 42.92 ppm, Ferrante et al., 1988a) in the C-3 hydroxylated *sn*-3

343 hydroxyarchaeol. A second possibility is that the hydroxyl group resides near the middle

344 of the chain, i.e., at a methylated C-atom adjacent to the head-to-head isopranyl linkage.

345   If this were the case, however, the six carbon resonances arising from the end groups

346   would show only three unique chemical shifts, instead of the six we observe. That is, the

347   symmetry-disrupting hydroxyl group must be near one end of the biphytanyl moiety in

348   order for the chemical shifts of the two ends to be distinct.

349   If the GDGT skeleton is a trans regioisomer, there will be two different C-3

350   positions (*sn*-2 and *sn*-3) in the molecule. Sprott et al. (1990) compared the ^{13}C NMR

351   shifts of archaeol, *sn*-2 and *sn*-3 hydroxyarchaeol. It was shown that in these three

352   compounds the *sn*-2 and *sn*-3 ether-bonded C-1 possessed different chemical shifts. There

353   are around 3 ppm difference between the *sn*-2 and *sn*-3 ether-bonded C-1 in *sn*-2

354   hydroxyarchaeol, 2 ppm difference in archaeol but nearly identical in *sn*-3

355   hydroxyarchaeol (see Fig. 4 in Sprott et al., 1990). Accordingly, if the hydroxyl group of

356   OH-GDGT-0 is located at the *sn*-2 O-biphytanyl at position C-3, there will be four

357   distinct C-1 signals. The analyzed compound generated three very similar ^{13}C NMR

358   shifts at 68.5~68.8 ppm and one at 69.9 ppm, therefore, we propose that the hydroxyl

359   group of this OH-GDGT-0 is located at the *sn*-3 O-biphytanyl moiety at position C-3.

360   Chemical shift prediction of the ^{13}C NMR shifts using ACD shift prediction

361   software was used as the starting point for the complete assignment of *sn*-3 OH-GDGT-0.

362   The observed chemical shifts of analyzed OH-GDGT-0 (Table 2) agrees well with the

363   values predicted by the ACD software and previously reported ^{13}C NMR shifts of

364   corresponding carbons in hydroxyarchaeol (Ferrante et al., 1988a) and GDGT-0

365   (Sinninghe Damsté, et al., 2002). The only exception is that a group of cross peaks in the

366   HSQC NMR spectrum showed a number of methylene groups with ^{13}C shifts near 30

367   ppm. The unusual methylene signals were assigned to the A4, the A4', and the B4' sites.

368   In previous works, the carbon shifts of C-4 methylene of hydroxyarchaeol (Ferrante et al.,

369   1988a) and GDGT-0 (Sinninghe Damsté, et al., 2002) were shown to be ~37 ppm, only

370   the methylene within a cyclopentane moiety gave ~30 ppm shift (e.g., the C-9 sites of

371   crenarchaeol, Sinninghe Damsté, et al., 2002). However, the possibility of sample

372   impurity had been ruled out by the purity test of our NMR sample with LC-MS.

373   Additional GC-MS analysis on ether cleavage released hydrocarbons of the NMR sample

374   detected only acyclic biphytane, which indicated a reduction of the tertiary hydroxyl

375   group during the ether cleavage reaction, but confirmed the acyclic biphytane structure of

376   OH-GDGT-0. Therefore, this mysterious 30 ppm methylene signal was tentatively

377   ascribed to a local conformational effect on the ends of the molecule caused by the

378   hydrophilic nature of the end groups relative to the bulk of the molecule plus the solvent

379   (benzene-d_6). It is also possible that the methylene groups with ^{13}C chemical shifts of 30

380   ppm are caused by the A16, A16', B16, and B16' sites, perhaps due to a hairpin-turn

381   conformation as opposed to a more elongated configuration. The appeal of this

382   interpretation is that the predicted ^{13}C chemical shift for the "16" carbon sites (34 ppm) is

383   closer to the observed methylene ^{13}C shifts (30 ppm) than the ^{13}C chemical shifts

384   predicted for the "4" sites (37-38 ppm). It was not possible to use the gCOSY or the

385   gHMBC spectra to resolve this question because of the extensive resonance overlap in

386   the 1.25-1.38 ppm shift region of the ^{1}H NMR spectrum.

387

388   *3.3. Occurrence of hydroxy-GDGT isomers*

389

390   Summons et al. (2002) reported the tentative structures of biphytane-N-ol where

391   N refers to C-atom 3 or 7 or 11 or 15, however, we observed isomers of hydroxy-GDGTs

392   with presumably different hydroxyl group positions. After acid hydrolysis, the TLE of

393   *Methanothermococcus thermolithotrophicus* yielded two compounds with identical MS[1]

394   and MS[2] spectra (compounds (3) and (4) in Fig. 4A and B). Compound (4) is the isolated

395   OH-GDGT-0 from marine sediment, so that the earlier eluting compound (3) could be an

396   isomer of OH-GDGT-0, in which the hydroxyl group is located at another terminal

397   tertiary C-atom in the biphytanyl chain (structures shown in Fig. 1 and 4). Theoretically,

398   if the hydroxylation only occurred on the C-3 position, two isomers can be formed for the

399   trans regioisomer OH-GDGT-0, four isomers for OH-GDGT-1, two for OH-GDGT-2

400   (assuming fixed cyclopentane ring position) and three for the 2OH-GDGT-0 (see Fig. 1

401   for the putative structures). In this work we only detected isomers of OH-GDGT-0, while

402   the presence of hydroxy-GDGT isomers with cycloalkyl moieties needs to be verified

403   with further study. In addition, hydroxy-GDGTs with more than two hydroxyl groups

404   may also exist.

405

406   *3.4. Identification of glycosidic hydroxy-GDGT*

407

408   A number of studies of IPLs in marine samples and archaeal cultures have

409   reported an unknown intact polar GDGT showing a neutral loss of a fragment that is

410   18 Da heavier than the dihexose moieties in the MS[2] spectrum (Sturt et al., 2004;

411   Schouten et al., 2008; Schubotz et al., 2009; Lipp and Hinrichs, 2009; Rossel et al., 2011;

412   Pitcher et al., 2011). Informally labeled H341-GDGT by Lipp and Hinrichs (2009), this

413   unknown lipid was a major archaeal IPL in a wide range of marine sediments (Lipp and

414   Hinrichs, 2009; Rossel et al., 2011) and was also observed in the TLE of the

415   Crenarchaeon *Candidatus 'Nitrosopumilus maritimus'* (Schouten et al., 2008; Pitcher et

416   al., 2011). In IPL fractions obtained by preparative LC of twelve marine sediment

417   samples, we detected IPLs with molecular weights 18 Da and 36 Da higher than mono-,

418   di- and triglycosyl GDGTs (Fig. 5).

419   Fragmentation behavior of H341-GDGT was tested under 100% and 25%

420   collision energy of the ion trap mass spectrometer (Fig. 6). With 100% collision energy,

421   the $[M+NH_4]^+$ ion of H341-GDGT (m/z 1660.4) generated only one dominant fragment

422   ion of m/z 1300 (m/z 1299.8 in Fig. 6a). However, with a more moderate 25% collision

423   energy a more detailed fragmentation behavior was revealed and this included the loss of

424   H_2O and sugar units (Fig. 6b). Together with the $[M+NH_4]^+$ ion of m/z 1660.4 there are

425   also fragment ions of m/z 1641.8, 1479.8, 1461.5, 1318.3 and 1300.0. The 180.6 Da

426   difference between m/z 1660.4 and m/z 1479.8 represents the loss of one hexose moiety

427   along with an adducted ammonium. The fragment ion of m/z 1318.3 was then assigned to

428   the core lipid of OH-GDGT-0, and 18 Da difference between m/z 1318 and 1300

429   represents the loss of H_2O. Therefore, this unknown IPL, H341-GDGT, was tentatively

430   identified as a hydroxy-GDGT bearing two glycosidic head groups. Likewise, all other

431   IPLs related to 1Gly-, 2Gly- and 3Gly-GDGT with mass differences of 18 Da and 36 Da

432   were tentatively identified as glycosidic mono- and dihydroxy-GDGTs (see molecular

433   structures provided in Fig. 5).

434   The relative abundances of these glycosidic hydroxy-GDGTs varied within

435   different samples and the distributions of intact polar GDGTs in analyzed samples are

436 shown in Table 1. 2Gly-OH-GDGT was generally the most common hydroxy-GDGT IPL,

437 which is consistent with earlier observations (cf. Lipp and Hinrichs, 2009; Schubotz et al.,

438 2009).

439

440 *3.5. Distribution and abundance of hydroxy-GDGT core lipids*

441

442   Hydroxy-GDGT core lipids occur widely and were present in all marine sediment

443 samples analyzed in this study. With our present analytical protocol it is difficult to gain

444 their absolute abundances due to their inherent dehydration during ionization. However,

445 the relative abundance of hydroxy-GDGTs compared to total GDGTs (100% x [hydroxy-

446 GDGTs]/[hydroxy-GDGTs + non-hydroxy-GDGTs]) was estimated with the APCI-MS

447 detection of both non-hydroxy-GDGT and hydroxy-GDGT core lipids in the same

448 analysis (data shown in Table 1). These ratios are based on the peak areas of $[M+H]^+$ ions

449 of GDGTs, $[M+H-18]^+$ ions of OH-GDGTs and $[M+H-36]^+$ ion of 2OH-GDGT-0, for

450 example, m/z 1298 for GDGT-2, OH-GDGT-1 and 2OH-GDGT-0 (Fig. 2). This

451 calculation does not provide the real abundance of hydroxy-GDGTs but a rough

452 estimation because the detected signal represents only ions that result from dehydration.

453 Nevertheless, the ratios are useful for comparing relative distributions between samples.

454 Using this approach for core lipids fractions from marine sediments, the estimated

455 abundances of hydroxy-GDGTs ranged from 1% to 8% of the total detected isoprenoid

456 GDGT core lipids, while 2OH-GDGT-0 generally comprises very small proportion

457 (Table 1). The widespread occurrence and remarkable abundance of hydroxy-GDGTs in

458 the marine environment renders these compounds an important target for future studies of

459 archaeal lipids.

460

461 *3.6. Source of hydroxy-GDGT*

462

463   In all analyzed marine sediments the ring composition of hydroxy-GDGT core

464 lipids differs markedly from that of non-hydroxylated GDGTs, suggesting that these two

465 lipid classes may be attributed to either distinct source organisms or production under

466 different environmental stresses. As it is shown in Fig. 2, GDGTs consist of 0-, 1-, 2-, 3-

467 and 5-ring structures (GDGT-4 was not taken into account because of the co-elution with

468 crenarchaeol and the interference of the $[M+H+2]^+$ signal of crenarchaeol) and with

469 GDGT-0 and crenarchaeol as the dominant components, however, there are only 0-, 1-

470 and 2-ring structures detected in hydroxy-GDGTs, and with the acyclic structure as the

471 most abundant component in both mono- and dihydroxy-GDGTs.

472   We also detected low abundant free core lipids of GDGT-0 and hydroxy-GDGT-0

473 in the TLE of *Methanothermococcus thermolithotrophicus*; the corresponding IPLs were

474 2Gly- and 3Gly-OH-GDGT (Table 1). Schouten et al. (2008) reported a 2Gly-GDGT

475 with a molecular mass elevated by 18 Da in *Candidatus 'Nitrosopumilus maritimus'*, a

476 marine group I crenarchaeon. We propose that this compound, observed as the same as

477 the H341-GDGT in a wide range of marine sediments, is 2Gly-OH-GDGT. The presence

478 of hydroxy-GDGTs in these two archaeal cultures showed that the biosynthesis of

479 hydroxy-GDGTs is not limited to either crenarchaeota or euryarchaeota (*M.*

480 *thermolithotrophicus*). However, the sources of hydroxy-GDGTs in marine sediments

481 may be as complex as the archaeal diversity (cf. Teske and Sørensen, 2008) and remain

482 to be constrained.

483

484 *3.7. Potential degradation of hydroxy-GDGTs during sample preparation*

485

486   Acid hydrolysis is a common approach to obtain core GDGTs from the IPLs or

487 biomass. We observed dehydrated by-products of hydroxy-GDGTs upon acid treatment.

488 With a simple experiment of acid hydrolysis on purified OH-GDGT-0 (Fig. 4D) and the

489 TLE of *Methanothermococcus thermolithotrophicus* (Fig. 4A and B), two major

490 dehydrated products resulting from hydrolyzed OH-GDGT-0 were detected (Fig. 4D).

491 The first dehydrated product, designated compound (1) in Fig. 4, which nearly co-eluted

492 with GDGT-1, showed the same molecular mass of 1300 Da as GDGT-1 and virtually

493 identical fragmentation behavior in MS^2. Compound (1) was assumed to be an acyclic

494 GDGT with one double bond as depicted in Fig. 4.

495   The second dehydration product, labeled as compound (2) in Fig. 4B and D, has

496 also the molecular mass of 1300 Da and reacts similarly in MS^2 as GDGT-1, but has a

497 higher polarity as indicated by its elevated retention time. It is questionable whether a

498 different double bond position, as could result from dehydration of OH-GDGT, would

499 cause this shift in retention time. We suggest that the compounds reported by Pitcher et al.

500 (2011) as GDGTs with head groups of 180 and 2 x 180 Da in marine sediment

501 enrichments were possibly glycosidic hydroxy-GDGTs. In their core lipid fractions

502 recovered from acid hydrolyzed lipid extracts, these authors observed both earlier and

503 later eluting 'isomers' of GDGT-1, -2 and -3. Considering the LC conditions used by

504   Pitcher et al. (2011), their later eluting 'isomers' could be the second dehydrated product

505   (compound (2)) and other analogues derived from OH-GDGT-0, -1 and -2.

506   Interestingly, we also noticed that compound (2) nearly co-elutes with H-shaped

507   GDGT-0. On the basis of its MS^2 spectrum, the structure of this compound cannot be

508   fully determined but based on its mass spectrometric fragmentation pattern we can

509   confidently exclude that it is a H-shaped structure. H-shaped GDGT-0 was detected in

510   only one of our marine sediment samples, ODP201 1229A 22H1 (Fig. 4E and F). The

511   same molecular mass of 1300 Da, and co-elution of compound (2) and H-shaped GDGT-

512   0 points to the danger of identifying H-shaped GDGT only on the basis of retention time

513   and molecular mass, if the lipid extracts were acid hydrolyzed.

514   In analogy to hydroxy-GDGTs, hydroxyarchaeol with its tertiary hydroxyl group

515   apparently degraded rapidly after *in-situ* production in methane-laden marine sediment

516   (Aquilina et al., 2010). Due to the labile nature of tertiary alcohols (cf., Nishihara and

517   Koga, 1991; Hinrichs et al., 2000), degradation of hydroxy-GDGT to its unsaturated

518   analogues in sedimentary environments needs to be evaluated and considered in future

519   work.

520

521   **4. Conclusion**

522

523   We identified OH-GDGT-0 on the basis of its mass spectrometric behavior and

524   NMR analysis. 2D-NMR analysis indicated hydroxylation at C-3 of the *sn*-3 biphytanyl

525   moiety. Combined with the tentative identification of monohydroxy-GDGTs with one

526   and two rings and 2OH-GDGT-0, this study has extended the known diversity of archaeal

527 lipids in marine sediments. These widespread hydroxy-GDGTs are subject to in-source

528 dehydration during the commonly applied protocols for LC-APCI-MS analysis of

529 GDGTs with the resulting dehydrated fragment showing highly similar mass spectral

530 fragmentation as the corresponding cycloalkyl-bearing molecular ion with identical mass;

531 this behavior may result in erroneous identification of the hydroxylated GDGTs. We also

532 identified the corresponding IPLs, comprising of a series of glycosidic hydroxy-GDGTs

533 that had been reported as unknowns in several previous studies.

534   Based on the occurrence of hydroxy-GDGTs in twelve widely distributed marine

535 sediments, we suggest that these lipids are as ubiquitous and widespread in marine

536 sediments as their non-hydroxylated analogues. In the samples analyzed in this study the

537 relative abundance of hydroxy-GDGT core lipids is substantial, up to 8% of total

538 isoprenoidal core GDGTs.

539   In marine sediments hydroxy-GDGTs contain up to two rings with the acyclic

540 compound being the most abundant component. Their biological sources in the

541 environment remain to be constrained. However, we identified hydroxy-GDGTs in one

542 methanogen, *Methanothermococcus thermolithotrophicus*, and based on our

543 interpretation of previously published findings (Schouten et al., 2008), we suggest that

544 2Gly-OH-GDGT is also present in *Candidatus 'Nitrosopumilus maritimus'*.

545 Consequently, it appears that both crenarchaeota and euryarchaeota are able to synthesize

546 this compound. Both the biophysical significance of hydroxylated tetraether lipids and

547 their ecological role need to be clarified in future biological studies. Tentatively, the

548 hydroxylation of the biphytanyl moiety may result in enhanced membrane rigidity due to

24

549   the potential for hydrogen bonding between the C-3 hydroxyl proton and the *sn*-2 and *sn*-

550   3 glycerol O-atoms

551

570

571

572   **References**

573

574   Aquilina, A., Knab, N. J., Knittel, K., Kaur, G., Geissler, A., Kelly, S. P., Fossing, H.,
575   Boot, C. S., Parkes, R. J., Mills, R. A., Boetius, A., Lloyd, J. R., Pancost, R. D.
576   (2010) Biomarker indicators for anaerobic oxidizers of methane in brackish–
577   marine sediments with diffusive methane fluxes. *Org. Geochem.* 41, 414–426.

578   Biddle J. F., Lipp J. S., Lever M. A., Lloyd k. G., Sørensen K. B., Erson R., Fredricks H.
579   F., Elvert M., Kelly T. J., Schrag D. P., Sogin M. L., Brenchley J. E., Teske A.,
580   House C. H. and Hinrichs K. -U. (2006) Heterotrophic archaea dominate
581   sedimentary subsurface ecosystems off Peru. *Proc. Natl. Acad. Sci. USA* 103,
582   3846–3851.

583   Comita, P. B. and Gagosian, R. B. (1983) Membrane lipid from deep-sea hydrothermal
584   vent methanogen: a new macrocyclic glycerol diether. *Science* 23, 1329–1331.

585   DeLong, E. F., King, L. L., Massana, R., Cittone, H., Murray, A., Schleper, C. and
586   Wakeham, S. G. (1998) Dibiphytanyl ether lipids in nonthermophilic
587   crenarchaeotes. *Appl. Environ. Microbiol.* 64, 133–138.

588   De Rosa, M., Gambacorta, A., Nicolaus, B. and Bu'Lock, J. D. (1980) Complex lipids of
589   *Caldariella acidophila*, a thermoacidophilic archaebacterium. *Phytochemistry* 19,
590   821–825.

591   Ferrante, G., Ekiel, I., Patel, G. B. and Sprott, G. D. (1988a) A novel core lipid isolated
592   from the aceticlastic methanogen, *Methanothrix concilii* GP6. *Biochim. Biophys.*
593   *Acta* 963, 173–182.

594   Ferrante, G., Ekiel, I., Patel, G. B. and Sprott, G. D. (1988b) Structure of the major polar
595   lipids isolated from the aceticlastic methanogen, *Methanothrix concilii* GP6.
596   *Biochim. Biophys. Acta* 963, 162–172.

597   Gonthier, I., Rager, M. -N., Metzger, P., Guezennec, J. and Largeau, C. (2001) A di-O-
598   dihydrogeranylgeranyl glycerol from *Thermococcus* S 557, a novel ether lipid,
599   and likely intermediate in the biosynthesis of diethers in Archæa. *Tetrahedron*
600   *Lett.* 42, 2795–2797.

601   Heathcock, C. H., Finkelstein, B. L., Jarvi, E. T., Radel, P. A. and Hadley, C. R. (1988)

602   1,4- and 1,5- stereoselection by sequential aldol addition to α, β- unsaturated

603   aldehydes followed by Claisen rearrangement. Application to total synthesis of

604   the vitamin E side chain and the archaebacterial C_{40} diol. *J. Org. Chem.* 53, 1922–

605   1942.

606   Hinrichs, K. -U., Hayes, J. M., Sylva, S. P., Brewer, P. G. and De Long, E. F. (1999)

607   Methane-consuming archaebacteria in marine sediments. *Nature* 398, 802–805.

608   Hinrichs, K. -U., Summons, R. E., Orphan, V., Sylva, S. P. and Hayes, J. M. (2000)

609   Molecular and isotopic analysis of anaerobic methane-oxidizing communities in

610   marine sediments. *Org. Geochem.* 31, 1685–1701.

611   Hopmans, E. C., Schouten, S., Pancost, R. D., van der Meer, M. T. J. and Sinninghe

612   Damsté, J. S. (2000) Analysis of intact tetraether lipids in archaeal cell material

613   and sediments by high performance liquid chromatography/atmospheric pressure

614   chemical ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* 14, 585–

615   589.

616   Kates, M., Sastry, P. S. and Yengoyan L. S. (1963) Isolation and characterization of a

617   diether analog of phosphatidylglycerophosphate from *Halobacterium cutirubrum*.

618   *Biochim. Biophys. Acta* 70, 705–707.

619   Koga, Y., Nishihara, M., Morii, H. and Akagawa-Matsushita, M. (1993) Ether Polar

620   Lipids of Methanogenic Bacteria: Structures, Comparative Aspects, and

621   Biosyntheses. *Microbiol. Rev.* 57, 164–182.

622   Kuypers, M. M. M., Blokker, P., Erbacher, J., Kinkel, H., Pancost, R. D., Schouten, S.

623   and Sinninghe Damsté, J. S., (2001) Massive expansion of marine Archaea during

624   a mid-Cretaceous oceanic anoxic event. *Science* 293, 92-95.

625   Langworthy, T. A. (1977) Long-chain diglycerol tetraethers from Thermoplasma

626   acidophilum. *Biochim. Biophys. Acta* 487, 37–50.

627   Lipp, J. S. and Hinrichs, K. -U. (2009) Structural diversity and fate of intact polar lipids

628   in marine sediments. *Geochim. Cosmochim. Acta* 73, 6816–6833.

629   Liu, X. -L., Lipp, J. S. and Hinrichs, K. -U. (2011) Distribution of intact and core GDGTs

630   in marine sediments. *Org. Geochem.* 42, 368-375.

631 Liu, X. -L., Lipp, J. S., Schröder, J. M., Summons, R. E. and Hinrichs, K. -U. (2012)
632   Isoprenoidal glycerol dialkanol diethers: a series of novel archaeal lipids in
633   marine sediments. *Org. Geochem.* 43, 50-55

634 Nemoto, N., Shida, Y., Shimada, H., Oshima, T. and Yamagishi, A. (2003)
635   Characterization of the precursor of tetraether lipid biosynthesis in the
636   thermoacidophilic archaeon Thermoplasma acidophilum. *Extremophiles* 7, 235–
637   243.

638 Nishihara, M. and Koga, Y. (1991) Hydroxyarchaetidylserine and hydroxyarchaetidyl-
639   *myo*-inositol in *Methanosarcina barkeri*: polar lipids with a new ether core
640   portion. *Biochim. Biophys. Acta* 1082, 211–217.

641 Pitcher, A., Hopmans, E. C., Mosier, A. C., Park, S. -J., Rhee, S. -K., Francis, C. A.,
642   Schouten, S. and Sinninghe Damsté. J. S. (2011) Core and Intact Polar Glycerol
643   Dibiphytanyl Glycerol Tetraether Lipids of Ammonia-Oxidizing Archaea
644   Enriched from Marine and Estuarine Sediments. *Appl. Environ. Microbiol.*
645   doi:10.1128/AEM.02758-10

646 Rossel, P. E., Lipp, J. S., Fredricks, H. F., Arnds, J., Boetius, A., Elvert, M. and Hinrichs,
647   K. -U. (2008) Intact polar lipids of anaerobic methanotrophic archaea and
648   associated bacteria. *Org. Geochem.* 39, 992–999.

649 Rossel, P. E., Elvert, M., Ramette, A., Boetius, A. and Hinrichs, K. -U. (2011) Factors
650   controlling the distribution of anaerobic methanotrophic communities in marine
651   environments: evidence from intact polar lipids. *Geochim. Cosmochim. Acta* 75,
652   164–184.

653 Schouten, S., Hopmans, E. C., Bass, M., Boumann, H., Standfest, S., Könneke, M., Stahl,
654   S. A. and Sinninghe Damste, J. S. (2008) Intact membrane lipids of "*Candidatus*
655   Nitrosopumilus maritimus", a cultivated representative of the cosmopolitan
656   mesophilic group I Crenarchaeota. *Appl. Environ. Microbiol.* 74, 2433–2440.

657 Schouten, S., Middelburg, J. J., Hopmans, E. C. and Sinninghe Damsté, J. S. (2010)
658   Fossilization and degradation of intact polar lipids in deep subsurface sediments:
659   A theoretical approach. *Geochim. Cosmochim. Acta* 74, 3806–3814.

660 Schubotz, F., Wakeham, S. G., Lipp, J. S., Fredricks, H. F. and Hinrichs, K. –U. (2009)
661   Detection of microbial biomass by intact polar membrane lipid analysis in the

662   water column and surface sediments of the Black Sea. *Environ*. *Microbiol*. 11,

663   2720–2734.

664 Sinninghe Damsté, J. S., Schouten, S., Hopmans, E. C., van Duin, A. C. T. and

665   Geenevasen, J. A. J. (2002) Crenarchaeol: the characteristic core glycerol

666   dibiphytanyl glycerol tetraether membrane lipid of cosmopolitan pelagic

667   crenarchaeota. *J*. *Lipid Res*. 43, 1641–1651.

668 Sprott, G. D., Ekiel, I. and Dicaire, C. (1990) Novel, Acid-labile, Hydroxydiether Lipid

669   Cores in Methanogenic Bacteria. *J*. *Biol*. *Chem*. 265, 13735-13740.

670 Sturt, H. F., Summons, R. E., Smith, K., Elvert, M. and Hinrichs, K. -U. (2004) Intact

671   polar membrane lipids in prokaryotes and sediments deciphered by high-

672   performance liquid chromatography/electrospray ionization multistage mass

673   spectrometry-new biomarkers for biogeochemistry and microbial ecology. *Rapid*

674   *Commun*. *Mass Spectrom*. 18, 617–628.

675 Summons, R. E., Embaye, T., Jahnke, L. L. and Baumgartner, M. (2002) Molecular

676   Signatures of Methanogens in Cultures and Environmental Samples. *American*

677   *Geophysical Union, Fall Meeting 2002*, abstract #B51B-0723.

678 Teske, A. and Sørensen, K.B. (2008) Uncultured archaea in deep marine subsurface

679   sediments: have we caught them all? *The ISME J*. 2, 3–18.

680 Tornabene, T.G. and Langworthy, T.A. (1979) Diphytanyl and dibiphytanyl glycerol

681   ether lipids of methanogenic archaea. *Science* 203, 51–53.

682

683

Figure captions

685

686 **Figure 1.** Molecular structures of major compounds mentioned in this report. Numbered

687 molecular structure of isolated compound (OH-GDGT-0 isomer-1) shows the carbon

688 labels used in Table 2. The occurrence of hydroxy-GDGT isomers is based on the

689 assumption that hydroxylation only occurred on the C-3 position of biphytane with fixed

690 cyclopentane ring position.

691

692 **Figure 2**. HPLC-APCI-MS chromatogram generated by the Agilent 6130 MSD single

693 quadrupole mass spectrometer, showing GDGTs, GDDs (Liu et al., 2012) and hydroxy-

694 GDGTs (compound a, b, c and d) in the core lipid fraction of GeoB 12806-2. Dehydrated

695 ions of monohydroxy-GDGTs (compound a, b, c) and dihydroxy-GDGT (compound d)

696 are displayed together with GDGTs in a separate mass window of extracted ion

697 chromatogram (EIC). The peak of 2OH-GDGT-0 is highlighted with the dashed circle.

698

699 **Figure 3**. MS^2 mass spectra from qTOF in centroid mode showing the major fragment

700 ions of selected precursor m/z 1298 (a), and MS^1 mass spectra in profile mode showing

701 the 'in-source' dehydration of OH-GDGT-1 and 2OH-GDGT-0 (b).

702

703 **Figure 4**. HPLC-APCI-MS chromatogram, generated by the Agilent 6130 MSD single

704 quadrupole mass spectrometer in SIM mode, showing the detection of compounds in

705 hydrolyzed TLE of *Methanothermococcus thermolithotrophicus*: (A) base peak

706 chromatogram (BPC) and (B) EIC, m/z 1300; in purified OH-GDGT-0: (C) BPC; in acid

707 hydrolyzed OH-GDGT-0: (D) EIC, m/z 1300 and in sample ODP201 1229A 22H1 core

708 lipid fraction: (E) BPC, (F) EIC, m/z 1300. Compounds (1) and (2) are two major

709 dehydration products of OH-GDGT-0, (3) and (4) are two isomers of OH-GDGT-0.

710 Possible molecular structures of (1), (3) and (4) are provided while the structure of

711 compound (2) remains ambiguous.

712

713 **Figure 5**. Density maps generated by ESI-ion trap-MS analysis showing the detection of

714 glycosidic GDGTs and hydroxy-GDGTs in the IPL fractions of samples IODP 311

715   1327C 10H5, ODP 201 1226B 10H3 and M76/1 GeoB 12807-2. Tentative molecular

716   structures of compounds are provided.

717

718   **Figure 6.** MS^2 mass spectra of 2Gly-OH-GDGT-0 in sample ODP 201 1229D 4H4

719   generated by ESI-ion trap-MS analysis. At 100% collision energy (upper panel), the

720   neutral loss of two hexose moieties together with the hydroxyl group (as H_2O) from the

721   biphytanyl moiety, and under 25% collision energy (lower panel) the successive losses of

722   hexoses and the biphytanyl-derived hydroxyl group.

723

724

725

Table 1. Samples analyzed. Detailed information on the twelve marine sediment samples has been published in Liu et al., (2011). Ratio of hydroxy-GDGT vs. the total core GDGT was calculated based on the detection of hydroxy-GDGT with APCI-MSD single quadrupole mass spectrometer.

730

Samples		OH-GDGT/total Core GDGT (%)	2OH-GDGT/total Core GDGT (%)	Identified Intact Polar GDGTs
M76/1 Namibia Margin	GeoB 12806-2	6.0	0.06	1Gly-, 2Gly-GDGT; 1Gly-OH-GDGT
	GeoB 12807-2	6.1	n.d.	1Gly-, 2Gly-GDGT; 1Gly-, 2Gly-OH-GDGT
ODP201 Peru Margin	1229D 4H4	7.2	0.11	2Gly-GDGT; 2Gly-, 3Gly-OH-GDGT
	1229A 22H1	5.3	0.06	2Gly-GDGT
ODP201 Equatorial Pacific	1226B 10H3	4.7	1.72	2Gly-, 3Gly-GDGT; 2Gly-, 3Gly-OH-GDGT; 2Gly-2OH-GDGT
	1226E 20H3	2.2	0.25	1Gly-, 2Gly-GDGT; 2Gly-OH-GDGT
ODP204 Hydrate Ridge	1250D 6H5	5.8	0.26	2Gly-GDGT; 2Gly-OH-GDGT
	1250D 12H5	6.3	0.37	2Gly-, 3Gly-GDGT; 2Gly-, 3Gly-OH-GDGT
IODP311 Cascadia. Margin	1327C 10H5	7.9	0.47	2Gly-, 3Gly-GDGT; 2Gly-, 3Gly-OH-GDGT; 3Gly-2OH-GDGT
	1327C 13X6	6.1	0.25	2Gly-GDGT; 2Gly-OH-GDGT
ODP160 Mediterranean Sapropel	966C 5H02	2.7	n.d.	2Gly-, 3Gly-GDGT; 2Gly-, 3Gly-OH-GDGT
	966C 7H04	1.2	0.07	2Gly-GDGT; 2Gly-OH-GDGT
Methanothermococcus thermolithotrophicus		14.3	n.d.	1Gly-, 2Gly, 3Gly-GDGT; 2Gly-, 3Gly-OH-GDGT

731

732   (n.d. not detected)

733

734

735

736 **Table 2.** ^{13}C- and ^{1}H-NMR data of OH-GDGT-0

737

Label	^{1}H shift (ppm)	^{13}C shift (ppm)	Label	^{1}H shift (ppm)	^{13}C shift (ppm)
A1	3.44, 3.56	68.5	A1'	3.38, 3.40	69.9
A2	1.41, 1.67	37.5	A2'	1.38, 1.67	37.0
A3	1.65	30.1	A3'	1.65	30.1
A4	1.33, 1.37	30.1	A4'	1.33, 1.37	30.1
A5	1.35, 1.38	24.9	A5'	1.35, 1.38	24.9
A6	1.21, 1.42	37.8	A6'	1.21, 1.42	37.8
A7	1.50	33.1	A7'	1.50	33.1
A8	1.21, 1.42	37.8	A8'	1.21, 1.42	37.8
A9	1.35, 1.48	24.9	A9'	1.35, 1.38	24.9
A10	1.21, 1.42	37.8	A10'	1.21, 1.42	37.8
A11	1.50	33.1	A11'	1.50	33.1
A12	1.21, 1.42	37.8	A12'	1.21, 1.42	37.8
A13	1.35, 1.48	24.9	A13'	1.35, 1.38	24.9
A14	1.21, 1.42	37.8	A14'	1.21, 1.42	37.8
A15	1.47	33.4	A15'	1.47	33.4
A16	1.25, 1.42	34.7	A16'	1.25, 1.42	34.7
A17	0.94	19.9	A17'	0.94	19.9
A18	0.99	20.0	A18'	0.99	20.0
A19	0.99	20.0	A19'	0.99	20.0
A20	0.98	19.6	A20'	0.98	19.6
B1	3.48, 3.50	68.8	B1'	3.50, 3.64	68.5
B2	1.57, 1.73	40.3	B2'	1.41, 1.65	37.5
B3	-	71.7	B3'	1.65	30.1
B4	1.49, 1.56	43.3	B4'	1.33, 1.37	30.1
B5	1.49, 1.53	21.8	B5'	1.35, 1.48	24.9
B6	1.22, 1.42	38.2	B6'	1.21, 1.42	37.8
B7	1.50	33.1	B7'	1.50	33.1
B8	1.21, 1.42	37.8	B8'	1.21, 1.42	37.8
B9	1.35, 1.48	24.9	B9'	1.35, 1.48	24.9
B10	1.21, 1.42	37.8	B10'	1.21, 1.42	37.8
B11	1.50	33.1	B11'	1.50	33.1
B12	1.21, 1.42	37.8	B12'	1.21, 1.42	37.8
B13	1.35, 1.48	24.9	B13'	1.35, 1.48	24.9
B14	1.21, 1.42	37.8	B14'	1.21, 1.42	37.8
B15	1.47	33.4	B15'	1.47	33.4
B16	1.25, 1.42	34.7	B16'	1.25, 1.42	34.7
B17	1.21	27.1	B17'	0.99	20.0
B18	0.99	20.0	B18'	0.99	20.0
B19	0.99	20.0	B19'	0.99	20.0
B20	0.98	19.6	B20'	0.98	19.6
C1	3.54, 3.59 (1.43 OH)	62.4	C1'	3.67, 3.72 (2.05 OH)	63.0
C2	3.33	79.1	C2'	3.51	79.2
C3	3.36, 3.41	71.5	C3'	3.43, 3.51	71.7

738

739 Carbon labels in table refer to the numbered molecule in Fig. 1.

740

741
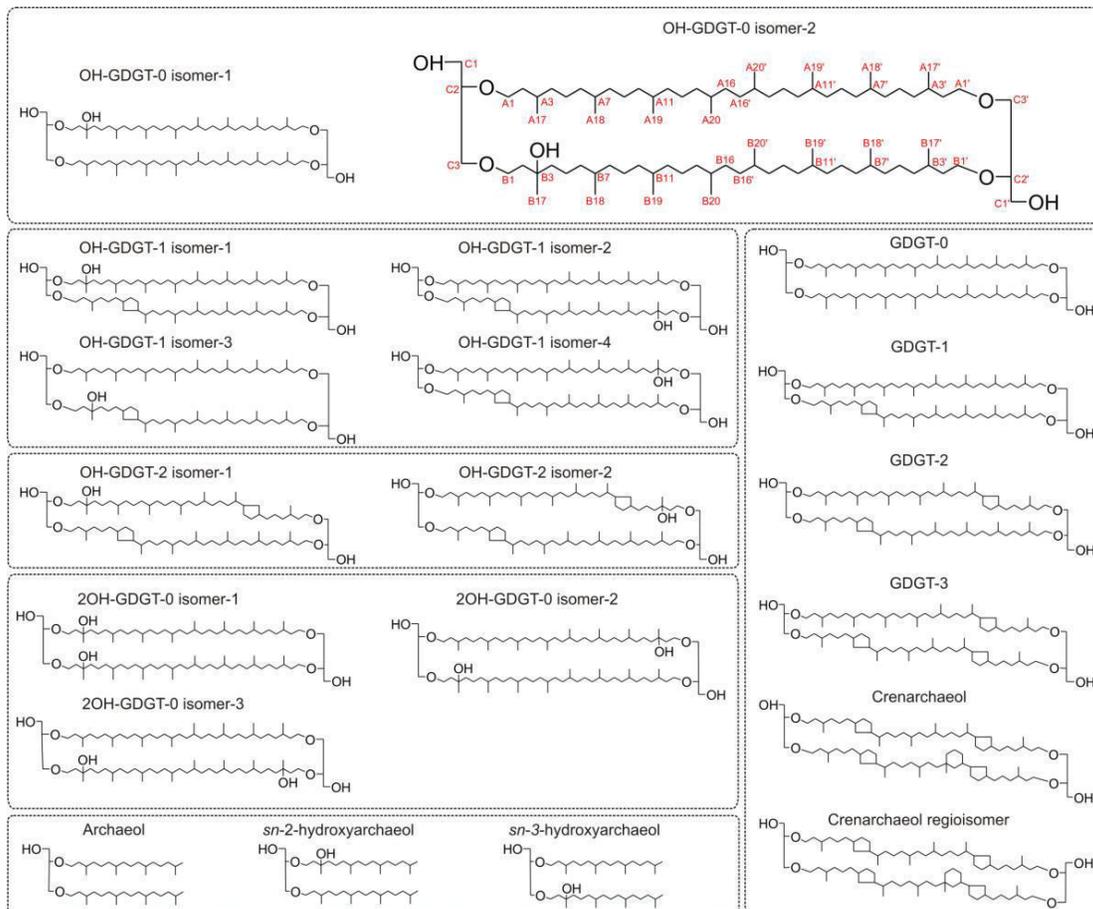
742
743
744 **Figure 1**.
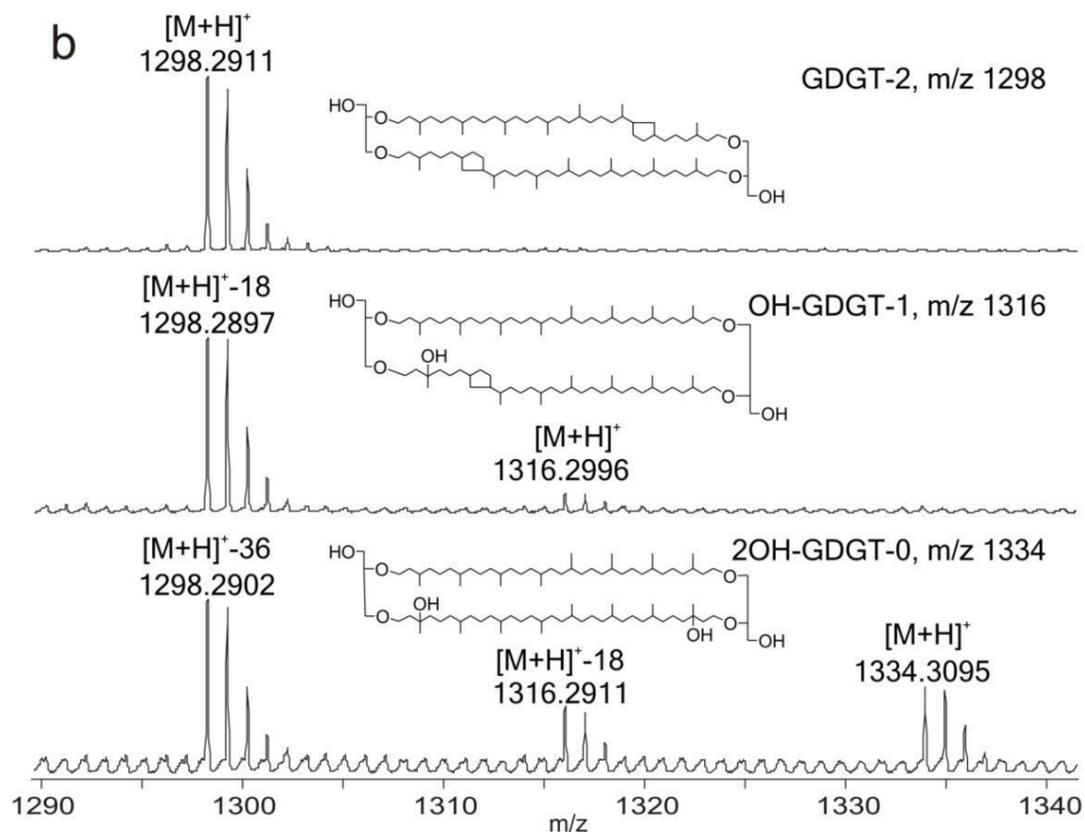745
746
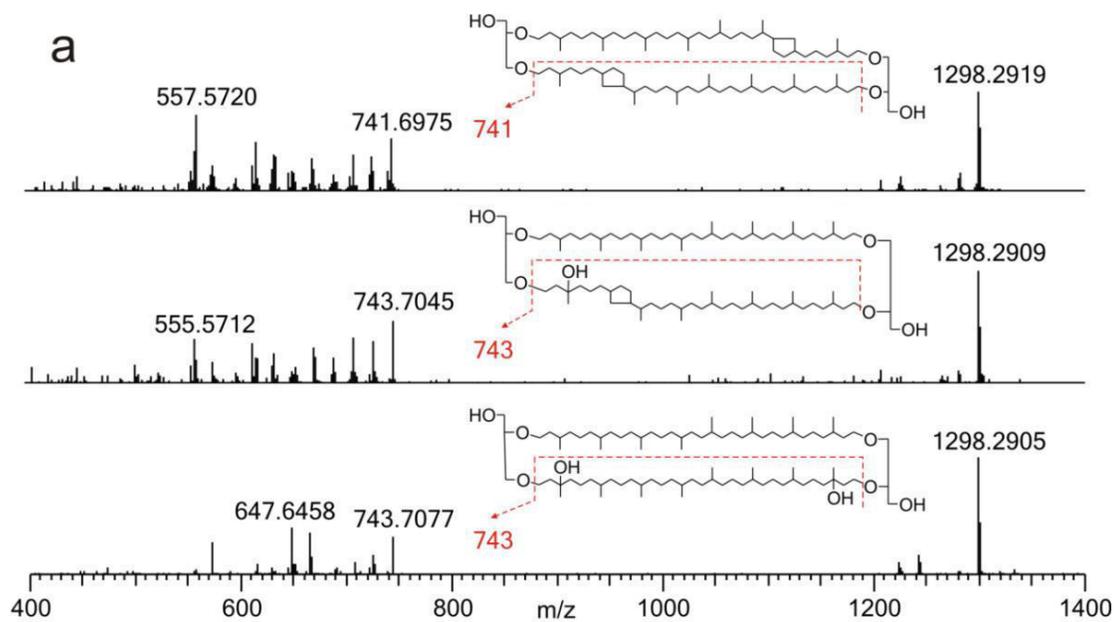
747

748

Figure 2.

749
750
751

752

753

Figure 3.

755

36

756

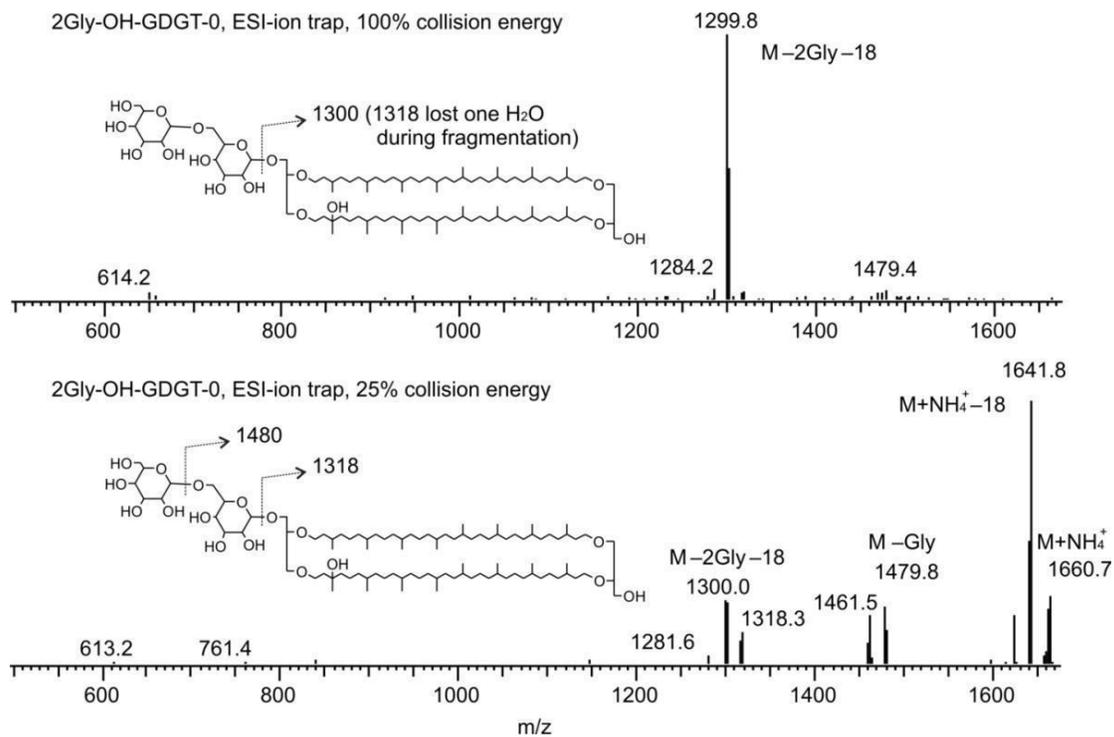**Figure 4**.

757

758

759
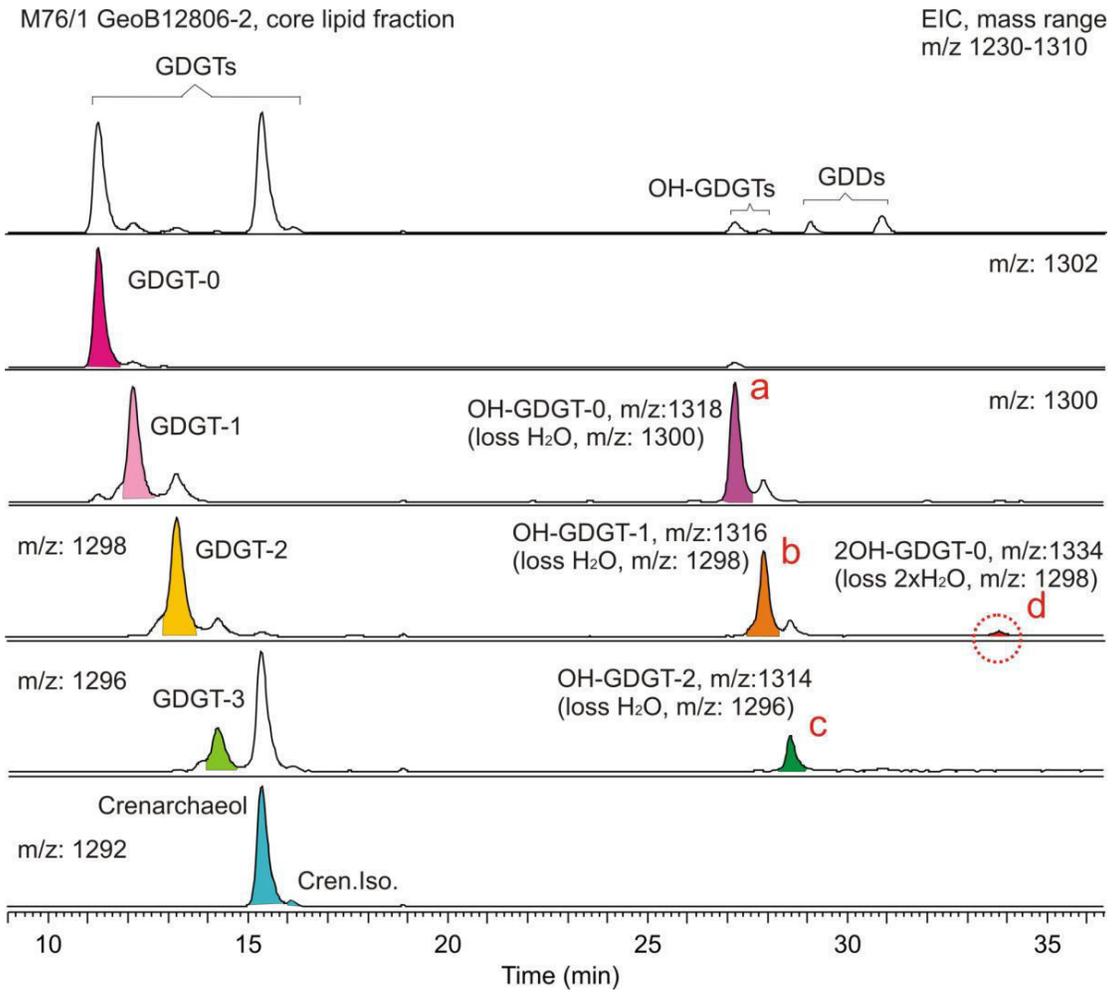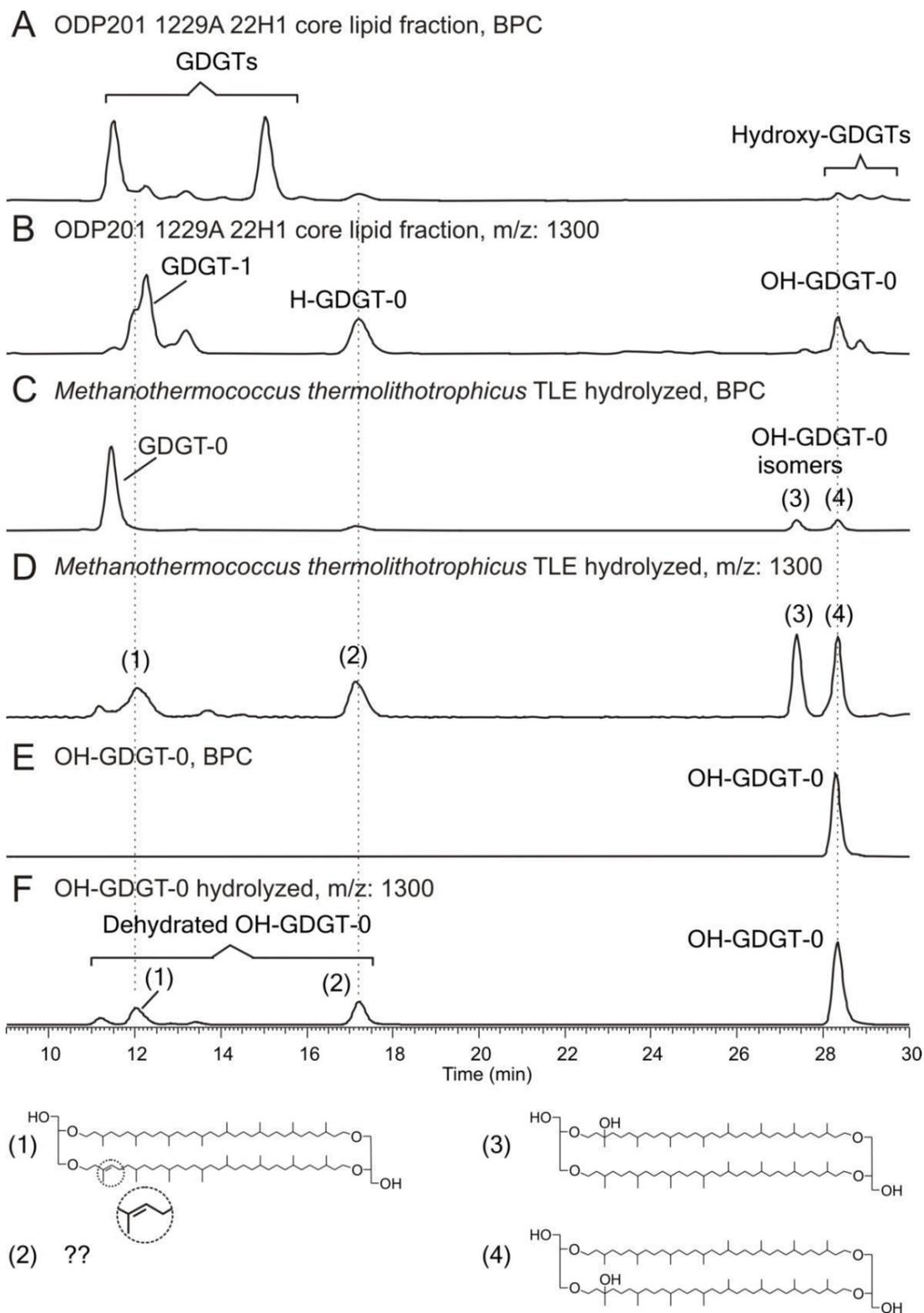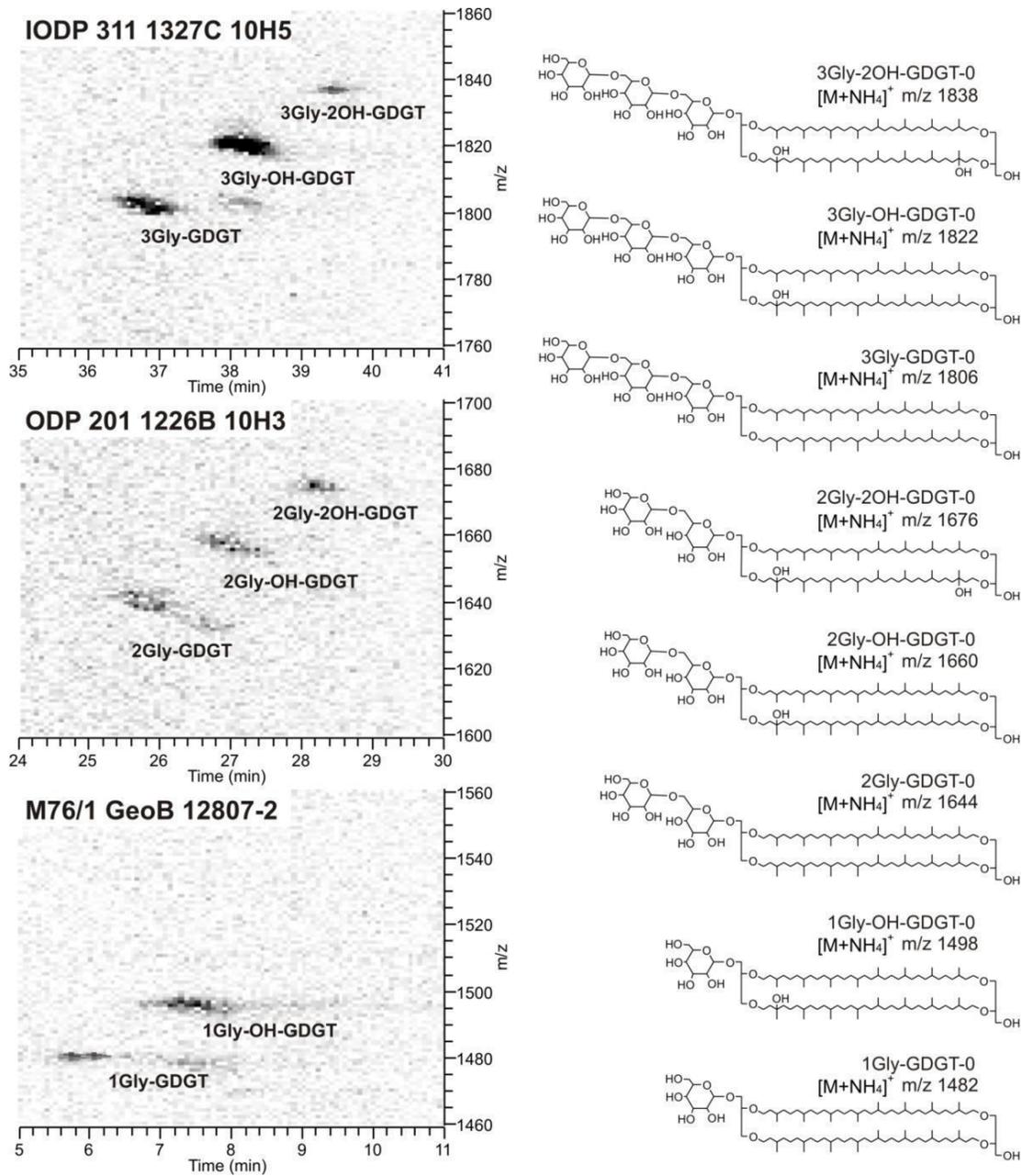
760
761

762

763 **Figure 5**.
764

765

766

767 **Figure 6.**

768

Table 1. Samples analyzed. Detailed information on the twelve marine sediment samples has been published in Liu et al., (2011). Ratio of hydroxy-GDGT vs. the total core GDGT was calculated based on the detection of hydroxy-GDGT with APCI-MSD single quadrupole mass spectrometer.

Samples		OH-GDGT/total Core GDGT (%)	2OH-GDGT/total Core GDGT (%)	Identified Intact Polar GDGTs
M76/1 Namibia Margin	GeoB 12806-2	6.0	0.06	1Gly-, 2Gly-GDGT; 1Gly-OH-GDGT
	GeoB 12807-2	6.1	n.d.	1Gly-, 2Gly-GDGT; 1Gly-, 2Gly-OH-GDGT
ODP201 Peru Margin	1229D 4H4	7.2	0.11	2Gly-GDGT; 2Gly-, 3Gly-OH-GDGT
	1229A 22H1	5.3	0.06	2Gly-GDGT
ODP201 Equatorial Pacific	1226B 10H3	4.7	1.72	2Gly-, 3Gly-GDGT; 2Gly-, 3Gly-OH-GDGT; 2Gly-2OH-GDGT
	1226E 20H3	2.2	0.25	1Gly-, 2Gly-GDGT; 2Gly-OH-GDGT
ODP204 Hydrate Ridge	1250D 6H5	5.8	0.26	2Gly-GDGT; 2Gly-OH-GDGT
	1250D 12H5	6.3	0.37	2Gly-, 3Gly-GDGT; 2Gly-, 3Gly-OH-GDGT
IODP311 Cascadia. Margin	1327C 10H5	7.9	0.47	2Gly-, 3Gly-GDGT; 2Gly-, 3Gly-OH-GDGT; 3Gly-2OH-GDGT
	1327C 13X6	6.1	0.25	2Gly-GDGT; 2Gly-OH-GDGT
ODP160 Mediterranean Sapropel	966C 5H02	2.7	n.d.	2Gly-, 3Gly-GDGT; 2Gly-, 3Gly-OH-GDGT
	966C 7H04	1.2	0.07	2Gly-GDGT; 2Gly-OH-GDGT
Methanothermococcus thermolithotrophicus		14.3	n.d.	1Gly-, 2Gly, 3Gly-GDGT; 2Gly-, 3Gly-OH-GDGT

(n.d. not detected)

1 **Table 2.** ^{13}C- and ^{1}H-NMR data of OH-GDGT-0

2

Label	^{1}H shift (ppm)	^{13}C shift (ppm)	Label	^{1}H shift (ppm)	^{13}C shift (ppm)
A1	3.44, 3.56	68.5	A1'	3.38, 3.40	69.9
A2	1.41, 1.67	37.5	A2'	1.38, 1.67	37.0
A3	1.65	30.1	A3'	1.65	30.1
A4	1.33, 1.37	30.1	A4'	1.33, 1.37	30.1
A5	1.35, 1.38	24.9	A5'	1.35, 1.38	24.9
A6	1.21, 1.42	37.8	A6'	1.21, 1.42	37.8
A7	1.50	33.1	A7'	1.50	33.1
A8	1.21, 1.42	37.8	A8'	1.21, 1.42	37.8
A9	1.35, 1.48	24.9	A9'	1.35, 1.38	24.9
A10	1.21, 1.42	37.8	A10'	1.21, 1.42	37.8
A11	1.50	33.1	A11'	1.50	33.1
A12	1.21, 1.42	37.8	A12'	1.21, 1.42	37.8
A13	1.35, 1.48	24.9	A13'	1.35, 1.38	24.9
A14	1.21, 1.42	37.8	A14'	1.21, 1.42	37.8
A15	1.47	33.4	A15'	1.47	33.4
A16	1.25, 1.42	34.7	A16'	1.25, 1.42	34.7
A17	0.94	19.9	A17'	0.94	19.9
A18	0.99	20.0	A18'	0.99	20.0
A19	0.99	20.0	A19'	0.99	20.0
A20	0.98	19.6	A20'	0.98	19.6
B1	3.48, 3.50	68.8	B1'	3.50, 3.64	68.5
B2	1.57, 1.73	40.3	B2'	1.41, 1.65	37.5
B3	-	71.7	B3'	1.65	30.1
B4	1.49, 1.56	43.3	B4'	1.33, 1.37	30.1
B5	1.49, 1.53	21.8	B5'	1.35, 1.48	24.9
B6	1.22, 1.42	38.2	B6'	1.21, 1.42	37.8
B7	1.50	33.1	B7'	1.50	33.1
B8	1.21, 1.42	37.8	B8'	1.21, 1.42	37.8
B9	1.35, 1.48	24.9	B9'	1.35, 1.48	24.9
B10	1.21, 1.42	37.8	B10'	1.21, 1.42	37.8
B11	1.50	33.1	B11'	1.50	33.1
B12	1.21, 1.42	37.8	B12'	1.21, 1.42	37.8
B13	1.35, 1.48	24.9	B13'	1.35, 1.48	24.9
B14	1.21, 1.42	37.8	B14'	1.21, 1.42	37.8
B15	1.47	33.4	B15'	1.47	33.4
B16	1.25, 1.42	34.7	B16'	1.25, 1.42	34.7
B17	1.21	27.1	B17'	0.99	20.0
B18	0.99	20.0	B18'	0.99	20.0
B19	0.99	20.0	B19'	0.99	20.0
B20	0.98	19.6	B20'	0.98	19.6
C1	3.54, 3.59 (1.43 OH)	62.4	C1'	3.67, 3.72 (2.05 OH)	63.0
C2	3.33	79.1	C2'	3.51	79.2
C3	3.36, 3.41	71.5	C3'	3.43, 3.51	71.7

3

4 Carbon labels in table refer to the numbered molecule in Fig. 1.

5

6

7

8