From ether to acid: A plausible degradation pathway of glycerol dialkyl glycerol tetraethers

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.
From ether to acid: a plausible degradation pathway of glycerol dialkyl glycerol tetraethers

Xiao-Lei Liu a, b, *, Daniel Birgel c, Felix J. Elling a, g, Paul A. Sutton d, Julius S. Lipp a,
Rong Zhu a, §, Chuanlun Zhang e, Martin Köneke a, Jörn Peckmann c, f, Steven J.
Rowland d, Roger E. Summons b, Kai-Uwe Hinrichs a

a Organic Geochemistry Group, MARUM Center for Marine Environmental Sciences and
Department of Geosciences, University of Bremen, 28359 Bremen, Germany

b Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of
Technology, 77 Massachusetts Avenue, Cambridge, MA 02139-4307, USA

c Institute of Geology, University of Hamburg, 20146 Hamburg, Germany

d School of Geography, Earth and Environmental Sciences, University of Plymouth,
Plymouth, Devon PL4 8AA, U.K.
e State Key Laboratory of Marine Geology, Tongji University, Shanghai 200092, China

f Department of Geodynamics and Sedimentology, Center for Earth Sciences, University
of Vienna, 1090 Vienna, Austria

§ Present address, Harvard University, 20 Oxford St, Cambridge, MA 02138, USA
§ Present address, ETH Zurich, Universitätsstrasse 16, 8092 Zürich, Switzerland

* Corresponding author, E-mail: xlliu@mit.edu
Abstract

Glycerol dialkyl glycerol tetraether (GDGT) are ubiquitous microbial lipids with extensive demonstrated and potential roles as paleoenvironmental proxies. Despite the great attention they receive, comparatively little is known regarding their diagenetic fate. Putative degradation products of GDGTs, identified as hydroxyl and carboxyl derivatives, were detected in lipid extracts of marine sediment, seep carbonate, hot spring sediment and cells of the marine thaumarchaeon *Nitrosopumilus maritimus*. The distribution of GDGT degradation products in environmental samples suggests that both biotic and abiotic processes act as sinks for GDGTs. More than a hundred newly recognized degradation products afford a view of the stepwise degradation of GDGT via 1) ether bond hydrolysis yielding hydroxyl isoprenoids, namely, GDGTol (glycerol dialkyl glycerol triether alcohol), GMGD (glycerol monobiphytanyl glycerol diether), GDD (glycerol dibiphytanol diether), GMM (glycerol monobiphytanol monoether) and bpdiol (biphytanediol); 2) oxidation of isoprenoidal alcohols into corresponding carboxyl derivatives and 3) chain shortening to yield C$_{39}$ and smaller isoprenoids. This plausible GDGT degradation pathway from glycerol ethers to isoprenoidal fatty acids provides the link to commonly detected head-to-head linked long chain isoprenoidal hydrocarbons in petroleum and sediment samples. The problematic C$_{80}$ to C$_{82}$ tetraacids that cause naphthenate deposits in some oil production facilities can be generated from H-shaped glycerol monoalkyl glycerol tetraethers (GMGTs) following the same process, as indicated by the distribution of related derivatives in hydrothermally influenced sediments.
1. INTRODUCTION

Lipid biomarkers have the potential of providing valuable information regarding the composition of ancient ecosystems and paleoenvironmental conditions throughout most of the sedimentary record (e.g., Peters et al., 2004). For most of the frequently used biomarkers, such as steroids, hopanoids, pigments and their derivatives, the post-depositional structural transformations are rather well constrained (e.g., Mackenzie et al., 1982; Innes et al., 1997; Peters et al., 2004; Brocks and Schaeffer, 2008) and contribute to the foundation of the geological biomarker concept that links geomolecules to their biological precursors. An exception is the intensively studied group of isoprenoidal glycerol dialkyl glycerol tetraether (GDGT) lipids produced by Archaea. Despite their prominent use as molecular proxies for the reconstruction of paleoenvironmental conditions (e.g., Ingalls and Pearson, 2013; Schouten et al., 2013) and, in their intact polar form, for the ecology of extant archaeal communities (e.g., Lipp and Hinrichs, 2009; Liu et al., 2011; Meador et al., 2015; Yoshinaga et al., 2015), our understanding of their post-depositional behavior is fragmentary.

Three principal archaeal lipid categories are observed in environmental and geological samples: (i) intact polar lipids (IPL) as building blocks of the cellular membrane, consisting of a glycerolalkylether backbone and a polar headgroup which, in most instances is glycosidic (Sturt et al., 2004, Lipp and Hinrichs, 2009), (ii) the corresponding core glycerol alkyl ethers derived from hydrolytic cleavage of the polar headgroups on timescales of days to several tens of millennia, depending on depositional conditions and
enzymatic activity (Harvey et al., 1986; Xie et al., 2013), and (iii) degradation products of core glycerol alkyl ethers that occur as hydrocarbons (Moldowan and Seifert, 1979), alcohols (Schouten et al., 1998; Saito and Suzuki, 2010) and carboxylic acids (Meunier-Christmann, 1988; Schouten et al., 2003; Birgel et al., 2008a).

The second group, i.e., the core lipids, is among the most extensively studied biomarker class in the last decade (Ingalls and Pearson, 2013; Schouten et al., 2013), in particular the GDGT derivatives. GDGTs accumulate in cold and moderately heated aquatic sediments with seemingly little molecular alteration and remain intact in sediments over tens of millions of years (e.g., Kuypers et al., 2001). GDGTs, including the bacterial non-isoprenoidal types (Weijers et al., 2009; Liu et al., 2012b), are among the most prominent lipids in marine sediments and soils. Their ubiquity and abundance result from both the widespread distribution of their producing, largely uncultured, microbes and their relatively high recalcitrance caused by the ether-linkages. Within the domain Archaea, GDGTs are taxonomically widely distributed and probably produced by members of all phyla (Pearson and Ingalls, 2013; Schouten et al., 2013).

Thermal diagenesis (Rowland, 1990) and hydrous pyrolysis (Pease et al., 1998) experiments have shown the generation of isoprenoidal hydrocarbons from fresh archaeal cultures. The exact fate of GDGTs is not clear, but they seem to be the most plausible precursors of a wealth of compounds of putative archaeal origin found in thermally mature formations. These compounds include head-to-head linked C\textsubscript{32} to C\textsubscript{40} isoprenoid hydrocarbons in petroleum samples (Moldowan and Seifert, 1979), biphytane diols
(Schouten et al., 1998; Saito and Suzuki, 2010) and biphytanediacids (Meunier-Christmann, 1988; Birgel et al., 2008a) in recent sediments and rock samples of possibly diagenetic and/or biogenic origin; however these diols or diacids have never been detected in any archaeal cell extracts. Another group of recently discovered, widespread compounds includes a series of glycerol ether derivatives, the glycerol dibiphytanol diethers (GDDs; Knappy et al., 2012; Liu et al., 2012a). Although the occurrence of core GDDs and their glycosidic intact polar lipids in archaeal cell extracts suggests that they play a role in archaeal lipid biosynthesis (Liu et al., 2012a; Meador et al., 2014), a diagenetic contribution of these lipids in natural settings cannot be ruled out (e.g., Yang et al., 2014).

Another conspicuous compound series of putative archaeal origin are the so-called ‘H-shaped’ or ‘ARN’ C_{80}-C_{82} isoprenoidal tetracarboxylic acids found in certain petroleum types (Lutnaes et al., 2006, 2007) and believed to contribute significantly to the problematic naphthenate deposits formed during oil processing (e.g., Baugh et al., 2004; 2005). The archaeal lipids that are structurally related to the C_{80} tetraacids are ‘H-shaped’ glycerol monoalkyl glycerol tetraethers (H-GMGTs) found in thermophilic archaeal taxa (Morii et al., 1998; Schouten et al., 2008a), although H-GMGT-0 may also have non-thermophilic origins (Schouten et al. 2013). Whether the C_{80} tetraacids are degradation products of H-GMGTs or actually bio-surfactants directly synthesized by Archaea living in the crude oil remains ambiguous (Lutnaes et al., 2006, 2007). To date, isoprenoidal tetracarboxylic acids have not been detected in archaeal cells.
Based on the distribution patterns of newly identified series of GDGT degradation products in sedimentary samples (hydrothermally overprinted sediments from the Guaymas Basin and a hot spring in China, Miocene seep carbonates, and marine subsurface sediments), cell extracts and hydrolysis experiments, here we construct a precursor-product reaction network from GDGTs to alcoholic and carboxylic acid biphytane derivatives. Additionally, through the identification of five types of putative intermediates in the sediment from Guaymas Basin, we provide strong support for the hypothesis that isoprenoidal C$_{80}$ to C$_{82}$ tetraacids are derived from step-wise degradation of ‘H-shaped’ GMGTs.

2. MATERIALS and METHODS

2.1. Sample collection and preparation

*N. maritimus* strain SCM1 was grown aerobically at 28 °C and pH 7.5 in 8.5 l HEPES-buffered Synthetic Crenarchaeota Medium (1.5 mM NH$_4$Cl; Könneke et al., 2005; Martens-Habbena et al., 2009). The medium was inoculated with 5% of a mid-logarithmic phase pre-culture of *N. maritimus*. Biomass was harvested early and late in the growth phase as well as early and late in the stationary phase (one batch for each time-point) using cross-flow filtration (Elling et al., 2014). Purity of the culture was checked daily by phase contrast microscopy. Growth was monitored by measuring nitrite formation photometrically (Stickland and Parsons, 1972) and by counting 2% formaldehyde-fixed, SYBR Green I stained cells (Lunau et al., 2005) at the beginning
and the end of the experiment. Lipids were extracted from each batch following a modified Bligh and Dyer protocol as described previously (Sturt et al., 2004; Elling et al., 2014).

The Marmorito seep carbonate samples (Marmorito limestone; see also Peckmann et al., 1999) were taken close to the village of Marmorito, in the Monferrato hills close to Torino, Italy. The Marmorito limestone is composed of dolomite and calcite, and is embedded in Miocene strata chiefly consisting of siliciclastic sediments that were deposited in a marine shelf environment. Rock samples from the vicinity Marmorito, comprising not only the so-called Marmorito limestone, are among the early examples where methane-seepage, fossil chemosynthetic benthic communities, molecular fossils, and methane-related carbonate precipitation have been described (e.g., Clari et al., 1988, 1994; Peckmann et al., 1999; Thiel et al., 1999). Apart from the characteristic molecular fossils of the consortium mediating the anaerobic oxidation of methane, biomarkers of aerobic methanotrophic bacteria were also found (Peckmann et al., 1999; Birgel and Peckmann, 2008). A detailed list of compounds identified in the Marmorito limestone and a description of the applied decalcification procedure can be found in Birgel and Peckmann (2008) and references therein. Lipid extraction was performed as described in Birgel et al. (2006).

The Guaymas Basin sediment sample was retrieved during the RV Atlantis cruise AT15-56 to the Guaymas Basin, Gulf of California, Mexico, during Alvin dive 4568 (November 22 to December 6 2009, 27° 00.449’ N, 111° 24.347’ W). The sample came from an oil-
impregnated hydrothermally active area, where sedimentary temperature steeply
increased from ~3 °C to ~100 °C within 35 cm (Gutierrez et al., 2015). The sample was
surface sediment (0-4 cm depth) with the highest temperature at time of sampling
reaching ~12 °C. However, due to the dynamic nature of the hydrothermal activity and
the upward flux of fluids in the Guaymas Basin (Pearson et al., 2005), it is likely that the
sample has been previously heated to higher temperatures and/or contains extractable
organic matter formed in deeper layers at higher temperatures. Wet sediment (15-20 g)
was extracted using the modified Bligh and Dyer method (Sturt et al., 2004). To
minimize the heavy background of oil contaminants, an aliquot of the total lipid extract
(TLE) was cleaned with the Hybrid SPE®-Phospholipid cartridges before LC-MS
analysis according to the method described by Zhu et al. (2013a).

Two further sediment samples, including a marine subsurface sediment (Leg201-1227;
Hole 1227A, mixed from five samples: 1227A-2H2-65-75cm, 1227A-2H5-83-93cm,
1227A-3H2-55-65cm, 1227A-11H2-118-128m, 1227A-13H3-0-15cm; spanning from
8.1-113.6 m below sea floor) and a hot spring sediment (T-15, see sample description in
SI) were also extracted with the Bligh and Dyer method for lipid analysis.

2.2. Lipid analysis

For lipid analysis by normal phase liquid chromatography (NP-LC), an aliquot of TLE of
each sample was dissolved in n-hexane/isopropanol (99.5:0.5 v/v) for injection.
Compound separation was performed on a Dionex Ultimate 3000 RS UHPLC system
Thermo Scientific, Bremen, Germany) at 50 °C, following the recently developed tandem column protocol (Becker et al., 2013) using two ACQUITY UPLC® BEH Hilic Amide columns (2.1 x 150 mm, 1.7 µm, Waters). Solvent gradient was programmed for a constant flow rate of 0.5 mL min⁻¹ and a linear increase from 3% B to 20% B in 20 minutes, and then linearly to 50% B at 35 minutes, after then up to 100% B at 45 min, holding for 6 minutes, finally back to 3% B for 9 minutes to re-equilibrate the column, where A was n-hexane and B was n-hexane/isopropanol (90:10). Detection was achieved with a Bruker Maxis accurate-mass quadrupole time-of-flight (qTOF) mass spectrometer (Bruker Daltonik, Bremen, Germany) coupled to the UHPLC via an atmospheric pressure chemical ionization (APCI) interface run in positive ion mode. APCI source parameters were as follows: corona current 3500 nA, nebulizer gas pressure 5 bar, drying gas flow 8 L min⁻¹, drying gas (N₂) temperature 160 °C, vaporizer temperature 400 °C. The scan range was 150-2000 m/z at a rate of 2 Hz. Lipids were identified based on accurate mass (better than 1 ppm), retention times and diagnostic fragments and under consideration of general GDGT mass spectral features (e.g., Liu et al., 2012b), and quantified by measurement of [M+H]⁺ responses, with a extraction window of individual ion chromatograms of ± 0.01 m/z units.

For the detection of carboxylic acid derivatives, reversed phase (RP) LC-MS (cf. Zhu et al., 2013b) was applied with the same LC system using an ACE3 C₁₈ column (2.1 × 150 mm, 3 µm; Advanced Chromatography Technologies Ltd., Aberdeen, Scotland) coupled with a guard cartridge and maintained at 45 °C, and the same mass spectrometer (qTOF) equipped with an electrospray ionization (ESI) source and operated in positive mode.
(Bruker Daltonik, Bremen, Germany). An aliquot of TLE of each sample was dissolved in methanol prior to injection. Separation of compounds was achieved isocratically with 100% eluent A for 10 min, followed by a rapid gradient to 24% B in 5 min, and then a slow gradient to 65% B in 55 min at a flow rate of 0.2 mL min$^{-1}$, where the eluent A was methanol/formic acid/14.8 M NH$_3$aq (100:0.04:0.10, v/v/v) and B was 2-propanol/formic acid/14.8 M NH$_3$(aq) (100:0.04:0.10, v/v/v). After each run, the column was washed with 90% B for 10 min and subsequently re-equilibrated with 100% A for another 10 min. The ESI-MS conditions were set as capillary voltage 4500 V, nebulizing gas (N$_2$) pressure 0.8 bar, and dry gas (N$_2$) 4 L min$^{-1}$ at a temperature of 200 °C.

3. RESULTS

3.1. Structural elucidation of products of GDGT breakdown

Table 1 provides an overview of 19 distinct compound classes observed in this study. Structural assignments are based on formulae established from accurate mass measurements, chromatographic behavior and tandem mass spectrometry (MS$^3$) (Fig. 1, 2 and 3 and S1a-d contain data showing detailed MS$^2$ fragmentation patterns). Many compounds remain only tentatively identified. Identifications of the hydroxyl derivatives, including GDGTol (glycerol dialkyl glycerol triether alcohol), GMGD (glycerol monobiphytanyl glycerol diether), GDD (glycerol dibiphytanol diether), GMM (glycerol monobiphytanol monoether) and bpdiol (biphytanediol), were additionally confirmed by examination of products released after acidic degradation of purified GDGT-0.
(compound compositions are given in Fig. S2). The identification of the H-tetrol was additionally supported by the co-injection of an H-tetrol mixture synthesized by reduction of the corresponding acids, which were isolated and characterized by NMR spectroscopy previously (Lutnaes et al., 2007; Fig. S3, preparation described in SI). C₃₉ isoprenoids were assigned based on chromatographic behavior, molecular formulae generated from accurate mass measurements and characteristic fragment ions (MS² fragmentation patterns are shown in Fig. S1a). For example, the two isomers of GDGTol-0 (highlighted with green circles in Fig. 3e) eluted earlier than those of C₄₀/₃₉-GDGTA-0 (two pink solid line circles in Fig. 3e). In addition, GDGTol-0 and C₄₀/₃₉-GDGTA-0 also have measurably different molecular masses, [M+H]⁺ of m/z: 1320.3312 and 1320.2930, respectively and these afford calculated formulae of C₈₆H₁₇₅O₇⁺ for GDGTol-0 and C₈₅H₁₇₁O₈⁺ for C₄₀/₃₉-GDGTA-0. To add weight to this assignment, the fragment ion of m/z: 651.6245 in the MS² of C₄₀/₃₉-GDGTA-0 compared to m/z: 665.6391 of GDGT-0 represents the loss of methylene from the biphytane (Fig. S1a).

3.2. Terminally hydroxylated biphytanyl derivatives

Hydrolysis of the glycerol ether bonds in isoprenoidal GDGT generates five types of biphytane-based alcohols, with and without the glycerol backbones, namely, GDGTol, GMGD, GDD, GMM and bpdiol (compound structures are given in Fig. S2). For example, we demonstrate their structural relationships to acyclic caldarchaeol, GDGT-0, with a parallel glycerol configuration, a distribution observed in a seep carbonate sample from Marmorito (Fig. 1a and b). Constitutional isomers, which are labeled as peak ‘a’
and ‘b’ in the chromatogram of GDGTol, GMGD and GMM in Fig. 1, represent different combinations of sn-2 or sn-3 glycerol ether bonding. Degradation products of the ring-containing GDGTs are more numerous, due to different possible arrangements of a ring on the biphytanyl chains. A detailed isomeric study of GDGTol, GMGD and GMM and their implications will be discussed in our following works. As for H-GMGTs, which have two biphytane chains linked by a bis-methylene C-C bond at position C_{20} (Lutnaes et al., 2006; 2007) the cleavage of two glycerol units results in a C_{80} H-tetrol, instead of two biphytanediols (bpdiols). These hydroxyl derivatives of GDGT were either all, or at least some, present in our sample set comprising extracts of marine sediment, seep carbonate, hydrothermal vent, hot spring sediment and archaeal cell extract (Table 1).

### 3.3. Terminally-carboxylated biphytanyl derivatives

Carboxylic acids, corresponding to each of hydroxyl derivatives mentioned above, could logically be generated by oxidation of each terminal alcohol of the biphytane. These were also detected in the analyzed samples and are subsequently referred to as GDGTA for the carboxyl analogue of GDGTol and GMMA for GMM (Fig. 1a, b and Table 1). For isoprenoids possessing two primary hydroxyl groups, such as GDD and bpdiol, both mono- and dicarboxyl analogues were detected, with the monocarboxyl derivative of GDD termed GDDA, the dicarboxyl derivative GDDAA, as well as biphytane mono- and diacids. In the case of GMGD, which does not possess a primary biphytane-bound hydroxyl group, no acid derivatives were detected. In the case of H-tetrol, which contains four primary hydroxyl groups on its two linked bpdiols, we detected mono-, di-, tri- and...
tetraacids in an oily sediment sample collected from a hydrothermal vent site in Guaymas Basin (Fig. 2).

Under APCI conditions, the ionization patterns of biphytane monoacid/monool (bpmonoacid/ol) and biphytanediacids (bpdiacid) differed from bpdiols. A protonated molecular ion \([M+H]^+\) was usually the major ion in mass spectra of bpdiols, while the acids produced more complex mass spectra in our analyses. Dehydrated molecular ions and unknown adducts were formed during the ionization of carboxyl derivatives (Fig. 1b and S1d). The main adduct ions detected were \([M+H+42]^+\) for bpmonoacid/ol and \([M+H+84]^+\) for bpdiacids (Fig. S1d).

### 3.4. C₃₉ and C₃₈ isoprenoids

In addition to the series of hydroxylated and carboxylated C₄₀ biphytane derivatives, we identified some with shortened alkyl chains such as the C₃₉ analogues. For example, in most analyzed samples analogues with one methylene unit less than GDGTol and GDGTA, here termed C₄₀/₃₉-GDGTol and C₄₀/₃₉-GDGTA (Arabic numbers represent the carbon number of the two isoprenoidal chains in the molecule), always co-occurred at lower abundance with their C₄₀/₄₀ analogues (Fig. 3b, e). C₃₉ derivatives were also detected as analogues of GDD, bpdiol and GMM (Fig. 3a, c and d). Additionally, we detected signals that we attribute to the pseudo-homologue of C₃₈ isoprenoids, as GDGTol and GDD (Fig. 3a and b). Within the group of GDDs reduced by two C-atoms, we observed both C₃₉/₃₉ and C₄₀/₃₈ derivatives (Fig. 3a). Three isomers of the acyclic and
four isomers of crenarchaeol-related GDD were detected (Fig. 3a). In previous studies, δ^{13}C depleted C_{39} head-to-head linked isoprenoids were detected in Cretaceous (Sandy et al., 2012) and Carboniferous seep carbonates (Birgel et al., 2008b); those compounds presumably represent the hydrocarbon derivatives of the C_{39} functionalized compounds found in this study. Interestingly, no derivatives of C_{80} H-tetrol and tetraacid with reduced carbon chains were detected in our samples.

4. DISCUSSION

4.1. The occurrence of GDGT degradation derivatives in environmental samples

The distributions of major degradation derivatives of GDGT-0 and H-GMGT-0 were compared in four representative environmental samples (Fig. 4), which include modern to late Miocene marine subsurface sediment (Leg201-1227), modern hot spring (T-15) and hydrothermally heated sediment (Guaymas Basin 4568), and the Miocene Marmorito seep carbonate. Distinct patterns of degradation derivatives in these four types of samples from different environments and of different age (modern sediments and ancient carbonate rock) reflect variable degrees of degradation and preservation.

4.1.1. Distributions of regular GDGT degradation products in environmental samples

The distributions of GDGT degradation products across our sample set suggest that the biological sources of GDGTs, as well as the debris depositional histories, influence the
The diagenetic trajectory of GDGTs. By way of example, the degradation derivatives of GDGT-0 are around twice as abundant as their precursor in the marine sediment sample of Leg 201-1227, in which bpdiol-0 comprises over half of all of the detected degradation products (Fig. 4). However, the GDGT degradation derivatives in Miocene seep carbonates exhibited a lower overall relative abundance, but with a greater variety and portion of labile components, such as the carboxyl derivatives (Table 1 and Fig. 4). Regular isoprenoidal GDGTs (as compared to H-GMGTs) preserved in marine sediments are primarily derived from planktonic archaea, dwelling in the water column. In contrast, those in seep carbonates receive a larger in-situ contribution from benthic communities engaged in anaerobic oxidation of methane (AOM), as confirmed by low $\delta^{13}$C values of biphytanic diacids (cf. Birgel et al., 2008a). The higher proportion of degradation derivatives compared to their GDGT precursors in deep subsurface sediment apparently results from a mild to moderate degradation process, but degradation lasted longer than at the Marmorito seep site and was favored by extensive transportation. The extraordinarily good preservation of labile compounds in seep carbonates can be attributed to significant in-situ GDGT production combined with co-eval carbonate formation and resulting early lithification within methane seepage systems (e.g., Peckmann et al., 1999; Birgel et al., 2008a), rather than representing a time-integrated pattern with input from various sedimentary or sedimentary/planktonic archaea (cf. Feng et al., 2014; Birgel et al., 2008a). The relative abundance of degradation products in Guaymas basin sediment is over 40% of all GDGT-derived compounds (Fig. 4). Both biodegradation and thermal diagenesis can be potential sources of degradation products, due to the presence of both active
microbial communities and high temperature hydrothermal fluids in Guaymas basin sediments (Teske et al., 2014; Gutierrez et al., 2015). In contrast, there are only low abundances of GDD and bpdiol (< 10%) detected as GDGT degradation products in the hot spring sediment (Fig. 4).

In addition to the contrasting patterns of the different environments, there is a compositional discrepancy between each class of GDGT product within the same sample. For example, previous studies that documented the presence of bpdiols and bpdiacids in environmental samples have shown that the ring distributions and carbon isotopic compositions differ for biphytanes released from coexisting GDGTs (Schouten et al., 1998; Birgel et al., 2008a; Saito and Suzuki, 2010). The Marmorito limestone samples (Fig. 5; Birgel et al. (2008a) provide results on hydrocarbons and bpdiacids), show distinct ring distribution patterns for GDGT, bpdiol, bpmonoacid/ol and bpdiacid; GDGTs are dominated by GDGT-0 and crenarchaeol, while the relative abundance of tricyclic biphytane derivatives derived from crenarchaeol gradually decreases for the hydroxyl to carboxyl products. Multiple inputs, combined with selective preservation could cause such distributional differences. The ring distribution of GDGTs preserved in sediments may reflect a mixed contribution from both planktonic and benthic species and is frequently dominated by compounds from planktonic sources (Wuchter et al., 2005; Huguet et al., 2007; Lengger et al., 2012). Planktonic and sedimentary archaeal communities could contribute different lipids. Further, compared to biphytanyl products derived from benthic species within the sediment, those derived from water column will have experienced a very different transportation history. This might explain why the
tricyclic diacid derived from planktonic archaea are rare, while the acyclic, mono- and bicyclic diacids with origins from sedimentary methanotrophic archaea, are more common and more $^{13}$C-depleted in the seep carbonate (isotopic data published in Birgel et al., 2008a).

4.1.2. The degradation of H-GMGTs to H-tetrols and H-tetraacids

H-GMGT-0 and its hydroxylated degradation products are generally present in the samples from low-temperature and hydrothermal environments (Table 1 and Fig. 4), while compounds with multiple cycloalkyl groups and additional methylations are restricted to the hydrothermal samples alone. In low-temperature samples, H-GMGT-0 is far more abundant than H-GMD-0 (acyclic H-shaped glycerol monoalkyl diether) and tetrol-0 (Fig. 4). The apparently different patterns of H-GMGT-0 and its derivatives, compared to GDGT-0 and its highly abundant degradation products, point to contributions of distinct species and/or different diagenetic pathways.

We detected H-tetrol and H-tetraacids, and their partially oxidized ‘H-shaped’ mono-, di- and triacid intermediates, in the oily Guaymas Basin sediment (Fig. 2). The detection of these intermediates likely implies enzymatic modification of existing ‘H-shaped’ GDGTs by living organisms. H-tetraacids accumulated in naphthenate deposits in oil well infrastructure and pipelines and in crude oils, have distributions usually dominated by 4-8 cyclopentyl rings and 1 or 2 additional methyl substitutions (Lutnaes et al., 2007; Sutton & Rowland, 2014). Although H-GMGT-0 is known to be present in a broad range of
mesophilic environments (Schouten et al., 2008a), those with multiple cyclizations and methylations have been only detected in hyperthermophilic archaeal species (Schouten et al., 2008b; Knappy et al., 2011; Liu et al., 2012b). H-GMGTs existed in both hot spring and Guaymas Basin sediments analyzed. However, H-tetraacids were only found in the oil-impregnated Guaymas Basin sediment (Table 1). In addition to the diagenetic contribution, Lutnaes et al. (2006) also speculated that these tetraacids might be produced by thermophilic, oil-degrading archaea as biosurfactants to facilitate their metabolism. Additional studies would be required to test this rigorously. Given the occurrence of biphytane, C_{39} and smaller head-to-head linked isoprenoids in the geological record, C_{80} based H-shaped isoprenoidal hydrocarbons might also exist. Whilst these are not detectable with conventional GC-MS, they would be amenable to high temperature GC-MS (cf. Sutton & Rowland, 2012).

4.2. The occurrence of GDGT degradation derivatives in archaeal cell extracts

In a previous study concerning the characterization of GDD, Liu et al. (2012a) observed the existence of core GDD in the cell extracts of the methanogen Methanothermococcus thermolithotrophicus. More recently, Meador et al. (2014) reported detection of both core and monoglycosidic GDD (1G-GDD) in a culture of the planktonic ammonia oxidizer N. maritimus. Here we extend these observations and report newly identified biphytane derivatives in fresh biomass collected at different growth phases of cell extracts of the archaean N. maritimus. Core lipids accounted for 8.2-15% of the total detected lipids during different growth phases (Elling et al., 2014). We selected the most
abundant component, crenarchaeol, to illustrate the distribution of its hydroxyl and carboxyl derivatives in different growth phases. GDD-Cren. and bpdiol-Cren. were the only detected hydroxyl components during early exponential growth (Fig. 6); they jointly accounted for less than 1% of all crenarchaeol-based derivatives. In the late growth phase, however, the relative abundances of GDD-Cren. and bpdiol-Cren. were almost doubled, while the acid (GDDA-Cren.) and a C$_{39}$ component (C$_{40/39}$-GDD-Cren.) emerged as well. GDD-Cren., depleted by two methylene units (C$_{39/39}$ and C$_{40/38}$-GDD-Cren.), together with GDGToI-Cren. were only observed in the stationary phase.

Although these derivatives occurred in cell extracts, they cannot be simply attributed as intermediates of GDGT biosynthesis. For example, the carboxyl and C$_{39}$ based derivatives, such as GDDA and C$_{40/39}$-GDD, are more likely further oxidized products of GDD. For example, archaeal biphytanyl moieties, synthesized via the mevalonic acid (MVA) or methyerythritol phosphate (MEP) pathways, should have a carbon number that is a multiple of five. Accordingly, the most plausible formation pathway for the C$_{38}$ and C$_{39}$ moieties is oxidation and decarboxylation of C$_{40}$ precursors. Formation of C$_{38}$ and C$_{39}$ moieties via degradation of C$_{40}$ biphytanyl moieties is also consistent with the higher relative abundance of GDDA and C$_{40/39}$-GDD during the late stationary phase (Fig. 6). The loss of one methylene unit represents an $\alpha$-oxidation step. A similar well-known enzymatic $\alpha$-oxidation of isoprenoids is that of phytanic acid (C$_{20}$) to pristanic acid (C$_{19}$) by a wide range of organisms (e.g., Rontani and Volkman, 2003; Jansen and Wanders, 2006, and other studies cited therein). As in phytanic acid, the C$_{3}$ methyl group in biphytanic acid (C$_{40}$) derivatives prevents an initial $\beta$-oxidation mechanism; instead,
these compounds undergo \( \alpha \)-oxidation to yield \( C_{39} \) based carboxyl isoprenoids. In such a scenario the production of GDD and GDGTol consisting of \( C_{38} \) derivatives would require two successive \( \alpha \)-oxidation steps. This is inconsistent with the reported \( \alpha \)-oxidation of phytanic acid, which is followed by hydroxylation and then \( \beta \)-oxidation (e.g. Jansen and Wanders, 2006, and other studies cited therein). Therefore, elucidation of the degradation process leading to a \( C_{38} \) isoprenoid requires further study. Furthermore, \( \alpha \)- as well as \( \beta \)-oxidation would only result in carboxyl derivatives, and could not explain the occurrence of hydroxyl analogues, such as \( C_{40/39} \)-GDD and \( C_{40/39} \)-GDGTol (Fig. 3). Although the detection of carboxyl derivatives and \( C_{38} \) and \( C_{39} \) based isoprenoids in a metabolic active culture of \textit{N. maritimus} may imply an intracellular modification of existing \( C_{40} \) based lipids, probably as a result of cell senescence, their contribution is less than 2% of the core lipid fraction, or approximately 0.01% of the entire lipidome. The exact mechanisms responsible for the remarkable proportions (up to 70% in the marine subsurface sediment, Leg201-1227, Fig. 4) of degradation products in environmental samples remain unresolved. The presence of degradation products in archaeal cell extracts, as well as their increase towards later growth and stationary phases, suggests that some are formed rapidly and probably via enzymatic catalysis. On the other hand, the high proportion and diversity of degradation products in the hydrothermally influenced settings also leaves room for an additional role of abiotically-mediated chemical degradation reactions, at least for some of the speculated steps leading via catagenesis from GDGTs to biphytanyl hydrocarbons (cf. Rowland, 1990).

4.3. Evidence of analogous behavior of non-isoprenoidal GDGTs
Non-isoprenoidal GDGTs, such as the hybrid isoprenoidal/branched GDGT (IB-GDGT) and branched GDGTs, are known to be widely distributed in various environments, although their exact structures and biological source(s) remain unknown (Liu et al., 2012b). The degradation pathways described here are not limited to archaeal GDGTs. For example, in one of the seep carbonate samples, which contains high abundances of branched and IB-GDGTs, there are non-isoprenoidal GDDs, GMMs and diols detected. These possibly represent degradation products of the branched and IB-GDGTs in the same deposit (Fig. S4 and S5). Our analytical methods also reveal the presence of carboxyl derivatives of these lipids. Non-isoprenoidal GDGTol, GDGTA, GMGD, and their corresponding products with reduced carbon chains (loss of one or two methylene units) were not identified, however, and we attribute this to their low overall abundance in the analyzed samples.

5. CONCLUSIONS

As with other lipid classes, the intact archaeal tetraethers released from the cells of defunct archaea into various depositional settings, are subjected to diagenesis in sediments. We detected three major classes of GDGT degradation products comprising biphytanyl molecules with terminal hydroxyl, terminal carboxyl and shortened carbon chains. A hypothetical scheme of the GDGT degradation pathway to rationalize such a pathway is illustrated in Fig. 7. The labile polar head groups of intact GDGTs, as they occur in living cells, are initially lost to produce the more recalcitrant core GDGTs. Hydrolysis of the different ether bonds then, as we suggest, results in discrete series of
hydroxyl derivatives composed of one or two glycerol and biphytanol units. Oxidation of each terminal hydroxyl functional group may then generate related carboxyl products. Elimination of the C₁ carbon via α-oxidation and subsequent β-oxidation steps, will likely convert the C₄₀ biphytanyl based compounds into shortened isoprenoids, such as C₄₀/3₉-GDGTA and C₄₀/3₉-GDGTol. Further chemical or biological alteration of these intermediates may over geological time, result in the C₄₀, C₃₉ and smaller head-to-head linked isoprenoidal hydrocarbons reported to occur in petroleum and marine deposits (e.g. Moldowan and Seifert, 1979; Birgel et al., 2008b).

A multitude of diagenetic processes will lead to the degradation of GDGTs in sediments. However, we also observed the accumulation of C₃₉ based carboxyl isoprenoids in the later growth and stationary phases of an archaeal cell culture. This implies that enzymatic pathways for degradation of GDGTs also exist, very likely, as a response to substrate limitation, senescence or cell lysis. Additional studies will be required to study this facet in greater detail.

The detection of H-GMGT, H-GMD, H-tetrols and their further oxidized carboxyl intermediates in the oil-contaminated sediments of Guaymas Basin (Fig. 2) elucidated a formation pathway from H-GMGT to H-tetraacids under anoxic conditions. Metagenomic data for samples from the Guaymas Basin and especially sediment with oil impregnation may help to elucidate the origin of H-tetraacids further.
In various geological settings these hydroxyl and carboxyl derivatives co-occur with their GDGT precursors, including both isoprenoidal and non-isoprenoidal GDGTs. Our analysis of their distributions in environmental samples and archaeal cell extracts represents an initial effort to document the possible diagenetic pathways of GDGTs and to bring about the same level of understanding that we have for steroids and triterpenoids (e.g., Peters et al., 2004).
ACKNOWLEDGEMENTS

X.-L.L. was funded by European Union's Seventh Framework Programme – “Ideas” Specific Programme, ERC grant agreement No. 247153 (Advanced Grant DARCLIFE; PI K.-U.H.), the Simons Foundation Collaboration on the Origins of Life (SCOL).

Cultivation experiments and lipid analyses at the University of Bremen were funded by the Deutsche Forschungsgemeinschaft through the Gottfried Wilhelm Leibniz Prize awarded to K.-U.H. (Hi 616-14-1). SJR and PAS were funded by a European Research Council Advanced Investigators Award to SJR for project OUTREACH (agreement no. 228149).
REFERENCES


Peckmann, J., Thiel, V., Michaelis, W., Clari, P., Gaillard, C., Martire, L., Reitner, J. (1999) Cold seep deposits of Beauvoisin (Oxfordian; southeastern France) and Marmorito (Miocene; northern Italy): microbially induced authigenic carbonates. *Int.


<table>
<thead>
<tr>
<th>Compounds</th>
<th>Abbreviation</th>
<th>Structural illustration</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>glycerol dialkyl glycerol tetraether</td>
<td>GDGT</td>
<td><img src="image" alt="Glycerol Dialkyl Glycerol Tetraether" /></td>
<td>+ + + + +</td>
</tr>
<tr>
<td>glycerol dialkyl glycerol triether alcohol</td>
<td>GDGToL</td>
<td><img src="image" alt="Glycerol Dialkyl Glycerol Triether Alcohol" /></td>
<td>+ + + n.d. +</td>
</tr>
<tr>
<td>glycerol dialkyl glycerol triether acid</td>
<td>GDGTA</td>
<td><img src="image" alt="Glycerol Dialkyl Glycerol Triether Acid" /></td>
<td>+ + n.d. n.d. n.d.</td>
</tr>
<tr>
<td>glycerol dibiphytanol diether</td>
<td>GDD</td>
<td><img src="image" alt="Glycerol Dibiphytanol Diether" /></td>
<td>+ + + + +</td>
</tr>
<tr>
<td>glycerol dibiphytanol diether monoacid</td>
<td>GDDA</td>
<td><img src="image" alt="Glycerol Dibiphytanol Diether Monoacid" /></td>
<td>+ + + n.d. +</td>
</tr>
<tr>
<td>glycerol dibiphytanol diether diacid</td>
<td>GDDAA</td>
<td><img src="image" alt="Glycerol Dibiphytanol Diether Diacid" /></td>
<td>n.d. + n.d. n.d. n.d.</td>
</tr>
<tr>
<td>glycerol monobiphytanyl glycerol diether</td>
<td>GMGD</td>
<td><img src="image" alt="Glycerol Monobiphytanyl Glycerol Diether" /></td>
<td>+ + + n.d. n.d.</td>
</tr>
<tr>
<td>glycerol monobiphytanol monoether</td>
<td>GMM</td>
<td><img src="image" alt="Glycerol Monobiphytanol Monoether" /></td>
<td>+ + + n.d. n.d.</td>
</tr>
<tr>
<td>glycerol monobiphytanol monoether acid</td>
<td>GMMA</td>
<td><img src="image" alt="Glycerol Monobiphytanol Monoether Acid" /></td>
<td>n.d. + n.d. n.d. n.d.</td>
</tr>
<tr>
<td>biphytanic diol</td>
<td>bpdiol</td>
<td><img src="image" alt="Biphytanic Diol" /></td>
<td>+ + + + +</td>
</tr>
<tr>
<td>biphytanic monoacid</td>
<td>bpmonoacid/ol</td>
<td><img src="image" alt="Biphytanic Monoacid" /></td>
<td>+ + n.d. n.d. n.d.</td>
</tr>
<tr>
<td>biphytanic diacid</td>
<td>bpdiacid</td>
<td><img src="image" alt="Biphytanic Diacid" /></td>
<td>n.d. + n.d. n.d. n.d.</td>
</tr>
<tr>
<td>H-shaped glycerol monoalkyl glycerol tetraether</td>
<td>H-GMGT</td>
<td><img src="image" alt="H-Shaped Glycerol Monoalkyl Glycerol Tetraether" /></td>
<td>+ + + + n.d.</td>
</tr>
<tr>
<td>H-shaped glycerol monoalkyl diether</td>
<td>H-GMD</td>
<td><img src="image" alt="H-Shaped Glycerol Monoalkyl Diether" /></td>
<td>+ + + + n.d.</td>
</tr>
<tr>
<td>H-shaped C_{80} tetrol</td>
<td>H-tetrol</td>
<td><img src="image" alt="H-Shaped C_{80} Tetrol" /></td>
<td>+ + + + n.d.</td>
</tr>
<tr>
<td>H-shaped C_{80} monoacid</td>
<td>H-monoacid</td>
<td><img src="image" alt="H-Shaped C_{80} Monoacid" /></td>
<td>n.d. n.d. + n.d. n.d.</td>
</tr>
<tr>
<td>H-shaped C_{80} diacid</td>
<td>H-diacid</td>
<td><img src="image" alt="H-Shaped C_{80} Diacid" /></td>
<td>n.d. n.d. + n.d. n.d.</td>
</tr>
<tr>
<td>H-shaped C_{80} triacid</td>
<td>H-triacid</td>
<td><img src="image" alt="H-Shaped C_{80} Triacid" /></td>
<td>n.d. n.d. + n.d. n.d.</td>
</tr>
<tr>
<td>H-shaped C_{80} tetraacid</td>
<td>H-tetraacid</td>
<td><img src="image" alt="H-Shaped C_{80} Tetraacid" /></td>
<td>n.d. n.d. + n.d. n.d.</td>
</tr>
</tbody>
</table>
Table 1. Compound classes discussed in the paper. Illustrated structures show only the acyclic biphytane derivatives. Constitutional isomers are not included. Representative samples showing the distribution of GDGT and degradation products in various environment settings and cell. ‘+’ compound detected, ‘n.d.’ compound not detected.
Density maps and extracted ion chromatograms (EICs) of NPLC-APCI-qTOF showing the detection of isoprenoidal GDGT-0 and its hydroxylated and carboxylated derivatives with two (Fig. 1a) or one biphytane units (Fig. 1b) in the Marmorito seep carbonate. Multiple isomers were observed for GDDA, GDGTol, GDGTA, GMM and GMGD and are labeled as ‘a’ and ‘b’. The isomeric composition of GMGD, ‘a’ and ‘b’, may provide insights regarding the regioisomerism of GDGTs and will be subject of a future report.
Density maps and EICs of NPLC-APCI-qTOF showing the detection of isoprenoidal GDGT-0 and hydroxylated and carboxylated derivatives with two (Fig. 1a) or one, biphytane units (Fig. 1b) in the Marmorito seep carbonate. Multiple isomers were observed for GDDA, GDGTol, GDGTA, GMM and GMGD and are labeled as ‘a’ and ‘b’. The isomeric composition of GMGD, ‘a’ and ‘b’, may provide insights regarding the regioisomerism of GDGTs.
Fig. 2 Density maps and EICs derived from RPLC-ESI-qTOF analysis showing the detection of H-GMGTs and their hydroxylated and carboxylated derivatives in a hydrothermal sediment from the Guaymas Basin. Stepwise oxidation of the four hydroxyl groups in the H-tetrols possibly generated C_{80} mono-, di-, tri- and H-tetraacids. Mono- and di-methylated H-GMGTs, H-GMDs and corresponding C_{81}, C_{82} H-tetrols and H-tetraacids were also detected. Molecular structures of the intermediates are suggested. Isomers of monoacid, diacid and triacid may exist but could not be separated with the present LC methods.
Fig. 3 Density maps of NPLC-APCI-qTOF showing in one marine sediment (Leg 201-1227) the detection of GDD, GDGTol, biphytane diol, GMM and their co-occurring further degradation derivatives with C_{39} and C_{38} based isoprenoids, (a-d). Highlighted are the most dominant components including the GDGT-0 (open circle) and crenarchaeol-related (rectangle) compounds. Different compound classes are also color coded. C_{40}
hydroxyl derivatives are in green, C₃⁹ in blue and the C₃⁸ related in grey, C₄₀ carboxyl derivatives are in orange and C₃⁹ in pink. Multiple isomers occur due to different ether bonding and ring arrangements. The occurrence of C₄₀/₃⁹-GDGTA was shown in one sample of Marmorito seep carbonate (e). The molecular ion of C₄₀/₃⁹-GDGTA-0 and GDGTol-0 gives similar but distinguishable masses, m/z: 1320.2930 and 1320.3312, respectively. Under APCI, dehydrated ions of carboxyl derivatives occurred (highlighted ions with dashed lines).
Fig. 4 Relative abundances of diagenetic derivatives of GDGT-0 (left panel) and H-GMGT-0 (right panel) in four representative environmental samples.
Fig. 5 Extracted ion chromatograms from NPLC-APCI-qTOF analysis, showing the distribution of GDGT, bpdiol, bmonoacid/ol and bpdiacids in a seep carbonate, Marmorito. Peaks of crenarchaeol and its related derivatives are shaded; crenarchaeol derivatives have decreased abundances in the acid fractions.
Fig. 6 The relative abundance of core crenarchaeol and its related hydroxyl, carboxyl and C_{39} and C_{38} based derivatives in different growth phases of *N. maritimus*. 
Fig. 7 Hypothetical scheme showing the suggested diagenetic pathways of GDGTs. GDGT-0 and its related IPL and degradation derivatives are used as an example. In the flow chart, R, R’ and R’’ represent different alkyl chains. Intermediate components, such as GDDA, bpmonoacid/ol are not included. The formation of C_{38} based isoprenoids is not clear and thus is labeled with ‘?’. 
Supplementary information

Sample preparation

1. Hot spring sediment T-15

Sediment was collected from the bottom of a geothermal wells located in Ruidian (Jinze hot spring: N 25.44138°, E98.46004°, Elevation:1740m. Tengchong County, Yunnan Province, China), with the water pH of 6.71 and temperature of 80.6 °C. Roughly 5 g of freeze-dried sediment was extracted in an ultrasonic bath with methanol (MeOH; twice), methanol-dichloromethane (DCM; 1:1, v/v; twice) and finally DCM (twice). All supernatants were collected in a flask and completely dried under N₂. The total lipid extract was fractionated over a pre-activated silica gel chromatography into apolar (alkane) and polar (intact polar GDGTs and GDGT core lipids) fractions using n-hexane-DCM (9:1, v/v) and DCM-MeOH (1:1, v/v) as eluents respectively. The polar fraction was used for the lipid analysis in this study.

2. H-tetrol standards

An oilfield calcium naphthenate deposit was Soxhlet extracted sequentially (dichloromethane/n-hexane, 1/1, v/v, 300 mL, 8h; toluene/acetone, 3/2, v/v, 250 mL, 6h; propan-2-ol/dichloromethane, 4/1, v/v, 250mL, 6h) to remove interstitial oil and the residue dried (overnight, 70°C) before acidification by heating (70°C, overnight) in a capped and sealed vial with hydrochloric acid (3 mL) and cyclohexane (3 mL). The acidified organic fraction was recovered by extraction into diethylether (DEE; 3 x 5 mL) by mixing (vortex 10 s) and centrifugation (2500 rpm/3 min). Solvent was removed from
the decanted supernatant under blow-down (N₂, 70°C) prior to dilution in DEE/0.1% ammonia (8 mL), loading on a pre-conditioned (1% aqueous ammonia, 20 mL; water, 20 mL; DEE, 10 mL) SAX solid phase extraction cartridge (Sigma-Aldrich Company Ltd., Dorset, UK; DSC-SAX, 12 mL, 2 g) and sequentially eluting the cartridge with 20 mL volumes of DEE, dichloromethane and DEE/2% formic acid (FA). Solvent was removed from the DEE/2% FA fraction (‘acid fraction’) under blow-down (N₂, 70°C) and an aliquot per-trimethylsilylated with BSTFA/1% TMCS (Sigma-Aldrich Company Ltd., UK; ca. 50 µL, 70°C, 1 hr) before dilution in cyclohexane for analysis using high temperature gas chromatography (HTGC; Fig. SX1), or per-methylation for infusion electrospray ionization/mass spectrometry (ESI/MS; Fig. SX2). An aliquot of the isolated acid fraction was diluted in DEE for infrared spectroscopy (Fig. SX3).

The isolated acid fraction was reduced to alcohols by treatment with lithium aluminum hydride (LithAl; Sigma-Aldrich Company Ltd., UK; 1M in DEE). Dried acid fraction was dissolved in a small volume (ca. 1 mL) dry DEE (sodium wire) and transferred with washings to a small three-neck flask with Teflon flea and fitted with a Subaseal, calcium chloride drying tube condenser and a glass stopper, all over a magnetic stirring block. LithAl solution (ca. 3 mL) was pumped into the flask under nitrogen through the Subaseal whilst stirring, additional dry DEE (ca. 5mL) was added through the condenser to break up formed solids, before placing a bowl of warm water under the flask for 10 minutes. After the reaction vessel had cooled, wet DEE (prepared by mixing DEE and water in separating funnel and drawing off aqueous phase) was added (10 mL, drop-wise at first) to hydrolyse unreacted hydrides, followed by drop-wise addition of sulphuric
acid (10% aqueous; ca. 1 mL) to destroy remaining LithAl. The contents of the flask were transferred with washings (ca. 3 mL, 10% $\text{H}_2\text{SO}_4$ aq) to a separating funnel and the organic phase (‘alcohol fraction’) collected after washing with water (3 x 3 mL). Solvent was removed from the alcohol fraction by blow-down ($\text{N}_2$, 70°C) and aliquots prepared for HTGC, ESI/MS and IR spectroscopy as above. All solvents were HPLC grade (Rathburn Chemicals Ltd., Walkerburn, UK or Fisher Scientific UK Ltd., Loughborough, UK) or LC/MS grade (Chromasolv®; Sigma-Aldrich Company Ltd., Dorset, UK), water was Elga Maxima (18.2 mΩ; Elga Ltd., Buckinghamshire, UK).

Reduction of the tetraacids to tetrols was monitored using HTGC (Fig. SX1), ESI/MS (Fig. SX4 and SX5) and IR spectroscopy (Fig. SX3). The HTGC system comprised an Agilent 6890 GC fitted with cool-on-column inlet (+3°C track oven mode; 0.5 µL manual injection), flame ionization detector (435°C) and Varian VF-5ht Ultimetal column (15 m x 0.25 mm x 0.1 µm) with helium carrier gas (1 ml min$^{-1}$, constant flow) and oven programme from 40 – 430°C at 10°C min$^{-1}$ with 10 min hold.

Infusion ESI/MS was carried out in positive ionization mode using a Finnigan Mat LCQ™ (ThermoFinnigan, San Jose, CA, USA) with ESI interface. Samples were diluted in propan-2-ol/10mM ammonium acetate and infused at 3 µL min$^{-1}$ with a Hamilton (Reno, CA, USA) 1725N (250 µL) syringe using the built-in syringe pump. Mass spectral data were acquired (and averaged over 1 minute) and processed using Xcalibur software. Typical instrument parameters were: source voltage (±) 4.5 kV; capillary voltage (±) 60 V; capillary temperature 200 °C; nitrogen sheath gas flow rate 24 (arbitrary units).
Instrumental parameters were optimised on the most abundant ion using the autotune function.

Infrared spectroscopy was carried out using a Bruker Alpha Platinum ATR (Bruker Optik GmbH, Ettlingen, Germany) by measuring 32 sample scans (resolution 4 cm\(^{-1}\); transmittance) and recording data between 4000 – 375 cm\(^{-1}\). Background comprised 32 scans without sample.

Figure SX1. High temperature gas chromatograms of trimethylsilylated tetraacids obtained from an oilfield deposit (upper) and lower the trimethylsilylated tetraols obtained following LithAl reduction (lower).
Figure SX2. Averaged mass spectrum from HPLC-electrospray ionization (+ve) mass spectrometry of per-methylated esters of tetraacids obtained from an oilfield deposit (nR refers to the number of cyclopentyl rings in the molecule)
Figure SX3. FTIR transmittance spectra of tetraacids isolated from an oilfield deposit (lower) and tetraols produced from the tetraacids by LithAl reduction (upper).
Figure SX4. Infusion electrospray ionization (+ve) mass spectrum of underivatised tetraols obtained from the LithAl reduction of tetraacids from an oilfield deposit (nR refers to the number of cyclopentyl rings in the molecule).
Figure SX5. Infusion electrospray ionization (+ve) mass spectrum of per-trimethylsilylated tetraols obtained from the LithAl reduction of tetraacids from an oilfield deposit (nR refers to the number of cyclopentyl rings in the molecule).

3. Acid hydrolysis of GDGT-0 standard

GDGT-0 standard was isolated with a semi-preparative LC protocol (as described in Zhu et al., 2014) from acid-hydrolyzed biomass of *Archaeoglobus fulgidus*. Strong acid hydrolysis was then performed with 10% methanolic HCl and GDGT-0 standard under 70 °C for 96 hours. After dried with a N₂ flow the treated sample was dissolved in *n*-hexane for LC-APCI-MS analysis.
Reference:

Figure S1a. MS$^2$ fragmentation patterns supporting the identification of GDGT-0, GDGTol-0 and GDGTA-0 in Fig. 1a, and C$^{40/39}$-GDGTol-0 and C$^{40/39}$-GDGTA-0 in Fig. 3e.
Figure S1b. MS² fragmentation patterns supporting the identification of GDD-0, GDDA-0, and GDDAA-0 in Fig. 1a.
**Figure S1c.** MS\(^1\) ions supporting the identification of GMM-0, GMMA-0 and GMGD-0 in Fig. 1b.
Figure S1d. MS$^1$ ions supporting the identification of bpdiol-0, bpmonoacid/ol-0 and bpdiacid-0 in Fig. 1b.
Figure S2

Major hydroxyl derivatives released from GDGT-0 by chemical degradation, a mild ether cleavage conducted by adding 1000ng GDGT-0 into 1mL of 10% HCl in methanol, and
heated to 70 °C for 96 hours. The composition of GMGDs indicated a nearly 1:1 mixture of parallel and anti-parallel GDGT-0.

Figure S3
Density maps of NPLC-APCI-qTOF showing the occurrence of H- GDGTs and their degradation derivatives in the hot spring sediment, T-15. The identification of H-tetrols are confirmed by their similar chromatographic behavior with added standards (green color text).
Figure S4

Extracted ion chromatograms of NPLC-APCI-qTOF showing the occurrence of branched GDGTs and derivatives in Marmorito seep carbonate.
Figure S5

Extracted ion chromatograms of NPLC-APCI-qTOF showing the occurrence of IB-GDGTs and derivatives in Marmorito seep carbonate.
<table>
<thead>
<tr>
<th>Compounds</th>
<th>Abbreviation</th>
<th>Structural illustration</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>glycerol dialkyl glycerol tetraether</td>
<td>GDGT</td>
<td><img src="image" alt="GDGT结构" /></td>
<td>+ + + + + +</td>
</tr>
<tr>
<td>glycerol dialkyl glycerol triether alcohol</td>
<td>GDGTol</td>
<td><img src="image" alt="GDGTol结构" /></td>
<td>+ + + n.d. +</td>
</tr>
<tr>
<td>glycerol dialkyl glycerol triether acid</td>
<td>GDGTA</td>
<td><img src="image" alt="GDGTA结构" /></td>
<td>+ + n.d. n.d. n.d.</td>
</tr>
<tr>
<td>glycerol dibiphytanol diether</td>
<td>GDD</td>
<td><img src="image" alt="GDD结构" /></td>
<td>+ + + + + +</td>
</tr>
<tr>
<td>glycerol dibiphytanol diether monoacid</td>
<td>GDDA</td>
<td><img src="image" alt="GDDA结构" /></td>
<td>+ + + n.d. +</td>
</tr>
<tr>
<td>glycerol dibiphytanol diether diacid</td>
<td>GDDAA</td>
<td><img src="image" alt="GDDAA结构" /></td>
<td>n.d. + n.d. n.d. n.d.</td>
</tr>
<tr>
<td>glycerol monobiphytanyl glycerol diether</td>
<td>GMGD</td>
<td><img src="image" alt="GMGD结构" /></td>
<td>+ + + n.d. n.d.</td>
</tr>
<tr>
<td>glycerol monobiphytanol monoether</td>
<td>GMM</td>
<td><img src="image" alt="GMM结构" /></td>
<td>+ + + n.d. n.d.</td>
</tr>
<tr>
<td>glycerol monobiphytanol monoether acid</td>
<td>GMMA</td>
<td><img src="image" alt="GMMA结构" /></td>
<td>n.d. + n.d. n.d. n.d.</td>
</tr>
<tr>
<td>biphytanic diol</td>
<td>bpdiol</td>
<td><img src="image" alt="bpdiol结构" /></td>
<td>+ + + + + +</td>
</tr>
<tr>
<td>biphytanic monoacid</td>
<td>bpmonoacid/ol</td>
<td><img src="image" alt="bpmonoacid/ol结构" /></td>
<td>+ + n.d. n.d. n.d.</td>
</tr>
<tr>
<td>biphytanic diacid</td>
<td>bpdiacid</td>
<td><img src="image" alt="bpdiacid结构" /></td>
<td>n.d. + n.d. n.d. n.d.</td>
</tr>
<tr>
<td>H-shaped glycerol monoalkyl glycerol tetraether</td>
<td>H-GMGT</td>
<td><img src="image" alt="H-GMGT结构" /></td>
<td>+ + + + n.d.</td>
</tr>
<tr>
<td>H-shaped glycerol monoalkyl diether</td>
<td>H-GMD</td>
<td><img src="image" alt="H-GMD结构" /></td>
<td>+ + + + n.d.</td>
</tr>
<tr>
<td>H-shaped C₈₀ tetrol</td>
<td>H-tetrol</td>
<td><img src="image" alt="H-tetrol结构" /></td>
<td>+ + + + n.d.</td>
</tr>
<tr>
<td>H-shaped C₈₀ diacid</td>
<td>H-diacid</td>
<td><img src="image" alt="H-diacid结构" /></td>
<td>n.d. n.d. + n.d. n.d.</td>
</tr>
</tbody>
</table>
Marmorito seep carbonate, NPLC-APCI-qTOF

- **B-GDGT-1022**: m/z: 1036.0253
- **B-GDGT-1036**: m/z: 1050.0410
- **B-GDGT-1050**: m/z: 1022.0097
- **B-GDD-966**: m/z: 965.9835
- **B-GDD-980**: m/z: 979.9991
- **B-GDD-994**: m/z: 994.0148
- **B-GMM-530**: m/z: 529.5190
- **B-GMM-544**: m/z: 543.5347
- **B-GMM-558**: m/z: 557.5510
- **B-diol-455**: m/z: 455.4823
- **B-diol-469**: m/z: 469.4979
- **B-diol-483**: m/z: 483.5141
Marmorito seep carbonate, NPLC-APCI-qTOF

IB-GDGT-1162 $\text{EIC, } m/z: 1162.1662$

IB-GDGT-1176 $m/z: 1176.1818$

IB-GDGT-1190 $m/z: 1190.1975$

IB-GDD-1106 $m/z: 1106.1400$

IB-GDD-1120 $m/z: 1120.1556$

IB-GDD-1134 $m/z: 1134.1713$

IB-diol-525 $m/z: 525.5605$

IB-diol-539 $m/z: 539.5762$

IB-diol-553 $m/z: 553.5918$