

ARTICLE



<https://doi.org/10.1038/s41467-021-21480-1>

OPEN

Unravelling the structural complexity of glycolipids with cryogenic infrared spectroscopy

Carla Kirschbaum  ^{1,2}, Kim Greis ^{1,2}, Eike Mucha ², Lisa Kain ³, Shenglou Deng ⁴, Andreas Zappe ¹, Sandy Gewinner ², Wieland Schöllkopf  ², Gert von Helden  ², Gerard Meijer  ², Paul B. Savage ⁴, Mateusz Marianski  ^{5,6}, Luc Teyton ³ & Kevin Pagel  ^{1,2}✉

Glycolipids are complex glycoconjugates composed of a glycan headgroup and a lipid moiety. Their modular biosynthesis creates a vast amount of diverse and often isomeric structures, which fulfill highly specific biological functions. To date, no gold-standard analytical technique can provide a comprehensive structural elucidation of complex glycolipids, and insufficient tools for isomer distinction can lead to wrong assignments. Herein we use cryogenic gas-phase infrared spectroscopy to systematically investigate different kinds of isomerism in immunologically relevant glycolipids. We show that all structural features, including isomeric glycan headgroups, anomeric configurations and different lipid moieties, can be unambiguously resolved by diagnostic spectroscopic fingerprints in a narrow spectral range. The results allow for the characterization of isomeric glycolipid mixtures and biological applications.

¹ Institut für Chemie und Biochemie, Freie Universität Berlin, Berlin, Germany. ² Fritz-Haber-Institut der Max-Planck-Gesellschaft, Berlin, Germany.

³ Department of Immunology and Microbiology, Scripps Research, La Jolla, CA, USA. ⁴ Department of Chemistry and Biochemistry, Brigham Young University, Provo, UT, USA. ⁵ Department of Chemistry and Biochemistry, Hunter College, The City University of New York, New York, NY, USA. ⁶ The PhD Program in Chemistry, Graduate Center, The City University of New York, New York, NY, USA. ✉email: kevin.pagel@fu-berlin.de

Glycolipids are amphiphilic biomolecules that are omnipresent in the cell membranes of all kinds of organisms ranging from bacteria to humans¹. Playing key roles in cellular interactions and signal transduction, they are essential for the development and function of multicellular organisms^{2–4}. Furthermore, immune responses can be modulated by α -linked glycolipid antigens such as α -galactosyl ceramides (GalCer). In case of microbial infections, they trigger the activation of natural killer T-cells (NKT), a T cell subset sitting at the interface between innate and adaptive immunities^{5–8}. α -GalCer was first isolated as an antitumor agent from marine sponge⁹ and was thought to be produced exclusively by bacteria and porifera¹⁰. In mammalian cells, however, only β -isomers were detected—which raised the question how NKT cells are triggered in mammals^{5,11}. At the end of a controversial debate lasting for more than a decade, the presence of low-abundant, endogenous α -GalCer in mammalian cells was finally revealed using a combination of biological, enzymatic, and immunological assays^{5,12} and was later confirmed by direct biochemical evidence¹³. On the other hand, established analytical techniques failed to detect α -GalCer in the presence of highly abundant, completely non-antigenic β -GalCer, which caused severe confusion in the meantime¹⁴.

The cumbersome search for endogenous antigens of NKT cells in mammals illustrates a general issue in glycolipid research: the lack of techniques to accurately analyze glycolipids and to synthesize isomerically pure standards^{10,15}. Classical (tandem-) mass spectrometry (MS) workflows are sensitive but unable to distinguish the anomeric configuration of GalCer; nuclear magnetic resonance (NMR) yields comprehensive stereochemical information but requires comparably large sample amounts, and cannot ensure the detection of low-abundant isomers in mixtures. Hundreds of different glycosphingolipids were identified in nature based upon sugar heterogeneity, without taking into account the structural diversity of lipid moieties³. However, as illustrated by the example of GalCer, a biological function is often related to one specific isomer, and minute alterations of the glycolipid structure can completely eradicate its function. Isomer distinction is thus a highly relevant issue requiring novel analytical approaches.

Here we investigate consistent sets of synthetic glycolipid isomers (Supplementary Table 1) using cryogenic gas-phase infrared (IR) spectroscopy in helium nanodroplets¹⁶. In this technique, protonated or sodiated glycolipids are generated by nano-electrospray ionization, mass-to-charge selected, pre-cooled by buffer gas cooling (80 K) and then captured in superfluid helium nanodroplets. The latter function as IR-transparent cryostats with an internal temperature of 0.4 K¹⁷. Upon the resonant absorption of an IR photon by the ion inside a droplet, the vibrational energy is rapidly dissipated by evaporative cooling. After the absorption of multiple photons and repeated cycles of helium evaporation, the bare ion is released from the droplet and detected by MS. IR spectra are generated by monitoring the ion yield while scanning the wavenumber range of interest. The tunable, high intensity IR radiation is provided by the Fritz Haber Institute free-electron laser (FHI FEL)¹⁸. The technique allows to distinguish not only between α -GalCer and β -GalCer but also between different isomeric glycan headgroups and different lipid moieties. The identification and relative quantification of particular glycolipid isomers in mixtures is demonstrated using synthetic 2-component, 3-component, and 4-component mixtures and two biological lipid extracts from mice.

Results

α -galactosylceramide and β -galactosylceramide. The study was initiated by investigating α -GalCer and β -GalCer (Fig. 1a). This

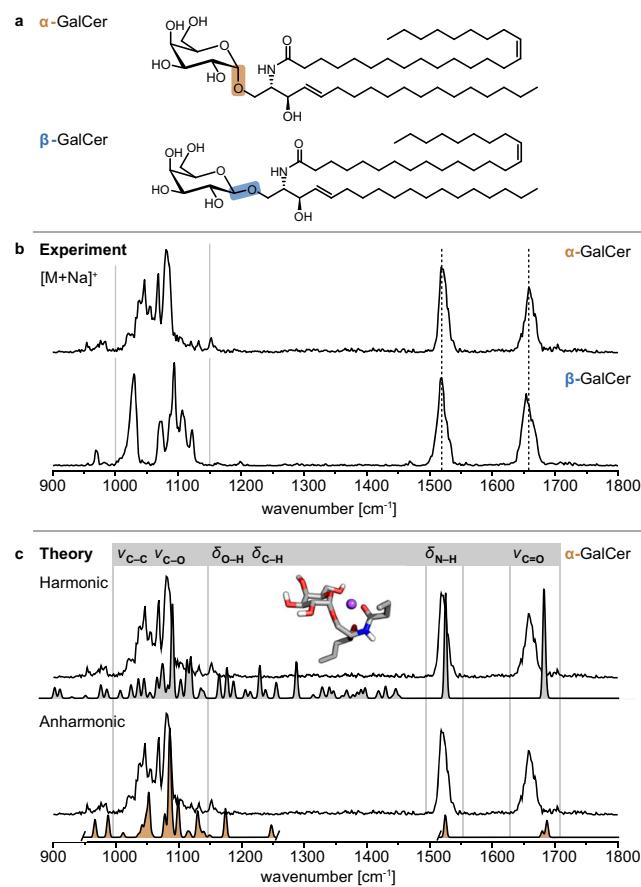


Fig. 1 Structures and IR spectra of α -GalCer and β -GalCer (d18:1/24:1 (15Z)). **a** The stereoisomers α -GalCer and β -GalCer differ by the configuration of the glycosidic bond. **b** Sodium adducts of α -GalCer and β -GalCer yield unique fingerprints in the 1000–1150 cm^{-1} region. **c** The assignment of vibrational bands is based on computed vibrational spectra of the lowest-energy conformer of $[\alpha\text{-GalCer}+\text{Na}]^+$ with the lipid chains trimmed to four heavy atoms. The anharmonic spectrum (orange) provides a clear match with the experimental fingerprint. Source data are provided as a Source Data file.

pair of stereoisomers is not distinguishable by ion mobility-mass spectrometry (IM-MS) (Supplementary Table 2), and tandem MS relies on relative ion intensities at different collision energies to resolve α -GalCer and β -GalCer¹³. In contrast, cryogenic gas-phase IR spectroscopy probes the ion's structure directly, involving the stereochemistry of the glycosidic bond¹⁶. The resulting IR spectra therefore feature distinct spectroscopic signatures in the 1000–1150 cm^{-1} region for α -GalCer and β -GalCer $[\text{M}+\text{Na}]^+$ (Fig. 1b) and $[\text{M}+\text{H}]^+$ ions (Supplementary Fig. 3).

The theoretical spectra of the lowest-energy adducts of $[\alpha\text{-GalCer}+\text{Na}]^+$ and $[\beta\text{-GalCer}+\text{Na}]^+$ with truncated lipid chains were derived using harmonic approximation (Fig. 1c and Supplementary Fig. 22) and revealed that this diagnostic fingerprint region is composed of C–O and C–C stretching vibrations (ν) of the sugar ring. However, while the absorption frequencies of the non-diagnostic N–H bending vibration (amide II) and the C=O stretching vibration (amide I) are in good agreement with the experimental values, the shape of the spectra in the diagnostic 1000–1150 cm^{-1} region differs. Reattachment of the lipid chains to a conformer of $[\alpha\text{-GalCer}+\text{Na}]^+$ resulted in only minor changes in the diagnostic region of the theoretical spectrum (Supplementary Fig. 23). Instead, the mismatch originates from the harmonic molecular potential¹⁹ derived using an approximate density

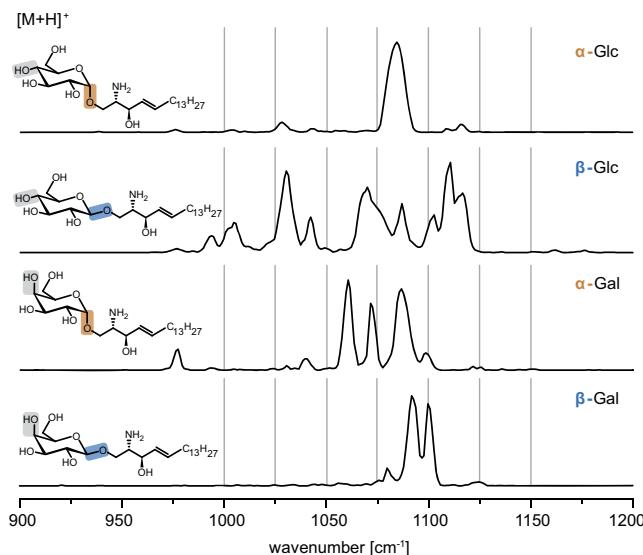


Fig. 2 Spectroscopic fingerprints of protonated isomeric Gal- and Glc sphingosines. The highly resolved absorption patterns of each permutation are diagnostic for both the monosaccharide and the configuration of the glycosidic bond. Source data are provided as a Source Data file.

functional²⁰, and including anharmonic effects in the theoretical IR spectrum improves the match between the spectra in the fingerprint region (Fig. 1c and Supplementary Fig. 24)²¹. The region between 1150–1450 cm⁻¹ is dominated by C–H and O–H bending vibrations (δ) and shows a very low intensity in the experimental spectra. In summary, the spectral signatures of α -GalCer and β -GalCer demonstrate that the assignment of the anomeric configuration can be accomplished exclusively based on the narrow, merely 200 cm⁻¹ wide fingerprint region.

Stereoisomeric monosaccharides. The ability to distinguish α -GalCer and β -GalCer entailed a more systematic study of isomeric glycolipids, starting with glycosylsphingosines as the simplest possible glycolipids and then gradually increasing size and complexity. The glycosylsphingosines α -Gal and β -Gal sphingosine are the primary degradation products of α -GalCer and β -GalCer formed by the enzymatic removal of the fatty acyl chain^{5,22}. It was recently shown that α -Gal sphingosine shows antigenic activity towards NKT cells despite the missing lipid chain²². In addition to α -Gal and β -Gal sphingosine, the corresponding glucose (Glc) epimers were investigated, as either Glc or Gal are typically linked as first sugars in mammalian glycolipids¹. Glc and Gal sphingosine are distinguishable after offline⁻²³ or online⁻²⁴ modification by tandem-MS but their distinction relies only on relative intensity differences of generated fragments. Without modification, tandem-MS and IM-MS provide no stereochemical information (Supplementary Fig. 2). In contrast, gas-phase IR spectra of the protonated species yield diagnostic, baseline-resolved absorption patterns in the fingerprint region (Fig. 2). The spectra are unique for each combination of monosaccharide (Glc or Gal) and anomeric configuration (α or β). Some absorption bands are so unique that the corresponding structure could be distinguished from the others by only monitoring the absorption at one specific wavenumber, for example 1065 cm⁻¹ for α -Gal sphingosine. Besides the diagnostic fingerprint region between 1000–1150 cm⁻¹, the spectra display only weak absorption bands associated with the umbrella motion of NH₃⁺ between 1400–1500 cm⁻¹ (Supplementary Fig. 4).

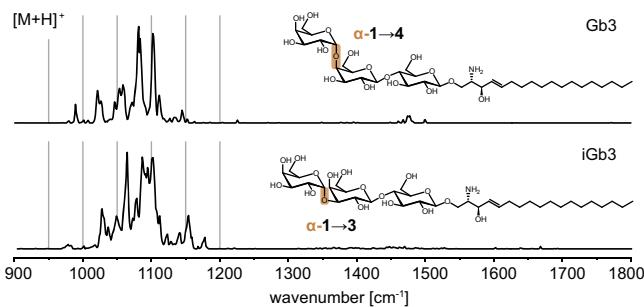


Fig. 3 IR spectra of isomeric Gb3-sphingosine and iGb3-sphingosine. The regiosomeric trisaccharides yield distinguishable and well-resolved absorption patterns in the fingerprint region. Source data are provided as a Source Data file.

Regiosomeric trisaccharides. With increasing glycan size, the number of possible glycan isomers rises exponentially²⁵ while the spectral resolution deteriorates¹⁶. To test the influence of the glycan size on the informational content of the IR spectra, globotriose (Gb3) was selected as a common, naturally occurring trisaccharide headgroup containing two Gal and one Glc unit. The ability to resolve subtle structural differences in the trisaccharide was tested by including iso-Gb3 (iGb3), the first reported endogenous NKT cell antigen^{26,27}. The chemical structures of Gb3 and iGb3 differ by the connectivity between the two Gal building blocks (1 → 3 vs. 1 → 4) (Fig. 3). This difference in the molecular geometry causes slightly different ion mobilities of sodiated Gb3-sphingosine and iGb3-sphingosine (Supplementary Table 2); however, the individual arrival time distributions of the isomers are not separated in a mixture. Cryogenic IR spectroscopy allows for a much clearer isomer distinction. Even though the IR spectra of Gb3-sphingosine and iGb3-sphingosine are more congested than the spectra of monosaccharide headgroups, they are still well-resolved, and the unique fingerprints demonstrate that the connectivity between monosaccharide building blocks can be determined by IR spectroscopy. Finally, the glycolipid size was further increased by replacing sphingosine by ceramide. The IR spectra of protonated and sodiated α -Gb3Cer (d18:1/26:0) display distinct absorption bands but the spectra are more congested, suggesting that the size limit for glycolipids is almost attained (Supplementary Fig. 5).

Lipid residues. So far, the investigated glycolipid structures were restricted to sphingolipids based on sphingosine. However, the sphingolipid backbones in nature are not exclusively relying on sphingosine, and a smaller number of glycolipids in mammals are not at all based on a sphingolipid—but a glycerol backbone^{1,3}. Several glycerolipids bearing α -linked Gal²⁸ or Glc²⁹ headgroups were for example identified as bacterial ligands of NKT cells. The influence of different lipid moieties on the IR spectra is shown on the example of α -Gal attached to sphingosine, phytosphingosine, ceramide (d18:1/24:1(15Z)) and diacylglycerol (14:0/14:0) (Fig. 4). The formal addition of water to the C=C double bond of mammalian sphingosine to generate its plant analog phytosphingosine does not significantly alter the fingerprint region. The frequencies of the three main bands are identical, whereas a weak absorption at 950 cm⁻¹ is only visible in the spectrum of α -Gal sphingosine. Ceramide, however, yields a significantly different absorption pattern in the diagnostic fingerprint region and leads to the appearance of characteristic amide vibrations above 1450 cm⁻¹. The spectrum of α -Gal diacylglycerol displays a less defined fingerprint region and more visible C–H bending vibrations. The ester groups yield additional C=O stretching vibrations above 1700 cm⁻¹. Calculation revealed substantial mixing

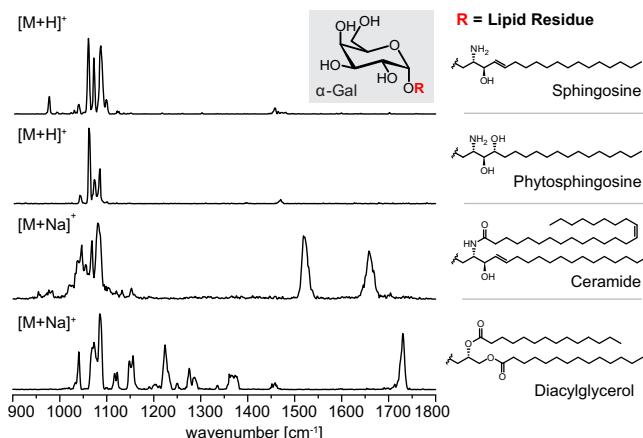


Fig. 4 Influence of different lipid moieties on the IR spectra of α -Gal lipids. The exchange of sphingosine for phytosphingosine leads to subtle differences in the spectral fingerprint, whereas ceramide and diacylglycerol considerably alter the fingerprint region. The amide- and ester groups yield additional bands beyond 1450 cm^{-1} . Source data are provided as a Source Data file.

of the stretching of the two carbonyl groups resulting in in-phase and out-of-phase vibrational modes (Supplementary Fig. 25). The two modes are, however, not resolved in the experimental spectrum. Overall, these examples highlight the fact that, despite their exceptional resolution, cryogenic IR spectra are challenging to deconvolute using known increments. As a result, routine analyses will require spectral libraries containing distinct glycolipid reference data.

Glycolipid mixtures. Reference spectra of glycolipid standards can allow for studying more complex isomeric mixtures and estimating molar ratios. To test the utility of this approach, a proof-of-concept study on several glycolipid mixtures was performed. Three different aspects were addressed: (1) variation of mixing ratios in binary synthetic mixtures to evaluate the dependence of the absorption intensities on the relative concentrations and to determine the limit of detection, (2) deconvolution of more complex synthetic mixtures composed of up to four isomeric glycolipids, and (3) application to biological lipid extracts.

At first, binary mixtures of α -GalCer and β -GalCer with defined mixing ratios were investigated (Supplementary Fig. 9). The experimental spectra were compared with theoretical spectra obtained by weighting and averaging the two reference spectra of pure α -GalCer and β -GalCer according to their mixing ratios. Even though this simple mathematical approach assumes a strictly linear decrease of intensity with decreasing relative concentration, the theoretical spectra model the experimental spectra well (Supplementary Fig. 9d-h). This finding implies that the relative intensities scale roughly linearly with the molar ratio over a wide range of mixing ratios. The respective contributions of the pure compounds to the mixtures were also quantitatively determined with an exceptional accuracy by non-negative matrix factorization (NMF, Supplementary Figs. 10–11)^{30,31}. This factorization method deconvolutes the spectra of isomeric mixtures into the spectra of α -GalCer and β -GalCer (basis vectors), and their relative contribution to each of the mixtures (weighting factors). The weighting factors were found to be accurate within an error range of less than 5%. NMF is thus a well-suited method for spectral deconvolution of binary glycolipid mixtures, provided that the abundance of the minor isomer is not much below 5%. In accordance with this limit of reliable detection, α -GalCer could be

detected and quantified in a 5:95 ($\alpha:\beta$) mixture but was undetectable in a 1:99 mixture. This detection limit for minor species in isomeric mixtures is within the same order of magnitude as that of NMR spectroscopy³².

More complex ternary and quaternary mixtures of isomeric glycolipids were investigated to assess if spectral deconvolution is still possible at increasing spectral congestion. For this purpose, the four isomers α -Gal/Glc and β -Gal/Glc phytosphingosine were mixed in any possible combination of 2-component, 3-component, and 4-component mixtures with equal concentrations. NMF correctly retrieves which isomer is present in which of the 11 mixtures (Supplementary Figs. 12–14). Due to the increased complexity of the mixtures (four instead of two possible compounds), the mixing ratios predicted by NMF are not as exact as in the case of binary GalCer mixtures but still sufficiently accurate.

Having established the utility of IR spectroscopy and NMF for the identification and relative quantification of isomers in synthetic glycolipid mixtures, the method was applied to biological lipid extracts. Two lipid extracts **1** and **2** were prepared from cells of α -galactosidase (GLA) and α -glucosidase (GAA) knockout mice, respectively^{33,34}. After reversed-phase HPLC separation, glycosylceramides (monoisotopic mass = 809.7 amu) were detected in both samples by MS and MS/MS (Supplementary Fig. 16). Isobaric phosphatidylcholines were removed by treatment with NaOH (Supplementary Figs. 17–18), and the remaining saponified glycosylceramides were investigated without further purification by IR spectroscopy in the diagnostic fingerprint region and in the amide region (Supplementary Fig. 19). The successful removal of phosphatidylcholines was confirmed by the absence of characteristic ester carbonyl stretching vibrations between $1700\text{--}1800\text{ cm}^{-1}$. To assign the isomers present in the biological samples, IR spectra of α -GlcCer and β -GlcCer (d18:1/24:1(15Z)) standards were recorded (Supplementary Fig. 8). Both biological samples display very similar spectroscopic fingerprints with β -GlcCer as predominant isomer (Fig. 5). However, the results from NMF clearly indicate that extract **2**, contrary to extract **1**, also contains a considerable fraction of α -GlcCer (Supplementary Figs. 20, 21). This finding agrees with the fact that extract **2** was obtained from GAA mice, which lack an enzyme cleaving alpha-glucosidic bonds. In Folch extract **1** from GLA mice, the presence of α -GalCer could, however, not be confirmed. In fact, only GlcCer but not GalCer was reliably detected in the biological samples. In general, the accuracy of weighting factors obtained by NMF is higher in the case of synthetic mixtures than for deconvolution of biological mixtures. This can be partly attributed to the lower signal-to-noise (*s/n*) ratio of the spectra caused by a lower glycolipid concentration in the biological extracts. The spectrum of extract **1** displays a higher *s/n* ratio than the spectrum of extract **2**, which agrees with a higher MS signal intensity (Supplementary Fig. 17 vs. Supplementary Fig. 18).

The exemplary investigation of two biological lipid extracts demonstrates that MS-based IR spectroscopy can provide informative spectra despite low sample concentration and interferences from the biological matrix, while requiring only few basic purification steps. As the number of possible monoglycosyl lipid isomers underlying a certain *m/z* peak is restricted (usually Glc or Gal), the isomers in question can be unambiguously identified with the help of a small set of reference spectra. The sensitivity of the technique is sufficient to identify certain changes in the isomer distribution, as shown by the example of GAA mice. In contrast to NMR spectroscopy—the gold-standard for direct structure assignment of molecules in solution—IR spectroscopy is furthermore compatible with the small glycolipid quantities typically found in biological samples.

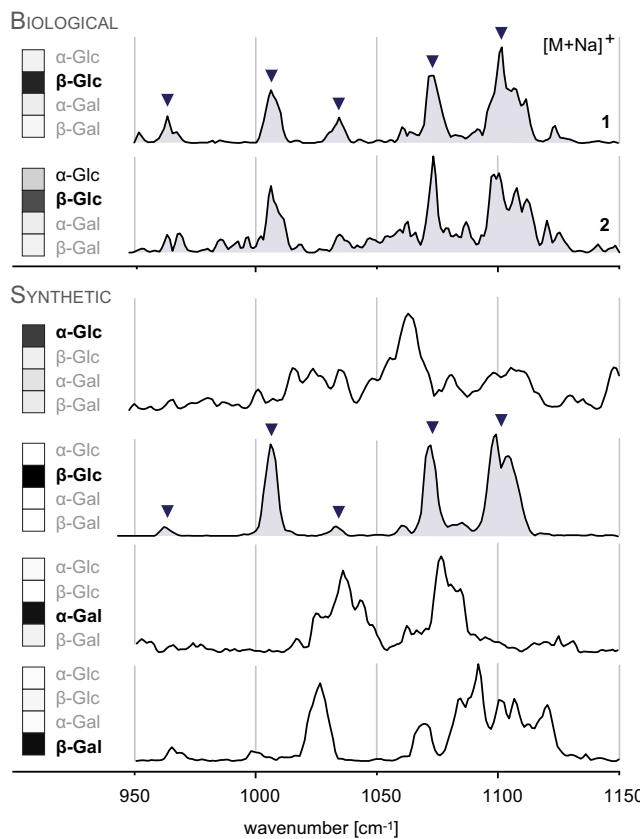


Fig. 5 IR spectra of sodiated glycosylceramides (m/z 832.7) from biological lipid extracts and from synthetic α -Glc/GalCer and β -Glc/GalCer (d18:1/24:1(1Z) standards. The five main absorption bands found in both spectra of biological glycolipids are equally present in the spectrum of β -GlcCer. In addition, α -GlcCer contributes to the spectrum of Folch extract **2** isolated from spleen of GAA knockout mice. The isomer contributions obtained from spectral deconvoluted by NMF are depicted in the left panel. Source data are provided as a Source Data file.

Assuming sample concentrations of 0.1–1 mM required for NMR spectroscopy vs. 0.01–0.1 mM for IR spectroscopy, and sample volumes of 1 mL and 10 μ L for a complete measurement, respectively, NMR spectroscopy requires a 100–1000 fold larger total sample amount (0.1–1 μ mol) than MS-based IR spectroscopy (0.1–1 nmol). Furthermore, and again contrary to NMR, IR spectroscopy does not require pure samples; the quality of IR spectra is not impaired by impurities from biological matrices present in solution, because the glycolipid of interested is isolated in the gas phase by a mass-to-charge filter prior to measurement.

Discussion

In conclusion, this comprehensive spectroscopic study demonstrates the potential of cryogenic gas-phase IR spectroscopy for the characterization of glycolipid isomers. The technique overcomes substantial analytical difficulties previously expressed with reference to immunological studies, where “the anomeric identity of the isolated compound could not be probed directly by MS given that α -anomers and β -anomers are isobaric species or by NMR because quantities were so limiting”⁵. Using IR spectroscopy, all investigated structural features including anomeric configurations, regiosomeric and stereoisomeric glycan headgroups and different lipid classes were unambiguously resolved. Substantial advantages of the MS-based detection scheme over existing structure-sensitive techniques, such as NMR spectroscopy are indeed a 100–1000-fold lower sample consumption and

tolerance towards non-isobaric impurities, which pave the way for straightforward biological applications without extensive sample purification and enrichment. The practical application is further facilitated by the narrow spectral range covering only 200 cm^{-1} , in which most of the structural information is condensed. Both the high sensitivity and short scan time allowed for the characterization of low-abundant glycolipids from biological lipid extracts and enabled the monitoring of changes in the isomer distribution. The deconvolution of spectra of glycolipid mixtures requires reference spectra from synthetic glycolipids; however, as the number of potential glycolipid isomers in biology is limited, this limitation will resolve over time. Much more complex biological glycan headgroups might become accessible by fragmentation and subsequent spectroscopic interrogation of the generated fragments. Furthermore, gas-phase IR spectroscopy has the potential to become more widely applicable in the future, when tagging spectroscopy techniques that require less powerful, commercially available benchtop light sources are used.

Methods

Sample preparation. β -GlcCer, β -Gb3 sphingosine, and β -iGb3 sphingosine were purchased from Avanti Polar Lipids (Alabaster, USA). Synthesis routes of the investigated glycosyl (phyto-)sphingosines⁵ and α -Gb3Cer³⁵ were described previously. α -GalCer and β -GalCer³⁶, α -GlcCer³⁶, α -Gal diacylglycerol³⁷ and β -Gal diacylglycerol³⁸ were synthesized by following published procedures and adapting the lipid residues. 100 μ M and 10 μ M solutions of each glycolipid were prepared for obtaining IR spectra and ion mobility data, respectively. β -Gb3- and β -iGb3 sphingosine were dissolved in pure methanol. α -Gb3Cer was dissolved in dimethyl sulfoxide and diluted with methanol. The other glycolipids were dissolved in dimethyl sulfoxide (1–15 ms) and diluted in a 1:1 (v:v) mixture of acetonitrile and chloroform to obtain 1 mM stock solutions. Prior to measurements, the stock solutions were diluted in a 2:2:1 (v:v) mixture of acetonitrile, methanol and water. All solvents were purchased from Sigma-Aldrich. The solutions were stored at -32°C .

IM-MS and tandem MS. Drift tube ion mobility-mass spectrometry (DT-IM-MS) and tandem mass spectrometry (MS/MS) were performed on a modified Synapt G2-S HDMS instrument (Waters Corporation, Manchester, UK) containing a drift tube instead of the commercial traveling wave cell³⁹. Glycolipid solutions (10 μ M) were prepared as described in the previous section. In addition to protonated and sodiated glycolipids, silver adducts were also investigated. Silver adduction was shown to enable isomer distinction in several lipids by IM⁴⁰, and silver ion chromatography is commonly employed to separate lipids due to the preference of Ag^+ ions to coordinate carbon–carbon double bonds in hydrocarbon chains^{41,42}. Silver adducts were prepared by mixing a 17 mM solution of $\text{Ag}[\text{PF}_6]$ in acetonitrile with 100 μ M glycolipid solutions in a ratio of 1:10. Ions were generated by nano-electrospray ionization and drift times were converted into collision cross sections (CCS) using the Mason–Schamp equation⁴³. The measurements were repeated on three different days. The double standard deviation of the individual measurements is in all cases $\leq 1\%$ of the absolute CCS. MS/MS spectra were obtained by collision-induced dissociation (CID) in the trap cell.

Computational methods. The conformational space of sodiated glycolipids was sampled using Maestro⁴⁴ relying on force field molecular dynamics and CREST⁴⁵. To save computational time and render the conformational search tractable, the lipid chain was truncated to feature the glycan moiety and relevant functional groups of the lipid chain. The sampling of sodiated α -GalCer and β -GalCer was performed using low-mode molecular dynamics sampling in Maestro and Amber* force field and the resulting structures were reoptimized in FHI-aims⁴⁶ using PBE + vdW^{TS}^{47,48} dispersion-corrected density-functional approximation (DFT) and light basis set settings. This method showed chemical accuracy for a large carbohydrate benchmark set^{49–51}. The sampling of sodiated α -Gal and β -Gal diacylglycerol was done using CREST with GFN2-xTB⁵² and default settings. A series of low-energy conformers below a threshold of 20 kJ mol^{-1} were reoptimized for each glycolipid in Gaussian 16 Rev. A.03⁵³ at PBE0 + D3/6-311 + G(d,p) level of theory⁵⁴ and using ultrafine grid settings. Harmonic frequencies were computed at the same level of theory and scaled by a factor of 0.965. For each glycolipid, the lowest-energy structure, which also yields the best spectral match, is shown (Supplementary Figs. 22 and 25). To determine whether truncation of the lipid chains affects the IR signature, full lipid chains were added to one of the low-energy conformers of truncated α -GalCer. The force field-based conformational search was repeated with restraints on atoms of the truncated parentation. A compact structure with multiple Van der Waals contacts between the lipid chains and the galactose moiety was selected, and reoptimized using the same DFT level of theory and followed by derivation of a harmonic spectrum (Supplementary Fig. 23). In addition, anharmonic spectra of four low-energy conformers of sodiated α -GalCer

were derived using the GVPT2 method implemented in Gaussian 16 Rev. B01^{21,55}. These calculations were performed at the same PBE0 + D3/6-311 + G(d,p) level of theory and using ultrafine grid settings for modes 68–88 (1000–1200 cm⁻¹ region) and 128–130 (amide II, amide I, and C=C vibrations, respectively), which correspond to the vibrations in the experimental window (Supplementary Fig. 24). The resulting anharmonic spectrum was shifted by a constant factor of 20 cm⁻¹.

Synthetic glycolipid mixtures. Synthetic isomeric mixtures of GalCer were prepared by mixing 100 μm solutions of α-GalCer and β-GalCer with different mixing ratios: 50:50, 75:25, 90:10, 95:5, and 99:1 (β:α). Mixtures of Glc/Gal phytosphingosines were prepared by mixing 100 μm solutions of the pure isomers to obtain any possible combination of 2-component mixtures (1:1), 3-component mixtures (1:1:1), and a 4-component mixture (1:1:1:1). IR spectra of the protonated glycolipids (GalCer *m/z* 810.7; Glc/Gal phytosphingosine *m/z* 480.4) were recorded in the diagnostic fingerprint region (1000–1150 cm⁻¹) for the pure isomers and mixtures. The individual measurements were usually performed during one day, and a constant laser focus was applied to reduce variations of the laser fluence. The experimental spectra of GalCer with different mixing ratios were compared to simulated spectra, which were generated by averaging the spectra of the pure isomers with the expected ratios (1:1, 1:3, 1:9, 1:19, and 1:99) using the averaging function in OriginPro 2020 (OriginLab Corporation). Both the resulting simulated spectrum and the experimental spectrum were then normalized to a surface area of 1 in the region from 1000 to 1150 cm⁻¹ and superposed (Supplementary Fig. 9d–h).

Non-negative matrix factorization (NMF). NMF factorizes an input matrix into two matrices containing the basis vectors and weighting factors, respectively. In contrast to other factorization methods such as principal component analysis, NMF forces all matrix elements to be non-negative—which is an inherent property of IR data—and therefore only allows for additive combinations of single components^{30,56}. In the present work, the input matrix contains experimental IR spectra of glycolipid mixtures, which are deconvoluted into the component spectra and their relative abundance. In the case of binary GalCer mixtures, the input matrix contains five spectra of isomeric mixtures as well as the spectra of the pure isomers. The output is a matrix containing the two spectra of α-GalCer and β-GalCer multiplied by a matrix containing the relative contribution of each isomer to the mixtures. Before applying NMF, the *x*-axis of the experimental spectra was binned into 76 data points from 1000 to 1150 cm⁻¹ (2 cm⁻¹ steps) in OriginPro 2020 using the 1D binning application. The obtained input matrices were normalized before applying NMF (Supplementary Table 3). The input matrix of Glc/Gal phytosphingosine mixtures contains 15 spectra comprising the four single component spectra, six 2-component spectra, four 3-component spectra, and one 4-component spectrum. All input spectra were binned into 76 data points from 1000 to 1150 cm⁻¹ and normalized following the procedure described above (Supplementary Table 4). The input matrix for deconvolution of biological lipid extracts contains four spectra of sodiated α-Glc/GalCer and β-Glc/GalCer standards and two spectra of sodiated biological glycolipids from extracts 1 and 2. The spectra were binned into 100 datapoints from 952 to 1150 cm⁻¹ (2 cm⁻¹ steps) and normalized (Supplementary Table 6). The factorization was carried out using the NMF python module sklearn.decomposition.NMF⁵⁷ with the following arguments: init = “random”, random_state=0, max_iter=2000, alpha=1, and n_components=2 or 4 (number of isomers contained in the input matrix). The output weighting factors were subsequently converted into percentages. Slight deviations of the predicted weighting factors from the actual mixing ratios, as observed in the case of monoglycosyl phytosphingosines, can be partly ascribed to the shortcoming of NMF that the factorization result is not unique³¹. The single component spectra of each of the four isomers exhibit slightly different absolute intensities. Consequently, a low intensity of a specific absorption band in the mixture spectra can indicate either a low relative abundance of the respective isomer or a higher abundance but low absolute intensity. Because the absolute intensities of the pure spectra are not perfectly modeled by NMF, the relative abundance of some isomers is systematically overestimated (see α-Gal in Supplementary Fig. 13), whereas others are underestimated (see β-Gal in Supplementary Fig. 13); however, within well-acceptable limits.

Folch's extraction, chromatography and hydrolysis. Biological lipid extracts were prepared by standard extraction procedures^{33,34}. Thymi and spleen from GLA and GAA knockout mice (Jackson Laboratory, California, USA) were harvested from 8 to 12-week-old mice in compliance with ethical regulations and with the approval of the Institutional Animal Care and Use Committee (protocol # 09-0057-4). The samples were frozen at -80 °C until use. Eight samples were pooled for a Folch's extraction that proceeded in four successive steps after homogenization of the tissue using a Polytron: 2:1, 1:1, 1:2 chloroform/methanol, and finally 1:1:1 chloroform/methanol/H₂O mixture. The four 20 mL fractions were pooled and lyophilized. The extraction was repeated on the dry pellet before a final lyophilization. The crude lipid extract was purified by reversed-phase HPLC using a Dionex Ultimate 3000 LC system. A Supelco C18 column (2.1 mm × 250 mm, 5 μm) at a constant temperature of 60 °C was used for lipid separation. The mobile phase consisted of 70% isopropanol/22% water/8% methanol, and the system was operated at a flow rate of 0.4

mL min⁻¹. The detection was carried out with a UV detector at 205 nm. A total of four measurements with an injection volume of 4 μL were carried out for each sample. Several fractions were collected and examined for the presence of glycolipids by MS and MS/MS. Prior to IR spectroscopy, the dried fractions were redissolved in 100 μL acetonitrile/methanol/H₂O (2:2:1), and phosphatidylcholines were removed by adding 1.5 μL of a 1 M aqueous NaOH solution per 50 μL glycolipid solution. IR spectra of *m/z* 832.7 were recorded when the hydrolysis was completed after 1–2 h (monitored by the disappearance of carbonyl stretching vibrations).

Data availability

The authors declare that the data supporting the findings of this study are available within the paper and its supplementary information files. Source data are provided with this paper.

Received: 29 May 2020; Accepted: 26 January 2021;

Published online: 22 February 2021

References

1. Schnaar, R. L. & Kinoshita, T., *Essentials of Glycobiology* Ch. 11 (Cold Spring Harbor Laboratory Press, 2015).
2. Yamashita, T. et al. A vital role for glycosphingolipid synthesis during development and differentiation. *Proc. Natl. Acad. Sci. USA* **96**, 9142–9147 (1999).
3. Merrill, A. H. Jr. Sphingolipid and glycosphingolipid metabolic pathways in the era of sphingolipidomics. *Chem. Rev.* **111**, 6387–6422 (2011).
4. Stoffel, W. & Bosio, A. Myelin glycolipids and their functions. *Curr. Opin. Neurobiol.* **7**, 654–661 (1997).
5. Kain, L. et al. The identification of the endogenous ligands of natural killer T cells reveals the presence of mammalian alpha-linked glycosylceramides. *Immunity* **41**, 543–554 (2014).
6. Bendelac, A., Savage, P. B. & Teyton, L. The biology of NKT cells. *Annu. Rev. Immunol.* **25**, 297–336 (2007).
7. Kawano, T. et al. CD1d-restricted and TCR-mediated activation of valpha14 NKT cells by glycosylceramides. *Science* **278**, 1626–1629 (1997).
8. Mattner, J. et al. Exogenous and endogenous glycolipid antigens activate NKT cells during microbial infections. *Nature* **434**, 525–529 (2005).
9. Natori, T., Koezuka, Y. & Higa, T. Agelasphins, novel α-galactosylceramides from the marine sponge Agelas mauritianus. *Tetrahedron Lett.* **34**, 5591–5592 (1993).
10. Vartabedian, V. F., Savage, P. B. & Teyton, L. The processing and presentation of lipids and glycolipids to the immune system. *Immunol. Rev.* **272**, 109–119 (2016).
11. Stanic, A. K. et al. Defective presentation of the CD1d1-restricted natural Va14Ja18 NKT lymphocyte antigen caused by beta-D-glucosylceramide synthase deficiency. *Proc. Natl. Acad. Sci. USA* **100**, 1849–1854 (2003).
12. Kain, L. et al. Endogenous ligands of natural killer T cells are alpha-linked glycosylceramides. *Mol. Immunol.* **68**, 94–97 (2015).
13. Brennan, P. J. et al. Structural determination of lipid antigens captured at the CD1d-T-cell receptor interface. *Proc. Natl. Acad. Sci. USA* **114**, 8348–8353 (2017).
14. Brennan, P. J. et al. Invariant natural killer T cells recognize lipid self antigen induced by microbial danger signals. *Nat. Immunol.* **12**, 1202–1211 (2011).
15. Farwanah, H. & Kolter, T. Lipidomics of glycosphingolipids. *Metabolites* **2**, 134–164 (2012).
16. Mucha, E. et al. Glycan fingerprinting via cold-ion infrared spectroscopy. *Angew. Chem. Int. Ed.* **56**, 11248–11251 (2017).
17. Toennes, J. P. & Vilesov, A. F. Superfluid helium droplets: a uniquely cold nanomatrix for molecules and molecular complexes. *Angew. Chem. Int. Ed.* **43**, 2622–2648 (2004).
18. Schöllkopf, W. et al. The new IR and THz FEL Facility at the Fritz Haber Institute in Berlin. *Proc. SPIE* **9512**, 95121L (2015).
19. Panek, P. T. & Jacob, C. R. Anharmonic theoretical vibrational spectroscopy of polypeptides. *J. Phys. Chem. Lett.* **7**, 3084–3090 (2016).
20. Howard, J. C., Enyard, J. D. & Tschumper, G. S. Assessing the accuracy of some popular DFT methods for computing harmonic vibrational frequencies of water clusters. *J. Chem. Phys.* **143**, 214103 (2015).
21. Bec, K. B. & Huck, C. W. Breakthrough potential in Near-Infrared Spectroscopy: spectra simulation. A review of recent developments. *Front. Chem.* **7**, 48 (2019).
22. Deng, S. et al. Psychosine variants as antigens for natural killer T cells. *Chem. Sci.* **8**, 2204–2208 (2017).
23. Pham, H. T. & Julian, R. R. Characterization of glycosphingolipid epimers by radical-directed dissociation mass spectrometry. *Analyst* **141**, 1273–1278 (2016).

24. Chao, H.-C. & McLuckey, S. A. Differentiation and quantification of diastereomeric pairs of glycosphingolipids using gas-phase ion chemistry. *Anal. Chem.* **92**, 13387–13395 (2020).
25. Laine, R. A. A calculation of all possible oligosaccharide isomers both branched and linear yields 1.05×10^{12} structures for a reducing hexasaccharide: the Isomer Barrier to development of single-method saccharide sequencing or synthesis systems. *Glycobiology* **4**, 759–767 (1994).
26. Anderson, B. L., Teyton, L., Bendelac, A. & Savage, P. B. Stimulation of natural killer T cells by glycolipids. *Molecules* **18**, 15662–15688 (2013).
27. Zhou, D. et al. Lysosomal glycosphingolipid recognition by NKT cells. *Science* **306**, 1786–1789 (2004).
28. Kinjo, Y. et al. Natural killer T cells recognize diacylglycerol antigens from pathogenic bacteria. *Nat. Immunol.* **7**, 978–986 (2006).
29. Kinjo, Y. et al. Invariant natural killer T cells recognize glycolipids from pathogenic Gram-positive bacteria. *Nat. Immunol.* **12**, 966–974 (2011).
30. Lee, D. D. & Seung, H. S. Learning the parts of objects by non-negative matrix factorization. *Nature* **401**, 788–791 (1999).
31. Huang, Z., Zhou, A. & Zhang, G. Non-negative matrix factorization: a short survey on methods and applications. *Comput. Intell. Intell. Syst.* 331–340 (2012).
32. Hofmann, J., Hahm, H. S., Seeberger, P. H. & Pagel, K. Identification of carbohydrate anomers using ion mobility–mass spectrometry. *Nature* **526**, 241 (2015).
33. Folch, J., Lees, M. & Sloane Stanley, G. H. A simple method for the isolation and purification of total lipides from animal tissues. *J. Biol. Chem.* **226**, 497–509 (1957).
34. Bligh, E. G. & Dyer, W. J. A rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiol.* **37**, 911–917 (1959).
35. Yin, N. et al. Alpha anomers of iGb3 and Gb3 stimulate cytokine production by natural killer T cells. *ACS Chem. Biol.* **4**, 199–208 (2009).
36. Chaudhary, V. et al. Synthesis of fungal glycolipid asperamide B and investigation of its ability to stimulate natural killer T cells. *Org. Lett.* **15**, 5242–5245 (2013).
37. Du, W., Kulkarni, S. S. & Gervay-Hague, J. Efficient, one-pot syntheses of biologically active alpha-linked glycolipids. *Chem. Commun.* **23**, 2336–2338 (2007).
38. Manzo, E., Ciavatta, M. L., Pagano, D. & Fontana, A. An efficient and versatile chemical synthesis of bioactive glyco-glycerolipids. *Tetrahedron Lett.* **53**, 879–881 (2012).
39. Allen, S. J., Giles, K., Gilbert, T. & Bush, M. F. Ion mobility mass spectrometry of peptide, protein, and protein complex ions using a radio-frequency confining drift cell. *Analyst* **141**, 884–891 (2016).
40. Maccarone, A. T. et al. Characterization of acyl chain position in unsaturated phosphatidylcholines using differential mobility–mass spectrometry. *J. Lipid Res.* **55**, 1668–1677 (2014).
41. Morris, L. J. Separations of lipids by silver ion chromatography. *J. Lipid Res.* **7**, 717–732 (1966).
42. Dobson, G., Christie, W. W. & Nikolova-Damyanova, B. Silver ion chromatography of lipids and fatty acids. *J. Chromatogr. B* **671**, 197–222 (1995).
43. Revercomb, H. E. & Mason, E. A. Theory of plasma chromatography/gaseous electrophoresis. *Rev. Anal. Chem.* **47**, 970–983 (2002).
44. Schrödinger Release 2019 (Maestro, Schrödinger, LLC, 2019).
45. Pracht, P., Bohle, F. & Grimme, S. Automated exploration of the low-energy chemical space with fast quantum chemical methods. *Phys. Chem. Chem. Phys.* **22**, 7169–7192 (2020).
46. Blum, V. et al. Ab initio molecular simulations with numeric atom-centered orbitals. *Comput. Phys. Commun.* **180**, 2175–2196 (2009).
47. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
48. Tkatchenko, A. & Scheffler, M. Accurate molecular van der Waals interactions from ground-state electron density and free-atom reference data. *Phys. Rev. Lett.* **102**, 073005 (2009).
49. Marianski, M., Supady, A., Ingram, T., Schneider, M. & Baldauf, C. Assessing the accuracy of across-the-scale methods for predicting carbohydrate conformational energies for the examples of glucose and alpha-maltose. *J. Chem. Theory Comput.* **12**, 6157–6168 (2016).
50. Mucha, E. et al. Unravelling the structure of glycosyl cations via cold-ion infrared spectroscopy. *Nat. Commun.* **9**, 4174 (2018).
51. Marianski, M. et al. Remote participation during glycosylation reactions of galactose Building Blocks: direct evidence from cryogenic vibrational spectroscopy. *Angew. Chem. Int. Ed.* **59**, 6166–6171 (2020).
52. Bannwarth, C., Ehlert, S. & Grimme, S. GFN2-xTB-An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput.* **15**, 1652–1671 (2019).
53. Frisch, M. J. et al. *Gaussian 16, Rev. A.03* (Gaussian, Inc., 2016).
54. Adamo, C. & Barone, V. Toward reliable density functional methods without adjustable parameters: the PBE0 model. *J. Chem. Phys.* **110**, 6158–6170 (1999).
55. Barone, V. Anharmonic vibrational properties by a fully automated second-order perturbative approach. *J. Chem. Phys.* **122**, 14108 (2005).
56. Thomas, D. A. et al. Probing the conformational landscape and thermochemistry of DNA dinucleotide anions via helium nanodroplet infrared action spectroscopy. *Phys. Chem. Chem. Phys.* **22**, 18400–18413 (2020).
57. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

Acknowledgements

C.K. is grateful for financial support by the Fonds der Chemischen Industrie and Studienstiftung des deutschen Volkes. K.G. thanks the Fonds National de la Recherche (FNR), Luxembourg, for funding the project GlycoCat (13549747). L.T. and P.B.S. acknowledge funding by the National Institute of Allergy and Infectious Diseases (RO1AI123130). M.M. acknowledges funding by the Army Research Office (W911NF2010271). We furthermore thank Prof. Daniel A. Thomas for advice on NMF-related questions.

Author contributions

S.D., L.K., and P.B.S. synthesized and prepared glycolipids; P.B.S., G.M., G.v.H., L.T., and K.P. designed and conceived the experiments; C.K., E.M., and K.G. performed the experiments; A.Z. purified biological lipid extracts by HPLC; S.G. and W.S. operated the free-electron laser; K.G. and M.M. performed the theoretical calculations; all authors co-wrote the manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-21480-1>.

Correspondence and requests for materials should be addressed to K.P.

Peer review information *Nature Communications* thanks Antonio Molinaro, Christian W. Huck and the other anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021, corrected publication 2021