Structure of Bolaamphiphile Sophorolipid Micelles Characterized with SAXS, SANS, and MD Simulations
Sabine Manet, Anne-Sophie Cuvier, Claire Valotteau, Giulia C. Fadda, Javier Perez, Esra Karakas, Stéphane Abel, Niki Baccile

To cite this version:
Sabine Manet, Anne-Sophie Cuvier, Claire Valotteau, Giulia C. Fadda, Javier Perez, et al.. Structure of Bolaamphiphile Sophorolipid Micelles Characterized with SAXS, SANS, and MD Simulations. Journal of Physical Chemistry B, American Chemical Society, 2015, 119 (41), pp.13113-13133. 10.1021/acs.jpcb.5b05374. hal-01281684

HAL Id: hal-01281684
https://hal.sorbonne-universite.fr/hal-01281684
Submitted on 3 Feb 2017

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

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Structure of Bolaamphiphile Sophorolipid Micelles Characterized with SAXS, SANS and MD Simulations

Sabine Manet,1 Anne-Sophie Cuvier,1 Claire Valotteau,1 Giulia C. Fadda,2 Javier Perez,3 Esra Karakas,4,5 Stéphane Abel,5 and Niki Baccile1*

[2] Laboratoire Léon Brillouin, LLB, CEA Saclay, F-91191 Gif-sur-Yvette Cedex, France
[5] Institute for Integrative Biology of the Cell (I2BC), Commissariat à l’Energie Atomique et aux Energies Alternatives (CEA), Centre National de la Recherche Scientifique (CNRS), Université Paris-Sud, CEA-Saclay, F-91191 Gif-sur-Yvette, France

* Corresponding author : Laboratoire de Chimie de la Matière Condensée de Paris, Collège de France, 11, Place M. Berthelot, F-75005, Paris, France. Tel. +33 1 44 27 15 44 ; Email : niki.baccile@upmc.fr

Abstract

The micellar structure of sophorolipids, a glycolipid bolaamphiphile, is analyzed using a combination of Small Angle X-ray Scattering (SAXS), Small Angle Neutron Scattering (SANS) and Molecular Dynamics (MD) simulations. Numerical modeling of SAXS curves shows that micellar morphology in the non-charged system (pH < 5) is made of prolate ellipsoids of revolution with core-shell morphology. Opposed to most surfactant systems, the hydrophilic shell has a non-homogeneous distribution of matter: the shell thickness in the axial direction of the ellipsoid is found to be practically zero, while it measures about 12 Å at its cross-section, thus forming a “coffee bean” like shape. The use of contrast matching SANS experiment shows that the hydrophobic component of sophorolipids is actually distributed in
a narrow spheroidal region in the micellar core. These data seem to indicate a complex
distribution of sophorolipids within the micelle, divided into at least three domains: a pure
hydrophobic core, a hydrophilic shell and a region of less defined composition in the axial
direction of the ellipsoid. To account for these results, we make the hypothesis that
sophorolipid molecules acquire various configurations within the micelle including bent and
linear, crossing the micellar core. These results are confirmed by MD simulations which do
show the presence of multiple sophorolipid configurations when passing from spherical to
ellipsoidal aggregates. Finally, we also used Rb\(^+\) and Sr\(^{2+}\) counterions in combination with
Anomalous SAXS experiments to probe the distribution of the COO\(^-\) group of sophorolipids
upon small pH increase (5 < pH < 7), where repulsive intermicellar interactions become
important. The poor ASAXS signal shows that the COO\(^-\) groups are rather diffused in the
broad hydrophilic shell rather than at the outer micellar/water interface.

Keywords
Self-assembly; Biosurfactants; Glycolipid; Starmerella bombicola; ASAXS; Bolaform

Introduction
Surfactants are compounds employed in a very large panel of applications in which
amphipathic interfaces are made compatible. The formation of micelles above a critical
micellar concentration is one of the most common aspects among amphiphilic molecules in
water and a thorough description of their structure is one of the first steps when studying the
properties of new compounds.\(^1\) If the aggregation of classical cationic (e.g., alkylammonium
salts),\(^2\)\(^-\)\(^4\) anionic (e.g., alkylsulfates salts) or non-ionic (e.g., block copolymers)\(^5\) surfactants
into micelles has been studied, both experimentally and theoretically, for decades, less
conventional compounds are much less looked at. Among these, bolaform lipids\(^6\) constitute a
class of interesting molecules with two functional end-groups. If the end-groups are
chemically equivalent, one speaks of symmetric bola amphiphiles (bolas), otherwise one refers
to asymmetric bolas. The atypical distribution of amphiphilic and lyophilic regions within
the same molecule is often responsible of interesting aggregation phenomena which lead to
the formation of micelles, bicelles,\(^7\) vesicles,\(^8\) cones,\(^9\) fibers,\(^10\) tubes,\(^11,12\) etc... Many
significant systems have been reviewed in 2004 by Fuhrhop,\(^6\) while Fariya \textit{et al.}\(^13\) have
recently reviewed bola amphiphiles in the context of pharmaceutical applications. If a large
amount of work has been dedicated to the synthesis and characterization of bola systems,
much less has been devoted to a thorough structural study. In the context of micellar objects
composed of bolaamphiphiles, advanced structural considerations of the micellar packing and bola distribution are rare. Indeed, Nagarajan predicted since 1987 the formation and structure of micelles composed of bolaamphiphiles basing his hypotheses on the analysis of the packing parameter. Interestingly, he predicted the formation of spherical, cylindrical and discoidal micelles, in which he imagined that the bolaform compound has various configurations: elongated, which induces transradial crossing of the end-groups; folded, under which the end-groups lie both on the same side. If Yan et al. have proceeded with macroscopical verification of Nagarajan’s assumptions by observing a pH/cosolvent-driven tube-to-vesicle-to-micelle transition in disodium phenyl-1,4-bis(oxyhexanoate) solutions, the actual bola distribution within the micelle was never elucidated. Additional insights in the structural composition of bola-composed micelles have been discussed in Ref. mainly using Small Angle Neutron Scattering (SANS) and/or self-diffusion Nuclear Magnetic Resonance (NMR) experiments. However, in practically all cases, these experiments have been used to describe the micellar shape and size. Only few authors, with little consensus in their conclusions, have tried to determine the molecular distribution of the bolaamphiphile within the micellar aggregate. Caponetti et al., for instance, have used advanced modelling analyses of SANS spectra to study the bola N-aza-18-crown-6 ether compound. If Davey et al. and Shinde et al. agree on the fact that their respectively studied asymmetric bolaamphiphiles, alcohol derivatives of alkylammonium and oleate salts, form uniform spherical/cylindrical micelles in which both polar groups are in contact with the solvent (water), they do not agree on the molecular arrangements. Davey proposes an elongated cross-micellar conformation, while Shinde proposes a bent configuration. Interestingly, these recent studies disagree with older works in which it was proposed that only one of the two polar groups is actually located at the micellar/solvent palisade.

Such a lack of deep characterization of bola-micelles, and disagreement on their structure and molecular conformation, is quite odd because of the importance in terms of applications for many bola systems. In fact, the chemical nature of end-groups in bolas can be as wide as imagination can allow: polymerizable, complexing, pH-responsive, fullerenic, chiral, glycosidic, functions are just some examples of the broad variety that one can find in the literature. The range of applications of bolas in general goes from one-pot porous material synthesis to supramolecular materials with temperature-driven ferroelectric-paraelectric transitions, ion-channel supramolecular membranes, metallo-hydrogels for dye adsorption and water purifier, viscoelastic fluids, biomedical carriers and much more. For this reason, it is of paramount importance to characterize the spatial distribution
and accessibility of end-functions within the supramolecular aggregates and, in particular, within micelles. Laying one of the end-groups within the interior of the micellar aggregate can be of relevance for drug-delivery systems and conception of nanoreactors. Whether charged end-groups concentrate at the micellar surface or not can be of relevance in nanomaterial synthesis procedure in which charge matching drives the material formation, such as mesostructured oxides or layer-by-layer approaches, but it can also have drastic relevance in gelling properties, for instance. The distribution and accessibility of biologically-relevant moieties like peptides or carbohydrates is important for biomedical applications like gene-antigene recognition for cancer treatment. Even in the system studied by Caponetti,\textsuperscript{16} in which authors looked at metal-binding affinity of the crown ether end-groups, it is necessary to know the distribution of the crown ether groups and their accessibility in view of metal ion removal in solution.

For these reasons, in the present work, we try to set up a combination of complementary analytical approaches based on advanced small angle scattering techniques in order to prove both indirectly (by way of model functions) and directly the spatial distribution of a glycolipid asymmetric bolaamphiphile. We study here the micellar structure of sophorolipids, a yeast-derived bolaform glycolipid composed of sophorose (glucose $\beta(1,2)$) attached to the C17 atom of oleic acid via an ether bond, the carboxylic group being free of access at the opposite side of the molecule (Figure 1).\textsuperscript{31} This compound is part of the glycolipid family and the presence of sophorose-COOH end-groups make it, \textit{de facto}, a bolaform compound with pH-responsive properties,\textsuperscript{32} which confer to it a double neutral/anionic or neutral/neutral nature.\textsuperscript{33} Recent works on the self-assembly of the acidic open chain form of sophorolipids have shown its ability to form assemblies such as fibers\textsuperscript{34,35} and micelles.\textsuperscript{32,36} In the latter case, our group has shown that the micellar charge can be adjusted using pH.\textsuperscript{33} However, no information on the charge distribution could be obtained.

Considering the fact that this compound is easily accessible and considering its importance in various fields like detergency,\textsuperscript{37} cosmetics,\textsuperscript{38} nanoscience,\textsuperscript{39} disinfection,\textsuperscript{40} medicine,\textsuperscript{41} it becomes a valuable candidate to use for a detailed study of the sophorose/COOH end-groups distribution within the micelle. To do so, we combine several small-angle scattering techniques: Small Angle X-ray Scattering (SAXS) is used to precisely define the micellar morphology and its size; Anomalous-SAXS (ASAXS), a technique that probes the distribution of counterions around a charged macroion, is employed to follow the distribution of the carboxilate end-group in the presence of specific ASAXS probe counterions, Rb$^+$ and Sr$^{2+}$; contrast matching Small Angle Neutron Scattering (SANS) is used to visualize the
distribution of the hydrophobic fraction of the molecule. This approach shows that, in the mild acidic pH region (pH< 5), the micelle is likely to be described as an ellipsoid of revolution having a “coffee bean” like morphology. In this structure, the hydrophobic region only occupies a small inner spheroidal core. The outer shell in the central region of the ellipsoid is mainly made of sophorose and COOH groups. The electron density distribution in the axial direction is more difficult to define and it probably reflects a more complex distribution of matter. We identify at least two different configurations of the sophorolipid molecules: bent and elongated, with possibility of crossing the ellipsoid section. We also show that these hypotheses, based on experimental work, are confirmed by Molecular Dynamics (MD) simulations specifically performed in this work. Upon formation of COO\(^-\) groups in the neutral pH range (5 < pH < 7), micelles become negatively charged: ASAXS experiments seem to indicate that the carboxylates are rather diffused in the broad hydrophilic shell rather than localized at the micelle/solvent palisade. To the best of our knowledge, such a complex analysis has never been carried neither on glycolipid-based nor on any other bolaamphiphile compound.

**Experimental**

Acidic sophorolipids (SL) have been prepared from a commercial batch of a sophorolipid mixture (Sopholiance, Soliance, France; sophorolipids are derived from rapeseed oil, batch number: 11103A, dry content: 60 ± 6 %) using alkaline hydrolysis to convert the lactonic/acidic mixture into a fully acidic sophorolipid compound. Extraction and purification have been performed using the method N°2 described in ref. 42, to which one should refer for a typical \(^1\)H solution NMR spectrum of the compound used. As a general observation on purity, one should note that our compound has a very low content of residual fatty acids (typically less than 5 mol% by \(^1\)H NMR). Nevertheless, biobased glycolipids in general are known to contain residual congeners. In the case of sophorolipids, mainly composed of a C18:1 fatty acid tail attached to sophorose, congeners generally represent a low fraction of the actual compound, generally being C16 and/or C18:2 fatty acids variations. A typical HPLC chromatogram was reported in ref. 43, showing that C18:1 does constitute the large majority of our sample (> 90 %), thus being in agreement with the \(^1\)H NMR data. Congeners are also detected but in the present work we make the hypothesis that, given the low amount, their influence on the self-assembly of C18:1 sophorolipid is negligible.

Sophorolipid solutions are generally prepared by simply dispersing the compound in water (50 mg/mL ≡ 80 mM). If needed, a slightly warm (30-35°C) ultrasonic bath is used for few
seconds to help solubilize the sample. After solubilization, the solutions are clear and stable
over time. Specific sample preparation is detailed in each sub-section below. MilliQ quality
water was used for the experiments and all solutions are freshly prepared. RbCl, SrCl₂, RbOH
and Sr(OH)₂ have been purchased at Sigma-Aldrich.

**Small Angle X-ray Scattering (SAXS).** The sample composition is presented in Table 1 and for
all of them the concentration of SL is 50 mg/mL. Rb⁺ and Sr²⁺ have been used instead of the
more classical Na⁺ and Ca²⁺ ions due to their activity in the ASAXS experiments, presented
later. Each sample is prepared by mixing the salt in water; this solution is eventually split in
half, one half is used for background acquisition and the indicated amounts of SL are
introduced in the second half.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Compound</th>
<th>[Sr²⁺] mM</th>
<th>[Rb⁺] mM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>RbCl</td>
<td>-</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>RbCl</td>
<td>-</td>
<td>50</td>
</tr>
<tr>
<td>3</td>
<td>RbCl</td>
<td>-</td>
<td>100</td>
</tr>
<tr>
<td>4</td>
<td>RbOH</td>
<td>-</td>
<td>40</td>
</tr>
<tr>
<td>5</td>
<td>RbOH</td>
<td>-</td>
<td>80</td>
</tr>
<tr>
<td>6</td>
<td>SrCl₂</td>
<td>5</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>SrCl₂</td>
<td>25</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>SrCl₂</td>
<td>50</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>Sr(OH)₂</td>
<td>10</td>
<td>-</td>
</tr>
<tr>
<td>10</td>
<td>Sr(OH)₂</td>
<td>20</td>
<td>-</td>
</tr>
</tbody>
</table>

These experiments have been done at the SWING beamline of the SOLEIL synchrotron at
Saint-Aubin (France). Sample-to-detector distance was set to 2526 mm and energy of the X-
ray beam was E= 14.6 keV. The experimental environment was provided at the beamline and
consisted of a flow-through quartz capillary which it was possible to fill using a syringe. The
X-ray beamline hits the same capillary position, which does not change between background
and sample acquisitions, thus making the background subtraction highly accurate. For the
background, we used the same solution used to prepare a given sophorolipid sample, as also
mentioned above. Absolute scaling was calibrated on the signal of water. All data have been
divided by the sample transmitted intensity. Data treatment (mask generation, integration) was
done using the Foxtrot software provided at the beamline.

*Modelling SAXS data.*
The analysis of the SAXS profiles for the determination of micelle size and shape parameters has been made with the models displayed by default in the SASview software, provided free of charge on the developer’s website. The general equation treating the scattered intensity versus the momentum transfer q is (Eq. 1)

\[ I(q) = \frac{\text{scale}}{V} (\rho - \rho_{\text{solv}})^2 P(q) S(q) + \text{bkg} \]  

Eq. 1

where, scale is the volume fraction, V is the volume of the scatterer, \( \rho \) is the scattering length density (SLD) of the object, \( \rho_{\text{solv}} \) is the SLD of the solvent, \( P(q) \) is the form factor of the object, \( S(q) \) is the structure factor and bkg is a constant accounting for the background level.

**Form factor**

The analytical expressions of the \( P(q) \) discussed below are directly implemented in the SASview software and their analytical form is provided in Ref. 45. The \( \chi^2 \) test is employed to characterize the quality of the fit, the lower value indicating the best fit. \( \chi^2 \) is defined as the Pearson's cumulative test statistic,

\[ \chi^2 = \sum_{i=1}^{N\text{pts}} \frac{(O_i - E_i)^2}{E_i} \]

where \( O_i \) is an observed measured value; \( E_i \) is an expected (theoretical) value, asserted by the null hypothesis. The \( \chi^2 \) is then normalized by the number of points in the distribution, \( \chi^2/N\text{pts} \). Some fit parameters can be simply estimated from the sample composition and have been fixed for the fitting process. Among them, the volume fraction of scattering objects has been set at 0.05, reflecting the SL mass concentration of 50 mg/mL (80 mM) used for all samples.

The core and solvent SLD’s have been fixed and calculated using Eq. 2:

\[ \rho = \frac{\sum_{i} Z_i r_e}{V_M} \]  

Eq. 2

where \( Z_i \) is the atomic number of the \( i^{th} \) of \( j \) atoms in a molecule of molecular volume \( V_M \) and \( r_e \) is the classical electron radius or Thomson scattering length (2.8179 x \( 10^{-15} \) m).

Sophorolipids are mainly composed of a C18 aliphatic chain with one unsaturated C=C bond (position C9,10), one COOH group (position C1), the CH in position C17 covalently bonded to sophorose and a CH3 in position C18 (Figure 1). Considering the fact that C17 and C18 atoms are located at the sophorose/aliphatic frontier, it is a difficult choice to decide which is their exact location. For this reason, we simplify our model and make the assumption that the hydrophobic core starts at the C16 carbon, neglecting the contribution of the C17 and C18 positions. Under these conditions, we define the molecular volume, \( V_M \), of the hydrocarbon...
chain to be $13 V_{CH_{2,\text{mic}}} + 2 V_{CH}$, where $V_{CH_{2,\text{mic}}}$ is the volume of one CH$_2$ group inside a micelle, 27.5 Å$^3$, and $V_{CH}$ is the volume of a CH group, 22.0 Å$^3$. These estimations provide a SLD core value, constant for all fitting procedures in this study, of $8.3 \times 10^{-6}$ Å$^{-2}$ (294 e/$\text{nm}^3$). The solvent SLD has been calculated taking into account the salt or base concentration influencing the molecular volume. The resulting SLD values are very close and lie between 9.4 and $9.5 \times 10^{-6}$ Å$^{-2}$, (334 and 337 e/$\text{nm}^3$, see full list in Table S1 in the Supporting Information). For core-shell models, the shell SLD is always a variable parameter. We make the hypothesis that the shell is most likely composed of hydrated sophorose; in this case, the shell SLD value should lie between the solvent and hydrated sophorose SLD values, roughly between 1.0 and $1.2 \times 10^{-5}$ Å$^{-2}$, (355 to 425 e/$\text{nm}^3$). As a result, the electron density profile from the center of the micelle to the solvent can be drawn as a hat-like shape, in which the shell SLD is the highest value. The choice of the form factor has been determined by comparing the best fits on the basis of the lowest $\chi^2/N\text{pts}$ among different model functions, as critically discussed on Page S2 in the Supporting Information and shown in Figure S1. We find that a core-shell prolate ellipsoid of revolution, schematized in Figure 2, best describes our SAXS data.

**Structure Factor**

The structure factor, $S(q)$, which quantifies the intermicellar interactions/correlations, is included in our fitting process using the “Hayter-MSA-Structure” model displayed in the SASview software. This accounts for a repulsive screened Coulombic intermicellar interaction potential, as previously used on similar systems. $S(q)$ has only been used for samples containing RbOH (samples 4,5) and Sr(OH)$_2$ (samples 9,10) according to the previously discussed assumption that deprotonation of the COOH introduces negative charges at the micellar/solvent palisade, thus being responsible of the broad interaction peak observed in small angle neutron scattering spectra, and reproduced in this work. $S(q)$ has been set equal to 1 for base-free samples only containing salts (samples 1-3 and samples 6-8 in Table 1). Six parameters are needed to compute the Hayter and Penfold structure factor: the dielectric constant, the volume fraction, the effective radius of interacting objects, the temperature, the surface charge of the micelle, and the salt concentration, used to calculate the ionic strength of the solution, which in turn is used to compute the Debye screening length. As this model is made for monovalent ions, we modified the salt concentration parameter of the samples, otherwise fixed and equal to the experimental concentration, with divalent ions.
so that the resulting Debye length is correct. The volume fraction of the scatterers is calculated from \( V_{\text{mic}} \) and \( n_{\text{mic}} \) (\( \eta = V_{\text{mic}}n_{\text{mic}} \)), defined more precisely later on in this same section, and inserted in the Hayter-MSA-Structure routine with \( R_E \), effective radius of interacting objects (\( R_E = \frac{1}{2} \left( \frac{6\eta}{\pi n_{\text{mic}}} \right)^{1/3} \)), see Ref. 48) as fixed parameters. The temperature, dielectric constant and salt concentration are also fixed parameters so that the micelle charge is the only refinable parameter for the calculation of \( S(q) \).

**Fit strategy using a core shell prolate ellipsoid of revolution model and a Hayter-MSA structure factor**

In order to reduce the number of parameters used for the fitting process, we used a customized version of the core-shell prolate ellipsoid form factor proposed in the 3.0.0 version of SASview. In particular, we introduce some molecular constraints inspired by the work of Hayter and Penfold.\(^4\) By making some assumptions described thereafter, only 5 fitting parameters are needed: \( N_{\text{agg}} \) (aggregation number), \( R_C \) (equatorial core radius, see Figure 2), \( \rho_S \) (scattering length density of the shell, see Figure 2), \( X_C \) (axial core ratio, see Figure 2) and \( z \), the surface charge of the micelle. A small background correction (<0.005 cm\(^{-1}\)) has also been applied. In the Hayter and Penfold model, if \( N_{\text{agg}} \) is greater than the one expected to accommodate a sphere, the radius of which is \( R_c = l_c \) (length of the fully extended hydrocarbon chain of the surfactant), then the object is allowed to become elliptical so to accommodate the molecular excess. This model has been adapted to the sophorolipid surfactants as follows. For a given \( N_{\text{agg}} \), the spherical hard core micelle radius \( R_{hc} \) is defined as:

\[
R_{hc} = \left[ \frac{3N_{agg} \left( 13V_{CH_{2,mic}} + 2V_{CH} \right)}{4\pi} \right]^{1/3} \quad \text{Eq. 3}
\]

If \( R_{hc} > l_c \), \( R_C \) is set to \( l_c \), and a parameter \( \alpha \) is introduced, defined as the fraction of \( \text{CH}_2 \) inserted in the dry core of the micelle:

\[
\alpha = \frac{V_C - 2V_{CH}}{N_{agg} 13V_{CH_{2,mic}}} \quad \text{Eq. 4}
\]

with \( V_C = \frac{4}{3} \pi R_c^3 \). If \( R_{hc} < l_c \), \( R_C \) is set to \( R_{hc} \) and \( \alpha = 1 \).
Sophorolipids are bolaform amphiphiles with two different polar heads (carboxylic acid and sophorose sugar moiety) connected to a 18:1 carbons lipid chain (Figure 1), which reduces to a C16:1 chain if one makes the hypothesis that C_{17} (CH) and C_{18} (CH₃) do not contribute to $l_c$, which in this case can be estimated to be around 10 Å, half the length of the 16 carbons chain. We first try to use this value of $l_c$ as $R_{test}$, but the results obtained were not reliable, leading to high $\chi^2/Npts$ values, so we let this constraint to relax, and $R_C$ and $X_C$ have been set as refinable parameters. Recently, Penfold et al.³⁶ have made the same choice when analyzing the SANS profiles of sophorolipid micelles, arguing that the particular alkyl chain geometry of the sophorolipid introduces some uncertainty into what value the inner radius should be constrained to.

Once $\alpha$ is known, the number of water molecules per sophorolipid in the shell, $n_{w/SL}$, and consequently the composition of the shell, can be calculated using the shell scattering length density $\rho_s$ as a refinable parameter, assuming that the shell hydration should not be a fixed parameter, because it depends on many effects like the number of hydrated CH₂ or the micelle surface charge. From Eq. 5 we calculate the molar volume, $V_{m,hg}$, and hence the SLD, $\rho_{hg}$, of the total dry headgroup. Composition of the dry headgroup is not easy to estimate; it certainly includes sophorose, as stated earlier, and most likely the COOH group, but it can also include the C_{17} (CH) and C_{18} (CH₃) atoms, previously excluded from the core region of the micelle, as well as the fraction of the CH₂ aliphatic chain excluded by the $\alpha$ parameter presented in Eq. 4. This is not surprising as hydration of the CH₂ groups close to the hydrophilic head was reported before both experimentally and predicted theoretically using MD simulation. In all cases, the ionic species coming from the added salts or bases are neglected. The general expressions for the headgroup volume and SLD are then:

$$V_{m,hg} = \sum_i V_i + 13(l - \alpha) V_{CH_2,w}$$

Eq. 5

$$\rho_{hg} = \frac{\sum_i \rho_i V_i + 13(l - \alpha) \rho_{CH_2} V_{CH_2,w}}{V_{m,hg}}$$

Eq. 6

where $i$ denotes the sophorose, carboxylate and the aliphatic moieties, CH, CH₃ and possibly CH₂ groups. $V_{CH_2,w}$ is the molar volume of hydrated CH₂ smaller than $V_{CH_2,mic}$, taken as 26.7Å³ while $\rho_{CH_2,w}$ is the SLD of the hydrated CH₂. The volume fraction of water in the
shell $x_{v,w}$ and the number of water molecules per sophorolipid in the shell $n_{w/SL}$ are obtained by considering the fraction of water molecules ($n_{w/SL}$) needed to equalize the calculated $\rho_{hg}$ (Eq. 6) and the experimental $\rho_{shell}$ (obtained from the fit). This is summarized in Eq. 7 and Eq. 8:

$$x_{v,w} = \frac{\rho_{shell} - \rho_{hg}}{\rho_w - \rho_{hg}}$$

Eq. 7

$$n_{w/SL} = \frac{x_{v,w} V_{m,hg}}{(1-x_{v,w}) V_{m,w}}$$

Eq. 8

with $V_{m,w}$ the molar volume of water (29.9 Å$^3$) and $\rho_w$ the water SLD taken as 9.4 x $10^{-6}$ Å$^{-2}$.

As it did not seem straightforward to evaluate the ionic composition and concentration within the shell, the presence of ions coming from added salts, acid or base has been neglected. However, the same fitting process applied using $\rho_{solv}$ instead of $\rho_w$ in Eq. 7 did not show a deep impact of the ionic composition of the shell on the results (results not shown).

The total micelle volume $V_{mic}$ is then calculated as the sum of all the components in the core and in the shell:

$$V_{mic} = N_{agg} \left( 13a V_{CH, mic} + 2 V_{CH} + V_{m,hg} + n_{w/SL} V_{m,w} \right)$$

Eq. 9

As $V_{mic}$ can also be defined as the volume of an ellipsoid described by the equatorial core radius $R_c$, the axial core ratio $X_C$, the axial shell ratio $X_S$ (see Figure 2) and the equatorial shell radius $R_S$ [$V_{mic} = \frac{4}{3} \pi R_s^3 \left( R_c X_C + (R_S - R_c) X_S \right)$], see Figure 2], the latter geometric parameter can then be obtained as the real root of this polynomial equation:

$$X_S R_s^2 + R_c (X_C - X_S) R_s^2 - 3 \frac{V_{mic}}{4\pi} = 0$$

Eq. 10

The core-shell prolate ellipsoid form factor is finally calculated by:

$$P(q) = n_{mic} \int_{0}^{j} \left| F(q, \mu) \right|^2 d\mu + \text{background}$$

$$F(q, \mu) = V_c (\rho_c - \rho_S) \left[ \frac{3j_l(x_c)}{x_c} \right] + V_{mic} (\rho_S - \rho_{solv}) \left[ \frac{3j_l(x_S)}{x_S} \right]$$

Eq. 11
\[
\begin{align*}
\bar{q}_i(x) &= \frac{\sin x - x \cos x}{x^2} \\
x_c(q, \mu) &= q \left( R_c^2 X_c^2 \mu^2 + R_c^2 (1 - \mu^2) \right) \\
x_s(q, \mu) &= q \left( R_s^2 X_c^2 \mu^2 + R_s^2 (1 - \mu^2) \right)
\end{align*}
\]

Eq. 12

Eq. 13

Eq. 14

The intensity scattered by interacting micelles defined as 
\( I(q) = n_{mic} P(q) S(q) \) can be calculated by relating the micelle density \( n_{mic} \) to the known sophorolipid surfactant concentration [SL] (80 mM) and the refinable aggregation parameter \( N_{agg} \), with \( N_a \) being the Avogadro’s number:

\[
n_{mic} = \frac{[SL] N_a}{N_{agg}}
\]

Eq. 15

**Anomalous Small Angle X-ray Scattering (ASAXS).**

*Elements of theory.* Extraction of the counterion scattering contribution around sophorolipid micelles was done via recording the anomalous scattering contribution of rubidium and strontium cations. The anomalous effect is measured when the energy of the X-ray beam approaches the absorption edge of a given element. In this case, the scattering factor becomes a complex function and it is no more a constant. The contribution of the scattering factor far from the absorption edge is

\[
f_{saa} = f_0 - \rho_m V_{ion}
\]

Eq. 16

where \( f_0 \) equals the atomic number of the element, \( \rho_m \) is the electron density of water and \( V_{ion} \) is the volume of the counterion. Upon approaching of the edge, \( f_{ion} \) becomes

\[
f_{saa} = f_0 - \rho_m V_{sa} + f'(E) + if''(E)
\]

Eq. 17

where \( f'(E) \) and \( f''(E) \) are, respectively, the real and imaginary part of the energy-dependent scattering vector. Values for \( f'(E) \) and \( f''(E) \) are tabulated,\(^{54,55}\) but they can also be calculated from the experimental absorption spectrum.\(^{56}\) When approaching the absorption edge by tuning the synchrotron incident beam energy, \( f_{ion} \) is neither constant nor negligible anymore.
and the scattered intensity must take that into account. The resulting general expression for the scattering intensity then becomes

\[ I_0(q,E) = F_0^*(q) + 2f'(E)F_0(q)v(q) + \left[f''^2(E) + f''^2(E)\right]v^2(q) \]  

Eq. 18

where \( F_0^2(q) \) is the nonresonant intensity of the macroion (= micelle, polymer, etc...) measured far from the absorption edge; \( F_0(q) \) and \( v(q) \) are the nonresonant and resonant amplitudes respectively of the macroion and the counterion; \( v^2(q) \) is the pure resonant curve of the counterion measured at the absorption edge. A constant term due to fluorescence, affecting the intensity above the absorption edge, has been omitted. For a more extended discussion on the SAXS theory near the absorption edge of a given element, one can refer to references 57-59. From Eq. 18, one is generally interested into extracting the pure resonant curve \( v^2(q) \), which gives the scattering profile of the counterion alone. Since \( I_0(q,E) \) contains three independent variables, one cannot extract directly \( v^2(q) \). Two complementary methods are proposed in the literature to operate in such a way. In all cases, one must record the \( I_0(q,E) \) at various values of the incident beam energy, far and close to the absorption edge. The classical SAXS experiment described in the previous section must then be repeated by changing the value of \( E \). The first method, called the “matrix” method, consists of a numerical solution of \( I_0(q,E) \) needing at least three different energy values. The second approach, called the “fit” (or “Ballauf”) method, consists of neglecting the \( f''(E) \) term before the absorption edge and performing a quadratic interpolation of \( I_0(E) \) for each value of the scattering vector. For a detailed discussion on the problem of solving \( I_0(q,E) \), one can refer to Ref. 60-63. The experimental approach and data treatment used to treat our ASAXS experiments is detailed on Page S6 in the Supporting Information.

Contrast matching Small Angle Neutron Scattering (SANS). These experiments have been performed to look at the contribution of the micellar core only, thus matching the scattering of the sophorose headgroup (molecular formula: \( \text{C}_{12}\text{H}_{21}\text{O}_{11} \), molar mass: 341.30 g/mol) in the simple ion-free sophorolipid micellar system at 80 mM. For the SLD calculations, we used the bulk sophorose density (1.68 g/mL) and we considered the fact that labile COH groups could be exchange into COD. The estimated SLD for sophorose under these conditions is \( \sim 2.6 \times 10^{-6} \) Å\(^{-2}\), corresponding to a \( \sim46:54 = \text{D}_2\text{O}:\text{H}_2\text{O} \) mixture. These values are to be considered as an approximation because the exact sophorose density in sophorolipid is not
exactly known. Two samples have then been prepared at concentration of the sophorolipid of 50 mg/mL: a) in fully deuterated water (100:0 = D₂O:Η₂O) and b) in 46:54 = D₂O:Η₂O mixture.

SANS experiments have been performed on the PACE beamline at the Laboratoire Léon Brillouin (LLB) facilities at CEA-Saclay (France). 1 mm quartz cells have been used as sample holder. The low-q portion of the data was recorded using a sample-to-detector distance of 4565.95 mm and wavelength of 6 Å while the high-q portion was recorded using 866.14 mm and 6 Å. The scattering intensity is obtained from the determination of the number of neutrons in the incident beam and the detector cell solid angle. Data were corrected for the ambient background, empty cell scattering, neutron beam transmission and detector efficiency and normalized to the neutron beam flux to get the scattered intensity I(q) in absolute units according to ref. 65. Incoherent signal was substrated by measuring the background value at high-q values for both, sample-free, 100:0 = D₂O:Η₂O and 46:54 = D₂O:Η₂O mixtures. All data have been treated using the Pasinet 2 software⁶⁶ provided at the beamline free of charge. Pair distribution function analysis, P(r), was done using the SASView software package⁴⁴ using estimated Dmax values of 40 Å and 140 Å, respectively for the 46:54 and 100:0 = D₂O:Η₂O mixtures.

Molecular dynamics simulations.

All the MD simulation discussed in this paper were carried out with GROMACS (v4.6.6)⁶⁷,⁶⁸ and with the protonated acidic sophorolipid (SL) molecule presented in Figure 1. Since the SL ionization degree is found to be < 5 % (see the experimental section) and similar to Prasad et al.,⁶⁹ we only simulated the protoned form of the SL and without ions. As in a previous work⁶⁹ to model the surfactant, we used an “united-atom” force field based on the GROMOS53A6⁷⁰. In particular, the parameters of the sophorose group were taken from the GROMOS53A6 force field for carbohydrates (56A6CARBO) developed by Hansen et al.⁷¹. Concerning the alkyl chain (with the double bond) and the COOH groups, the parameters were taken from Kukol⁷² and the glutamic acid parameters available in the GROMOS53A6⁷⁰ force field, respectively. To be consistent with the GROMOS force field, we used the simple charge water model (SPC)⁷³ to model water with the SETTLE algorithm⁷⁴ to keep its geometry rigid.
We have to mention that by not taking into account in the simulation the effects of the pH and presence of the ions in the solution, our simulations may have some limitations. However, as we will see further, the simulation results are consistent with the reported experimental data and we think they provide a good model basis of the micelle as a function of their aggregation numbers. To have a large overview of the micelle characteristics as a function of the SL aggregation number, $N_{agg}$, we carried out 6 MD simulations with different SL $N_{agg}$ values (Table 2). Specifically, we chose 28 and 37, values obtained by Penfold et al.\textsuperscript{36} for mixture of sophorolipids (sample S4 in ref. 36: 52.2% non acetylated, 39% mono-acetylated and 8.5% di-acetylated) respectively at 5 mM and 30 mM; 56 and 68 adapted from the SAXS data in this work (Table 3, RbCl system). In addition to these simulations, we also performed two more simulations with $N_{agg} = 80$ and 112, to examine the effect of high $N_{agg}$ values of the SL aggregate structure. These two $N_{agg}$ values correspond to an intermediary point between 56 and 68 and a value estimated by Penfold.\textsuperscript{36} The effective concentrations of the simulated systems are in a range of 110 mM and 456 mM (given SL mass fraction, SL wt\% of 6.2 – 12.6 \%), larger than the experimental concentration (80 mM) to reduce the number of water molecules and, consequently, the computational cost of the simulations. To construct the system for simulation, we used a self-assembled approach, where the corresponding number of sophorolipids with a random orientation was inserted in a SPC water cubic box. All the systems were minimized, equilibrated and finally simulated at ambient conditions (P = 1.015 bar and T = 303 K). Specifically, the temperature and the pressure were controlled with the Bussi et al.\textsuperscript{75} thermostat ($\tau_T = 0.1$ ps) and the Parrinello-Rahman\textsuperscript{76,77} barostat ($\tau_p = 3.0$ ps and with a compressibility of 45 x 10\textsuperscript{-6} bar\textsuperscript{-1}). During the
equilibration stages, the SL molecules were harmonically restrained with a force constant of 1000 kJ.mol.nm\(^{-1}\) to equilibrate the solvent around the sophorolipids. SL and water molecules were coupled separately with two thermostatic baths as a standard rule. To integrate the equations of motions, we used a time step of 2 fs with the P-LINCS algorithm\(^{78}\) to restrain bond lengths to their equilibration values. Electrostatic interactions were treated with the reaction field approach\(^{79}\) with the non-bonded interactions evaluated with a twin range cutoff scheme, with a short- and long-range cutoff distances of 8 Å and 14 Å, respectively and an update frequency of 2 timesteps for the short-range pair list. To correct the truncation of electrostatic interactions outside the cutoff of 14 Å, a reaction field term\(^{79}\) corresponding to a relative dielectric permittivity of the SPC water (61)\(^{80}\) was added. Finally, the production simulations were performed during 110 – 162 ns with the atomic data collected every 2 ps for subsequent analysis.

Table 2 - Overview of the simulated systems. \(N_{agg}\), \(n_{H2O}\) and \(n_{atom}\) are the numbers of sophorolipid monomers, water, and atoms composing each systems. \(n_{Water}/SL\), SLwt\% [SL], \(L_{box}\) and \(t_{sim}\) are the number of water per sophorolipid molecule, the sophorolipid concentration, the total mass fraction, the box size and the simulation time (in ns), respectively.

<table>
<thead>
<tr>
<th>#MD</th>
<th>(N_{agg})</th>
<th>(n_{H2O})</th>
<th>(n_{atom})</th>
<th>(n_{Water}/SL)</th>
<th>SLwt% [SL]</th>
<th>(L_{box}) (Å)</th>
<th>(t_{sim}) (ns)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MD1</td>
<td>28</td>
<td>13600</td>
<td>42228</td>
<td>485.7</td>
<td>6.6</td>
<td>0.114</td>
<td>76.3</td>
</tr>
<tr>
<td>MD2</td>
<td>37</td>
<td>18000</td>
<td>55887</td>
<td>486.5</td>
<td>6.6</td>
<td>0.114</td>
<td>77.7</td>
</tr>
<tr>
<td>MD3</td>
<td>56</td>
<td>13600</td>
<td>43656</td>
<td>242.9</td>
<td>12.4</td>
<td>0.228</td>
<td>83.7</td>
</tr>
<tr>
<td>MD4</td>
<td>68</td>
<td>16500</td>
<td>52968</td>
<td>242.6</td>
<td>12.4</td>
<td>0.228</td>
<td>82.9</td>
</tr>
<tr>
<td>MD5</td>
<td>80</td>
<td>19400</td>
<td>62280</td>
<td>242.5</td>
<td>12.4</td>
<td>0.228</td>
<td>87.5</td>
</tr>
<tr>
<td>MD6</td>
<td>112</td>
<td>13600</td>
<td>46512</td>
<td>121.4</td>
<td>22.0</td>
<td>0.456</td>
<td>80.4</td>
</tr>
</tbody>
</table>

Results and Discussion

**Description of the “coffee bean” like prolate ellipsoid of revolution model**

As critically discussed on Page S2 in Supporting Information, the best shape model to fit SAXS data is the core-shell prolate ellipsoid of revolution function, which gives good \(\chi^2/Npts\) values below 15. The geometric parameters of this model are the equatorial core radius, \(R_c\), the axial core ratio \(X_C\) (\(X_C>1\) for prolate micelles), the equatorial shell thickness, \(T_S\), and the polar shell ratio in the axial direction, \(X_S\) (see Figure 2).
Figure 2 – Core-shell prolate ellipsoid of revolution model chosen to fit the SAXS data

For a constant shell thickness, $X_S = 1$, the fit gives a $\chi^2/Npts$ value around 13. However, the best fit ($\chi^2/Npts = 6.7$) is obtained for a slightly negative $X_S$ value ($X_S \sim -0.2$) corresponding to a small and negative polar shell thickness in the axial direction, $T_P$, of the core-shell ellipsoid of revolution model.

In Figure 3, we report a typical SAXS curve recorded on sample 7 (SrCl$_2$-25mM), which shows a series of fits using the same core-shell prolate ellipsoid of revolution form factor model and a variable $X_S$, allowed to vary between 0 and 1. It can be clearly seen that the bottom of the first minimum is best fitted with a negative $X_S$ value, a fact which has no clear
physical meaning, as it supposes that there is a small part of the hydrocarbon core which is in close contact with the aqueous solvent at the end-tips of the ellipsoid in the axial direction. However, as the best fit is always obtained for $X_S < 0$, we impose $X_S = 0$ for all samples’ profiles. We interpret this assumption as follows: the end-tips in the axial direction of the sophorolipid aggregates are somewhat difficult to describe and may consist of a quite disorganized assembly, containing COOH groups, sophorose, salt, water and part of the aliphatic chain.

It is important to note that whatever the absolute value of $X_S$, the trends obtained for the other variables is unchanged, as it can be seen from the comparison between the fit parameters for $X_S = 0$ presented in Table 3 (discussed later in more detail) and the fit parameters obtained for $X_S = 0.3$, shown in Table S2/Table S3 in the Supporting Information. Except the surprising feature of $X_S = 0$, all other variable parameters give realistic values: the equatorial core radius and shell thickness are around 7.7 and 12.3 Å respectively, and the shell SLD is $1.1 \times 10^{-5}$ Å$^{-2}$ (387 e/ nm$^3$), corresponding roughly to 73 % v/v of water in a simple water-sophorose composition of the shell (Table 4). The $X_C$ ratio of 9.4 characterizing the elongation of the ellipsoid is slightly high, if it compared with previous results$^{33}$ and will be discussed thereafter.

The resulting shape that best describes the sophorolipid micelles is very close to a “coffee bean” like, with a variable shell thickness from $T_S$ to 0, which is quite atypical with respect to the majority of both ionic, non-ionic and bolaform$^{16}$ surfactant systems, generally described by core-shell form factors with a homogeneous shell thickness.$^{19-21}$ Nevertheless, this peculiar shape is the best suited to fit at best all systems described in this work and a large number of complementary other systems.

All SAXS measurements were made at the same sophorolipid concentration, 80 mM, and at concentrations of salt and/or base in the range from 5 to 100 mM. Table 3 shows the model parameters ($N_{agg}, R_c, X_c, z, \rho_s$) for core-shell ellipsoidal micelles of the sophorolipids with either salt or a base at various concentrations, but also the calculated shell thickness ($T_S$) and volume fraction of CH$_2$ in the dry core ($\alpha$). Table 4 presents the volume fraction of water ($x_{v,w}$), and number of water molecules per sophorolipid ($n_{w/SL}$) in the shell; these have been indicated for three possible headgroup ($hg$) scenarios: $hg_1$ considers only sophorose and COOH (or COO$^{-}$); $hg_2$ includes the $hg_1$ hypothesis and the C$_{17}$ and C$_{18}$ atoms; $hg_3$ considers the $hg_2$ scenario as well as the contribution of CH$_2$ groups from the aliphatic chain.

18
excluded from the core (13(1-α), where α has been defined in Eq.4 and experimental values given in Table 3).

First of all, the results show that the equatorial core radius $R_C$, which corresponds to the dry part of the micelle, is relatively small, between 6.0 and 8.0 Å, a value which is the lower limit of what is acceptable for the model. In return, the thickness of the shell, between 11.0 and 12.5 Å, is quite large if this value is compared with those reported in the literature for similar systems studied by SANS (~8 Å, ~3 Å), a discrepancy that may be explained by the different contrasts inherent in each technique. At the same time, the value of $\alpha$ is contained in the range between 0.52 and 0.72 depending on the system and it mainly indicates that a significant proportion of the hydrocarbon chain is in an aqueous environment and that is the counterpart of the small core and large shell of these aggregates. This is expected for a bolaamphiphile, as quantitatively demonstrated here below.

Table 3 – Fit parameters obtained from the core-shell ellipsoid of revolution form factor using $X_s=0$ (defined in Figure 2). $N_{agg}$ aggregation number; $R_C$ equatorial core radius; $X_C$ axial core ratio; $T_S$ equatorial shell thickness; $\rho_s$ scattering length density of the shell; $\alpha$ fraction of CH$_2$ inserted in the dry core of the micelle (Eq. 4); $z$ charge per micelle.

<table>
<thead>
<tr>
<th>Sample</th>
<th>$N_{agg}$</th>
<th>$R_C$ (Å)</th>
<th>$X_C$</th>
<th>$T_S$ (Å)</th>
<th>$\rho_s \times 10^4$(Å$^2$)</th>
<th>$\alpha$</th>
<th>$z$</th>
</tr>
</thead>
<tbody>
<tr>
<td>RbCl-10mM</td>
<td>55</td>
<td>8.0</td>
<td>6.5</td>
<td>12.0</td>
<td>1.0741</td>
<td>0.70</td>
<td></td>
</tr>
<tr>
<td>RbCl-50mM</td>
<td>64</td>
<td>8.0</td>
<td>7.7</td>
<td>11.7</td>
<td>1.0768</td>
<td>0.72</td>
<td></td>
</tr>
<tr>
<td>RbCl-100mM</td>
<td>69</td>
<td>7.9</td>
<td>8.1</td>
<td>11.5</td>
<td>1.0829</td>
<td>0.69</td>
<td></td>
</tr>
<tr>
<td>SrCl$_2$-5mM</td>
<td>52</td>
<td>7.7</td>
<td>6.9</td>
<td>12.4</td>
<td>1.0608</td>
<td>0.71</td>
<td></td>
</tr>
<tr>
<td>SrCl$_2$-25mM</td>
<td>72</td>
<td>7.7</td>
<td>9.4</td>
<td>12.3</td>
<td>1.0611</td>
<td>0.72</td>
<td></td>
</tr>
<tr>
<td>SrCl$_2$-50mM</td>
<td>73</td>
<td>7.8</td>
<td>8.9</td>
<td>11.8</td>
<td>1.0751</td>
<td>0.69</td>
<td></td>
</tr>
<tr>
<td>RbOH-40mM</td>
<td>34</td>
<td>6.6</td>
<td>5.5</td>
<td>12.3</td>
<td>1.0618</td>
<td>0.55</td>
<td>6.8</td>
</tr>
<tr>
<td>RbOH-80mM</td>
<td>24</td>
<td>6.1</td>
<td>4.7</td>
<td>11.0</td>
<td>1.0707</td>
<td>0.53</td>
<td>9.9</td>
</tr>
<tr>
<td>Sr(OH)$_2$-10mM</td>
<td>50</td>
<td>7.2</td>
<td>6.8</td>
<td>12.6</td>
<td>1.0632</td>
<td>0.59</td>
<td>4.6</td>
</tr>
<tr>
<td>Sr(OH)$_2$-20mM</td>
<td>41</td>
<td>7.0</td>
<td>5.4</td>
<td>12.3</td>
<td>1.0785</td>
<td>0.52</td>
<td>5.0</td>
</tr>
</tbody>
</table>

Table 4 - $x_{w,v}$ volume fraction of water in the shell (Eq. 7) and $n_{w/SL}$ number of water molecules per sophorolipid (Eq. 8). Several compositions of the headgroup are considered: headgroup 1, $hg_1$ sophorose + COOH; headgroup 2, $hg_2$ sophorose + COOH + CH (C$_{17}$) + CH$_3$ (C$_{18}$); headgroup 3, $hg_3$ sophorose + COOH + CH (C$_{17}$) + CH$_3$ (C$_{18}$) + 13(1-α)CH$_2$. Grey-shaded area refers to the hypothesis best describing the shell composition.
Shell hydration vs. ion condensation

\( n_{w/SL} \) gives an estimate of the number of water molecules per sophorolipid in the shell. As shown in Table 4, this parameter strongly depends on the supposed composition of the dry headgroup. If one only considers the sophorose and COOH groups to be part of the shell (\( hg_1 \) hypothesis in Table 4), \( n_{w/SL} \) varies between 29 and 37. In the \( hg_2 \) scenario (\( hg_1 \) plus CH and CH\(_3\)), \( n_{w/SL} \) does not vary much, as it is comprised between 26 and 34, indicating that inclusion of CH and CH\(_3\) in the headgroup does not have a substantial impact on the shell hydration. Finally, if one includes the fraction of CH\(_2\) not included in the core (\( 13(1-\alpha) \)), \( hg_3 \) hypothesis in Table 4, the hydration is sensibly lower and \( n_{w/SL} \) now varies between 18 and 28. As first comment, we must state that these estimations should be taken with caution, because our SAXS model does not include the presence of ions in the shell, which can also be responsible for the variation of the shell SLD. The proportion of ions in the shell could be controlled in the fit through, for example, an adjustable parameter \( k= \) number of cations per sophorolipid in the shell. However, \( k \) cannot be determined without fixing the number of water molecules per sophorolipid, both parameters influencing the value of \( \rho_S \) in opposite ways: ionic species increase the value of \( \rho_S \), whereas water molecules decrease it. By neglecting the influence of ion concentration, the number of water molecules per sophorolipid is adjusted to account for both effects, so that an increase of \( \rho_S \) leads to a decrease of \( n_{w/SL} \) even if it can come in reality from an increase of the ion condensation. As a result, the number

<table>
<thead>
<tr>
<th>Sample</th>
<th>( x_{v,w} )</th>
<th>( n_{w/SL} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( RbCl-10mM )</td>
<td>0.70 0.65 0.54</td>
<td>32 29 23</td>
</tr>
<tr>
<td>( RbCl-50mM )</td>
<td>0.69 0.64 0.54</td>
<td>31 28 22</td>
</tr>
<tr>
<td>( RbCl-100mM )</td>
<td>0.68 0.63 0.51</td>
<td>29 26 20</td>
</tr>
<tr>
<td>( SrCl_2-5mM )</td>
<td>0.73 0.68 0.59</td>
<td>37 34 28</td>
</tr>
<tr>
<td>( SrCl_2-25mM )</td>
<td>0.73 0.68 0.59</td>
<td>37 34 28</td>
</tr>
<tr>
<td>( SrCl_2-50mM )</td>
<td>0.70 0.65 0.54</td>
<td>32 29 22</td>
</tr>
<tr>
<td>( RbOH-40mM )</td>
<td>0.73 0.68 0.53</td>
<td>37 34 24</td>
</tr>
<tr>
<td>( RbOH-80mM )</td>
<td>0.71 0.66 0.49</td>
<td>33 30 21</td>
</tr>
<tr>
<td>( Sr(OH)_2-10mM )</td>
<td>0.72 0.68 0.54</td>
<td>36 33 24</td>
</tr>
<tr>
<td>( Sr(OH)_2-20mM )</td>
<td>0.69 0.64 0.46</td>
<td>31 28 18</td>
</tr>
</tbody>
</table>
of water molecules per sophorolipid reported in this case can be interpreted as minimum values, even if the space filling requirement may also have an influence. This approximation will be confirmed by the ASAXS experiments, which will show that the amount of counterions in the shell is practically negligible, unless their concentration becomes important.

Secondly, values of $n_{w/SL}$ above ~25 are higher than expected from the literature about glycolipids and disaccharides. It is well-known that hydration properties of disaccharides with the same chemical formula as sophorose (maltose, sucrose and trehalose, $C_{12}H_{22}O_{11}$) but different structures, are different, depending on the solute-water interactions. Effects coming from the linkage type between the two glucose unit (sophorose is a glucose β(1,2) saccharide), the position of hydroxyl groups, the number of intracellular hydrogen bonds may influence the hydration properties of disaccharides and have been extensively studied. For example, the headgroup hydration number of maltoside surfactants is reported to be 8 in the case of SAXS studies and in Abel et al. with MD simulations. For the trehalose, which is known as being slightly more hydrated than maltose, Lupi et al. report values taken from literature that lie between 4 and 18. It is well-known that hydration numbers reported depend on the experimental method and/or assumption adopted. Ultrasonic measurements give values among the highest, and MD simulations for which the hydration number is defined as the number of H-bound water leads to similar values (around 15 to 18) suggesting that these experimental hydration numbers mainly refer to water molecules directly involved in H-bonds with the solute. In contrast, in their study on hydration and mobility of trehalose in aqueous solution, Revsbech Winther et al. used a geometric hydration number, identified as the number of water molecules required to cover the solute with one layer (47 water molecules per trehalose).

The values obtained in this work are for a micellar shell composed of a sophorose group, that is a disaccharide and the hydration properties of which, if they are reasonably expected to differ from maltose and trehalose, should still not be excessively high. For this reason, the $hg_3$ hypothesis provided in Table 4 is most likely to be real, as reasonable values are certainly the ones including the COO$^-$ (or COOH), the CH and CH$_3$ groups but also 13(1-α)CH$_2$ groups and an unknown, and variable, number of ions, depending on the sample composition. More than reasonable, $hg_3$ hypothesis is actually expected: if hydration numbers of COOH, CH$_2$ and CH$_3$ are reported to be small, 1.153, 0.999 and 0.586 respectively, it is expected the aliphatic chain portion close to the headgroup to be hydrated. Typical estimations (both experimental and theoretical) for alkylammonium salt and SDS
Surfactants are that up to 3 CH$_2$ groups from the polar head can be hydrated.$^{50-52}$ In this work, we find between 4 and 7 hydrated CH$_2$ groups per sophorolipid (calculated as 13(1-$\alpha$)CH$_2$, where $\alpha$ is given in Table 3); these values can reasonably be expected for a system containing two polar headgroups. On the other hand, hydration number of ions is higher and, depending on the ionic concentration, it can influence the SLD of the shell quite heavily: reported hydration numbers lie between 6 and 8 for Rb$^+$, between 4 and 8 for Na$^+$,$^{88}$ and between 7.3 to 10.3 for Sr$^{2+}$.$^{89}$ According to the ASAXS experiments reported in the next section and showing a negligible amount of counterions in the shell, one can consider the amount of hydration water to be also negligible.

If the contribution to the hydration from all species can reasonably explain the experimental values of $n_{w/SL}$ found here, one should still take these values with caution, given the wide range of variation of the theoretical values. Moreover, there is no indication that the micelle shell does not contain free water that is not accounted for in the hydration numbers reported. The shell hydration may also be strongly related to the conformation of sophorolipids within the micelle, depending on the molecular structure of the sophorolipid and its impacts on the micellar structure.

As we can see from Table 3, the shell thicknesses reported here are always greater than 10.5 Å and usually between 11.0 and 12.5 Å, a range of values which seems to be high compared to the size of the sophorose alone, estimated to be below 10 Å,$^{32}$ 7.3 Å in polysophorolipids,$^{90}$ or 9 Å for trehalose.$^{86}$ The shell thickness of SL micelles determined by SANS was also reported to be around 8 Å.$^{33}$ The quite thick shell found here may take into account the “roughness”$^{91}$, confirmed by the MD simulation in the next sections, of the apolar core and hydrated shell interface coming from the presence of CH$_2$ in the hydrated shell environment. We can estimate the typical separation distance between sophorose molecules in the shell from the sophorose molar concentration (0.93-1.34 mol.L$^{-1}$) to be around 3.5-5.1 Å, by using molecular parameters of trehalose (see Ref 86). This space, filled by the other components of the shell, can accommodate one water layer, and even if there are some alkyl groups in close contact with the sophorose, regarding the geometric hydration number of 47 for trehalose in water, the number of water molecules per sophorolipid in the shell presented here, which lie from 20 to 30, in the $hg_3$ hypothesis, does not seem outrageous.

Table 5: Area parameters; $A_{mic,wSL}$ = area per sophorolipid at the micelle/solvent interface; $A_{C,SL}$ = dry core/hydrated shell interface; $A_{alkyl,sopha}$ = area per sophorolipid between the alkyl chain and the sugar/carboxylate headgroup; $L$ = average length between the headgroups
### Table: Area per Component

<table>
<thead>
<tr>
<th>Sample</th>
<th>$A_{\text{C,SSL}}$ (Å²)</th>
<th>$A_{\text{alkyl/sopho}}$ (Å²)</th>
<th>$A_{\text{mic,w/SL}}$ (Å²)</th>
<th>L (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RbCl-10mM</td>
<td>75</td>
<td>104</td>
<td>196</td>
<td>3.76</td>
</tr>
<tr>
<td>RbCl-50mM</td>
<td>76</td>
<td>103</td>
<td>194</td>
<td>3.69</td>
</tr>
<tr>
<td>RbCl-100mM</td>
<td>74</td>
<td>102</td>
<td>187</td>
<td>3.49</td>
</tr>
<tr>
<td>SrCl$_2$-5mM</td>
<td>78</td>
<td>108</td>
<td>213</td>
<td>4.21</td>
</tr>
<tr>
<td>SrCl$_2$-25mM</td>
<td>79</td>
<td>106</td>
<td>208</td>
<td>4.19</td>
</tr>
<tr>
<td>SrCl$_2$-50mM</td>
<td>75</td>
<td>103</td>
<td>193</td>
<td>3.72</td>
</tr>
<tr>
<td>RbOH-40mM</td>
<td>71</td>
<td>124</td>
<td>219</td>
<td>4.04</td>
</tr>
<tr>
<td>RbOH-80mM</td>
<td>75</td>
<td>141</td>
<td>231</td>
<td>3.72</td>
</tr>
<tr>
<td>Sr(OH)$_2$-10mM</td>
<td>70</td>
<td>111</td>
<td>204</td>
<td>4.03</td>
</tr>
<tr>
<td>Sr(OH)$_2$-20mM</td>
<td>64</td>
<td>112</td>
<td>190</td>
<td>3.48</td>
</tr>
</tbody>
</table>

### Area per Sophorolipid

Another interesting piece of information is given by the analysis of the area per sophorolipid at the micelle/solvent interface ($A_{\text{mic,w/SL}}$), and at the dry core/hydrated shell interface ($A_{\text{C,SSL}}$, see Table 5). These are calculated using the classical formula for the surface area of a prolate ellipsoid divided by $N_{\text{agg}}$; the values of the semi-axes, as defined in Figure 2, are given in Table 3. Micellar structure of dodecyl maltoside has been described by Cecutti et al. showing that ellipticity of the micelle comes from the bulky hydrated maltose headgroups and their perpendicular conformation to the interface. Area per dodecyl maltoside at the micelle/solvent interface and at hydrophobic core/sugar headgroup interface are 87 and 50 Å², respectively. In the case of sophorolipids, however, one must keep in mind that there is one alkyl chain and two different polar heads per surfactant, leading to values lying from 187 to 230 Å² per sophorolipid at the micelle/water interface, and from 64 to 79 Å² per SL at the core/shell interface, depending on the sample composition. Another point to consider is that the core/shell interface in our case is located quite far from the sugar headgroup, considering the $\alpha$ value lying around 0.52 - 0.72. If the alkyl chains are perpendicular to the core/shell interface, we can estimate the area per sophorolipid between the alkyl chain and the sugar/carboxylate headgroup $A_{\text{alkyl/sopho}}$ to be the surface per SL of an ellipsoid of equatorial radius $R_c+6.5(1-\alpha)1.265$ and axial radius $R_cX_c+6.5(1-\alpha)1.265$ (noticing that this surface may be slightly overestimated as the influence of the $X_S$ parameter is neglected). In this case, areas reported lie between 102 to 141 Å² for the most ionized micelles, comparable to that reported by Chen et al. for SL at air-water interface (104 Å²).

23
Ionic and basic effects

According to the results in Table 3, some general observations can be made regarding the effect of salts on sophorolipids micelles. First, the salts cause an elongation of the micellar aggregates and subsequent increase of the aggregation number, while the diameter of the core remains constant. For instance, increasing RbCl from 10 mM to 100 mM, it produces an increase of $X_C$ from 6.5 to 8.1, whereas $R_C$ is practically 8 Å. However, it appears that the elongation due to the presence of salt has a limit, observed through $N_{agg}$ and $X_C$ values which increase rapidly between 10 and 50 mM of RbCl (5 and 25 mM of SrCl$_2$) and then seem to reach a plateau at higher concentration, around 73 for $N_{agg}$ and 8.5 – 9.0 for $X_C$. Another fairly clear effect of adding salt is the slight shell shrinkage and the concomitant dehydration observed through the shell thickness ($T_S$), the volume fraction of water in the shell ($x_{v,w}$) and the decreasing number of water molecules per sophorolipid ($n_{w/SL}$).

The shell SLD, $\rho_S$, is in the range between 1.06 and 1.09 x 10$^{-5}$ Å$^{-2}$, where the smaller value is systematically observed for the SrCl$_2$ salt system. Small differences between the two salts is also observed on the values of $R_C$, $T_S$, $x_{v,w}$ and $n_{w/SL}$. The sophorolipids micelles in the presence of SrCl$_2$ have slightly smaller cores, a shell slightly thicker and hydrated than in the presence of RbCl. These differences have no impact on the value of $\alpha$ that remains constant for all the salt systems, and quite high (around 0.69 - 0.72). The aggregation number $N_{agg}$ is slightly smaller for SrCl$_2$ at low salt concentration (52 vs 55), while at higher concentration the elongation of sophorolipids micelles is more pronounced with SrCl$_2$ than with RbCl ($X_C$ reached a value of 8.9 for SrCl$_2$ and 8.1 only for RbCl). Areas per sophorolipid and the average length between the headgroups, $L$, reflect these small differences, being slightly higher for SrCl$_2$ than for RbCl (Table 5).

Adding a base to sophorolipids micelles solutions considerably changes the physico-chemical parameters. First of all, one observes the appearance of intermicellar interactions, accounted here with the introduction of a structure factor described with a repulsive screened Coulombic intermicellar interaction potential. This is consistent with what it was previously reported. In terms of micellar size, the equatorial core radius is dramatically reduced, from 7.2 Å for the less basic system (Sr(OH)$_2$ – 10 mM) to 6.1 Å for the most basic system (80 mM RbOH) while the $X_C$ ratio is much smaller than those reported above, decreasing to values smaller than 6 for all [OH$^-$] higher than 40 mM. The aggregation number follows this trend, decreasing to 24 for RbOH 80 mM. $\alpha$ also undergoes a fall, reaching values around
0.52, which indicates that in basic medium, half of the total amount of CH\textsubscript{2} are located in an aqueous environment. These trends are consistent with an increase of the micelle curvature due to the increased electrostatic repulsion taking place between more ionized headgroups. The shell thickness is practically invariable (12.3 Å) for the less basic samples (RbOH-40 mM and Sr(OH)\textsubscript{2}-20 mM) and substantially decreases (11.0 Å) at high concentration of RbOH. Finally, Figure 4 shows the typical SAXS profiles of sophorolipids in the presence of SrCl\textsubscript{2} and Sr(OH)\textsubscript{2}, showing the experimentally found spectral differences and a pictorial imaging of the salt and base effect on the micellar size.

![SAXS spectra of sophorolipids in the presence of SrCl\textsubscript{2} and Sr(OH)\textsubscript{2}](image)

**Figure 4** – SAXS spectra of sophorolipids in the presence of SrCl\textsubscript{2} and Sr(OH)\textsubscript{2}. The pictures represent the ellipsoidal nature of the micelle with a $X_s \approx 0$, as commented before. The difference in longitudinal dimension among the ellipsoids is based on the data obtained from the best fits and presented in Table 3

**Counterion distribution in the salt systems studied by ASAXS**

ASAXS is a powerful method to put in evidence the distribution of counterions in soft matter, and in particular around polyelectrolytes and micelles.\textsuperscript{57-62} It was employed to describe the charge distribution around charged polymers,\textsuperscript{58,61} DNA and RNA\textsuperscript{93} and CTAB micelles.\textsuperscript{60,62} In this study, ASAXS is employed with the goal of revealing indirectly the distribution of the COO\textsuperscript{-} groups, if one assumes that its negative charge is neutralized by a counterion upon increasing the solution pH. In this case, the SAXS spectrum shows a broad scattering hump, which can be nicely fitted using a Screened Coulomb potential, as it was previously done in SANS experiments.\textsuperscript{33} The resulting negative charge of the micelles settles between -5 and -10 (Table 3) and it clearly increases upon increasing the amount of base for
both Rb\(^+\) and Sr\(^{2+}\). ASAXS can be used both in a qualitative and quantitative way\(^{62,63}\). If the former tells about the space distribution of counterions, the latter provides data on the amount of counterion in the proximity of charged macroions and even their binding degree. However, the quantification process can be very delicate\(^{60,61,94}\) and it was mainly used so far on either polyelectrolytes or robust micellar systems.
Figure 5 – a, c, e, g) Energy-dependent SAXS spectra for the sophorolipid – RbCl (10, 100 mM) and RbOH (40, 80 mM) systems. The energies at which the spectra have been recorded are indicated in the figure. b, d, f, h) $F^2(q)$ and $v^2(q)$ scattering profile of the cation (Rb$^+$ and Sr$^{2+}$) as obtained from the ASAXS treatment of the spectra. More experimental details are presented on Page S6 in the Supporting Information.

Figure 5 shows the typical scattering profile, $I(q)$, recorded for the RbCl (10, 100 mM) and RbOH (40, 80 mM) systems at the incident energy $E= 14599, 14999, 15179, 15195, 15199, 15201, 15203$ eV, corresponding to the $\Delta E_{Rb^+} = -600, -200, -20, -4, 0, +2, +4$ eV (treatment of background is shown in Figure S2 in the Supporting Information). The ASAXS results are only shown for the RbCl system but very similar conclusions can be drawn for the SrCl$_2$-containing samples, and for this reason they will not be reported here. As explained in the experimental section, fluorescence contribution has been subtracted for all spectra. At a first glance, spectra are very close to each other but a closer look in the inset of Figure 5a,c, for instance, indicate that $I(q)$ appreciably (with respect to the error bars) increases with increasing energy (colour sequence: black, red, green, blue...) for both the RbCl 10 mM and 100 mM systems. This effect is expected in the case of an anomalous effect upon varying energy through the edge and it demonstrates that Rb ions decorate the micellar outer surface. Nevertheless, variation in $I(q,E)$ is very mild; at 100 mM RbCl (Figure 5c), a concentration in counterions at which the anomalous effect was reported to be very important for polyelectrolytes and ionic surfactants systems,$^{62,63} I(q,E)$ only varies by a factor 1.04 when
measured far below and at the absorption edge $[I(0.03,14599)= 0.47 \text{ cm}^{-1}$ compared to $I(0.03,15199)= 0.49 \text{ cm}^{-1}$ at the Rb$^+$ edge], while expected variations should not be less than a factor 2. This is further demonstrated by the splitting of $I(q)$ into the $F_0^2(q)$ and $v^2(q)$ components, respectively, the nonresonant intensity of the macro-ion and the pure resonant contribution of the counterion, shown in Figure 5b,d. The splitting shows no noticeable effect on the 10 mM system using both the matrix and fit methods, and a mild anomalous effect on the 100 mM system, but only employing the matrix method. This last result is coherent with the higher salt amount present in this system, thus indicating that part of Rb$^+$ cation in solution decorate the micellar shell. This result helps better understanding the evolution of the shell SLD ($\rho_s$) commented in the SAXS section (Table 3). One can now affirm that the amount of salt in the shell is below the limit of detection of the ASAXS technique, which can reasonably be settled in the mM range. This shows that the very small increase in the $\rho_s$, found for the RbCl, but also the SrCl$_2$ systems, is most likely due to a true variation in the number of water molecules per sophorolipid. However, for the higher RbCl concentrations (e.g., 100 mM), one cannot exclude the presence of larger amounts of salt in the shell, a fact which may also, as commented before, account for the mild increase in $\rho_s$. Interestingly, in the presence of salt alone, the equilibrium pH of a sophorolipid solution is slightly below 5.

The corresponding experimental ionization degree ($\frac{[\text{COO}^-]}{[\text{COOH}]+[\text{COO}^-]}$), calculated from the titration curve (Figure S3 in the Supporting Information), is below 5 %. This means that at high salt content, counterions decorate the hydrophilic shell independently from the presence of carboxylate groups. In terms of quantification, the mild anomalous effect recorded does not allow a reasonable quantification of the amount of counterions-per-sophorolipid.

Upon addition of RbOH, carboxylic groups deprotonate into carboxylates, which carry their respective counterions, our hypothesis$^{33}$ being that the negative charges are settled at the micellar/water palisade. Under these assumptions, this effect should be detectable by ASAXS at a larger extent with respect to the base-free medium. The corresponding experiments in Figure 5e,g show, on the contrary, that $I(q)$ undergoes no sensible evolution with the incident energy, as all curves practically lie within a small range comprised in the error bars. The poor anomalous scattering effect is confirmed by the signal splitting using both the fit and matrix methods (Figure 5f,h), which do not provide a significant $v^2(q)$ component. Since similar results are obtained with Sr(OH)$_2$, these will not commented further. How to interpret this result? One of the main results so far is the fact that the bolaform morphology of
sophorolipids promotes the formation of “non-standard” micelles. By this we intend a micelle that does not have a homogeneous hydrophobic core well-separated from the hydrophilic corona with constant thickness. The size of micelles and the aggregation number sensibly decrease upon addition of a base (Table 3) and a partial solubilisation of sophorolipids in solution should not be excluded. Such a scenario could suggest that not all COO\(^{-}\) groups do actually settle in a narrow region at the micellar/water interface, but they could be randomly distributed in the thick shell region, being closer to the hydrophobic core. If resonant species (the counterions) penetrate deeper inside the micelle, the difference in terms of core-shell SLD at the element absorption edge will be attenuated, and one expects the ASAXS signal to be attenuated. This scenario could be suggested by the larger number of hydrated CH\(_2\) (about 6 per sophorolipid, for an \(\alpha\) parameter close to 0.5, Table 3) and not excluded with respect to old works on ionic/neutral bolaamphiphiles\(^{22,23}\) and according to which water can penetrate deep inside a bola’s micelle. However, a comparison with previous ASAXS studies is difficult because all of them have been carried out on either polyelectrolytes or micellar surfactant systems with a well-defined core-(charged)shell interface (e.g., CTAB). In both cases, there are no doubts about the localization of the counterions. Even if fluctuations in the counterion distribution is probably amplified in the case of sophorolipids due to a redistribution of the carboxylate groups away from the outer micellar layer, this phenomenon has been already reported to potentially affect ASAXS measurements on other systems. In micellar systems (TTAB), fluctuation of the counterion cloud was shown to affect the scattering intensity of the macroion (TTA\(^{+}\)) at \(q > 0.2\) Å\(^{-1}\). In a rod-like polyelectrolyte, it was shown that longitudinal fluctuations of the number density of the counterions affect SAXS data at \(q > 0.3\) Å\(^{-1}\) and they can actually be measured experimentally.\(^{59,61}\) Finally, fluctuations due to the polydispersity of the number of acrylic acid (PAA) chains on polystyrene (PS) was also shown to affect the pure resonant curve \(v^2(q)\) at \(q > 0.04\) Å\(^{-1}\). According to these works, it seems possible that, if fluctuations of the COO\(^{-}\) distribution occurs in sophorolipid micelles upon addition of a base, the corresponding \(v^2(q)\) profile will inevitably be affected. It seems that in all systems where the macroion (polyelectrolyte, micelle) charge density is well-defined and located at the surface, only the counterions number density fluctuations at the macroion surface should be considered. In this case, the \(v^2(q)\) profiles will be affected at (relatively) large \(q\) values above 0.2 Å\(^{-1}\), which, in the case of sophorolipid micelles, may not play an important role after all. On the contrary, if fluctuations of the macroion charge distribution are experienced, as seen in the case of PAA-PS system, the affected \(q\)-range may be much larger, starting at \(q < 0.1\) Å\(^{-1}\). In this specific case, the effect on the \(v^2(q)\) profile for sophorolipid micelles may not be
negligible. We then expect that if the COO\textsuperscript{−} distribution will occur at the micellar surface only, the $v^2(q)$ will be enhanced, while if the COO\textsuperscript{−} distribution will occur in the interior of the micelle, the $v^2(q)$ will be reduced. We believe that the second scenario does occur in the sophorolipid-based system upon addition of a base.

*Contrast matching by Small Angle Neutron Scattering*

In the SAXS section, we have shown that sophorolipid micelles at 80 mM adopt a non-conventional “coffee bean” like structure, the length of which is slightly influenced by the amount of salt or base. By “coffee bean” like structure we have intended the fact that the composition at the extremities in the axial direction of the micelle is hard to describe with a simple core-shell model and they are most-likely constituted by a mixture of aliphatic chain, sophorose and COOH groups. According to the ASAXS experiments, we could also put in evidence that the distribution of the COO\textsuperscript{−} groups is not only confined at the micellar surface but it is probably random in the thick hydrophilic region. Unravelling the distribution of the sophorolipid molecule inside a micelle can be approached through a direct experiment which consists of “visualizing” the distribution of the oleic acid, or the sophorose group by mean of contrast variation SANS. This experiment consists into matching the SLD of a given section of a particle, or molecule, via a well-chosen mixture of deuterated and hydrogenated solvent, the SLD of which can be tuned by playing on the relative ratio between the two. In the case of SL micelles, it is possible to look at the scattering of the oleic acid, or sophorose, alone by choosing the correct proportion of D\textsubscript{2}O and H\textsubscript{2}O. Sophorose headgroups can be matched with a 46% D\textsubscript{2}O, while oleic acid can be matched with 8% D\textsubscript{2}O. Here, we prefer the first approach because experiments with large amounts of H\textsubscript{2}O will suffer from the strong incoherent diffusion of hydrogen, inherent to the technique, and which are responsible for a critical increase in the background level. A complete match between the solvent and sophorose will give an idea of the spatial distribution of the pure aliphatic-containing core.

Figure 6 shows the typical SANS curve of a salt-free 80 mM solution of acidic SL in pure D\textsubscript{2}O, the description of which is given elsewhere\textsuperscript{33} and the geometry of which we now assume to be “coffee bean” like, as largely commented above. Without going into a further model-based analysis, we simply display the corresponding Indirect Fourier Transformation method (IFT), from which it is possible to evaluate the shape of an aggregate from the corresponding Pair Distribution function, $P(r)$\textsuperscript{96}. The $P(r)$ of a 80 mM solution of SL in D\textsubscript{2}O (Figure 6b) of course reflects the I(q) data and it is typical for highly elongated object, in agreement with the SAXS data presented above and previous data recorded on a similar
system.\(^\text{32}\) If the hypothesis of a homogeneous core-shell ellipsoid (the shell being of constant thickness and homogeneously composed of sophorose and water only) is correct, then one should expect, after matching the contrast between sophorose and the solvent, a SANS profile describing an ellipsoidal micelle being smaller in size (only the core will contribute). The SANS scattering profile of the contrast-matched (46% D\(_2\)O) SL solution is shown in Figure 6a (blue circles) and it presents two interesting features. First of all, the low-q intensity is decreased by a factor 100 with respect to the fully deuterated medium; this is a direct proof that the experiment was successful. In fact, according to Eq. 1, the intensity is directly proportional to the contrast between the object and the solvent. By matching the SLD between the sophorose and the solvent, the contrast only occurs between the core and the solvent and it is thus reduced, thus contributing to decrease the scattered intensity. The second point concerns the micellar shape and size after contrast matching. The \(P(r)\) function at 46% D\(_2\)O (Figure 6b) is typical of a much less elongated morphology, closer to a spheroid, if compared to the fully deuterated system. This suggests that a pure aliphatic core is mainly concentrated in a spheroidal region located at the center of the micelle.

![Figure 6](image)

**Figure 6** – a) SANS curves recorded on a 80 mM system of SL in 100/0 D\(_2\)O/H\(_2\)O (red squares) and 46/54 D\(_2\)O/H\(_2\)O (blue circles); b) Pair distribution function, \(P(r)\), issued from data in a).

This experiment gives credit to, and it actually refines, the sophorolipid micelle model deduced from SAXS analysis. At the equilibrium pH (< 5) and at 80 mM, the sophorolipid micelle should be considered as a prolate ellipsoid divided into three regions, as identified in Figure 7: Region I is composed of a fully aliphatic core and the morphology of which is practically spherical; Region II constitutes the end-tips in the axial direction of the ellipsoid and it is a less defined region composed of aliphatic, sophorose and COOH groups. Region I is directly identified by contrast matching SANS experiments while region II is deduced by the fitting SAXS data using a “coffee bean” like model. Finally, Region III is identified as
being the outer shell composed of sophorose, COOH groups, water and, possibly, counterions. Contribution from vicinal CH\textsubscript{x} groups should not be excluded, as detailed in the headgroup, \textit{hg}\textsubscript{3}, hypothesis of Table 4. As commented before, the contribution of salt plays a role on the ellipticity of the micelle, only but neither on its diameter nor its configuration. Furthermore, the ASAXS data show that the amount of salt in the sophorose surrounding is most likely negligible at least up to 100 mM. The orientation of the sophorolipid molecule within the micelle is sketched in Figure 7. Considering the typical micellar dimension with respect to the sophorolipid size, it seems highly probable that molecules can either cross the micelle from one end to the other, or bend in such a way that the COOH and sophorose groups belonging to the same molecule both lie in the same side of Region III. The addition of small amounts of base (pH range comprised between 5 and 7) promotes electrostatic repulsion among micelles induced by the deprotonation of the COOH into COO\textsuperscript{-}. According to the ASAXS data showing a non-existing anomalous effect in the presence of both Rb\textsuperscript{+} and Sr\textsuperscript{2+}, we believe that in the intermicellar interaction regime, the negative COO\textsuperscript{-} groups and their respective counterions do not seem to be distributed only at the external micellar-water palisade, as we previously hypothesized,\textsuperscript{32,33} but most likely to occupy the entire hydrophilic shell region.

Several authors have tried to study the structure of bolaform amphiphiles and in particular the one of glucose-containing units. The group of Vill has published several works on the micellar formation of bolaform glycolipids, mainly symmetric\textsuperscript{97,98} and/or in mixture with anionic or glycolipid surfactants,\textsuperscript{97,98} and they concluded that slightly ellipsoidal micelles could form, without going in detail in terms of the molecular distribution within the micelles. If similar works exist on other types of bolaform amphiphiles,\textsuperscript{10,14} the precise description of molecular distribution is always a challenging task. Nagarajan,\textsuperscript{21} basing his line of thought on geometrical considerations only, predicted in 1987 the self assembly behaviour of bola amphiphiles, mainly indicating the formation of spherical, cylindrical and vesicular objects. Interestingly, for spherical micelles he predicted a micellar radius equal to, or smaller, than half the full length of the hydrophobic chain. He also predicted interpenetration of the molecules and, for cylindrical micelles, local disorder at the end-tips in the axial direction of the cylinder. All these features, especially the disordered Region II shown in Figure 7, nicely confirm Nagarajan’s models. Caponetti \textit{et al}.,\textsuperscript{16} are the only ones who deeply discussed the distribution of \textit{N}-azo-18-crown-6 ethers in micelles. Even if they did not provide a tentative image of their system, they discussed some interesting parameters. They used a core-shell ellipsoid of revolution form factor and used a homogeneous shell to fit their data. They found that between 15 and 20 water molecules are contained (bound and free) the hydrophilic
headgroup but none enters the core. They also found that addition of salt had no major effect on the core radius and shell thickness, even if some differences are found between LiCl and NaCl due to the respective difference in terms for complexation by the crown ether macrocycle. More insight on the structure of bolaamphiphile micelles, including molecular conformation, was also provided by Davey\textsuperscript{19} and Shinde,\textsuperscript{20} who worked on asymmetric bolaamphiphiles, alcohol derivatives of alkylammonium and oleate salts, that is ionic/neutral bolas analogous in structure to sophorolipids in the neutral pH regime. Interestingly, both authors agreed on the fact that micelles are spherical/ellipsoidal objects and on the fact both hydrophilic headgroups come in contact with the solvent at the micelle outer palisade, in contrast with older claimings according to which only the ionic headgroup is located at the micelle/solvent frontier while the non-ionic group is settled in the micelle interior, into which water easily penetrates. However, they do not agree on the molecular arrangements: Davey proposes an elongated cross-micellar conformation while Shinde proposes a bent configuration. If compared to the studies presented above, our more complex “coffee bean” like model presented in Figure 7 could be considered as a confluence of Nagarajan’s\textsuperscript{21}, Davey’s\textsuperscript{19} and Shinde’s\textsuperscript{20} models together, in which the end-caps in the axial direction of the micelle are less defined and where the bolaamphiphile can adopt more than one configuration according to several parameters such as its position in the micelle itself, the aggregation number, the presence of salt, the ionic/neutral nature of one of the end-groups, the shell hydration state, etc… In order to verify the micellar model and sophorolipid distribution shown in Figure 7, we have used molecular dynamics simulations shown in the next section.

Figure 7 – Tentative model structure of an acidic sophorolipid micelle at its equilibrium pH (< 5). Region I= full aliphatic; Region II= mixed sophorose/water/COOH/aliphatic; Region III= sophorose/water/COOH. At 5 < pH < 7, the boundaries between the three regions are probably less defined and the COO\textsuperscript{−} groups occupy a broader volume in the hydrophilic region.
Molecular modelling by Molecular Dynamics.

In the following sections, we computed representative parameters (e.g. size, shape, surface properties, SL headgroup hydration and surfactant alkyl chain conformation) to characterize the micelle structure as a function of the SL aggregation number.

1. Aggregation process of the SL monomers into micelle. In Figure 8, we depict the SL aggregation process into a single aggregate for the MD2 simulation with 37 sophorolipids. The simulations results show that, in general, the aggregation process of the SL monomers into a single and stable micelle is relatively fast (~10 – 60 ns) and further analysis (not shown) also suggests that it occurs with two different time scales: one “fast” ($t_{sim} < 10$ ns) and one “slow” ($t_{sim} = 10 – 60$ ns) depending the system size as previously observed for others micellar systems (see, for instance, ref. 99 and references therein). For instance, in case of the MD2 simulation (Figure 8), it takes ~10 ns for that all SL monomers form small clusters with various sizes and shapes and ~55 ns for that the clusters form a single and stable aggregate. In the others systems, the aggregation time depends on the SL numbers and vary between 20 – 80 ns for the system with 27 and 68 SL, respectively. The stability of the aggregates can be also estimated by obtaining a constant radius of gyration, $R_g$, values of the aggregate (see next section) and agree with these values.

Figure 8 - Representative aggregation process of 37 SL monomers into a micelle vs. time for the MD2 run. The outermost and innermost glucose are colored in yellow and red, respectively, whereas the alkyl chain,
double bonds and COOH groups are in grey, cyan and blue colors, respectively. Water is not shown for visual clarity. The black line shows the box limits. The figures were drawn with PyMol.\textsuperscript{100}

II. Micelle shapes and sizes

In Figure S4 in the Supplementary Information, we depicted the final structure of the aggregate as a function of the SL aggregation numbers obtained from the simulations. With the increase of the $N_{agg}$ value, we observe a (quasi)sphere-to–(elongated) ellipsoid transition of the micelle shapes. At high $N_{agg}$ value (i.e. MD6 system), the SL monomers form a structure to a spheroidal rod and this result does not depends on the SL concentration, since we found a similar shape with another simulation (not shown) where SL concentration was set 2 times lower. The micelle shape transition is actually in very good agreement with SANS data obtained on acidic sophorolipids in water as a function of concentration.\textsuperscript{101}

To estimate more precisely the micelle shape changes during the course of the MD1-5 simulations, we employ a three-axis ellipsoid of revolution shape to model the micelles. The ellipsoid of mass $M_T$ is identified by three semi-axis $a_M$, $b_M$ and $c_M$ ($a_M > b_M > c_M$) parameters, and the following equations\textsuperscript{102}:

\begin{align*}
I_1 &= \frac{1}{5} M_T (a_M^2 + b_M^2) \\
I_2 &= \frac{1}{5} M_T (a_M^2 + c_M^2) \\
I_3 &= \frac{1}{5} M_T (b_M^2 + c_M^2)
\end{align*}

where $I_1$, $I_2$ and $I_3$ (with $I_1 > I_2 > I_3$) are the principal moments of inertia of the micelle. Please note that, with respect to the ellipsoid model in Figure 2 used to fit SAXS data, $a_M$ is associated to the axial ($XcR_c+T_P$) and $b_M$, $c_M$ to the equatorial $R_S$ dimensions.

The ratio between the major ($a_M$) and minor ($c_M$) semi-axis for the complete 28 – 80 SL micelle ($a_M$, $b_M$, $c_M$ and $(a/c)_M$) are reported in the 6\textsuperscript{th} column of Table 6. The computed micelle $a_M$, $b_M$ and $c_M$ values confirm that at low $N_{agg}$ the micelle is nearly spherical and that the micelle change to a (pronounced) tri-axial ellipsoid with the increase of $N_{agg}$. At $N_{agg} = 112$ the aggregate form a rod shape oriented along the $y$ axis and we may considered that the upper limit to have micelle is around 80 SL. Moreover, as noted previously\textsuperscript{84} for similar micellar systems and in agreement with our experimental results, the shape of the micellar hydrophobic core is always significantly more ellipsoidal that the overall micelle ($(a/c)_HC$ and $(a/c)_M$ in range of 1.48 – 3.22 and 1.30 – 2.33, respectively).
By subtracting the semi-axis lengths of the aggregate hydrophobic core from those of the whole micelle, one can also obtain an estimation of the average thickness of the polar outer layer, $l_{pl}$ (including the sophorose and the COOH groups, see last column of Table 6), which is nearly equivalent to the shell thickness ($T_S$) estimated from the SAXS data (Figure 2). We have to say that the estimation of the $l_{pl}$ values from the semi-axis is quite rough and strongly depends on the presence of alkyl chain carbon atoms at the aggregate surface (and $a_M$, $b_M$ and $c_M$ values) that can underestimate in the $l_{pl}$ calculations. The average polar layer thickness, $l_{pl}$, values change a little with the micelle shape and are in range of 4.9 – 5.1 Å for the aggregates with 28 - 80 SL monomers. In case of the aggregate with $N_{agg} = 112$, we estimate a $l_{pl}$ value of 3.3 Å with only the $b_M$ and $c_M$ semi-axis. The computed $l_{pl}$ values here are much smaller than the shell thickness values, $T_S$, obtained from the SAXS fit (>11.0 Å, see Table 3). $l_{pl}$ value reflects the average shell thickness, which, according to SAXS data presented above, seems to be highly anisotropic, because $T_S$ varies between ~11 Å (equatorial) and 0 Å (longitudinal), the average between these values, $<T_S>$, being in very good agreement with $l_{pl}$. The computed $l_{pl}$ values, are consistent with those obtained previously for N-dodecyl-β-maltoside (DDM) micelles\(^8\) where we found an average thickness for the α-maltose group (linear β(1→4) disaccharide) in range of 7.2 – 7.7 Å and consequently support the simulation results.

Table 6 - Aggregate size and shape. Values with M and HC subscripts were computed by including all the micelle atoms and those of the hydrophobic core, respectively. The radius of gyration, $R_g^M$, and the ellipsoid three semi-axis lengths (in Å) were computed from the inertia tensor (e.g., see ref. 102). $a_M$ and $c_M$ are the lengths and the radius of the rod, $l_{pl}$ is the average polar layer thickness (including the sophorose and COOH groups) (in Å). The statistical errors (maximum errors) are always lower than 0.3, 0.5, and 0.2 Å for $R_g$, semi-axis lengths, and polar layer thickness, respectively. *In case of the MD6 system, due to its rod shape the $(a/c)_M$ and $(a/c)_{HC}$ values were not computed and the radius of gyration was obtained for a cylindrical rod of $2b_M$ length (equivalent to the box size) and an average radius of $(a_M+c_M)/2$.

<table>
<thead>
<tr>
<th>System</th>
<th>$R_g^M$</th>
<th>$a_M$</th>
<th>$b_M$</th>
<th>$c_M$</th>
<th>$(a/c)_M$</th>
<th>$a_{HC}$</th>
<th>$b_{HC}$</th>
<th>$c_{HC}$</th>
<th>$(a/c)_{HC}$</th>
<th>$l_{pl}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MD1</td>
<td>15.5</td>
<td>22.2</td>
<td>19.9</td>
<td>17.6</td>
<td>1.30</td>
<td>17.3</td>
<td>15.0</td>
<td>12.1</td>
<td>1.48</td>
<td>5.1</td>
</tr>
<tr>
<td>MD2</td>
<td>17.3</td>
<td>27.3</td>
<td>21.2</td>
<td>17.5</td>
<td>1.56</td>
<td>23.0</td>
<td>16.3</td>
<td>12.0</td>
<td>1.95</td>
<td>4.9</td>
</tr>
<tr>
<td>MD3</td>
<td>20.3</td>
<td>34.3</td>
<td>22.2</td>
<td>20.0</td>
<td>1.68</td>
<td>30.9</td>
<td>17.3</td>
<td>14.9</td>
<td>2.09</td>
<td>4.4</td>
</tr>
<tr>
<td>MD4</td>
<td>22.3</td>
<td>38.5</td>
<td>24.0</td>
<td>20.5</td>
<td>1.88</td>
<td>35.7</td>
<td>18.5</td>
<td>14.8</td>
<td>2.41</td>
<td>4.7</td>
</tr>
<tr>
<td>MD5</td>
<td>24.4</td>
<td>44.7</td>
<td>25.0</td>
<td>19.2</td>
<td>2.33</td>
<td>41.7</td>
<td>20.5</td>
<td>12.9</td>
<td>3.22</td>
<td>4.5</td>
</tr>
</tbody>
</table>
To characterize the aggregate sizes, we computed the time evolution of $R_g$ of the 5 micellar systems starting from the beginning of the production runs. As shown in the Figure S5, after large fluctuations of the $R_g$ due to the SL monomers aggregation processes, the $R_g$ values stabilize after 20 – 90 ns depending on the number of SL monomers in the box. We consider that all the SL monomers form a single and stable aggregate when the micelle $R_g$ remain constant until the end of the simulation times. In the 2nd column of Table 6, we reported the average $R_g$ for MD1-5 systems computed from the last 40 ns of each simulation. The values are found between 15.5 – 24.4 Å for the aggregates with 28 and 80, respectively. The values are in relative agreement with data found in this study (18 - 30 Å from the geometrical parameters detailed in Table 3 or computed from the Penfold values36, and according to which the $R_g$ (where $R_g^2 = \frac{1}{5}(a_M^2 + b_M^2 + c_M^2)$) for an ellipsoid of revolution, with $a_M$, $b_M$, $c_M$ being the semi-axes of the ellipsoid obtained from the inertia tensor (see below). In case of the simulation with 112 SL, we obtained a $R_g \approx 23.4$ Å from the expression for a cylindrical rod 

$$\frac{R^2}{2} + \frac{L_{rod}^2}{12},$$

where $L_{rod}$ (here, $2b_M$) and $R$ are average lengths of the rod and its radius (i.e. $(a_M + c_M)/2 = 13.85$ Å), respectively.

### III. Aggregate surface contact properties and hydration

As shown in the MD snapshots depicted Figure S4, the surfactant different parts (i.e. sophorose headgroup and alkyl chain) are in contact with the solvent. To characterize more precisely the surface contacts shared between the aggregates and water, we computed the solvent accessible surface area (SASA) properties with two approaches. In the first one, we used the whole aggregate semi-axis lengths ($a_M$, $b_M$, and $c_M$) determined in the previous section and we calculated the SASA for an ellipsoid with a planar surface, $SASA_{SL}^e$, with the

Knud Thomsen's Formula: $4\pi(a_{sl}^e b_{sl}^e + a_{sl}^e c_{sl}^e + b_{sl}^e c_{sl}^e)\frac{1}{p}$ where $p \approx 1.6075$. Please note that this equation is valid for ellipsoids with three distinct semi-axes, so in case of the aggregate with 112 SL monomers that has a spherical rod shape, the $SASA_{SL}^e$ was estimated from the expression for an open cylinder (i.e. without the lateral areas): $2\pi R L_{rod}$, where $R$ and $L_{rod}$ are the characteristic dimensions of the rod defined just above. In the second approach, we
compute the SASA of the SL part in the aggregate (i.e. the headgroup, the double bond, the alkyl chain and COOH) by splitting the SL molecule into 6 parts (namely, the external and internal glucose ($SASA_{GLA}^e$ and $SASA_{GLB}^e$, respectively), the COOH group ($SASA_{COOH}^e$), the C7 alkyl chains separated by the C=C and bonded to the sophorose and COOH moieties ($SASA_{C7A}^e$ and $SASA_{C7B}^e$, respectively) and finally the two carbon atoms involved in the double bond ($SASA_{C=C}^e$). With these values, one can obtain the total SASA of each aggregate, $SASA_{SL}^e$, with $SASA_{GLA}^e + SASA_{GLB}^e + SASA_{COOH}^e + SASA_{C7A}^e + SASA_{C7B}^e + SASA_{C=C}^e = 100\%$ (Table S5).

These values were computed with the *trjVoronoi* program developed in our group (see ref. 99 and references cited therein). This program uses a Voronoi–Delauney tessellation algorithm to estimate the surface shared between the SL and water atoms (excluding hydrogen). In contrast to the first approach where we assume that aggregate is a planar triaxial ellipsoid, in the latter calculation the micelle surface irregularity is fully taken into account. This is why the computed $SASA_{SL}^e$ values are always greater than $SASA_{SL}^e$ ones. By comparing the $SASA_{SL}^e$ and $SASA_{SL}^e$ values reported in Table S5 and by visual inspection of the simulation snapshots in Figure S4, one can deduce that the aggregate surface is quite rough, as also estimated by the analysis of the SAXS data above. To quantify this, we compute the surface rugosity factor ($f_{rug} = \frac{SASA_{SL}^e}{SASA_{SL}^e}$) and we found that $f_{rug}$ values are always greater than 1.8 and slightly increase with SL aggregation number. For the aggregate with 112 SL monomers, we found a larger $f_{rug}$ value (3.95) probably because we underestimate the $SASA_{SL}^e$ by considering the aggregate shape as a spherical rod. For sake of comparison, the computed $f_{rug}$ values are found higher than those obtained in DDM micelles with similar size (1.6 – 2.0).$^{84}$

From the $SASA_{SL}^e$ values, one can also estimate the area per sophorolipid at the micelle/solvent interface, $A_{SL}^e$, with the expression $A_{SL}^e = \frac{SASA_{SL}^e}{n_{SL}}$. We obtained $A_{SL}^e$ values lying from 177.4 to 132.5 Å² with SL monomers between 28 and 80 in fair agreement (discrepancy estimated between 10% and 30%) with the experimental estimation given in Table 5. In case of the aggregate with 112 SL monomers the $A_{SL}^e$ is significantly lower (62.5 Å²).
Concerning the $\mathcal{S}_{\text{ASAS}}^{V \mu}$ of each SL components, the values for the MD1-6 systems are reported in the 2nd – 7th columns of Table S5. The low SASA values for the alkyl chain and the double bond moieties (~14 and < 1 % of the total aggregate surface, respectively) indicate that these groups are protected from the solvent by SL headgroup or are buried in the hydrophobic core of the aggregates. For the sophorose and the COOH groups, they represent ~78 % and 8 - 9 % of the $\mathcal{S}_{\text{ASAS}}^{V \mu}$ values and tend to slightly decrease with the increase of the SL aggregation number are probably the consequence of the decrease of the micelle curvature and micelle shape changes leading to a different accessibility of the sophorose and COOH solvent to water.

To characterize the aggregate hydration, we have also computed the number of “unique” water ($N_w$) in the first shell (with a cutoff < 4.0 Å)\(^{84}\) of whole aggregate, the sophorose, the COOH alkyl chain and double bond atoms (Table S6). The number of water molecules per sophorolipid is found to be ~27 for the smallest aggregate and it decreases to ~17 for the aggregate with 112. These results are in excellent agreement with the experimental data ($n_w$/SL), varying between 18 and 28 for the headgroup 3, \(h_g\_3\), hypothesis, reported in Table 4. Moreover, the $N_w$ values also confirm (as we noted from the corresponding $\mathcal{S}_{\text{ASAS}}^{V \mu}$ values), that the hydration of the sophorolipid is mainly due to the sophorose headgroup (as also found experimentally in the headgroup 1 hypothesis in Table 4) and at a less extent to COOH since these groups have 16.9 – 24.9 and 2.4 – 4.4 water in their first shells depending on the aggregate size. If small hydration values obtained for the alkyl chain and the C=C bond (3.1 – 5.0 and < 1, respectively), confirming the assumption made above about the accessibility of these groups to water, one should still observe that $N_w$ for the alkyl chain varies between 3 and 5, which are very close estimates to what it is experimentally found in the \(h_g\_3\) hypothesis presented in Table 4. The slight decrease of the sophorose headgroup hydration as a function of the SL aggregation number may be caused by the slight change of the sophorose headgroup conformation and/or interaction between adjacent headgroup at the micelle surface.

To obtain a more precise picture of the SL headgroup hydration as a function of the SL aggregate sizes, we also computed the average number of nearest waters for each sophorose and COOH oxygen atoms (Figure 1) or $< n_{Ox-Ow} >$ from the computed radial pair density functions ($g(Ox-Ow)$) of the sophorose, and COOH ($OC\_1A$ and $OC\_1B$)-oxygen and water-oxygen ($O_w$) atoms (Figure S6). To obtain $< n_{Ox-Ow} >$ values, we integrated the RDF functions until the first minimum at $r \approx 3.5$ Å after the first peak. No defined peaks at ~2.5 Å in the ($O\_1A$, $O\_1B$,
$O_{5A}$ and $O_{5B}$) RDFs indicate that these non-hydroxyl oxygen atoms are poorly hydrated by water in contrast to the sophorose OH oxygen. The computed $<n_{Ox-Ow}>$ values in Table S7 in the Supporting Information these observations. More specifically, the RDFs show that water molecules solvate primarily the hydroxyl oxygen (i.e. $O_{2A}$, $O_{3A}$, $O_{4A}$, $O_{6A}$, $O_{3B}$, $O_{4B}$ and $O_{6B}$) of the sophorose headgroup and the COOH oxygen atoms ($OC_{1A}$ and $OC_{1B}$). The only differences in RDFs between the different systems are for the amplitude of the first peak indicating variations of the number of hydration water molecules in the vicinity these oxygen atoms. The $<n_{Ox-Ow}>$ values for the sophorose hydroxyl oxygen atoms are close (1.8 – 2.4) between the six aggregates in contrast to the carboxyl ones where we observe a slight decrease of the $<n_{Ox-Ow}>$ values with the aggregate size (1.8 – 1.2 and 1.0 – 0.7). The hydrations of the ring oxygen ($O_{5A}$ and $O_{5B}$) of the glucose are around 0.3 and 0.1, respectively. This is close to the value found previously$^{84}$ for the same atoms in maltose (~0.3 water). Finally concerning the oxygen atoms involved in the glycosidic bond ($O_{1A}$) and the link between the headgroup and alkyl chain ($O_{1B}$), the $<n_{Ox-Ow}>$ values (0.6) indicate these atoms are shielded from the solvent by the headgroup, as noted previously.$^{84}$

IV. Conformation of the SL alkyl chain in the aggregates

Finally, we also examined the conformation of the sophorolipid as a function of the SL aggregation numbers by computing the distribution of the end-to-end distance between the first C1 and the C17 atoms, $P(d_{C1-C17})$, of the SL alkyl chain (Figure 1 and Figure S7 in the Supplementary Information). The figure shows a wide distribution of the C1-C17 end-to-end distances (10.5 – 13.5 Å) with the micelle size indicating a wide different conformation of the SL alkyl chain in the aggregates. The $d_{C1-C17}$ are smaller than extended oleic acid chain length estimated to be below 26 Å in the “extended” conformation$^{106,107}$ indicating that the SL alkyl chain are bent with the COOH pointing toward the sophorose headgroup or slightly extended in the micelles. Similar observation was found by Prasad et al.$^{69}$ with acidic (three double bonds) linoleic acid sophorolipids.
Figure 9 – Representative conformation of the sophorolipids in the aggregates with (a): 28, (b) 37, (c) 56, (d) 68 and (e) 80, (f) 112 monomers. Red, blue, green and yellow colours corresponds to the surfactant conformations where the sophorose and the COOH headgroups are located at the micelle surface with the rest of the surfactant in the hydrophobic core (red), the sophorolipid reside entirely at the micelle surface (blue), alkyl chain is extended and cross the hydrophobic core and with the sophorose and headgroups at the surface (magenta), where COOH in the hydrophobic core (green), where the overall sophorolipid are deeply buried in the hydrophobic core (yellow). The figures were drawn with PyMol.  

To illustrate this, we highlighted some representative conformations of the sophorolipids in each aggregate (Figure 9) according to their conformations and localizations and the model depicted in Figure 7. The Figures clearly show that the sophorolipid molecules adopt different conformations and localization in the micelles depending on their sizes. In particular, in the smallest micelle ($N_{agg} = 28$), which has a more spherical shape, there is only one
conformation (red-highlighted molecule in Figure (9(a-f) and region I in Figure 7) for the sophorolipid: the sophorose and the COOH headgroups are located at the micelle surface with the rest of the surfactant partially folded in the hydrophobic core. With the increase of the micelle size, the SL adopt other conformations, for instance, where the all the sophorolipid reside entirely at the micelle surface (blue-highlighted molecule in Figure 9(b-f) and region II in Figure 7), with the COOH in the hydrophobic core (green-highlighted molecule in Figure 9) or where the alkyl chain in an extended conformation that crosses the hydrophobic core and with the sophorose and headgroups at the surface (magenta-highlighted molecule in Figure 9(c-f) and region 3 in Figure 7)). In the aggregates with \(N_{agg} = 68, 80\) and 112, we also see that one sophorolipid where the overall sophorolipid is buried in the hydrophobic core (yellow-highlighted molecule in Figure 9(d-f)). Taking together the simulation results show that for spherical micelles, the bent conformation of SL is predominant, but as soon as the shape turns into an ellipsoid, various conformations can coexist. These results also put in evidence the fact that the micelle is not constituted by a homogeneous arrangement of sophorose at the surface and oleic acid in the core, but rather a more complex distribution of these chemical groups, which is in agreement with the model picture proposed in Figure 7, deduced from experimental data. In particular, simulation data, both conformational and statistical, support the idea of a disordered region, which was identified as region II in Figure 7, in which molecular conformation and water interpenetration occur, thus explaining the complexity of fitting SAXS data using a standard core-shell model and justifying SANS contrast matching data showing that a pure aliphatic core only exists in the inner micellar center.

**Conclusion**

This work shows a detailed analysis of the micellar structure composed of sophorolipids, an asymmetric bolaform glycolipid containing a pH-sensitive COOH group. If the micellar structure was previously reported to be of prolate ellipsoidal shape, we show here how the asymmetric bolaform compound settles within the micelle itself. Detailed modelling of SAXS spectra show an uneven distribution of matter within the ellipsoid. We find that the core-shell model is by far the most adapted to describe the micelle; however, if the hydrophobic core, the size of which is about 8 x 55 Å, actually corresponds to an ellipsoidal shape, the hydrophilic shell has a non-zero thickness (about 12 Å) only in the equatorial region of the micelle. The actual micellar model rather corresponds to a “coffee bean” like shape than an actual homogeneous ellipsoid. Contrast matching SANS was used to probe the actual
distribution of the hydrophobic components of the micelle. The experiment was run at 100% D$_2$O, and at 46% D$_2$O, at which one expects sophorose to be fully contrasted. Interestingly, the hydrophobic core has rather a globular morphology, rather than an elongated ellipsoid. Combining both SAXS and SANS experiments, we make the hypothesis that the ellipsoid of revolution can be divided into three regions: Region I, defining a fully aliphatic core in the micellar center; a less-defined Region II in the axial direction of the ellipsoid and composed of aliphatic, sophorose and COOH groups; Region III, identifying the outer shell composed of sophorose, COOH groups, water and possibly counterions and in which contribution from vicinal CH$_x$ groups should not be excluded. To account for such a complex structure, one must make the hypothesis that several configurations of sophorolipids are actually found within the micelle: both bent and elongated in the peripheral region but also crossing through the micelle itself. This picture is compatible with both the double polar nature of sophorolipids and the bent conformation of its oleic acid aliphatic chain.

To verify this model issued from experimental data, we carried out explicit MD simulations with 28, 37, 56, 68, 80 and 112 sophorolipid monomers. The simulation results show, in agreement, with scattering experiments, the micelle shape changes from a (quasi)sphere to an elongated ellipsoid that with the increase of the $N_{agg}$ value. At high $N_{agg}$ value (here, $N_{agg}$ = 112 SL), the aggregate forms a cylindrical rod, as previously reported in the literature for higher volume fractions. Moreover, all the aggregates present rough surface with the sophorose and COOH headgroups in direct contact with the solvent leading to a larger hydration of these groups compared to the alkyl chain moiety. Examination of the sophorolipid conformations also shows that when the micelle is small and nearly spherical (here with $N_{agg}$ = 28), the bent conformation of SL is predominant where the headgroup the sophorose and the COOH are located at the micelle surface with the alkyl chain partially folded in the hydrophobic core. With the increase of the micelle size, others SL conformations in the aggregate can coexist leading to a complex distribution of sophorolipids in the micelle surface and in a hydrophobic core, also in the agreement, with the model of structure deduced from the scattering experiments.

Finally, we have also investigated the sophorolipid configuration upon small increase in pH (< 7), when repulsive intermicellar interactions, due to the formation of COO$^-$ groups, characterize the small angle scattering spectra and, consequently, the micellar network. To do so, we have employed anomalous SAXS, consisting into analyzing the sample using an X-ray incident energy close to the X-ray absorption edge of a given element, in this case Rb$^+$ (up to 100 mM) and Sr$^{2+}$ (up to 20 mM) counterions, used to neutralize the negative COO$^-$ charges.
Contrarily to our expectations, the ASAXS effect near the edge was very mild, if actually non-existent, at concentrations in counterions at which it is supposed to be very strong, if we compared to ionic surfactant systems. We interpret these results by making the hypothesis that the negatively-charged COO− groups do not preferably distribute at the outer micellar/water palisade but rather throughout the micellar hydrophilic shell volume, thus reducing the scattering effect of the counterions alone.

Acknowledgements

The research leading to these results has received funding from the European Community’s Seventh Framework Programme (FP7/2007-2013) under Grant Agreement n° Biosurfing/289219. This work was, in part, granted access to the HPC resources of CCRT/CINES under the allocation x2014077322 (project number: gen7322) made by GENCI (Grand Equipement National de Calcul Intensif). SOLEIL Synchrotron (Saint Aubin, France) is kindly acknowledged for providing access to the SWING beamline.

Supporting Information Available

Page S1: Solvent SLD for all salt compositions. Page S2: Critical determination of the best fit model. Page S4: Fit parameters obtained for X_S = 0.3. Page S5: Molecular area values obtained for X_S = 0.3. Page S6: Experimental approach and data treatment in ASAXS experiments. Page S8: Experimental determination of the ionization degree. Page S9: Additional data obtained from molecular modelling by Molecular Dynamics. This information is available free of charge via the Internet at http://pubs.acs.org

References

1 Tanford, C. The hydrophobic effect; Wiley: New York, 1973
Pure Solution and Destabilizing Effects in Mixtures of Glycolipids, Bolaform Surfactants, . Chem. Comm
Intermediate Properties between Crystalline Nanofibers and Wormlike Micelles: Formation of Viscoelastic
64 Value obtained from the SciFinder® database
66 http://didier.lairez.fr/pasinet2/doku.php
73 Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F. Intermolecular Forces; Pullman, B. Ed.: Reidel
74 Dodrecht, 1981; p. 331


