# From lab to market: An integrated bioprocess design approach for new-to-nature biosurfactants produced by Starmerella bombicola

Lisa Van Renterghem, Sophie Roelants, Niki Baccile, Katrijn Uyttersprot, Marie Claire Taelman, Bernd Everaert, Stein Mincke, Sam Ledegen, Sam Debrouwer, Kristel Scholtens, et al.

**From lab to market: An integrated bioprocess design approach for new-to-nature biosurfactants produced by *Starmerella bombicola*[†]**

**Short running title:** New biosurfactants produced by *S. bombicola*

**Lisa Van Renterghem[1], Sophie L.K.W. Roelants[1,2], Niki Baccile[3], Katrijn Uyttersprot[4], Marie Claire Taelman[4], Bernd Everaert[2], Stein Mincke[5], Sam Ledegen[1], Sam Debrouwer[2], Kristel Scholtens[2], Christian Stevens[5] and Wim Soetaert[1]**

Ghent University, Centre for Industrial Biotechnology and Biocatalysis (InBio.be), Coupure Links 653, Ghent, Oost-Vlaanderen, BE 9000

Bio Base Europe Pilot Plant, Rodenhuizekaai 1, Ghent, Oost-Vlaanderen, BE 9042

Sorbonne Universités, UPMC Univ Paris 06, CNRS, Collège de France, Chimie de la Matière Condensée de Paris (UMR 7574), 4 Place Jussieu, Paris, Île-de-France, FR 75005

EOC Group, De Bruwaan 12, Oudenaarde, Oost-Vlaanderen, BE 9700

Ghent University, Sustainable Organic Chemistry and Technology, Coupure Links 653

Gent, Oost-Vlaanderen, BE 9000

Corresponding author:

Dr. ir. Sophie Roelants

Centre for Industrial Biotechnology and Biocatalysis (InBio.be), Coupure Links 653, Ghent, BE 9000

sophie.roelants@ugent.be

+329 264 6027

## Abstract

Glycolipid microbial biosurfactants, like sophorolipids (SLs), generate high industrial interest as 100% biobased alternatives for traditional surfactants. A well-known success story is the efficient SL producer *Starmerella bombicola*, which reaches titers well above 200 g/L. Recent engineering attempts have enabled the production of completely new types of molecules by *S. bombicola*, like the 'non-symmetrical bolaform (nsBola) SLs'. As classic SLs are mostly applied in eco-friendly detergents, the possible use of these bolaform SLs in detergent applications was evaluated by scaling up the production process (150 L) and evaluating the purified product. This paper shows that they can be used as green and non-irritant surfactants in for example (automatic) dishwasher applications. However, the limited chemical stability at higher pH values (> 6.5), due to the presence of an ester function in the biosurfactant molecule, is a major drawback that will most likely inhibit market introduction. An integrated bioprocess design was thus applied to resolve this issue. The strategy was to replace the fed fatty acids, responsible for the ester bond in nsBola SLs, with fatty alcohols, to generate so-called 'symmetrical bolaform (sBola) SLs', containing two instead of one glycosidic bond. This requires a change in the feeding strategy, but also the blocking the fatty alcohols from metabolizing/oxidizing through the suggested ω-oxidation pathway. Two putative fatty alcohol oxidase genes (*fao1* and *fao2*) were identified in the *S. bombicola* genome and deleted in the nsBola SL producing strain (Δ*at*Δ*sble*). Shake flask experiments for these new strains (Δ*at*Δ*sble*Δ*foa1 and* Δ*at*Δ*sble*Δ*foa2*) were performed to evaluate if the fed fatty alcohols were directly implemented into the SL biosynthesis pathway. Indeed, sBola SL production up to 20 g/L was observed for the Δ*at*Δ*sble*Δ*fao1* strain, while the Δ*at*Δ*sble*Δ*fao2* strain only produced nsBola SLs. The sBola SLs were purified and their symmetrical structure was confirmed by NMR. They were found to be significantly more stable at higher pH, opening up the application potential of the biosurfactant by enhancing its stability properties. 

## Keywords

*Starmerella*, biosurfactant, sophorolipid, fermentation and purification, strain engineering, application testing

## Introduction

Surfactants are important high-performance molecules, whose worldwide annual production amounts to more than 17 million tons today (Allied, 2016). About half of this production volume is used for household and laundry detergents, while the other half is employed in various industries; e.g. chemicals, mining, textile and paper, construction, food, pharmaceuticals, cosmetics etc. (Nitschke et al., 2007; Savarino et al., 2009; Sleiman et al., 2009). An emerging class of surfactants are the so-called biosurfactants, especially the ones produced by microorganisms. Sophorolipids are a well-known example of glycolipid microbial biosurfactants, and are composed of a disaccharide sophorose attached to a fatty acid chain (Fig. 1A-B). SLs are typically produced by certain yeast species of the *Starmerella* clade, for example *Starmerella bombicola* (Spencer et al., 1970). The latter non-pathogenic yeast displays a substantial industrial interest as it naturally produces very high amounts of SLs (> 200 g/L) (Daniel et al. 1998b; Gao et al., 2013). In the fermentation process, natural building blocks such as sugars and plant oils (Daniel et al., 1998a; Davila et al., 1997) or even waste or side streams (Daverey et al., 2011; Savarino et al., 2009) are employed. In this way, SLs (and other microbial biosurfactants) offer a renewable and 100 % biobased alternative to the traditional (petrochemically produced) surfactants (Mukherjee et al., 2006). This ecological advantage, combined with the rising awareness towards sustainability, clearly underpins the market potential of biosurfactants.

As over 20 major homologs are present in the wild type mixture of SLs, this makes it a very complex system to study. Therefore, the past years a lot of research has been performed on strain engineering of *S. bombicola* to standardize/uniformize the SL production (Saerens et al., 2011b; Roelants et al., 2016) or to reorient SL production towards completely new glycolipids (Saerens et al., 2011a; Roelants et al., 2013; Van Bogaert et al., 2016). One of these newly developed *S. bombicola* strains produces so-called “bolaform sophorolipids”. In contrast to the

structure of classic SLs (Fig. 1A-B), bolaform SLs consist of two sophorose units located on each side of the lipophilic alkyl chain (Fig. 1C). They were first discovered in minute amounts in the wild type SL mixture by Price et al. (2012). Shortly after this publication, Soetaert et al. (2013) succeeded in generating a strain that almost exclusively produces these bolaform SLs by deleting specific genes of the SL production pathway (Fig. S3, Van Bogaert et al., 2016).

Due to the unique structure of bolaamphiphiles, namely two identical hydrophilic head groups connected to the ends of a hydrophobic linker, they are promising for a range of applications. Bolaamphiphiles form monolayer membranes that generally possess less permeability compared to regular polar lipids (Fuhrhop and Wang, 2004; Puri et al., 2009), making them very efficient molecules for gene and drug delivery (Farija et al., 2015). For example, transport of the anti-HIV drug Zidovudine® to target organs (liver/spleen) is significantly improved by incorporating it as a bolaampohiphilic prodrug (Jin et al., 2010). Some other bolaamphiphiles act as carriers for vitamin B12 or C (Ambrosi et al., 2010). The -until today- sole-discovered natural bolaamphiphiles are found in the membranes of *archae* bacteria, enabling the bacteria to survive under extreme conditions. Unfortunately, these lipids are almost impossible to extract (< 4% yield), making them not (yet) industrially relevant (Chong 2010). Consequently, only synthetic bolaamphiphiles are currently commercialized. Bolaform SLs could represent an interesting biological alternative, possibly with different/additional functionality. However, besides such rather 'high-end' applications, the current market for industrially produced SLs is > 90 % dominated by its use in detergent applications. For this reason, this was the first field of application we chose for evaluation of the new bolaform SLs.

In this paper, the production and purification of bolaform SLs was investigated and scaled up. The functionality of the purified 'non-symmetrical bolaform (nsBola) SLs' was evaluated in a range of performance tests for detergent applications. These application tests showed that their limited chemical stability seriously hinders their market potential as green and non-irritant

surfactant for, amongst others, the cleaning industry. Therefore, further strain and process engineering was performed to overcome this limitation. The resulting 'symmetrical bolaform (sBola) SLs' were successfully purified and their putative structure was confirmed by NMR.

## Materials and Methods

### Strains and culture conditions used for glycolipid production

For the nsBola SL production, a *S. bombicola* strain deficient in its acetyltransferase (*Δat*) and lactone esterase (*Δsble*) genes was used (Soetaert et al., 2013, Van Bogaert et al., 2016). The 150 L scale fermentation using the *ΔatΔsble* strain was performed as described before for other *S. bombicola* strains (Baccile et al., 2017). During the fermentation, both the hydrophilic (glucose) and hydrophobic substrates (high oleic sunflower oil, HOSO) were added in batch. To control foam formation in the exponential growth phase and to stimulate SL production, 8.3 g/L of HOSO was present in the medium at inoculation. Glucose (60 wt %) was fed once when its concentration dropped below 60 g/L. HOSO was fed in daily shots of 10 g/L the first three days, and 5 g/L per day for the remainder of the fermentation, to avoid accumulation of oil.

*S. bombicola* fatty alcohol oxidase (*fao1* or *fao2*) knockout strains were obtained by integrating the *S. bombicola ura3* gene (Van Bogaert et al., 2008) under the regulatory control of its own promotor and terminator at the respective *fao1* or *fao2* loci in the *ΔatΔsble* strain (Van Bogaert et al. 2016). A detailed description of the strain construction is described in the Supplementary Materials. Three transformant colonies of each new strain were evaluated in terms of growth and production. An adapted version of the production medium described by Lang et al. (2000) was used in all experiments (glucose: 150 g/L; yeast extract: 4 g/L; sodium citrate tribasic dihydrate: 5 g/L; $NH_4Cl$: 1.5 g/L; $KH_2PO4$: 1 g/L; $K_2HPO4$: 0.16 g/L; $MgSO_4.7H_2O$: 0.7 g/L; NaCl: 0.5 g/L; $CaCl_2.2H_2O$: 0.27 g/L). For shake flask experiments, 5 mL precultures were set up for 24 h (30°C at 200 rpm), before transferring to shake flask level (4 % inoculation) at 30°C

and 200 rpm. After 48 h, 18 g/L 1-hexadecanol (C16:0-OH) or oleyl alcohol (C18:1-OH) was added. The growth experiments were stopped when glucose was depleted. Two volumes of ethanol were added to the broth and the resulting SL mixture was analyzed on HPLC-ELSD/UPLC-ELSD and LC-MS after removal of cellular debris (4500 rpm, 20 min) (for more information, see Analytical Techniques).

**Downstream processing (DSP) of SLs**

From the performed 150 L fermentation, the nsBola SLs were purified at small pilot scale as described by Baccile et al. (2017), where a different *S. bombicola* strain exclusively producing acidic SLs was employed. More precisely, ceramic microfiltration (0.45 µm) was applied to remove the yeast cells followed by a two-step ultrafiltration (50 and 2 kDa, respectively) to separate the bolaform SLs from water soluble impurities like residual sugar, salt, proteins etc. A final freeze-drying step was applied on the retentate of the second ultrafiltration to obtain a dry powder. As such, nsBola SLs were obtained, in which the two sophorose units are connected through a C18:1 fatty acid linker, originating from the fed substrate HOSO (mainly composed out of oleic acid). These molecules will be further referred to as (C18:1) nsBola SLs (Fig. 1C).

The sBola SLs were purified from the performed shake flask experiments. First, the yeast cells were removed from the ethanol/water mixture by centrifugation (20 min at 4500 rpm), and ethanol was evaporated using a rotary evaporator (40°C, <100 mBar) until only the water fraction remained. An alkaline hydrolysis (pH 12, 5 M NaOH, 37°C, 1 h) was performed to hydrolyze residual nsBola SLs. Afterwards, the pH was adjusted to its original value of 4 (5M HCl). Ultrafiltation was performed on lab scale as described by Roelants et al. (2016) to remove all residual hydrophilic impurities (salts, proteins, residual sugars) similarly as was done for nsBola SLs. Finally, lyophilization was applied. If necessary, preparative liquid chromatography (PLC) was applied to selectively purify the sBola SLs for NMR analysis (similar to the Thin Layer Chromatography (TLC) methodology described by Asmer et al.

(1988) in the Analytical Techniques). Depending on the fed substrate, the alkyl chain connecting the sophorose units was either C16:0 (derived from hexadecanol) or C18:1 (derived from oleyl alcohol), respectively further referred to as C16:0 and C18:1 sBola SLs (Fig. 1D and 1E, respectively).

**Analytical techniques**

*Monitoring of growth, glucose consumption, substrate concentration and SL production*

Growth (optical density (OD), colony forming units (CFU) and CDW (cell dry weight)) and glucose concentration were determined as described by Saerens et al. (2011b). Additionally, glucose consumption was followed up by using Ultra Performance Liquid Chromatography (Waters Acquity H-Class UPLC), coupled with an Evaporative Light Scattering Detector (Waters Acquity ELSD Detector) (UPLC-ELSD). An Acquity UPLC BEH Amide column (130 Å, 1.7µm, 2.1 x 100 mm) (Waters) was used at 35°C and an isocratic flow rate of 0.5 mL/min of 75% acetonitrile and 0.2% triethylamine (TEA) was applied (5 min/sample). For the ELS detection, the nebulizer was cooled to 15°C and the drift tube was kept at a temperature of 50°C. The linear range was between 0 and 5 g/L glucose, using a gain of 100 for ELS detection (Empower software).

Gas Chromatography with Flame Ionization Detection (GC-FID) (Trace GC 2000 series, Thermo Quest) as described by Roelants et al. (2016) was used to determine the residual fatty alcohol after purification. To this end an extraction was performed with a diethyl ether/n-hexane mixture (1:1). An internal standard of 0.1% n-hexadecanol was used.

Thin Layer Chromatography (TLC) was performed to easily follow-up residual substrate and SL production using the chloroform/methanol/water eluent (65/15/2, v/v/v) (Asmer et al., 1988). 1 µL of broth or standard was spotted on silica-coated aluminum TLC plates (Silica gel

60 $F_{254}$, 20 cm x 20 cm, VWR) and dried. After elution, the TLC plate was dried and submerged in a 10% $H_2SO_4$ solution. Spots were visualized using a heat gun.

Samples for SL analysis were prepared for Liquid Chromatography – Mass Spectrometry (LC-MS) and High Pressure Liquid Chromatography – Evaporative Light Scattering Detector (HPLC-ELSD) analysis as described by Saerens et al. (2011a) or UPLC-ELSD analysis as described below. An Acquity UPLC CSH C18 column (130Å, 1.7 µm, 2.1 mm x 50 mm) (Waters) and a gradient elution system based on 0.5% acetic acid in milliQ (A) and 100% acetonitrile (B) at a flow rate of 0.6 mL/min was used as follows: the initial concentration of 5% acetonitrile increases linearly until 95% during the first 6.8 min and then linearly decreases again to 5% during 1.8 min. Subsequently, 5% acetonitrile is maintained until the end of the run (10 min/sample). For the ELSD detection, the nebulizer was cooled until 12°C and the drift tube was kept at a temperature of 50°C, the gain was set to 200. To quantify the glycolipids, a serial dilution of purified product was used as an external standard. First, a 10 g/L solution was prepared, which was diluted two-fold in a stepwise manner, until at least five different solutions were obtained (5, 2.5, 1.25, 0.66, 0.33, 0.165 g/L, respectively).

*NMR analysis and structure characterization of new-to-nature glycolipids*

All $^{1}H$ and $^{13}C$ NMR spectra were recorded at 400 and 100.6 MHz, respectively, on a Bruker Avance III, equipped with $^{1}H$/BB z-gradient probe (BBO, 5 mm). DMSO-[$D_6$] was used as solvent, and as internal chemical shift standard (2.50 ppm for $^{1}H$ and 39.52 ppm for $^{13}C$). All spectra were processed using TOPSPIN 3.2. APT, $^{13}C$, COSY and HSQC spectra were acquired through the standard sequences available in the Bruker pulse program library. Custom settings were used for HMBC (32 scans), TOCSY (100 ms MLEV spinlock, 0.1 s mixing time, 1.27 s relaxation delay, 16 scans) and H2BC (21.8 ms mixing time, 1.5 s relaxation delay, 16 scans), according to literature (Petersen et al., 2006; Gheysen et al., 2008).

**Characterization and application test results of nsBola SLs**

*Solubility, CMC and surface tension determination*

Solubility in different solvents was assessed in a step-wise manner for nsBola SLs (batch B01, see Table SI), sBola SLs and compared with non-acetylated acidic SLs as a reference (batch T37, see Table SI). Heating up to 60°C (max.10 min) was performed to stimulate solubility of the compounds in water. The determination of the surface tension and critical micelle concentrations (CMCs) for the nsBola SLs was performed as described by Roelants et al. (2016).

*Chemical stability evaluation*

The stability of nsBola SLs in water in function of pH (2, 4, 7 and 10), temperature (4, 21, 37 and 50°C) and time (3 and 24 h, 1 week, 1, 3 and 6 months) was determined. The solutions were buffered in order to remain at constant pH, as possible degradation of the tested SLs could result in a drop in pH-value, and this is of course unwanted. More precisely, the pH 2 buffer was made by combining 50 mL of 0.4 M KCl and 13 mL of 0.4 M HCl, the pH 4 buffer by combining 33 mL 0.2 M citric acid and 17 mL 0.2 M sodium citrate, the pH 6 buffer by combining 100 mL 0.2 M $KH_2PO_4$ and 11.2 mL 0.2 M NaOH and the pH 10 buffer by combining 100 mL 0.2 M $NaHCO_3$ and 21.4 mL 0.4 M NaOH. All the buffers were sporadically tested over the course of the experiment to assess if the value was maintained. Fenoxyethanol (1%) was added to avoid microbial contamination. The rather well described lactonic SLs (batch T43, see Table SI) were included to compare the characteristics of the new bolaform compounds.

To compare the stability of the nsBola SL to the sBola SL compounds, an alkaline hydrolysis (pH 12, 5 M NaOH, 1 h, 37°C) was performed. After pH adjustment to its original value 4 (5 M HCl), UPLC-ELSD was performed to investigate the possible breakdown of (n)sBola SLs.

*Supramolecular assembly: Small Angle X-ray Scattering (SAXS)*

SAXS experiments were performed on the D02 and BM29 beamlines, at two different periods of the year, at the ESRF Synchrotron Facility (Grenoble, France). In both beamlines, the setup was optimized to record a signal in the $0.01 < q\ (nm^{-1}) < 10$ range. Experiments were reproduced on the two different beamlines, and no apparent variation occurs among the data. Beam energy and sample-to-detector distance were 12.5 keV and 2.9 m on BM29 and 12.6 keV and 1 m ID02. Acquisition time was 1 s and between 3 and 10 spectra-per-sample are generally acquired and averaged. Water and capillary are systematically measured before each experiment and subtracted. Samples are corrected for transmission and intensity is normalized on the signal of water ($0.016\ cm^{-1}$) to obtain an absolute scale (Schnablegger & Singh, 2017). Based on studies on acidic SLs (Baccile et al., 2012), the SAXS data were fitted using a standerd core-shell sphere form factor model using Sasview 3.1.2 software, where the fatty acid represents the core and sophorose the shell. The fit was performed by setting the volume fraction, the solvent scattering length density (SLD) at $9.4 \times 10^{-4}\ nm^{-2}$, the core SLD at $8.6 \times 10^{-4}\ nm^{-2}$. The latter is a value evaluated for oleic acid, corresponding to the core of the nsBola SL. The shell thickness was assumed to be homogeneous. The variable parameters are the shell SLD, the core radius and shell thickness. At high volume fractions, a typical hard sphere potential structure factor was used in the model. The hard sphere radius was fixed at $1.75 \pm 0.5$ nm.

*Foaming, wetting, emulsification, irritant, film forming and degreasing properties of nsBola SLs*

To assess possible applications of nsBola SLs in household applications, different tests were set up. Foaming properties were assessed using a Dynamic Foam Analyzer (DFA, Krüss) using 0,005 % solutions and results were expressed in average foam height (mm). The wetting properties were assessed using the Draves wetting test (ISO8022). The film forming properties were evaluated by measuring the static contact angle between a dried glass plate (immersed in

a 0.1 % surfactant solution and dried) and water. The *in vitro* Red Blood Cell (RBC) test (INVITTOX37/99) was performed to assess possible irritation of the molecules. The degreasing properties were assessed using a dish washing liquid formulation (Table IA) to degrease fat treated plates with known weight. The plates treated with a fat mixture were put into a beaker containing 0.05 % of the dish washing liquid in water (40°C, 1 min, 160 rpm) and the weight of the plates was determined after drying. The rinse aid properties were assessed using the formulation represented in Table IB. The formulation solutions were added to hard tap water, containing 0.5 g/L $CaCl_2$ and a blue dye, at a temperature of 50°C. Glass plates were put into the solution (1 min) and left to dry. Visual observation of the amount of droplet deposition onto the plates on the one hand, and quantification of the rinse aid performance on the other hand was performed. The latter was done by placing the plates into a certain volume of water to wash off the remaining dye and determining the absorbance. Besides the in-house batches of nsBola SLs, acidic and lactonic SLs (Roelants et al., 2016), a commercially available SL product called 'Sophoclean' (Soliance) (a mixture of lactonic and acidic SLs) was included as reference. In Table SI, the composition of the biosurfactant samples is represented. Additionally, other commercial surfactants were also included, for example SLES (sodium lauryl ether sulfate, Kao), but also chemically-derived alkyl polyglucosides (APGs): APG215 (C8-C16, BASF), APG425 (C8-C10, BASF).

## Results and discussion

### Production of non-symmetrical bolaform (nsBola) SLs

As mentioned in the introduction, Soetaert et al. (2013) generated a *S. bombicola* (*ΔatΔsble*) strain mainly producing new types of SLs, called bolaform SLs (Fig. 1C and S3). As these new bolaform SLs contain twice the amount of glucose per molecule compared to the wild type SLs (Fig. 1A-B), the optimal glucose concentration was first determined. Concentrations of 80, 100,

120, 150 and 160 g/L glucose were added to the production medium described by Lang et al. (2000) and SL production was compared. A glucose concentration of 150 g/L was found to be optimal for bolaform SL production, whereas 120 g/L was optimal for the wild type (results not presented). With this adapted production medium, the first fermentation was performed at 7 L scale and subsequently scaled-up to 150 L scale. An overview of the typical fermentation parameters is given in Fig. 2. NsBola SLs were produced up to a titer of 63 g/L, while its precursor, the non-acetylated acidic SLs, were produced up to 3 g/L. The latter compounds were mainly secreted during the exponential growth/early stationary growth phase, as the SL biosynthetic pathway is not fully induced yet. The overall productivity of the *ΔatΔsble* strain was 0.22 g/L.h of nsBola SLs. This value can definitely be improved in future research by means of process optimization, which is currently ongoing. The entire batch of nsBola SLs was purified similarly as described for acetylated acidic SLs (Baccile et al., 2017), using a microfiltration (0.45 micron) and a two-step diafiltration (50 and 2 kDa) (Fig. 3). The overall recovery yield and final degree of purity were equal to 65 % and 95 % respectively. The largest loss of product was observed in the 50 kDa ultrafiltration step (75 % yield), and this will be further improved towards industrial production, as yields of > 95 % in the different DSP unit operations should be put forward. Finally, 1.2 kg of purified nsBola SLs was freeze dried for characterization and application testing purposes.

**Characterization and application performance tests for nsBola SLs**

After obtaining this fairly large batch of nsBola SLs, properties (CMC, surface tension, supramolecular behavior, solubility etc.) were determined and compared to (bio)surfactant references. As can be concluded from Fig. 4A, the obtained CMC value for the nsBola SL is in line with values obtained for the traditional surfactant SLES. However, the CMC value of bolaform SLs is 2 and 6 times higher as that of the biosurfactants APG215 and commercial SL product Sophoclean, respectively. The solubility of bolaform SLs in water is very high and in

line with non-acetylated acidic SLs, they are soluble in concentrations over 500 g/L (see Table SII), in contrast to the poorly water soluble lactonic SLs. On the contrary, the solubility of the bolaform SLs in hydrophobic solvents like ethanol or ethyl acetate is very low, pointing towards their potential use in water-based formulations.

The supramolecular behavior of the nsBola SLs was investigated using SAXS and detailed information can be found in the Supplementary Materials. The concentration dependent profile is represented in Fig. S1. In summary, the bolaform SLs in water form a system of highly stable spherical micelles with an average radius of 1.79 ± 0.02 nm, almost independent of concentration. In fact, the bolaform SLs act rather as “frozen” rather than dynamic colloids. The latter is classically found in micellar systems, where an equilibrium is always present between the self-assembled particle and the free molecule. In classical surfactant systems, both the aggregation number and the micellar shape generally evolve with concentration. This does not seem to be the case here, as the spherical shape does not vary at any concentration value explored: concentration only seems to pack the existing micelles in solution to a higher volume fraction. This is quite surprising as very different behavior was observed for other types of SLs (Baccile et al., 2010; Baccile et al., 2012; Baccile et al., 2017).

As wild type SLs are mostly applied in household applications, e.g. in the ecological cleaning solutions of Ecover and Henkel, the performance/relevant properties of nsBola SLs for such applications were determined. Besides CMC values, the foaming, wetting, film forming, irritant, degreasing, rinse aid and anti-scaling properties are presented in Fig. 4. Low foaming potential, displayed by the nsBola SLs, Sophoclean and APG215 (Fig. 4B), is required for certain applications like automatic dishwasher applications. Next, bolaform SLs and acidic SLs display bad wetting properties (Fig. 4C), in contrast to the commercial SL product (Sophoclean). The presence of rapeseed methyl esters (RME) in the latter was shown to be at the base of this observation, as the addition of 0.1% of RME to the other types of SLs also

dramatically increased their wetting properties to similar levels as the Sophoclean product (results not shown). Good film forming properties are associated with the bolaform SLs (Fig. 4D), in line with the commercial biosurfactant samples. When looking at the irritant potential in Fig. 4E, bolaform SLs and acidic SLs can be classified as non-irritant, and score better in comparison to lactonic SLs or commercial surfactants like SLES. Moreover, bolaform SLs were found to be completely associated with the water phase (100 % in the water phase in an octanol/water system), indicating that these compounds display no bioaccumulation potential. The best results for nsBola SLs were obtained for grease removal (Fig. 4F) and rinse aid drying performance (Fig. 4G). As lower absorbance values correspond to a smaller amount of deposited dye on the glass plates, the decreased scale deposition on glass for bolaform SLs is shown compared to Sophoclean in Fig. 4H. For this last property, the nsBola SLs even scored better as a commercial rinse aid.

In conclusion, the nsBola SLs are non-irritant, have a good spreading onto the surface, show excellent grease removal properties and an improved drying performance. Combined, these properties are promising for use in household applications, such as automatic dishwashing or window cleaning. Bolaform SLs score well in comparison to commercial ‘green’ biosurfactants such as APGs and Sophoclean, but show inferior performance compared to classical surfactants (SLES) for a number of properties.

An important parameter for the application of a new (bio)chemical in a commercial product is its chemical stability, as this determines its shelf-life and functional stability. Therefore, this parameter was evaluated in terms of pH and temperature, and the results of the nsBola SLs and lactonic SLs are shown in Fig. 5 and Fig. S2, respectively. Unfortunately, the stability of the nsBola SLs is rather low at higher pH values (pH > 7). This can be a limitation when applying the molecules in formulations at higher pH values, like automatic dishwashing. The low stability at high pH values is caused by the presence of the ester bound (Fig. 1C) connecting

the second sophorose molecule to the acidic SL backbone. By applying an integrated bioprocess design, i.e. redesigning both the strain and the fed substrate, symmetrical bolaform SLs (sBola SLs, Fig. 1D-E) could be produced, which would alleviate this problem.

**Integrated bioprocess design approach: production of symmetrical bolaform (sBola) SLs**

As mentioned above, the ester bound of nsBola SLs gives rise to rather low chemical stability of the compound (Fig. 1C). As this ester function is derived from HOSO (i.e. fatty acids) as hydrophobic substrate, the feeding of the latter should be omitted to avoid the presence of an ester bound in the final molecule. Feeding the yeast with diols could be a possibility, however, no long chain α,ω-diols (C12 or longer) are readily commercially available. Another viable option is to feed the *ΔatΔsble* strain with fatty alcohols, as was done for the *S. bombicola* wild type by Brakemeier et al., 1998 (Fig. S4). If the CYP52M1 enzyme of *S. bombicola* (Van Bogaert et al., 2013) is be able to (sub)terminally hydroxylate the fatty alcohol, this would give rise to the corresponding α, ω or ω-1 diols *in situ*, which could as such be implemented into the SL production pathway. The latter would give rise to the production of sBola SLs. However, the latter proves to be not so straightforward, due to the presence of fatty alcohol oxidases (FAO) in *S. bombicola* (Hommel and Ratledge, 1990). The fed fatty alcohols will be (partially) oxidized to the corresponding fatty aldehydes and further on to fatty acids in the suggested ω-oxidation pathway (proven for *Candida maltosa* or *Candida tropicalis* by Cheng et al., 2005). Consequently, these fatty acids will be incorporated in the *ΔatΔsble S. bombicola* production pathway, giving rise to a mixture largely consisting of nsBola SLs (results not shown), which is unwanted. To tackle this, and to stimulate production towards the sBola SLs, the deletion of the above-mentioned *fao* genes of *S. bombicola* can be a viable option. Similar knockouts have already been performed in *C. maltosa* and *C. tropicalis,* and these strains still showed good viability (Cheng et al., 2005; Eirich et al., 2004).

Two putative FAO were identified from an in-house genomic database of *S. bombicola*. The first one (*fao1)* (GenBank: AB907775) was already proven to have alcohol oxidase activity by Takahashi et al. (2016). When this gene was blasted against the available *S. bombicola* genome, a second putative fatty alcohol oxidase was found, with 32 % identity to the FAO1 enzyme. This one could correspond to the one predicted by Hommel and Ratledge (1990), and will be further referred to as the (putative) *fao2* gene (GenBank MF431618).

The creation of the *Δfao* deletion strains was performed as shown in Fig. S5. The knockouts were generated using the *ura3* marker gene in the *S. bombicola* Δ*at*Δ*sble* strain background, after making this strain *ura3* negative again (Roelants et al., 2017). A detailed description of the strain construction is described in the Supplementary Materials.

An overview of the different strains generated in this work and some general characteristics are shown in Table II. Three successful colonies were randomly chosen of each *fao* knockout and their fitness and SL production was evaluated. After 48h, 18 g/L 1-hexadecanol (C16:0-OH) or oleyl alcohol (C18:1-OH) was added as hydrophobic substrate. Detailed results are presented in Fig. S6. The three colonies of both knockout strains did not show any distinct difference in terms of growth compared to the *ΔatΔsble* parental strain, as the obtained CFU values were similar (Fig. S6). However, the average glucose consumption for the *ΔatΔsbleΔfao1* strain was 0.42 g/L.h, which is significantly lower compared to the parental *ΔatΔsble* and *ΔatΔsbleΔfao2* strains (around 0.75 g/L.h). UPLC-ELSD and LC-MS analyses of the broths of the *ΔatΔsble, ΔatΔsbleΔfao1* and *ΔatΔsbleΔfao2* strains fed with hexadecanol or oleyl alcohol were performed (see Table SIV for an overview of all possible molecular masses). Molecular masses corresponding to sBola SLs (with C16:0 or C18:1 hydrophobic linker, respective masses 906 and 932, depending on the fed substrate) were indeed detected in the broth of the new *ΔatΔsbleΔfao1* strain. For the latter strain, the amount of produced nsBola SLs was roughly 3 times lower when hexadecanol was fed in comparison with oleyl alcohol. NsBola SLs were

detected in all fermentation broths. This can be explained by the fact that the CYP52M1 enzyme will also hydroxylate *de novo* produced fatty acids (mainly C18:0 and C18:1), giving rise to contaminating nsBola SLs and acidic SLs. Finally, also so-called 'alkyl SLs' (Fig. 6A), were found in the broths of the *Δfao1* mutants, as was also described by Brakemeier et al. (1998). However, their concentration was very low (1-2 g/L) in comparison with sBola SLs (20 g/L). Optimization towards improved production of these alkyl SLs will be described elsewhere. Minute amounts of sBola SLs were detected for the *ΔatΔsbleΔfao2* strain and *ΔatΔsble* parental strain. These two strains showed a very similar production profile, *i.e.* a mixture of mainly nsBola SLs and a small percentage of acidic SLs. Thus, it appears that the influence of the *fao2* deletion is not effective in terms of shifting the production towards a significant increase in sBola SL production. These results suggest that this putative alcohol oxidase is not involved in fatty alcohol oxidation, at least not in the oxidation of long-chain alcohols. Its exact role remains to be elucidated and will be investigated in future research.

To confirm the assumed increased stability of the sBola SL in comparison with the nsBola SL, symmetrical ones, an alkaline hydrolysis was performed. Indeed, for the nsBola SLs complete hydrolysis was observed, giving rise to acidic SLs and sophorose, whereas the sBola SLs remained intact. Related with this, the mixed nature of the products produced by the *ΔatΔsbleΔfao1* was also confirmed by applying alkaline hydrolysis: the sBola SLs remained intact, while the nsBola SLs were completely hydrolyzed (Fig. 7). The two remaining peaks both correspond to C18:1 sBola SLs (C18:1), but differ in the attachment of the second sophorose moiety to the hydrophobic linker, *i.e.* terminal versus subterminal linkage (Fig. 6C).

**NMR characterization of sBola SLs**

The sBola SLs were purified from both batches (fed with hexadecanol and oleyl alcohol) as explained in Materials and Methods, to enable NMR structure confirmation of the expected compounds (for more details, see Supplementary Materials). Their masses were already

confirmed by LC-MS analysis (906 and 932 g/mol for hexadecanol and oleyl alcohol based sBola SLs, respectively).

The results from NMR analysis for sBola SLs derived from hexadecanol (C16:0-OH) are summarized in Table SV and SVI. The numbering of the structure is confirmed by different 1D and 2D spectra (Fig. 6B and S7), which are included in the Supplementary Materials. In this respect, the chemical formula $C_{40}H_{74}O_{22}$ is confirmed by the integration of the proton NMR together with the signals present in $^{13}C$ and the ATP spectrum. To conclude, all NMR data support the results from previous observations (LC-MS, hydrolysis test) and confirm the structure as shown in Fig. 6B.

For the sBola SLs derived from oleyl alcohol (C18:1-OH), the NMR results are summarized in Table SVII and SVIII. NMR analysis suggests that a mixture of two compounds is present in a 50/50 ratio (Fig. 6C and S23). Both compounds have the same chemical formula, $C_{42}H_{76}O_{22}$, but differ in how one of both ends of the fatty alkyl chain is linked to the second sophorose unit. Depending on the preference of the CYP52M1 enzyme of *S. bombicola*, the hydroxylation of the fatty alcohol either took place terminally or subterminally (Fig. S4). Both compounds are equally present in the mixture, indicating that the CYP52M1 enzyme did not have a preference. To conclude, all NMR data support the results from previous observations (LC-MS, hydrolysis test) and confirm the structures as shown in Fig. 6C.

## Conclusions

In this article, production of bolaform SLs by the genetically engineered *S. bombicola ΔatΔsble* strain was successfully scaled up to the 150 L scale. Productivity, DSP yield and purity corresponded to 0.22 g/L.h, 75 % and 95 % respectively. Characteristics of these new molecules were assessed. Results point towards the use of bolaform SLs in mild hair or personal care products, degreasers or lubricants for cleaning applications, or the use in automatic or hand

dishwashing applications. Unfortunately, some limitations were discovered in terms of the chemical stability of these bolaform SLs, due to the ester functionality present in the molecules (therefore called 'non-symmetrical' (nsBola)). Applying an integrated bioprocess design approach i.e. feedback coupling towards the strain and process level, gave rise to the production of new-to-nature biosurfactants with two glycosidically linked sophoroses, i.e. symmetrical bolaform SLs. The structure of these novel compounds was confirmed by NMR. This was only successful for the *foa1* gene, which thus confirms that the corresponding enzyme is responsible for long chain alcohol oxidation in *S. bombicola*. The exact function of the other putative alcohol oxidase gene (*foa2*) remains to be elucidated. The sBola SLs indeed display an increased chemical stability and thereby greatly enhance the possible applications of bolaform SLs. Nevertheless, only part of the nsBola SLs could be redirected towards the sBola SLs, as nsBola SLs are still produced in substantial amounts by the new *S. bombicola* strain. Further strain improvement is necessary to completely shift the bolaform SL production towards the symmetrical compounds. An important remark is the influence of the fed alcohol in terms of terminal or subterminal hydroxylation, giving rise to different sBola SL structures. When hexadecanol was fed, only the subterminally hydroxylated compound was produced, whereas for oleyl alcohol, a 50/50 mixture of the respective subterminally and terminally hydroxylated sBola SL was produced. This is a reflection of the substrate specificity of the CYP52M1 enzyme.

In this article, we confirm that applying an integrated bioprocess design strategy (IBPD), i.e. considering the entire innovation chain, from genetic engineering through fermentation and downstream processing to final application testing, is key to develop new strains and processes for the industrial production and commercialization of new biosurfactants.

## Conflict of interest

The authors declare that there is no conflict of interest regarding the publication of this article.

## Acknowledgements

Stijn Verweire and Lien Saey are both acknowledged for their excellent work in the lab with strain engineering, bioreactor experiments, and analytics. Emile Redant is acknowledged for the development of the GC protocol for fatty alcohol quantification. This research was funded by IWT (innovation mandate 140917), FWO (International Mobility 23310), the European FP7 Project Biosurfing, (289219), the European FP7 Project IB2Market (111043), and European Horizon 2020 Bio-Based Industries (BBI) Consortium Project Carbosurf (669003).

## References


Allied Analytics LLP (2016). World Surfactants Market - Opportunities and Forecast 2014 - 2020. Portland, Oregon: Allied Market Research.

Ambrosi M, Fratini E, Alfredsson V, Ninham BW, Giorgi R, Lo Nostro P, Baglioni P. (2006). Nanotubes from a vitamin C-based bolaamphiphile. *J Am Chem Soc* 128, 7209-14.

Asmer HJ, Lang S, Wagner F, Wray V. (1988). Microbial production, structure elucidation and bioconversion of sophorose lipids. *J Am Oil Chem Soc* 65, 1460-1466.

Baccile N, Nassif N, Malfatti L, Van Bogaert INA., Soetaert W, Pehau-Arnaudet G, Babonneau F. (2010). Sophorolipids: a yeast-derived glycolipid as greener structure directing agents for self-assembled nanomaterials. *Green Chem* 12, 1564-1567.

Baccile N, Babonneau F, Jestin J, Pehau-Arnaudet G, Van Bogaert INA. (2012). Unusual, pH-induced, self-assembly of sophorolipid biosurfactants. *ACS Nano* 6, 4763–4776.

Baccile N, Babonneau F, Banat IM, Ciesielska K, Cuvier A-S, Devreese B, Everaert B, Lydon H, Marchant R, Mitchell CA, Roelants S, Six L, Theeuwes E, Tsatsos G, Tsotsou GE, Vanlerberghe B., Van Bogaert I.N.A., Soetaert W. (2017). Development of a cradle-to-grave approach for acetylated acidic sophorolipid biosurfactants. *ACS Sustainable Chem Eng* 5, 1186-1198.

Brakemeier A, Wullbrandt D, Lang S (1998). *Candida bombicola*: production of novel alkyl glycosides based on glucose/2-dodecanol. *Appl Microbiol Biotechnol* 50, 161-166.

Chong P. (2010). Archaebacterial bipolar tetraether lipids: Physico-chemical and membrane properties. *Chem Phys Lipids* 163, 253–265.

Cheng Q, Sanglard D, Vanhanen S, Liu HT, Bombelli P, Smith A, Slabas AR. (2005). *Candida* yeast long chain fatty alcohol oxidase is a c-type haemoprotein and plays an important role in long chain fatty acid metabolism. *Biochim Biophys Acta* 15, 192-203.

Conrath KJ. (2008). Rapid identification of common hexapyranose monosaccharide units by a simple TOCSY matching approach. *Chem Eur J* 14, 8869-8878.

Daniel HJ, Reuss M, Syldatk C. (1998a). Production of sophorolipids in high concentration from deproteinized whey and rapeseed oil in a two stage fed batch process using *Candida bombicola* ATCC 22214 and *Cryptococcus curvatus* ATCC 20509. *Biotechnol Lett* 20, 1153–1156.

Daniel HJ, Reuss M, Syldatk C. (1998b). Production of sophorolipids in high concentration from deproteinized whey and rapeseed oil in a two stage fed batch process using *Candida bombicola* ATCC 22214 and *Cryptococcus curvatus* ATCC 20509. *Biotechnol Lett* 20, 1153-1156.

Daverey A, Pakshirajan K, Sumalatha S. (2011). Sophorolipids production by *Candida bombicola* using dairy industry wastewater. *Clean Technol Environ Policy* 13, 481–488.

Davila AM, Marchal R, Vandecasteele JP. (1997). Sophorose lipid fermentation with differentiated substrate supply for growth and production phases. *Appl Microbiol Biotechnol* 47, 496-501.

Eirich LD, Craft DL, Steinberg L, Asif A, Eschenfeldt WH, Stols L, Donnelly MI, Wilson CR. (2004). Cloning and characterization of three fatty alcohol oxidase genes from *Candida tropicalis* strain ATCC 20336. *Appl Environ Microbiol* 70, 4872–4879.

Fariya M, Jain A, Dhawan V, Shah S, Nagarsenker MS. (2014). Bolaamphiphiles: A Pharmaceutical Review. *Adv Pharm Bull* 4, 483-91.

Fuhrhop JH, Wang T. (2004). Bolaamphiphiles. *Chem Rev* 104, 2901-37.

Gao RJ, Falkeborg M, Xu XB, Guo Z. (2013). Production of sophorolipids with enhanced volumetric productivity by means of high cell density fermentation. *Appl Microbiol Biotechnol* 97, 1103-1111.

Gheysen K, Mihai C, Conrath K, Martins J. (2008). Rapid identification of common hexapyranose monosaccharide units by a simple TOCSY matching approach. *Chem Eur J* 14, 8869-8878.

Hommel R, Ratlegde C. (1990). Evidence for two fatty alcohol oxidases in the biosurfactant producing yeast *Candida (Torulopsis) bombicola*. *FEMS Microbiol Lett* 70, 183-186.

Jin Y, Qi N, Tong L, Chen D. (2010). Self-assembled drug delivery systems. Part 5: self-assemblies of a bolaamphiphilic prodrug containing dual zidovudine. *Int J of Pharm* 386, 268-274.

Lang S, Brakemeier A, Heckmann R, Spockner S, Rau U. (2000). Production of native and modified sophorose lipids. *Chim Oggi* 18, 76–79.

Mukherjee S, Das P, Sen R. (2006). Towards commercial production of microbial surfactants. *Curr Trends Biotechnol* 24, 509–15.

Nitschke M, Costa S. (2007). Biosurfactants in food industry. *Trends Food Sci Technol* 18, 252–9.

Nuraje N, Bai H, Su K. (2013). Bolaamphiphilic molecules: Assembly and applications. *Prog Polym Sci* 38, 302–43.

Price NPJ, Ray KJ, Vermillion KE, Dunlap C, Kurtzman CP. (2012). Structural characterization of novel sophorolipid biosurfactants from a newly identified species of *Candida* yeast. *Carbohydr Res* 348, 33–41.

Puri A, Loomis K, Smith B, Lee JH, Yavlovich A, Heldman E, Blumenthal R. (2009). Lipid-based nanoparticles as pharmaceutical drug carriers: From concepts to clinic. *Crit Rev Ther Drug Carrier Syst* 26, 523-80.

Petersen BO, Vinogradovb E, Kayc W, Würtza P, Nyberga NT, Duusa JO, Sørensen OW. (2006). H2BC: A new technique for NMR analysis of complex carbohydrates. *Carbohydr Res* 341, 550-556.

Roelants SLKW, Saerens KMJ, Derycke T, Li B, Lin Y-C, Van de Peer Y. (2013). *Candida bombicola* as a platform organism for the production of tailor-made biomolecules. *Biotechnol Bioeng* 110, 2494–2503.

Roelants SLKW, Ciesielska K, De Maeseneire S, Moens H, Everaert B, Verweire S, Denon Q, Vanlerberghe B, Van Bogaert INA, Van der Meeren P, Devreese B, Soetaert W. (2016). Towards the industrialization of new biosurfactants: biotechnological opportunities for the lactone esterase gene from *Starmerella bombicola*. *Biotechnol Bioeng* 113, 550-559.

Saerens KMJ, Zang J, Saey L, Van Bogaert INA, Soetaert W. (2011a). Cloning and functional characterization of the UDP-glucosyltransferase UgtB1 involved in sophorolipid production by *Candida bombicola* and creation of a glucolipid-producing yeast strain. *Yeast* 28, 279–292.

Saerens KMJ, Saey L, Soetaert, W. (2011b). One-step production of unacetylated sophorolipids by an acetyltransferase negative *Candida bombicola*. *Biotechnol Bioeng* 108, 2923-2931.

Savarino P, Montoneri E, Bottigliengo S, Boffa V, Guizzetti T, Perrone DG. (2009). Biosurfactants from urban wastes as auxiliaries for textile dyeing. *Ind Eng Chem Res* 48, 3738–3748.

Schnablegger H, Singh Y. (2017). The SAXS Guide: Getting acquainted with the principles. 4th Edition. Graz, Austria: Anton Paar.

Sleiman JN, Kohlhoff S, Roblin PM, Wallner S, Gross R, Hammerschlag MR. (2009). Sophorolipids as antibacterial agents. *Ann Clin Lab Sci* 39, 60–63.

Soetaert W, Van Bogaert INA, Roelants SLKW. (2013). Methods to produce bolaamphiphilic glycolipids. Patent No. WO2015028278 A1.

Takahashi F, Igarashi K, Hagihara H. (2016). Identification of the fatty alcohol oxidase FAO1 from *Starmerella bombicola* and improved novel glycolipids production in an FAO1 knockout mutant. Appl Microbiol Biotechnol 100, 9519-9528.

Van Bogaert I, De Maeseneire S, Develter D, Soetaert W, Vandamme E. (2008). Development of a transformation and selection system for the glycolipid-producing yeast *Candida bombicola*. *Yeast* 25, 273-278.

Van Bogaert I, Holvoet K, Roelants SL, Li B, Lin YC, Van de Peer Y, Soetaert W. (2013). The biosynthetic gene cluster for sophorolipids: a biotechnological interesting biosurfactant produced by *Starmerella bombicola*. *Mol Microbiol* 88, 501-509.

Van Bogaert I, Buys D, Martins J, Roelants SLKW, Soetaert W. (2016). Synthesis of bolaform biosurfactants by an engineered *Starmerella bombicola* yeast. *Biotechnol Bioeng* 113, 2644–2651.

## List of figures

**Fig. 6** Confirmed new-to-nature biosurfactants produced by the *ΔatΔsbleΔfao1 S. bombicola* strain: (A) C16:0 alkyl SL, (B) C16:0 sBola SL (subterminal hydroxylation), (C) C18:1 sBola SL (both terminal and subterminal hydroxylation). Except for the alkyl SLs (which was proven by their mass using LC-MS), the new-to-nature structures were confirmed by NMR. In magenta and blue, the ether (glycosidic) and ester linkages are indicated, respectively.

**Fig. 7** HPLC-ELSD chromatogram from a broth sample of the *ΔatΔsbleΔfao1* strain fed with oleyl alcohol before (blue) and after alkaline hydrolysis (red). The eluting compounds at 1-2 min represent hydrophilic impurities as sugars, proteins etc.

**Table I** The composition of the self-prepared dish wash liquid (A) and rinse aid formulations (B) to respectively score the degreasing and anti-scaling/rinse aid properties of the (bio)surfactants.

| **A) Hand dishwashing liquid formulation** | **Weight percent (w/w)** |
|---|---|
| Biosurfactant/Surfactant (of interest) | 1.5 |
| Ammonium lauryl sulfate (28 %) | 32.1 |
| Disodium laureth sulfosuccinate (40 %) | 7.5 |
| Acrylates/Beheneth-25 Methacrylate Copolymer (30 %) | 6.9 |
| Adapt pH to 6.4-6.5 | |
| **B) Rinse aid formulation** | **Weight percent (w/w)** |
| Biosurfactant/Surfactant (of interest) | 5.2 |
| Citric acid | 15 |
| Ethanol (96 %) | 5 |
| CAPB (40 %) | 5 |
| Sodium 2-naphtalene sulfonate | 3 |
| Adapt pH to 2 | |

**Table II** Overview of different fatty alcohol oxidase (*fao*) *S. bombicola* knockout strains created in this article, and their expected glycolipid production when fed with long chain alcohols. To allow a comparison, the parental *ΔatΔsble* strain is also represented.

| ***S. bombicola* strain** | **Glucose consumption (g/L.h)** | **$\log(CFU)_{stat}$ (> 48h)** | **Expected glycolipid production when fed with long-chain fatty alcohols** | **Reference** |
|---|---|---|---|---|
| *ΔatΔsbleΔfao1* | 0.47 | 8.2 | mainly sBola SLs and minor nsBola SLs (*de novo* fatty acid) | This article |
| *ΔatΔsbleΔfao2* | 0.78 | 7.8 | | |
| *ΔatΔsble* | 0.73 | 8.0 | minute sBola SLs, mainly nsBola SLs and acidic SLs | Van Bogaert et al., 2016; This article |


| ***S. bombicola* strain** | **Glucose consumption (g/L.h)** | **$\log(CFU)_{stat}$ (> 48h)** | **Expected glycolipid production when fed with long-chain fatty alcohols** | **Reference** |
|---|---|---|---|---|
| *ΔatΔsbleΔfao1* | 0.47 | 8.2 | mainly sBola SLs and minor nsBola SLs (*de novo* fatty acid) | This article |
| *ΔatΔsbleΔfao2* | 0.78 | 7.8 | | |
| *ΔatΔsble* | 0.73 | 8.0 | minute sBola SLs, mainly nsBola SLs and acidic SLs | Van Bogaert et al., 2016; This article |

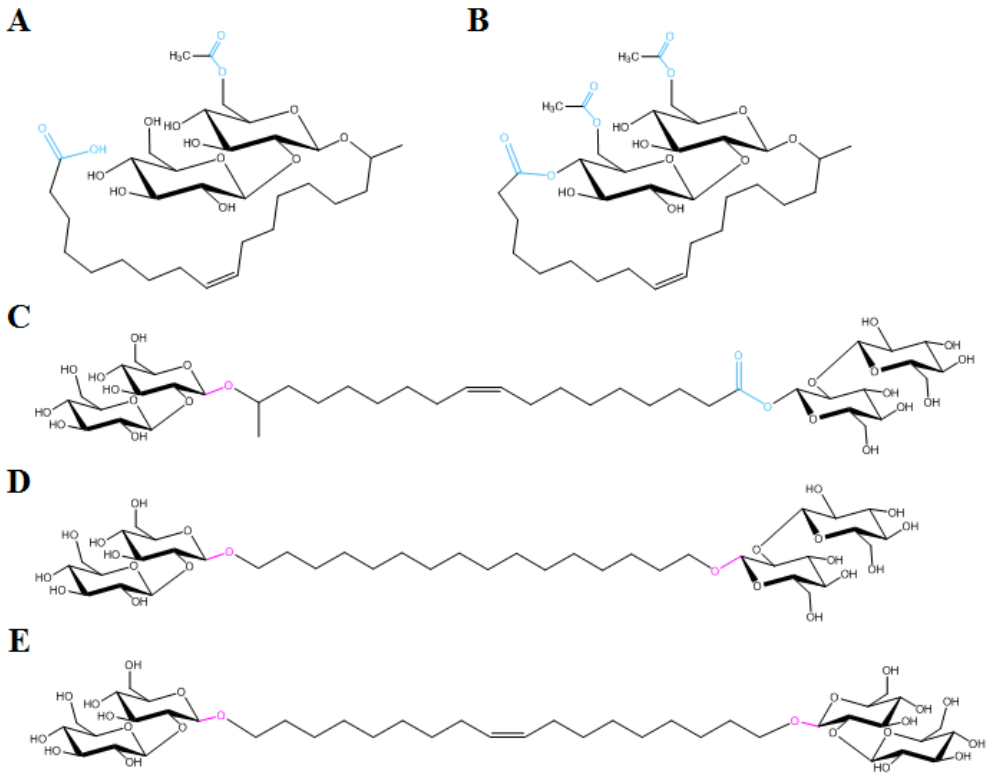



Figure 1

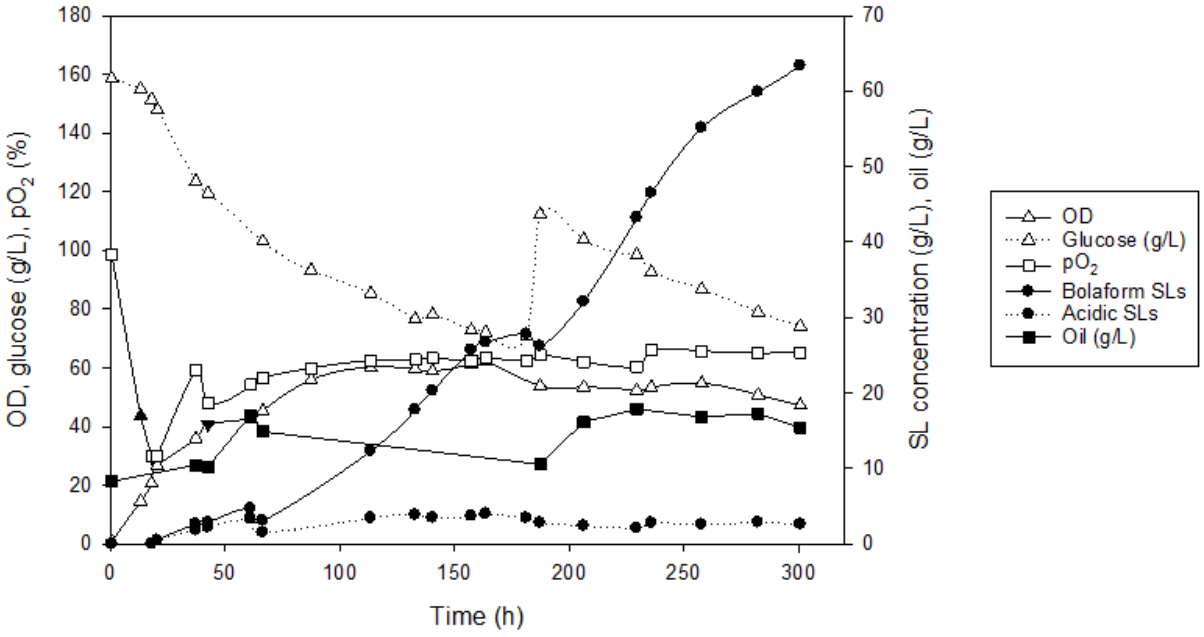

OD, glucose (g/L), pO2 (%)
SL concentration (g/L), oil (g/L)
Time (h)
OD
Glucose (g/L)
pO2
Bolaform SLs
Acidic SLs
Oil (g/L)

Figure 2

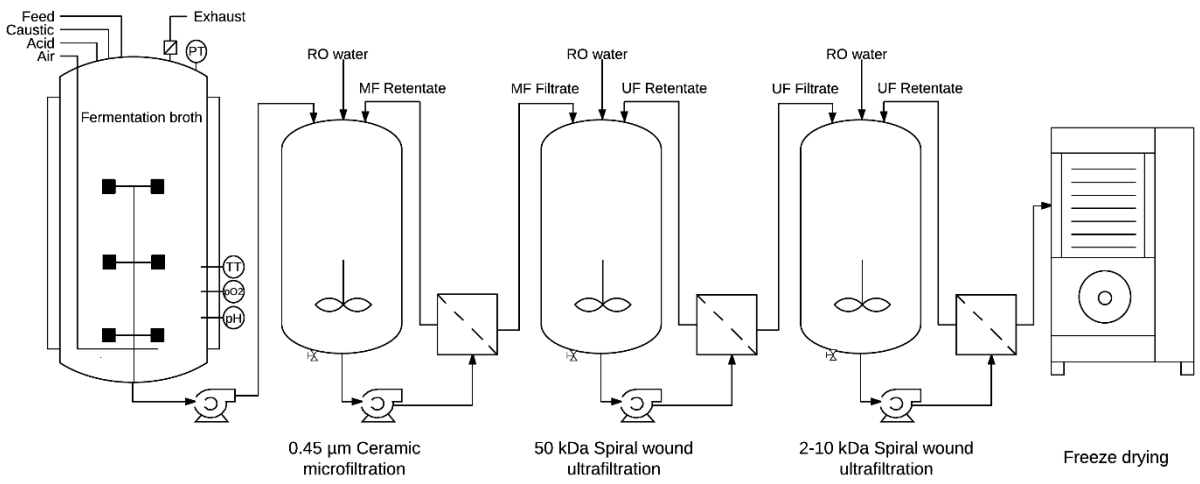

Feed
Caustic
Acid
Air
Exhaust
PT
Fermentation broth
TT
pO2
pH
RO water
MF Retentate
MF Filtrate
RO water
UF Retentate
UF Filtrate
RO water
UF Retentate
0.45 µm Ceramic microfiltration
50 kDa Spiral wound ultrafiltration
2-10 kDa Spiral wound ultrafiltration
Freeze drying

**Figure 3**

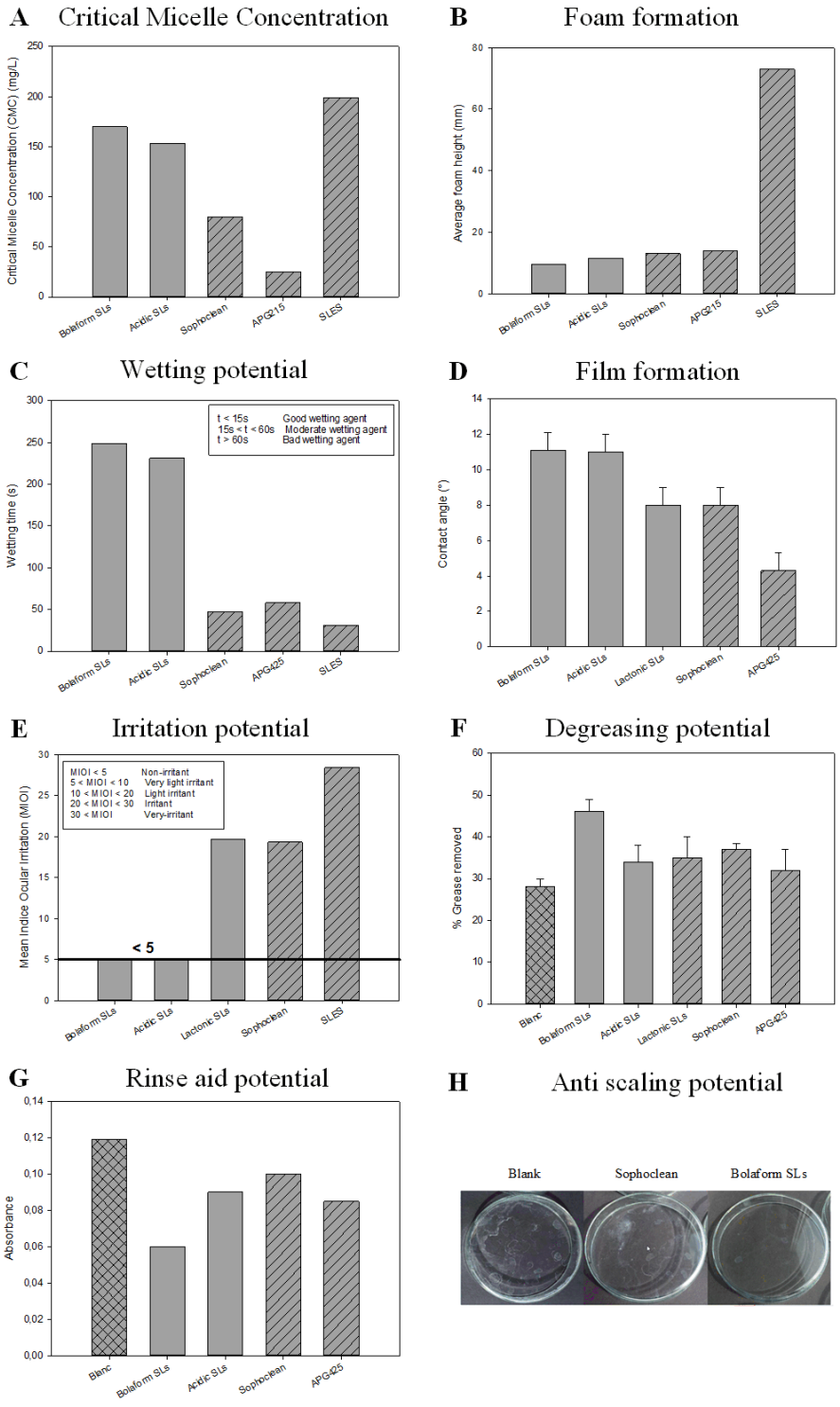

A Critical Micelle Concentration
Critical Micelle Concentration (CMC) (mg/L)
Bolaform SLs
Acidic SLs
Sophoclean
APG215
SLES
B Foam formation
Average foam height (mm)
C Wetting potential
Wetting time (s)
t < 15s Good wetting agent
15s < t < 60s Moderate wetting agent
t > 60s Bad wetting agent
APG425
D Film formation
Contact angle (°)
Lactonic SLs
E Irritation potential
Mean Indice Ocular irritation (MOI)
MIOI < 5 Non-irritant
5 < MIOI < 10 Very light irritant
10 < MIOI < 20 Light irritant
20 < MIOI < 30 Irritant
30 < MIOI Very-irritant
< 5
F Degreasing potential
% Grease removed
Blanc
G Rinse aid potential
Absorbance
H Anti scaling potential
Blank
Sophoclean
Bolaform SLs

Figure 4

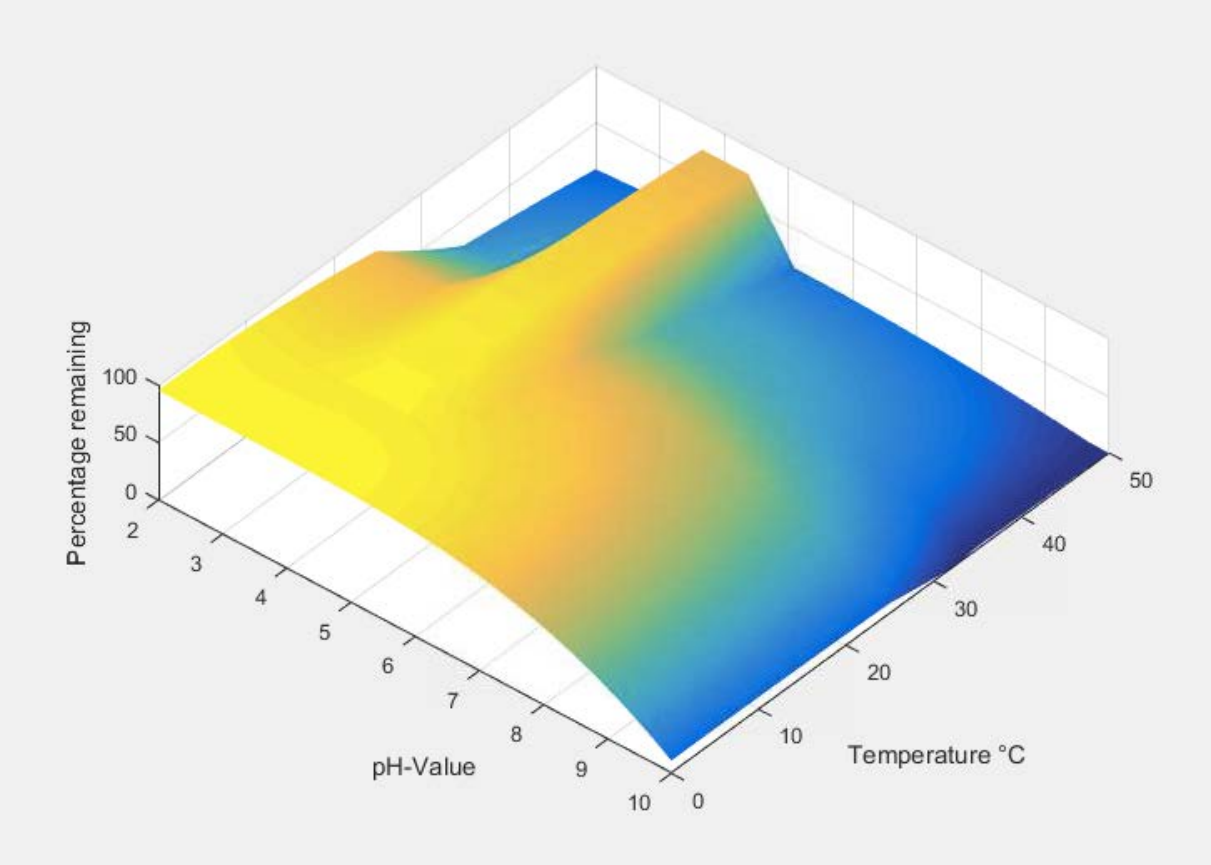

Percentage remaining
100
50
0
2
3
4
5
6
7
8
9
10
pH-Value
0
10
20
30
40
50
Temperature °C

Figure 5

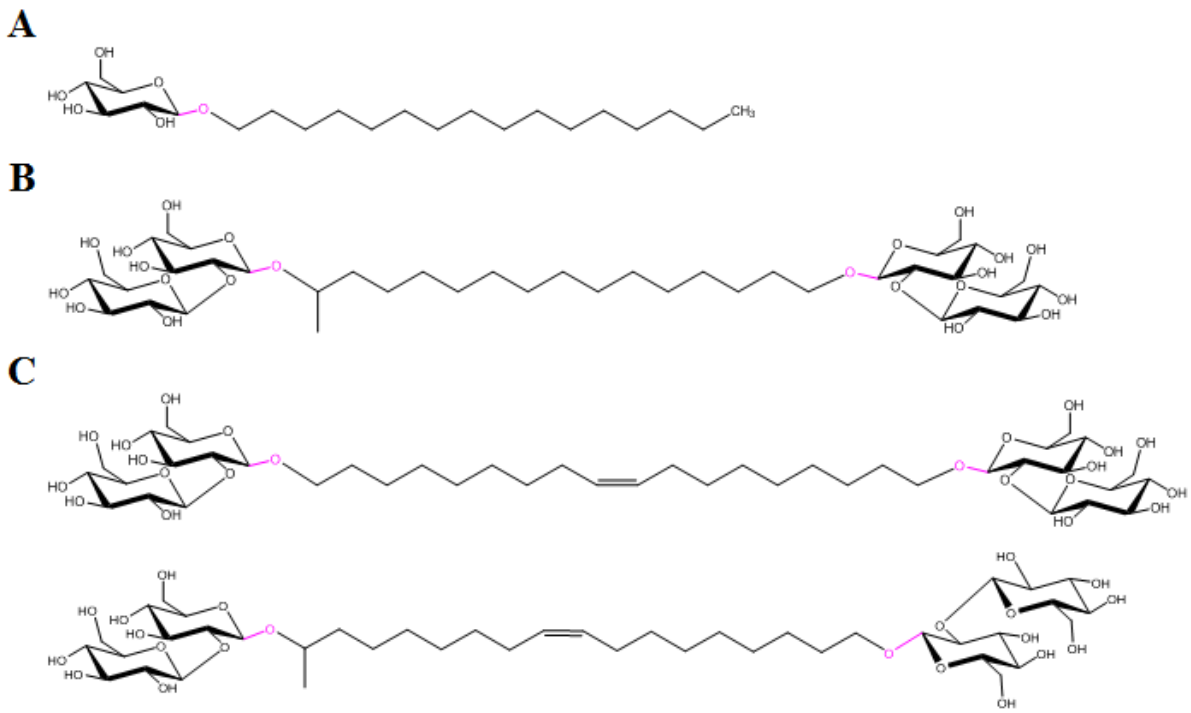



Figure 6

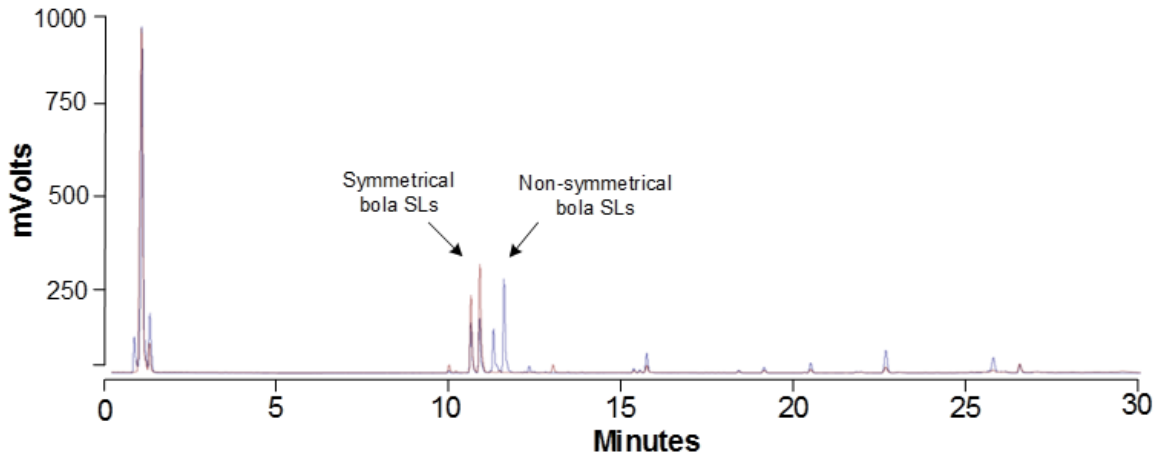

1000
750
500
250
mVolts
Symmetrical
bola SLs
Non-symmetrical
bola SLs
0
5
10
15
20
25
30
Minutes

Figure 7

## Supplementary materials

*Characterization and application potential of non-symmetrical bolaform (nsBola) SLs*

**Table SI** Composition of different batches of biosurfactant samples (Roelants et al., 2016), expressed in percentage. Sophoclean (Soliance) is a commercially available sophorolipid product, containing lactonic and acidic SLs.

| | **Acid T37** | **Lacton T43** | **Bola B01** | **Sophoclean** |
|---|---|---|---|---|
| nsBola SL | | | 84% | |
| Di-acetyl lactonic SL C18:2 | | 9% | | 30% |
| Diacetyl lactonic SL C18:1 | | 88% | | |
| Diacetyl lactonic SL C18:0 | | 2% | | |
| Non-acetylated acidic SL | 48% | | 10% | |
| Mono-acetylated acidic SLs | 28% | | | 58% |
| Diacetyl acidic SL | 19% | | | |
| Other (fatty acids,...) | 5% | 1% | 6% | 12% |

**Table SII** Solubility of new-to-nature biosurfactants in water, ethanol, isopropanol and ethyl acetate, expressed in g/L.

| | **Water** | **Ethanol** | **Isopropanol** | **Ethyl Acetate** |
|---|---|---|---|---|
| C18:1 nsBola SL | > 500 | < 10 | < 10 | < 1 |
| C18:1 sBola SL | > 500 | < 1 | < 1 | < 1 |
| C16:0 sBola SL | > 300* | < 1.5 | < 1 | < 1 |
| Non-acetylated acidic SLs | > 500 | < 10 | < 1 | < 1 |

* problems with strong gelation, making it impossible to test higher concentrations

*Supramolecular assembly characteristics of nsBola SLs*

The supramolecular assembly was assessed as mentioned in Materials and Methods section. In Fig. S1, the represented SAXS curves display typical features of micelles in solutions, whereas the oscillation at 3 $nm^{-2}$ identifies first oscillation of the micelles form factor. The spectra are sub-sequentially shifted at higher intensity when increasing the concentration, as expected. Fig. S1c and S1d show a series of selected SAXS curves recorded on both beamlines and the

respective fit, showing the good match between them. The evolution of the fit parameters is presented in Fig. S1e.

More precisely, from Fig. S1, one can observe the following:

1) The sphere form factor can be used throughout the concentration series, where no evolution towards either ellipsoids or cylinders is observed. This result, verified twice at different periods in time and on different beamlines, is very atypical. In fact, one would expect the elongation of the micelles at higher volume fractions, as this was experimentally found, and theoretically predicted, for acidic sophorolipids (Baccile et al., 2010; Manet et al., 2015). What is found for the nsBola SL (C18:1) is in fact a system of extremely stable spherical micelles, as found for instance in block copolymers with a strong difference in their hydrophilic/hydrophobic balance (e.g., Ps-b-PEO block copolymers) (Nicolai et al., 2010).

2) Opposed to what was found for acidic SLs (Manet et al., 2015), the hydrophilic shell is homogeneous.

3) The average value of the micellar radius (core + shell) as a function of concentration is 1.79 ± 0.02 nm, where the very small standard deviation (1.4 %) again testifies of the strong micellar stability. Individually, the shell thickness ranges between 1.1 nm and 1.3 nm, while the core radius lies between 0.5 nm and 0.7 nm, which are typical values for SL micelles (Manet et al., 2010). The expected size of the nsBola SL compound is expected to be about 3.7 nm, if one considers the length of the monounsaturated fatty acid (C1-C16, due to the subterminal glycosidic bond) to be about 1.7 nm according to the Tanford formula (for a bent *cis* conformation), and the size of sophorose to be about 1.0 nm, typical for disaccharides. It is then easy to observe the strong match between the micellar experimental radius (1.79 ± 0.02 nm) and half the nsBola SL length (1.85 nm), indicating an interpenetrated structure of the nsBola SL compound, as hypothesized by Nagarajan (1987) for these type of bolaform

surfactants. Finally, the average value of the shell SLD is 9.82 ± 0.05 x $10^{-4}$ $nm^{-2}$, which is in the range of hydrated sophorose and it reveals to be stable with concentration.

4) Above 150 mg/mL, it is not possible to fit the SAXS curves using a core-shell sphere model only, as an effect of the micellar packing density starts to become visible, as shown by the superimposition of the SAXS curves plateau below 1 $nm^{-1}$. A typical hard-sphere structure factor potential (Percus et al., 1958), commonly used for monodisperse spherical particles interacting through hard sphere (excluded volume) interactions, has been added to the fitting function with a hard sphere particle radius of 1.75 ± 0.05 nm. The effective volume fraction, reported in Fig. S1e, nicely increases from about 0.03 ± 0.02 to 0.14 when the concentration varies between 150 mg/mL and 378 mg/mL.

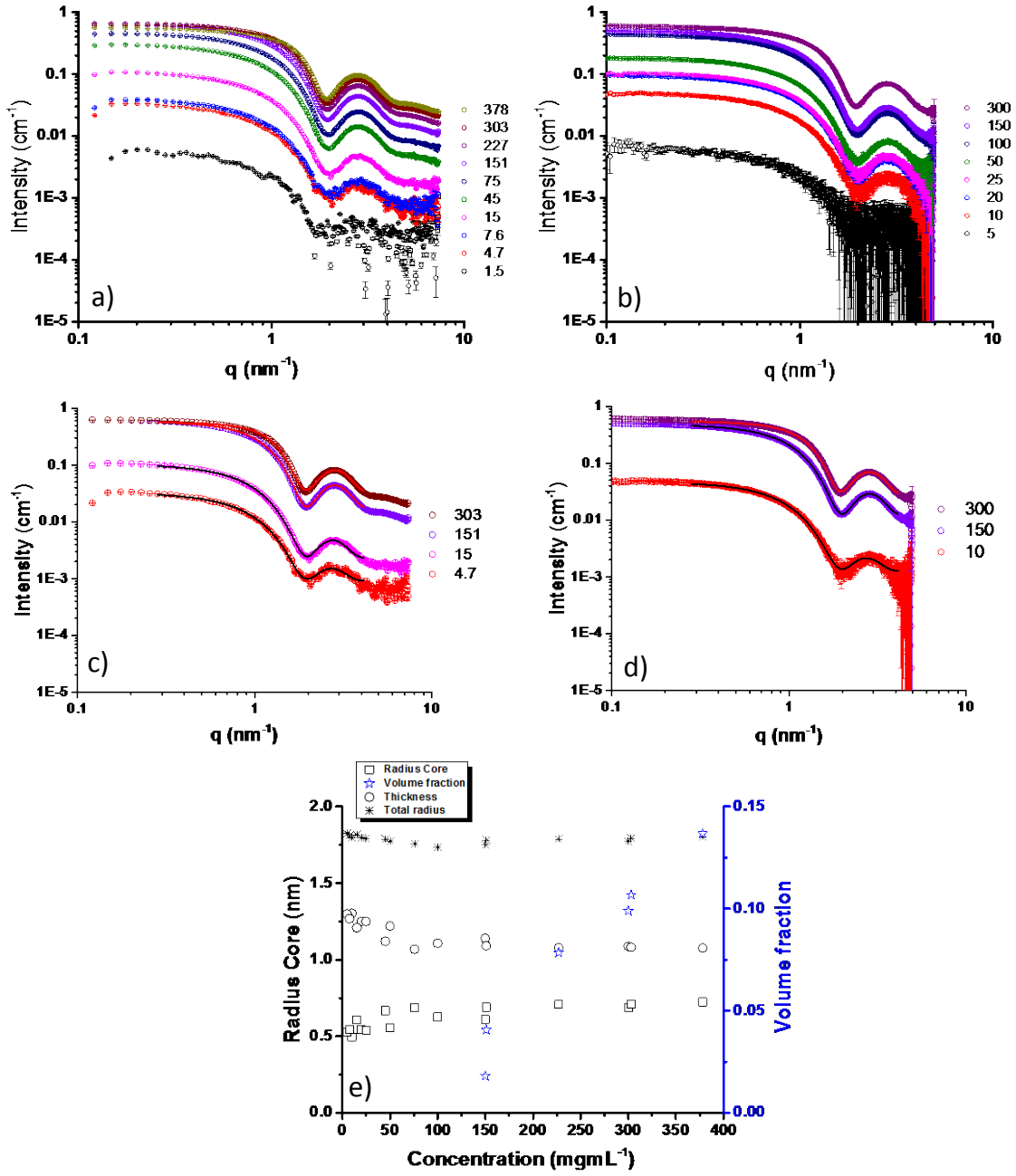



**Fig. S1** Supramolecular assembly study of nsBola SLs (C18:1). Respective SAXS data for different nsBola SL concentrations (mg/L) gathered at beamline ID02 (A) and repeated on the BM29 beamline (B) at the European Synchrotron Radiation Facility (ESRF; Grenoble, France). In (C) and (D), only a selected amount of SAXS curves from (A) and (B) are shown and their respective fit, respectively, showing the resemblance between the data gathered at the two different beamlines. In (E), the evolution of the fit parameters is presented.

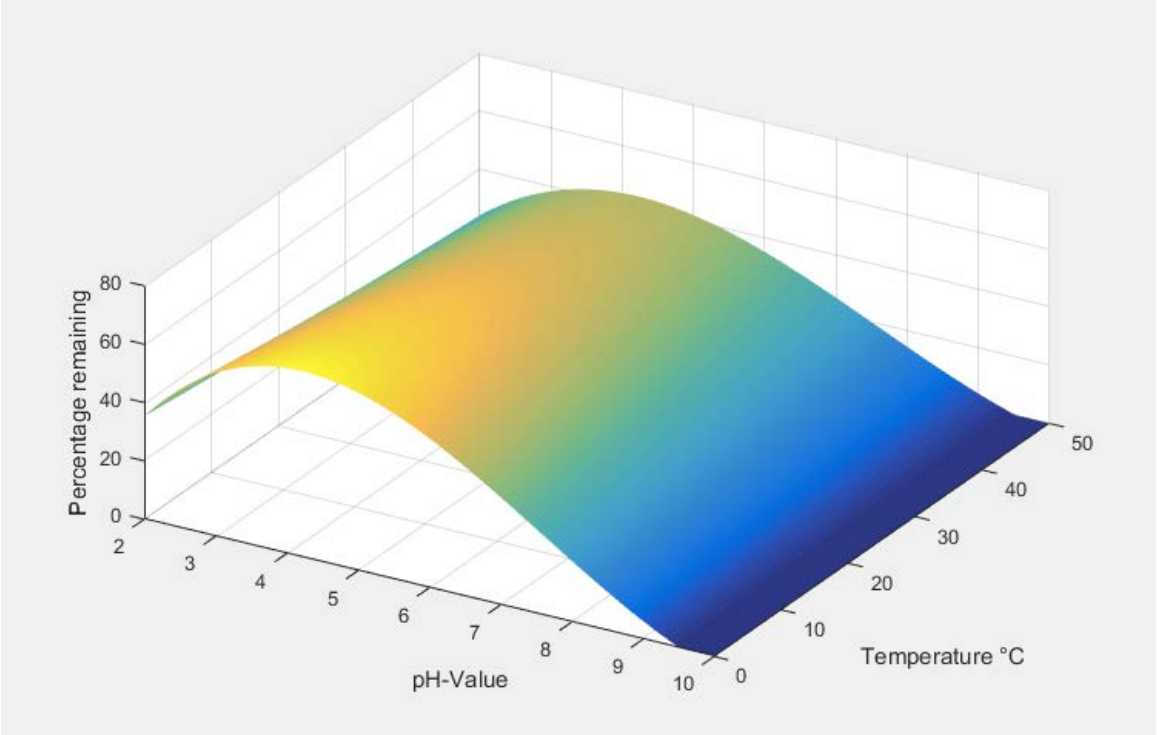



**Fig. S2** Chemical stability of lactonic SLs in function of pH and temperature after three months of incubation.

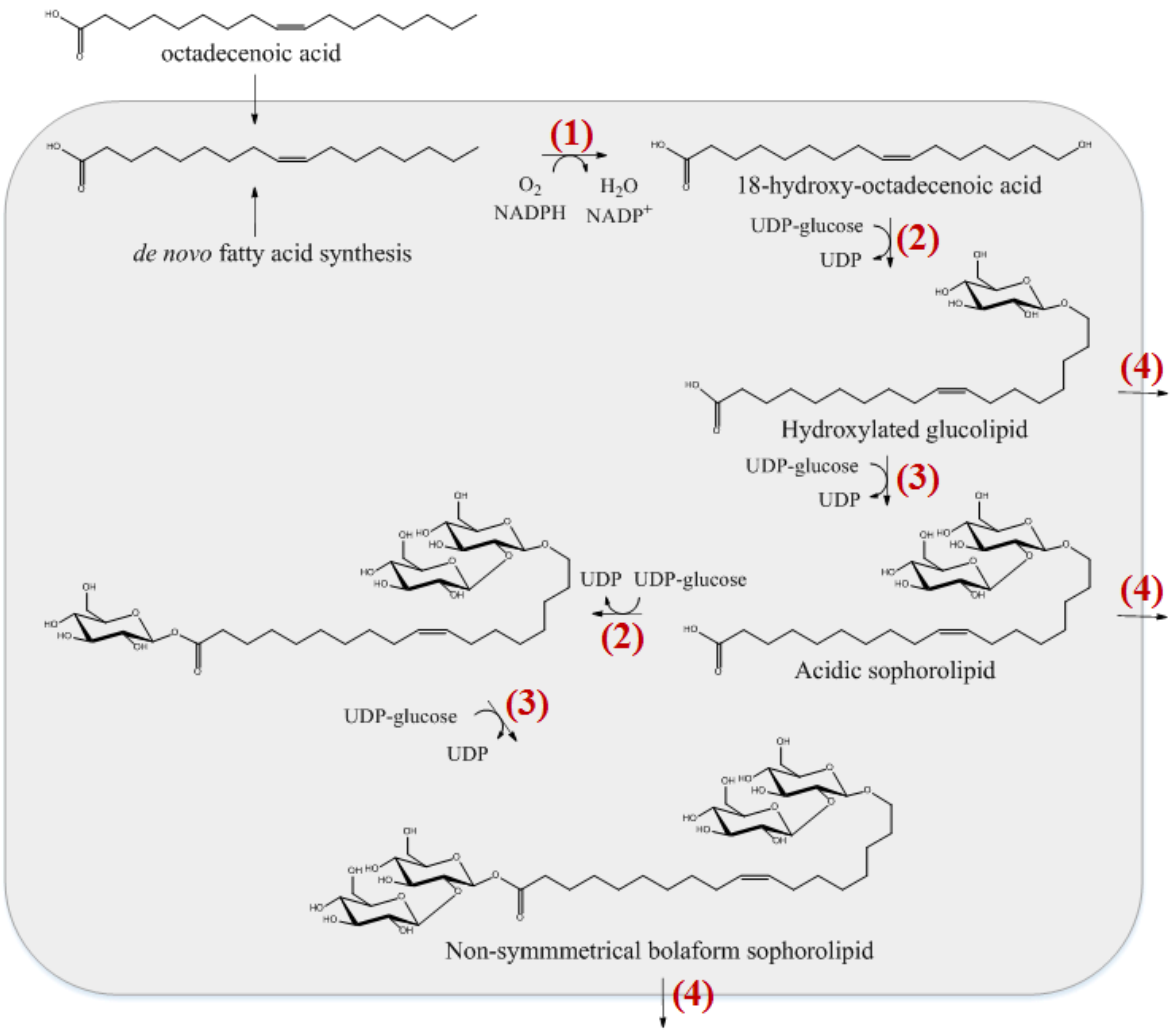



**Fig. S3** Production pathway of nsBola SLs in the *ΔatΔsble S. bombicola* strain. (1) Cytochrome P450 monooxygenase CYP52M1, (2) glucosyltransferase UGTA1, (3) glucosyltransferase UGTB1, (4) SL transporter MDR. The fatty acid is implemented by the CYP52M1, hydroxylating the latter to its corresponding subterminal or terminal hydroxy fatty acid (in the Fig., only terminal hydroxylation is presented). Next, both glucosyltransferases UGTA1 and UGTB1 sequentially add a glucose molecule on both side of the hydroxy fatty acid, giving rise to nsBola SLs. Finally, these compounds are transported out of the cell using the MDR transporter.

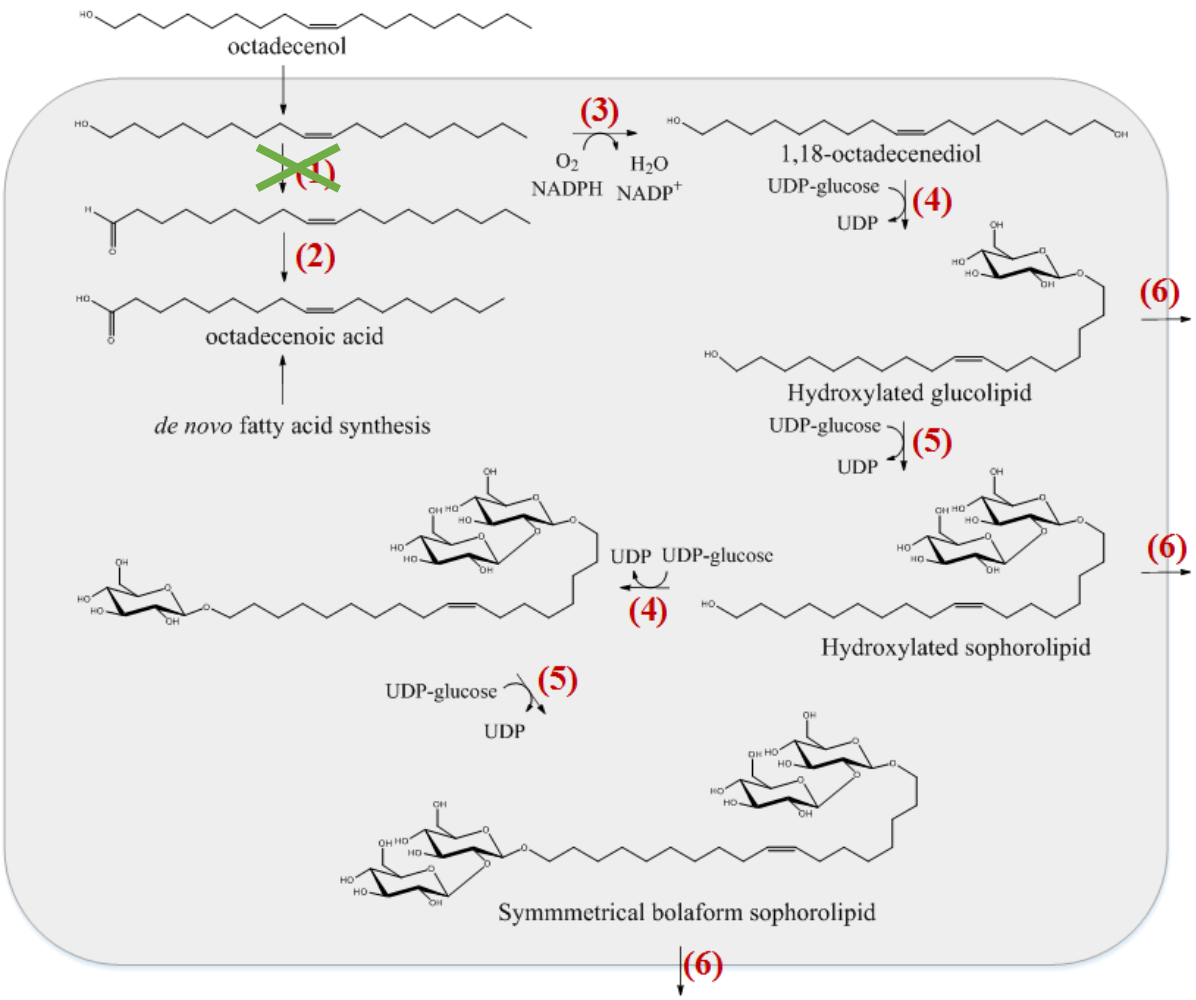



**Fig. S4** Production pathway of sBola SLs in the *ΔatΔsbleΔfao S. bombicola* strain. (1) Fatty alcohol oxidase FAO, (2) fatty aldehyde dehydrogenase, (3) cytochrome P450 monooxygenase CYP52M1, (4) glucosyltransferase UGTA1, (5) glucosyltransferase UGTB1 and (6) SL transporter MDR. By knocking out the FAO, the fatty alcohol is not (or partially) converted towards the corresponding fatty aldehyde and fatty acid. Hereby, the CYP52M1 enzyme can hydroxylate the fatty alcohol to the respective diol. Next, both glucosyltransferases UGTA1 and UGTB1 sequentially add a glucose molecule on each hydroxylated side of the diol, giving rise to sBola SLs. Finally, these compounds are transported out of the cell using the MDR transporter.

*Generation of the fao knockout strains*

General molecular techniques were employed as described by Green and Sambrook (2012). *Escherichia coli* DH5α cells were used in all cloning experiments and were grown in Luria-Bertani (LB) medium (1% tryptone, 0.5% yeast extract and 0.5% sodium chloride) supplemented with 100 mg/L ampicillin. Liquid *E. coli* cultures were cultivated at 37 °C on a rotary shaker (200 rpm). Linear deletion cassettes were generated from vector backbones cloned and maintained in *E. coli* based on the pGEM-T vector (Promega) and pJET (Thermo Fisher) and cloning steps are described below. All primer sequences can be found in Table SIII and a graphical representation of the strain construction is shown in Fig. S5.

To generate the first fatty alcohol oxidase (FAO1) deletion cassette, the *fao1* gene (GenBank AB907775) was amplified from isolated genomic DNA (Saerens, 2012) of *S. bombicola* with up- and downstream regions of 1000 bp using primerpair P1001/P1002. Subsequently, the genomic DNA fragment was cloned into the pJET vector (Thermo Fisher) using primerpair P1003/P1004, and checked with colony PCR using primer pairs P1023/P276. The coding sequence of the *fao1* gene was then replaced by the *ura3* selection marker (including its own promotor and viral tyrosine kinase (TK) terminator sequence) by primer pairs P1064/P1065 and P1062/P1063. The resulting vector of 6846 bp, checked with colony PCR using primer pair P265/P34, was used as a template to generate the *fao1* knockout cassette using primerpair P1020/P1021. The fragment of 3872 bp was used to transform the *S. bombicola* Δ*at*Δ*sble* strain (Van Bogaert et al., 2016). The latter parental strain had to be made *ura3* negative first by using so-called 'PT' cassettes (promotor/terminator of *ura3* gene) as described by Roelants et al. (2017), giving rise to a new *S. bombicola* strain, Δ*at*Δ*sble*Δ*ura3*. After successful transformation of the three strains with the *fao1* knockout cassette, the *ura3* positive colonies were selected on selective SD medium (YNB without AA: 6.7 g/L, agar: 20 g/L, glucose: 20 g/L, CSM-ura: 0.77g/L). Correct integration of the cassette was confirmed by colony PCR with

primer pair P1347/P34 upstream of the knockout cassette and primerpair P30/P1348 downstream of the knockout cassette. For the newly-created strain (Δ*at*Δ*sble*Δ*fao1*), three successful colonies were chosen.

Similar as for the *fao1* gene, the second alcohol oxidase gene (*fao2*) (GenBank MF431618) with up- and down-regions of 1000 bp was amplified from genomic DNA of *Starmerella bombicola* (Saerens, 2012) using primerpair P1371/P1372, respectively. Subsequently, the genomic DNA fragment was cloned into the pJET vector (Thermo Fisher) using primerpair P1373/P1374, and checked with colony PCR. The coding sequence of the *fao2* gene was then replaced by the *ura3* selection marker (including its own promotor and Ttk terminator sequences) using primer pairs P1378/P1377 and P1375/P1376. The resulting vector of 6530 bp was used as a template to generate the *fao2* knockout cassette using the primerpair P1371/P1372. The fragment of 3596 bp was used to transform the respective *ura3* auxotrophic Δ*at*Δ*sble*Δ*ura3 S. bombicola* strain (see above). After successful transformation of the strain with the *fao2* knockout cassette, the *ura3* positive colonies were selected on selective SD medium (YNB without AA: 6.7 g/L, agar: 20 g/L, glucose: 20 g/L, CSM-ura: 0.77g/L). Correct integration of the cassette was confirmed by colony PCR with primerpair P1576/P34 upstream of the knockout cassette and primerpair P30/P1396. For the newly-created strain (Δ*sble*Δ*at*Δ*fao2*), three successful colonies were chosen.

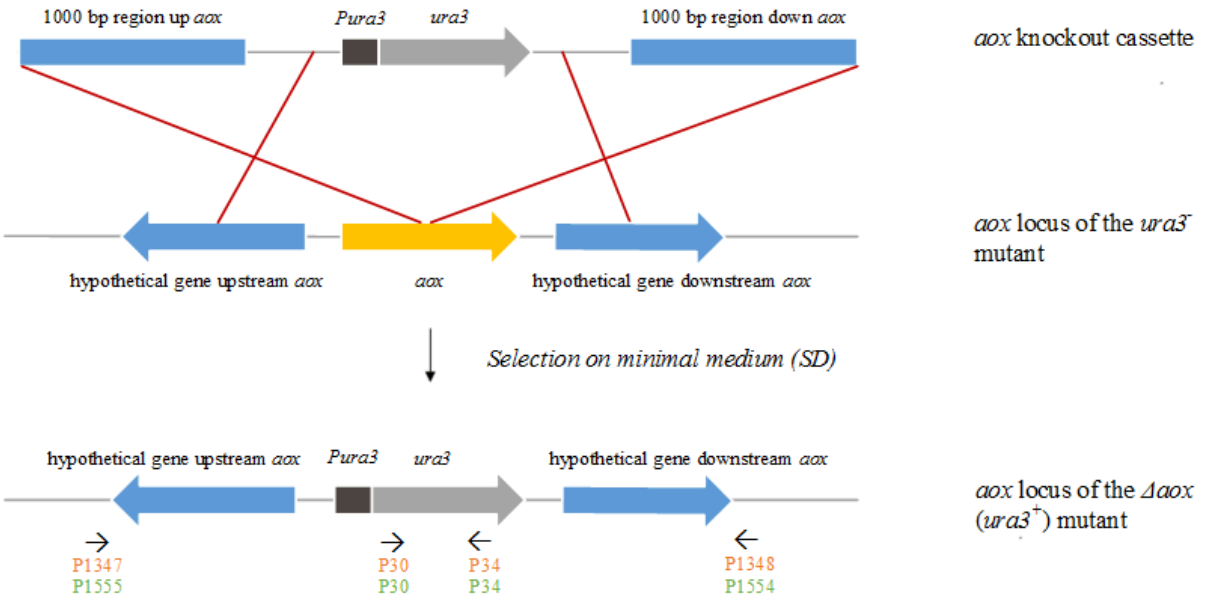



**Fig. S5** Overview of the genetic engineering strategy for knocking out the FAO in an *ura3*$^{-}$ *Starmerella bombicola* strain, creating a Δ*fao* (*ura3*$^{+}$) strain. The respective primerpairs for colony PCR to check correct integration of the *fao* knockout cassette colony PCR are indicated in orange (Δ*fao1*) and green (Δ*fao2*).

**Table SIII** List of primers used to generate the *fao1* and *fao2* deletion cassettes.

| **Primer** | **'5 sequence 3'** |
|---|---|
| P30_FOR_checkpromIN | AAGGCGGGCTGGAATGCATATCTGAG |
| P34_REV_checkcassIN | GATGTCGAATAGCCGGGCTGCTAC |
| P276_pJET_Rev | AAGAACATCGATTTTCCATGGCAG |
| P625_pJET_For | CGACTCACTATAGGGAGAGCGGC |
| P1001_FOR_FAO1_extgibpJET | CTCGAGTTTTTCAGCAAGATTGCCAAGTCGTTCAACACAG |
| P1002_REV_FAO1_extgibpJET | AGGAGATCTTCTAGAAAGATCTGAGACAGCAGCTTGTCAC |
| P1003_FOR_pJETextGib_FAO1 | GTGACAAGCTGCTGTCTCAGATCTTTCTAGAAGATCTCCTAC |
| P1004_REV_pJETextGib_FAO1 | CTGTGTTGAACGACTTGGCAATCTTGCTGAAAAACTCGAGCCATC |
| P1020_FOR_FAO1 | TGCCAAGTCGTTCAACACAG |
| P1021_REV_FAO1 | CTGAGACAGCAGCTTGTCAC |
| P1022_REV_checkFAO1KO | GCCTTGGCATTCAACATCTCAGGGAATC |
| P1023_FOR_checkFAO1KO | GCACGCCCTTAGCTTCAGAG |
| P1062_FOR_P*ura3*_extgib_upFAO1 | GACTGAGATGACGGAAGAGGCCCGAACATACCAGTTTCGC |
| P1063_REV_tTK_extgib_downFAO1 | AAGCTTAGTGAGATCCGCGTGAACAAACGACCCAACACCC |
| P1064_FOR_downFAO1_extgibtTK | GGGTGTTGGGTCGTTTGTTCACGCGGATCTCACTAAGCTTC |
| P1065_REV_upFAO1_extgibP*ura3* | GCGAAACTGGTATGTTCGGGCCTCTTCCGTCATCTCAGTC |

| | |
|---|---|
| P1347_FOR_FAO1_up_check | GCCAGTGCAACAAGTATGAG |
| P1348_REV_FAO1_down_check | GACCAGGCTAAACGCATCAC |
| P1371_FOR_upFAO2_FAO2_extgibpJET | CTCGAGTTTTTCAGCAAGATGAGCGCCCTTCATCAATGTC |
| P1372_REV_downFAO2_FAO2_extgibpJET | AGGAGATCTTCTAGAAAGATCTACACTGGAGGTGCATAGG |
| P1373_FOR_pJETextgib_downFAO2 | CCTATGCACCTCCAGTGTAGATCTTTCTAGAAGATCTCCTAC |
| P1374_REV_pJET_extgib_upFAO2 | GACATTGATGAAGGGCGCTCATCTTGCTGAAAAACTCGAGCC |
| P1375_FOR_pURA3_extgibupFAO2 | GGAAGAAACTGCTGCTCATCCCCGAACATACCAGTTTCG |
| P1376_REV_tTK_extgibdownFAO2 | GGATCTTCTCGCTGGCCTTAGAACAAACGACCCAACACC |
| P1377_REV_upFAO2_extgibpURA3 | CGAAACTGGTATGTTCGGGGATGAGCAGCAGTTTCTTCC |
| P1378_FOR_dFAO2_extgibtTK | GGGTGTTGGGTCGTTTGTTCTAAGGCCAGCGAGAAGATCC |
| P1396_REV_FAO2_checkIN_DOWN | CTGCCATTTTAGTTTGCTCAAGGTGTGTGTC |
| P1576_FAO2_check_out_fw | TAGCCAGATAGTCCAGACAG |

**Table SIV** Overview of molar masses (g/mole) of the glycolipids in the LC-MS spectra

| Type of SL | Acidic SLs | nsBola SLs | sBola SLs | Alkyl SL |
|---|---|---|---|---|
| Acetylation / Chain length | non | non | non | non |
| C18:0 | 624 | 948 | 934 | 594 |
| C18:1 | 622 | 946 | 932 | 592 |
| C18:2 | 620 | 944 | 930 | 590 |
| C16:0 | 596 | 920 | 906 | 566 |
| C16:1 | 594 | 918 | 904 | 564 |
| C16:2 | 592 | 916 | 902 | 562 |
| C14:0 | 568 | 892 | 878 | 538 |
| C14:1 | 566 | 890 | 876 | 536 |
| C14:2 | 564 | 888 | 874 | 534 |
| C12:0 | 540 | 864 | 850 | 510 |
| C12:1 | 538 | 862 | 848 | 508 |
| C12:2 | 536 | 860 | 846 | 506 |

*Extra data for shake-flask experiments with the newly-created Δfao strains*

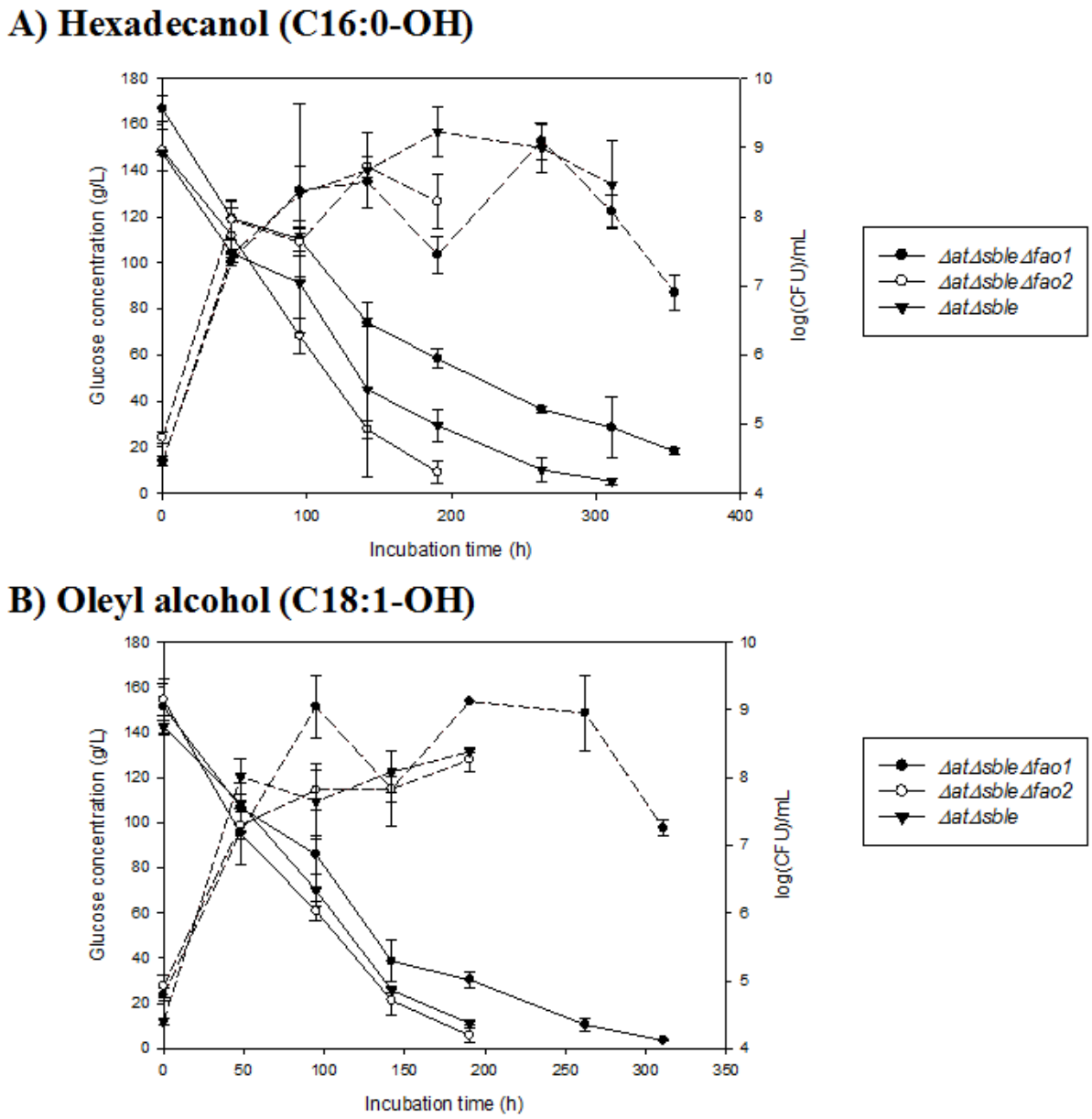



**Fig. S6** Glucose consumption (left, full line) and log(CFU)/mL (right, dashed line) in function of incubation time for the *ΔatΔsbleΔfao1* and *ΔatΔsbleΔfao2* strains as compared to the *S. bombicola ΔatΔsble* strain. All colonies of each strain showed identical behavior, so only one colony per strain is displayed. Hexadecanol (A) or oleyl alcohol (B) was added after 48h of incubation as hydrophobic substrate.

*NMR characterization of the symmetrical bolaform (sBola) SLs*

**1) Hexadecanol (C16:0-OH) based sBola SL**

The results from NMR analysis for sBola SLs derived from hexadecanol (C16:0-OH) are summarized in Table SV and SVI. The numbering of the structure is confirmed by different 1D and 2D spectra (Fig. 6B and S7), which are included in the Supplementary Materials. In this respect, the chemical formula $C_{40}H_{74}O_{22}$ is confirmed by the integration of the proton NMR together with the signals present in $^{13}C$ and the ATP spectrum.

The presence of the bolaform tetraglycolipid structure will be highlighted here, through a few key observations. First of all, the presence of four anomeric protons (4.25, 4.30, 4.36 and 4.38 ppm, Fig. S8) affirms that there are four sugar rings present. From the coupling constants of these anomeric protons (7.64, 7.68, 7.44 and 7.12 Hz, Fig. S8 zoom), it can be concluded that they are coupled via a β-connection. Second, the TOCSY matching approach (Gheysen et al., 2008), validates that each of the sugar moieties corresponds with a D-glucose pyranose sugar (Fig. S18). Third, the analysis of the HMBC combined with HSQC confirms that two sophorose units are present with a β-(1-2) linkage (Fig. S13).

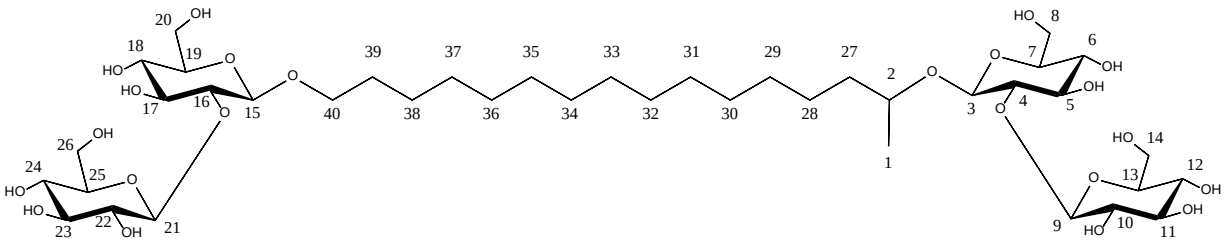



**Fig. S7:** Structure of the synthesized bolaform tetraglycolipid with numbering

**Table SV:** overview assignments $^{1}H$ chemical shifts

| δ $^{1}H$ (ppm) | Multiplicity | Integral | # Protons | Annotation |
|---|---|---|---|---|
| 1.12 | d | 3 | 3 | 1 |
| 1.17 - 1.44 | m | 22.79 | 23 | 27A, 28, 28, 30, 31, 32, 33, 34, 35, 36, 37, 38 |
| 1.44 - 1.55 | m | 2.93 | 3 | 39, 27B |

| | | | | |
|---|---|---|---|---|
| 2.95 - 3.02 | m | 2.06 | 2 | 10, 22 |
| 3.02 - 3.11 | m | 6.64 | 7 | 13, 25, 12, 6, 18, 19, 7 |
| 3.11 – 3.17 | m | 3.19 | 3 | 11, 23, 24 |
| 3.17 – 3.23 | m | 2.37 | 2 | 4, 16 |
| 3.32 – 3.38 | m | 2.30 | 2 | 5, 17 |
| 3.38 – 3.51 | m | 5.79 | 5 | 8A, 20A, 14A, 26A, 40A |
| 3.59 – 3.68 | m | 5.05 | 5 | 2, 8B, 20B, 14B, 26B |
| 3.70 – 3.78 | m | 1.49 | 1 | 40B |
| 4.25 | d | 1.24 (anomeric) | 1 | 15 |
| 4.30 | d | 1.04 (anomeric) | 1 | 3 |
| 4.36 | d | 2.20 (anomeric) | 1 | 21 |
| 4.38 | d | 2.20 (anomeric) | 1 | 9 |
| | | **Total:** | 60 | |

**Table SVI:** overview assignments $^{13}C$ chemical shifts

| $\delta$ $^{13}C$ (ppm) | Type of Carbon | # Carbons | Annotation |
|---|---|---|---|
| 21.35 | $CH_3$ | 1 | 1 |
| 24.65 | $CH_2$ | 1 | 28 |
| 25.56 | $CH_2$ | 1 | 38 |
| 29.06 – 29.38 | $CH_2$ | 10 | 29, 30, 31, 32, 33, 34, 35, 36, 37, 39 |
| 36.21 | $CH_2$ | 1 | 27 |
| 60.86 – 61.06 | $CH_2$ | 4 | 8, 14, 20, 26 |
| 68.72 | $CH_2$ | 1 | 40 |
| 69.76 – 69.99 | CH | 4 | 6, 12, 18, 24 |
| 74.91 | CH | 1 | 22 |
| 75.05 | CH | 1 | 10 |
| 75.95 | CH | 1 | 2 |
| 76.03 – 76.25 | CH | 4 | 5, 11, 17, 23 |
| 76.48 | CH | 1 | 7 |
| 76.61 | CH | 1 | 19 |
| 77.00 – 77.04 | CH | 2 | 13, 25 |
| 82.05 | CH | 1 | 4 |
| 82.32 | CH | 1 | 16 |
| 101.10 | CH (anomeric) | 1 | 3 |
| 101.35 | CH (anomeric) | 1 | 15 |
| | CH (anomeric) | 2 | 9, 21 |
| | **Total:** | 40 | |

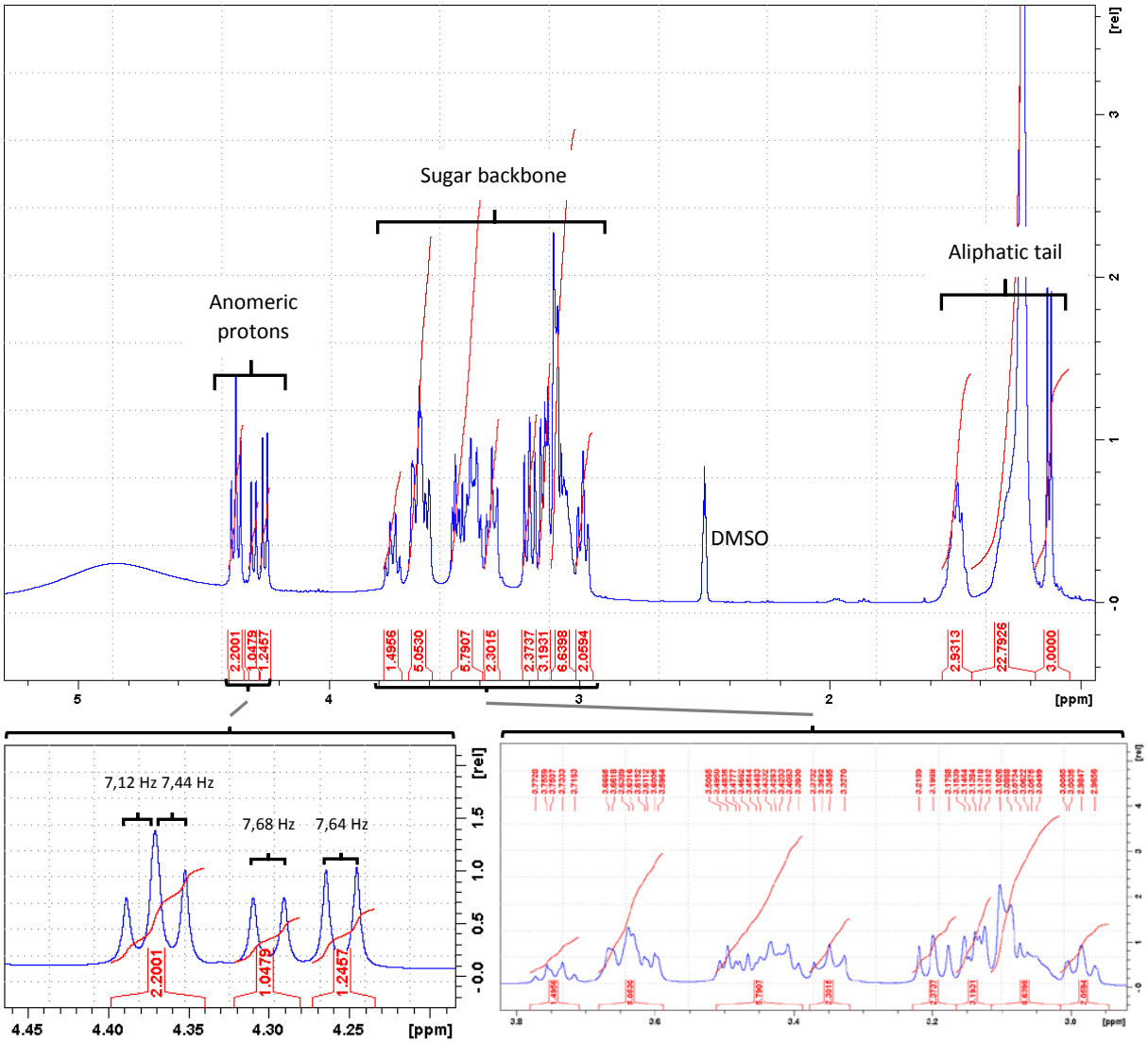



**Fig. S8:** $^{1}$H spectrum with zoom

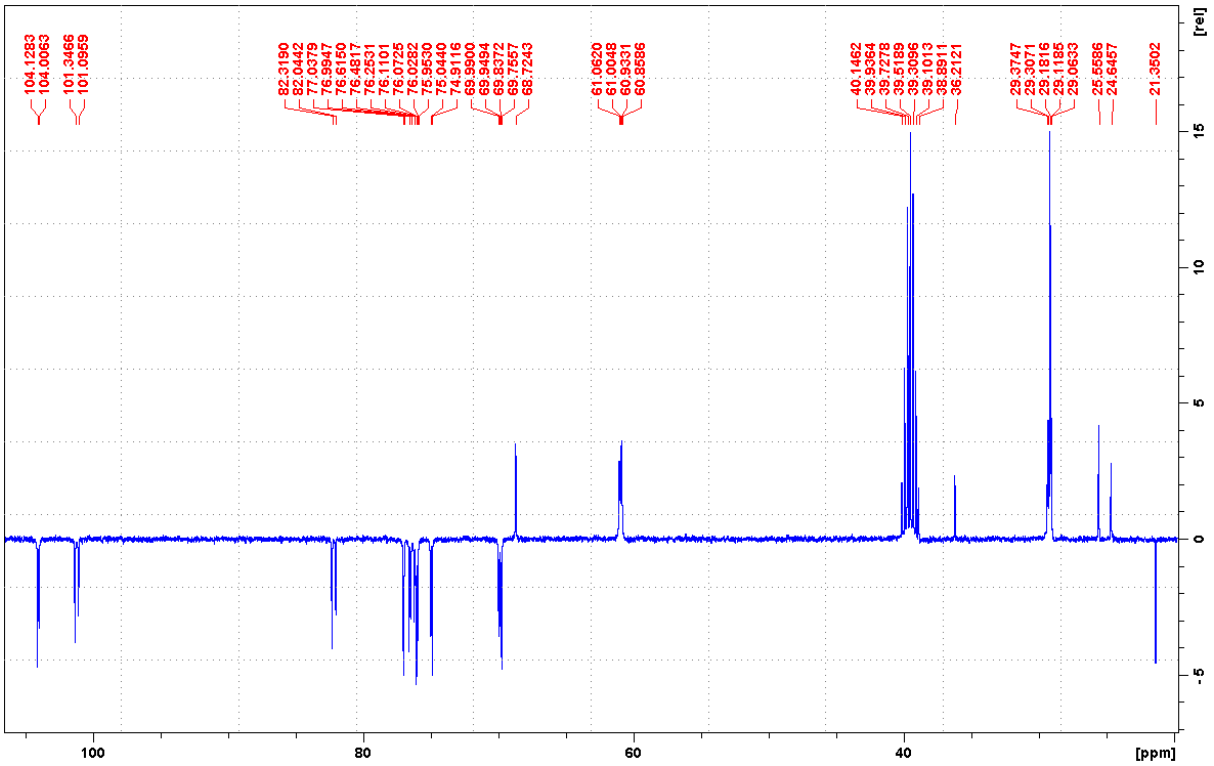



**Fig. S9:** APT $^{13}C$ spectrum ($CH_2$ carbons pointing upward and $CH/CH_3$ carbons pointing downward)

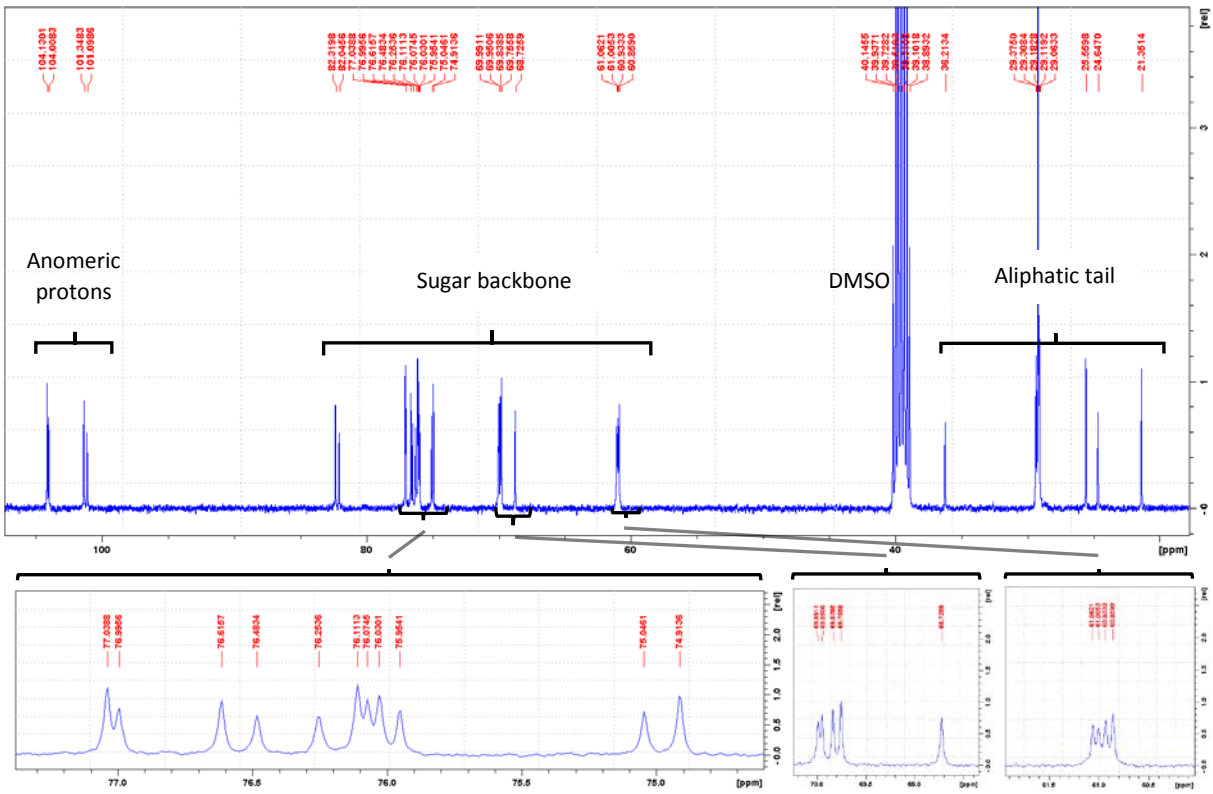



**Fig. S10:** $^{13}C$ spectrum with zoom

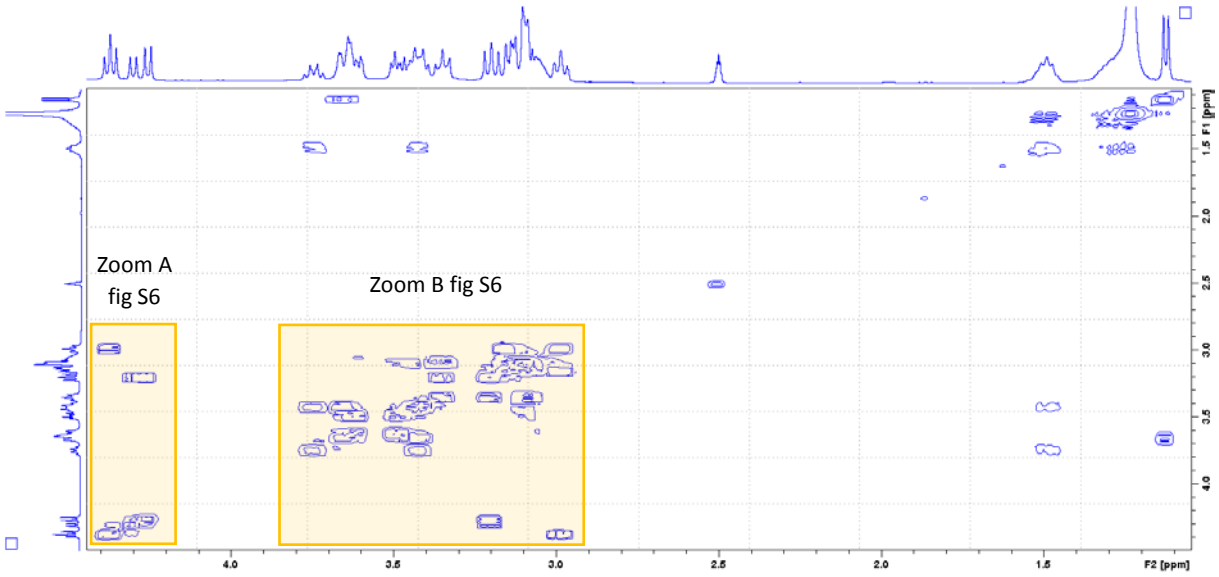



**Fig. S11:** COSY spectrum

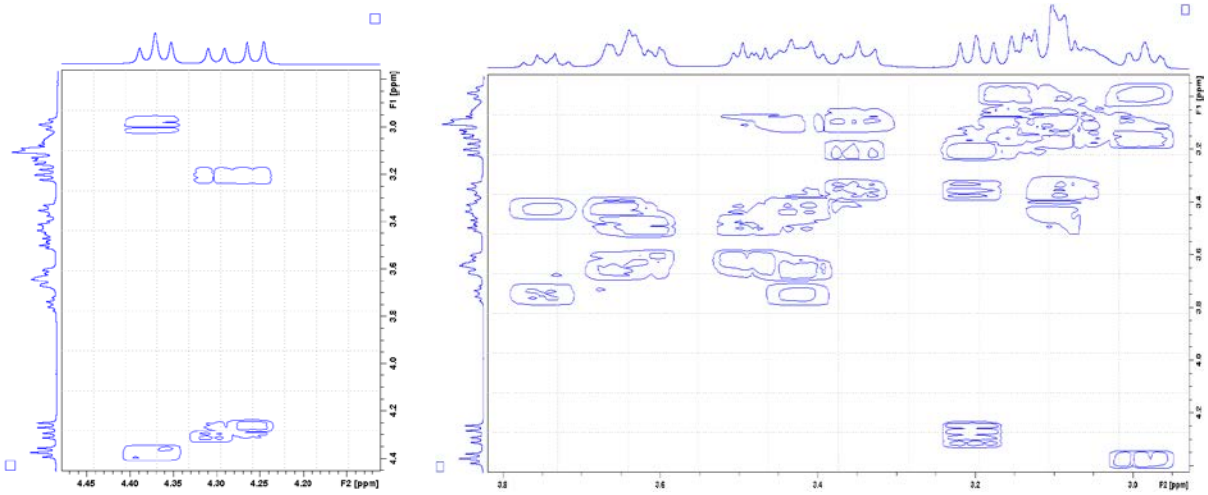



**Fig. S12:** COSY spectrum zoom A (left) and zoom B (right)


**Fig. S13:** HSQC (blue ($CH_2$)/ green (CH &$CH_3$)) + HMBC (red)

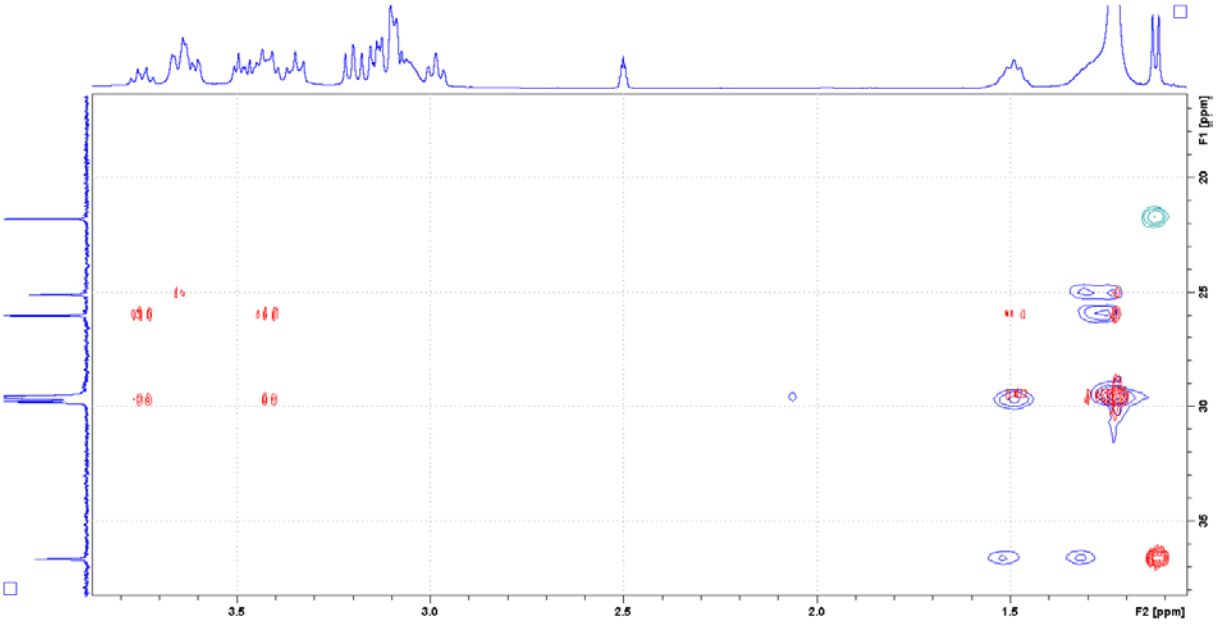



**Fig. S14:** HSQC/HMBC zoom


**Fig. S15:** HSQC/HMBC zoom

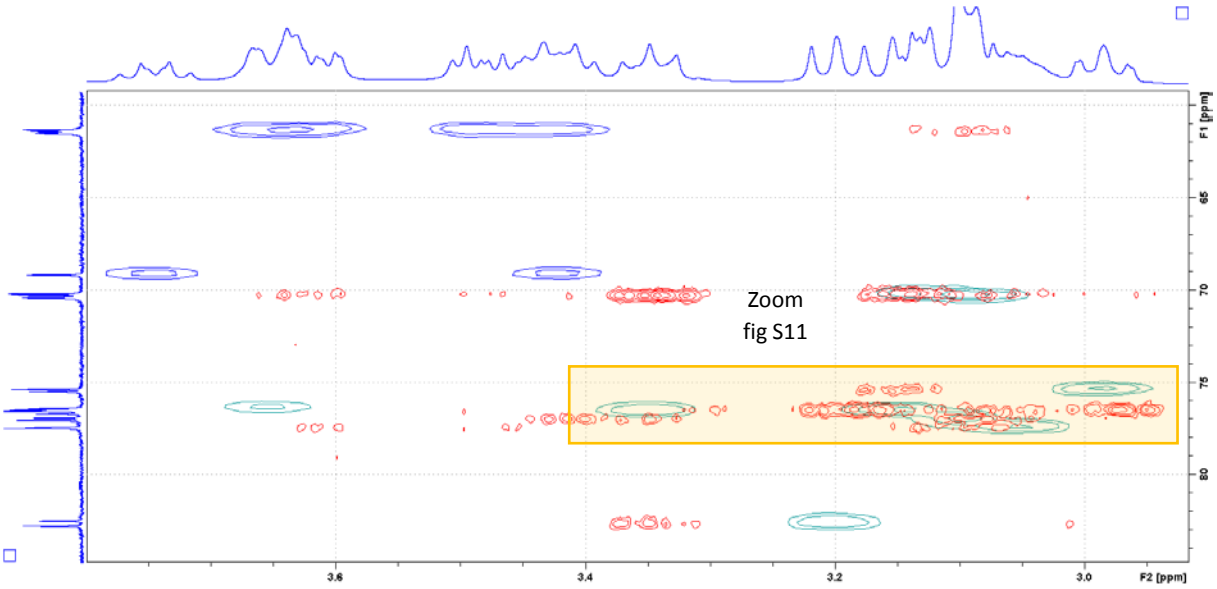



**Fig. S16:** HSQC/HMBC zoom

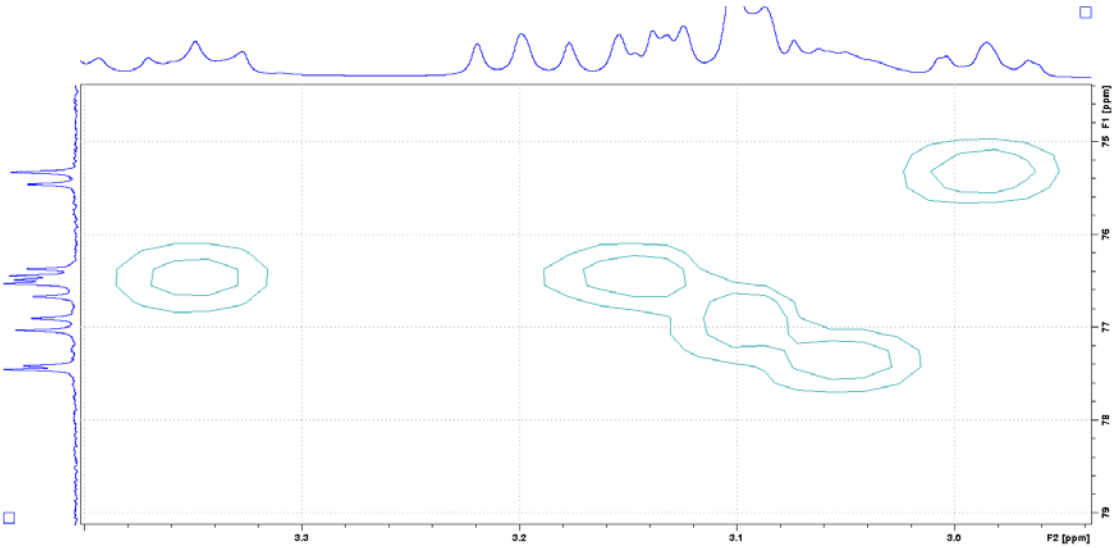



**Fig. S17:** HSQC zoom

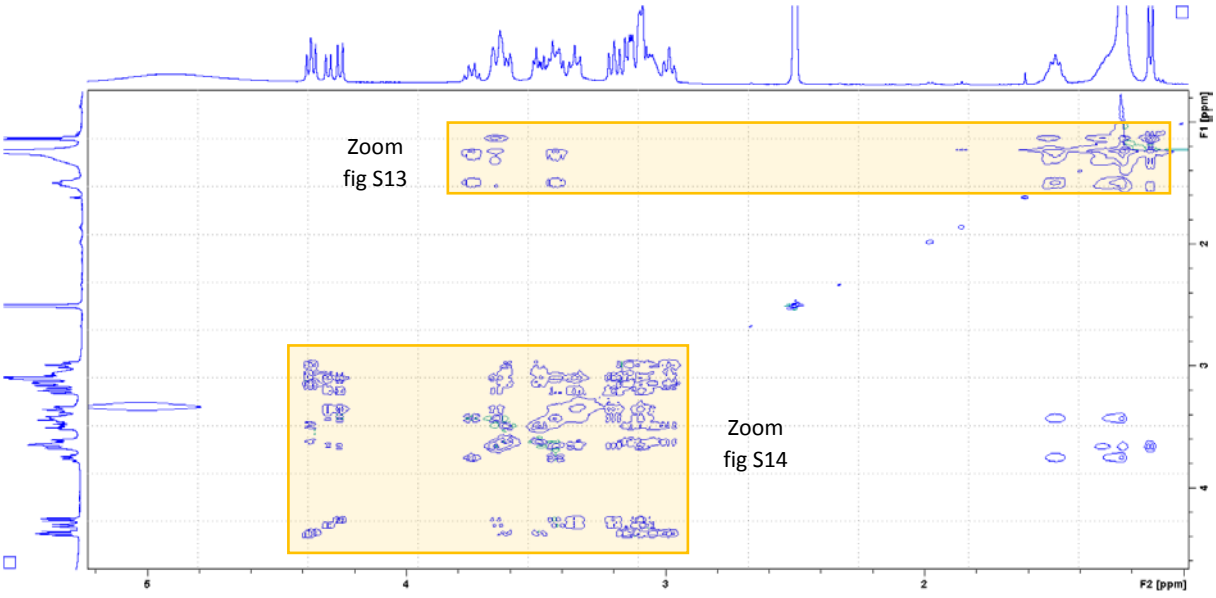



**Fig. S18:** TOCSY spectrum

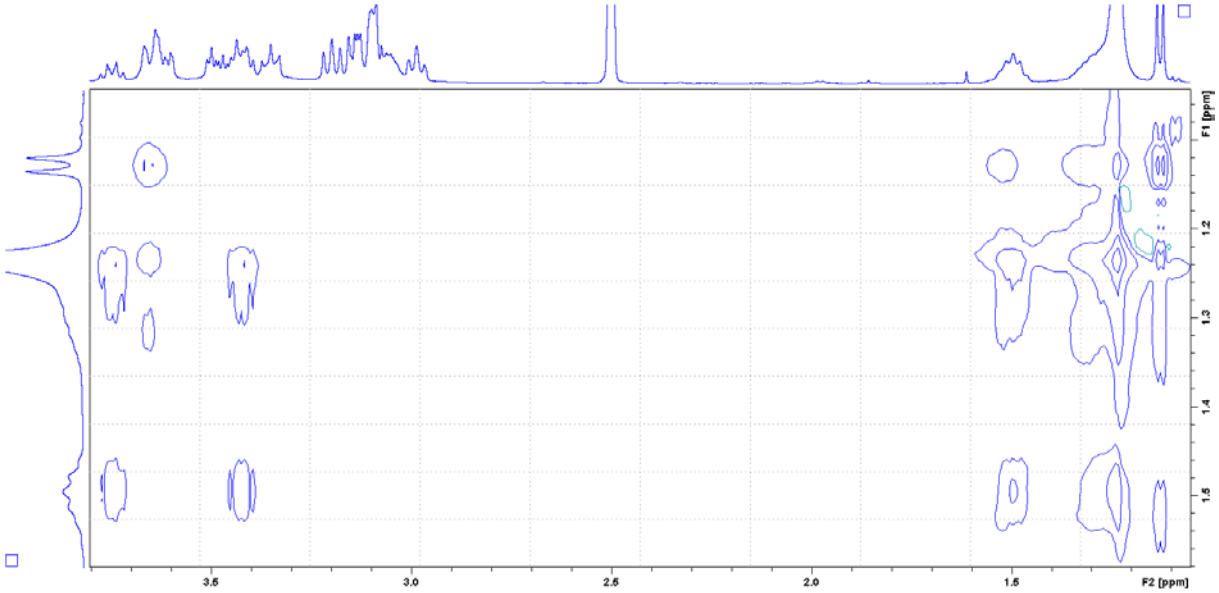



**Fig. S19:** TOCSY spectrum zoom

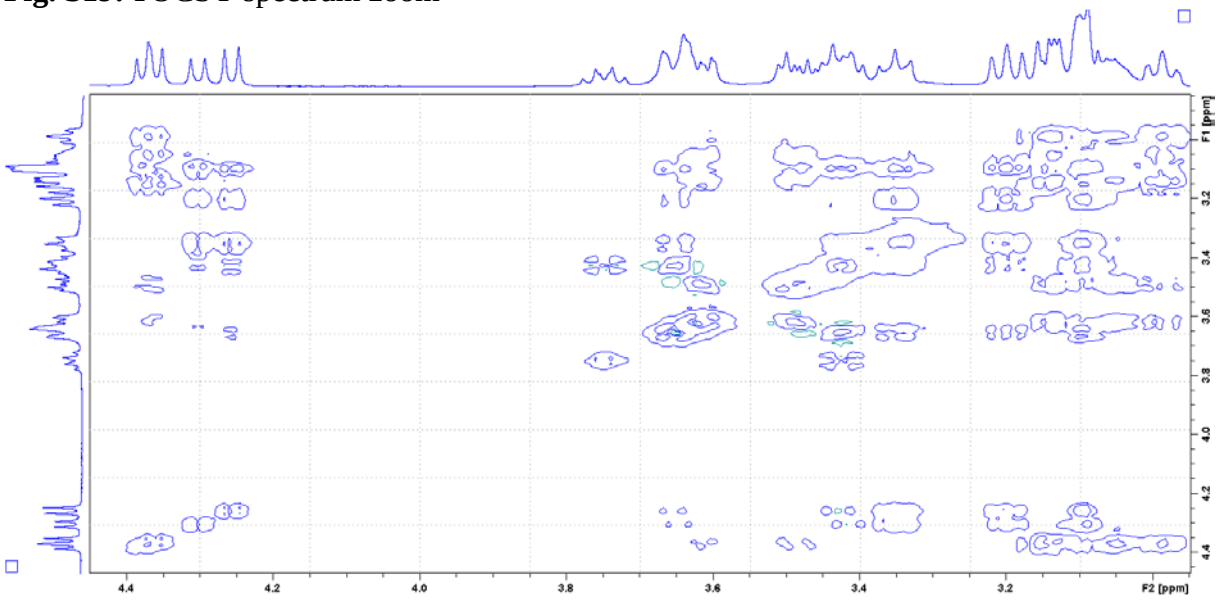



**Fig. S20:** TOCSY spectrum zoom


**Fig. S21:** H2BC (blue) + HSQC (red ($CH_2$)/ pink (CH & $CH_3$))

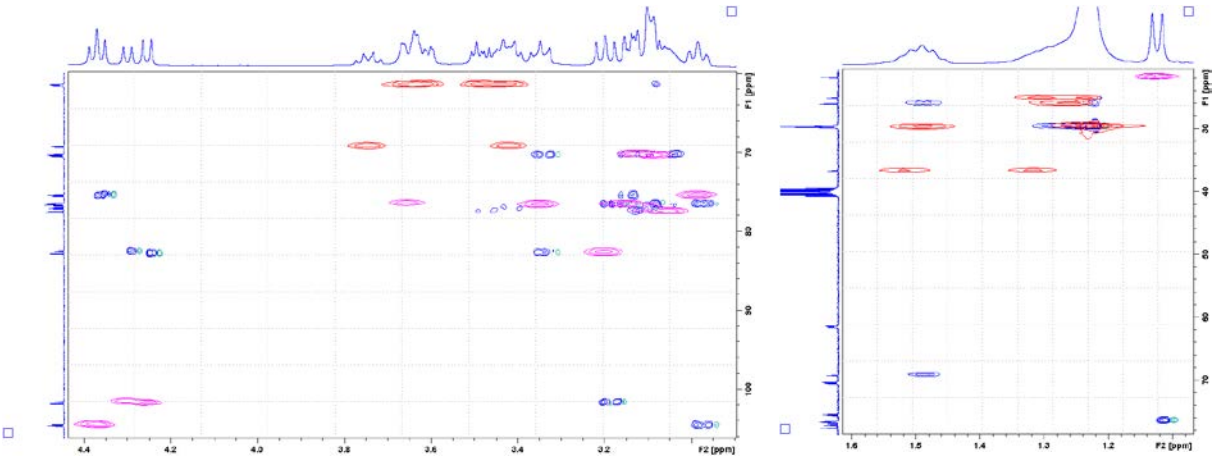
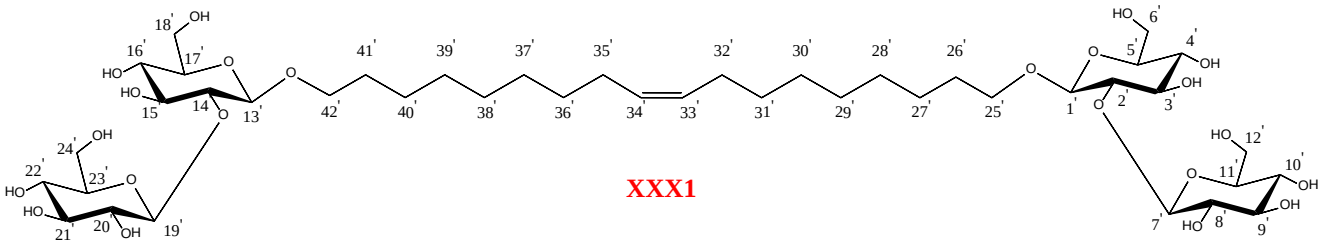



**Fig. S22:** H2BC/HSQC spectrum zoom A (left) and zoom B (right)

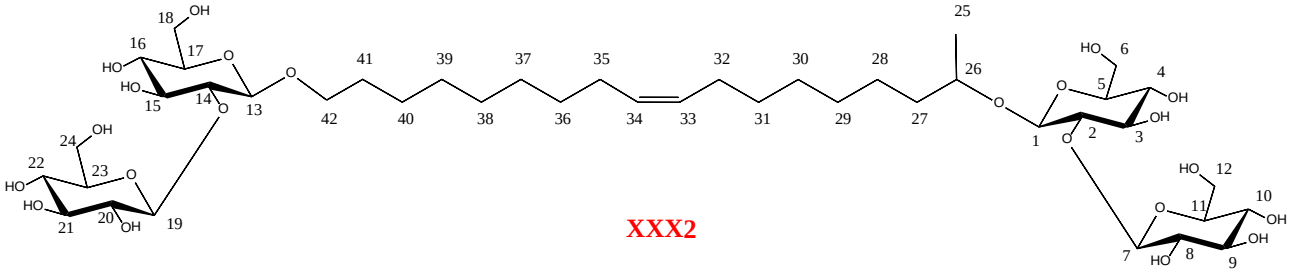
**2) Oleyl alcohol (C18:1-OH) based sBola SL**

For the symmetrical bolaform SLs derived from oleyl alcohol (C18:1-OH), the NMR results are summarized in Table SVII and SVIII. NMR analysis suggests that a mixture of two compounds is present in a 50/50 ratio (Fig. 6C and S23). Both compounds have the same chemical formula, $C_{42}H_{76}O_{22}$, but differ in how one of both ends of the fatty alkyl chain is linked to the second sophorose unit.

The presence of both bolaform tetraglycolipid structures will be highlighted here, through a few key observations. First of all, proton NMR is considered (Fig. S24). Both compounds have an alkene bond in the fatty acid chain (1H: 5.28 – 5.37 ppm, 13C: 129.67 ppm) for which the integration of the protons is calibrated to four (see Table SVII and SVIII for the $^{1}H$ and $^{13}C$ NMR spectra, respectively). This calibration shows that the integration of the methyl group ($^{1}H$: 1.13 ppm subterminal compound) equals approximately three. This indicates that both compounds are present in an equal ratio (50/50). Other integrations confirm this result as well. Furthermore, the protons from the hydroxyl groups of the sugar moieties are present in between 4.2 and 5.6 ppm. This range also includes the anomeric protons which overlap with some of the hydroxyl protons. Via $^{13}C$ NMR analysis, the presence of the four anomeric carbons is confirmed (Fig. S25). HSQC analysis shows that the hydroxyl protons do not couple to any carbon. From the COSY and H2BC spectrum, those hydroxyl protons could be assigned properly (Fig. S26 and S35). Second, the TOCSY matching approach (Petersen et al., 2006), validates that each of the sugar moieties corresponds with a D-glucose pyranose sugar (Fig. S32). Third, the analysis of the HMBC combined with HSQC confirms that sophorose units are present with a β-(1-2) connection (Fig. S28).



XXX1



XXX2

**Fig. S23:** Structures of the synthesized bolaform tetraglycolipid mixture with numbering

**Table SVII:** overview assignments $^{1}H$ chemical shifts

| δ $^{1}H$ (ppm) | Multiplicity | Integral | # Protons | Annotation |
|---|---|---|---|---|
| 1.13 | d | 3.29 | 3 | 25 |
| 1.2 – 1.34 | m | 46.09 | 39 | 27A, 27’, 28, 28’, 29, 29’, 30, 30’, 31, 31’, 36, 36’, 37, 37’, 38, 38’, 39, 39’, 40, 40’ |
| 1.44 – 1.55 | m | 8.34 | 7 | 26’, 27B, 41, 41’ |
| 1.92 – 2.03 | m | 8.03 | 8 | 32, 32’, 35, 35’ |
| 2.94 – 3.02 | m | 4.55 | 4 | 8, 8’, 20, 20’ |
| 3.02 – 3.17 | m | 22.47 | 20 | 4, 4’, 5, 5’, 9, 9’, 10, 10’, 11, 11’, 16, 16’ 17, 17’, 21, 21’, 22, 22’, 23, 23’ |
| 3.17 – 3.23 | m | 5.03 | 4 | 2, 2’, 14, 14’ |
| 3.27 – 3.55 | m | 76.67 | 15 + $H_2O$ | 3, 3’, 6A, 6A’, 12A, 12A’, 15, 15’, 18A, 18A’, 24A, 24A’ 25A’, 42A, 42A’ |
| 3.55 – 3.70 | m | 9.91 | 9 | 6B, 6B’, 12B, 12B’, 18B, 18B’, 24B, 24B’, 26 |
| 3.70 – 3.80 | m | 3.66 | 3 | 25B’, 42B, 42B’ |
| 4.22 – 4.40 | m | 13.38 | 12 | 1, 1’, 7, 7’, 13, 13’, 19, 19’, 12OH, 12OH’, 24OH, 24OH’ |
| 4.45 | t | 0.96 | 1 | 6OH |
| 4.52 | t | 3.53 | 3 | 6OH’, 18OH, 18OH’ |
| 4.89 | d | 4.40 | 4 | 10OH, 10OH’, 22OH, 22OH’ |
| 4.96 | d | 4.44 | 4 | 9OH, 9OH’, 21OH’, 21OH’ |
| 5.02 – 5.09 | m | 4.45 | 4 | 4OH, 4OH’,16OH, 16OH’ |
| 5.19 | d | 3.56 | 3 | 8OH |
| 5.22 | d | 0.85 | 1 | 8OH’, 20OH, 20OH’ |
| 5.28 – 5.37 | m | 4 | 4 | 33, 33’, 34, 34’ |

| | | | | |
|---|---|---|---|---|
| 5.45 | d | 3.56 | 3 | 3O<u>H</u> |
| 5.52 | d | 0.94 | 1 | 3O<u>H</u>', 15O<u>H</u>, 15O<u>H</u>' |
| | | **Total:** | 152 | |

**Table SVIII:** Overview assignments $^{13}C$ chemical shifts

| **δ $^{13}C$ (ppm)** | **Type of Carbon** | **# Carbons** | **Annotation** |
|---|---|---|---|
| 21.32 | $CH_3$ | 1 | 25 |
| 24.97 | $CH_2$ | 1 | 28 |
| 25.53 | $CH_2$ | 3 | 27', 40, 40' |
| 26.67 | $CH_2$ | 4 | 33, 33', 34, 34' |
| 28.71 – 29.28 | $CH_2$ | 18 | 26', 28', 29, 29', 30, 30', 31, 31', 36, 36', 37, 37', 38, 38', 39, 39', 41, 41' |
| 36.18 | $CH_2$ | 1 | 27 |
| 60.81 – 61.01 | $CH_2$ | 8 | 6, 6', 12, 12', 18, 18', 24, 24' |
| 68.69 | $CH_2$ | 3 | 25', 42, 42' |
| 69.69 – 69.90 | CH | 8 | 4, 4', 10, 10', 16, 16', 22, 22' |
| 74.90 – 75.04 | CH | 4 | 8, 8', 20, 20' |
| 75.91 | CH | 1 | 26 |
| 76.00 – 76.20 | CH | 8 | 3, 3', 9, 9', 15, 15', 21, 21' |
| 76.47 – 76.61 | CH | 4 | 5, 5', 17, 17' |
| 77.04 | CH | 4 | 11, 11', 23, 23' |
| 82.07 – 82.32 | CH | 4 | 2, 2', 14, 14' |
| 101.08 | CH (anomeric) | 1 | 1 |
| 101.32 | CH (anomeric) | 3 | 1', 13, 13' |
| 104.13 | CH (anomeric) | 4 | 7, 7', 19, 19' |
| 129.67 | CH (alkene) | 4 | 33, 33', 34, 34' |
| | **Total:** | 84 | |

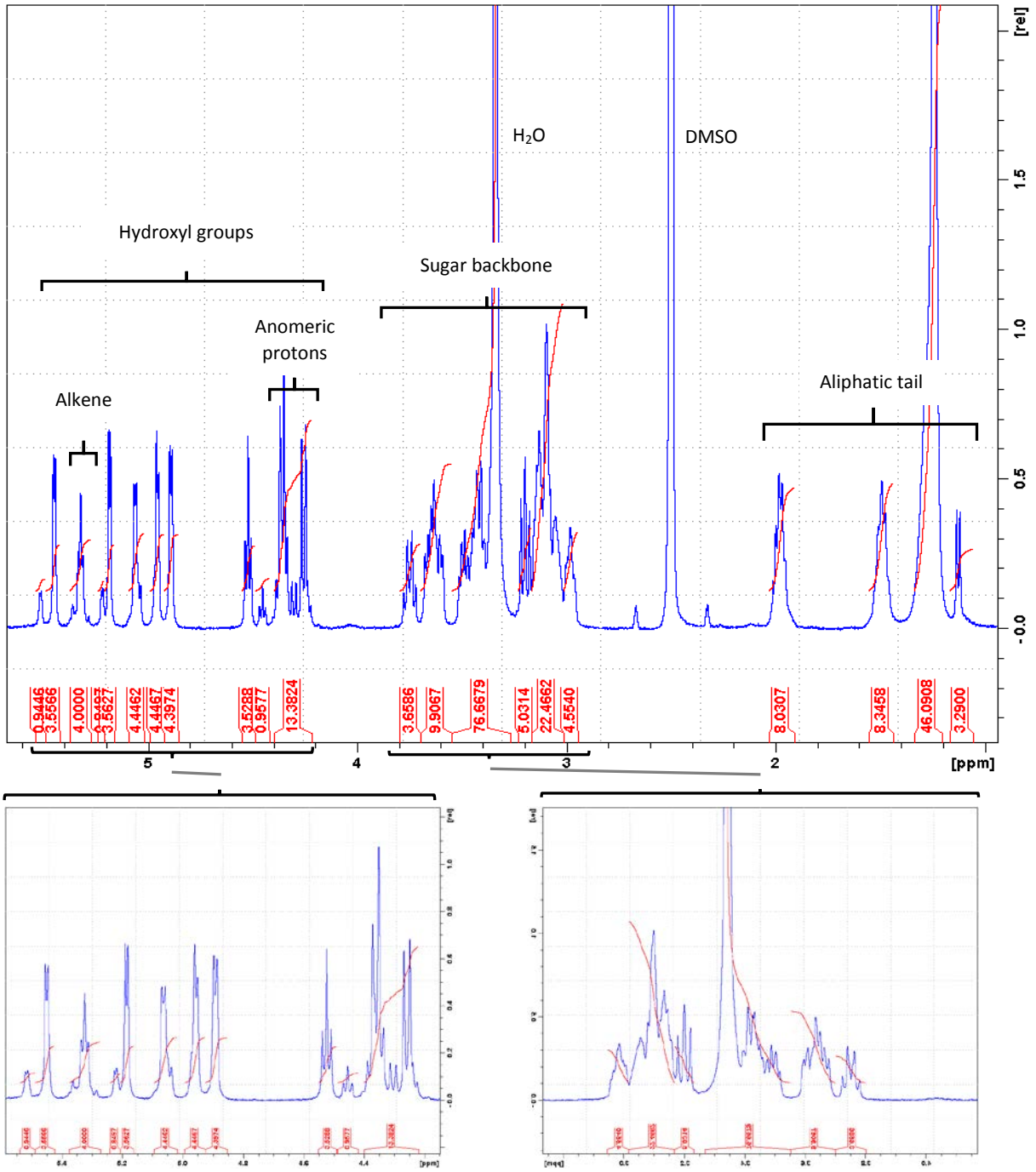



**Fig. S24:** $^{1}$H spectrum with zoom

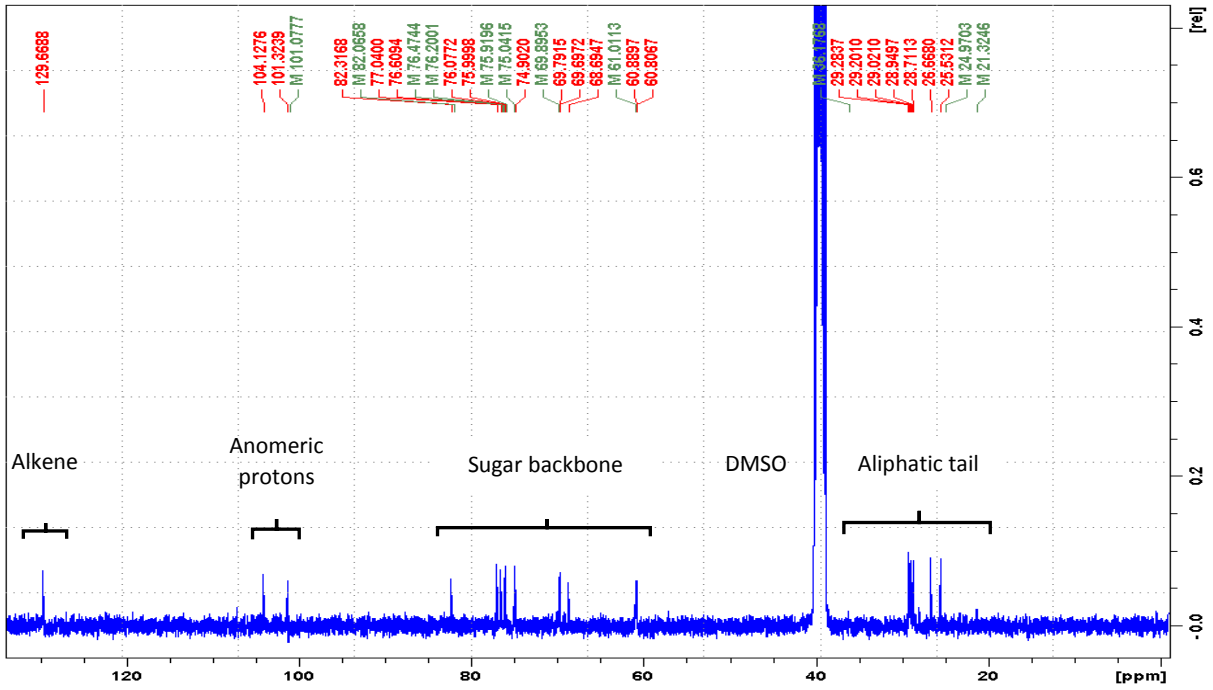



**Fig. S25:** $^{13}C$ spectrum

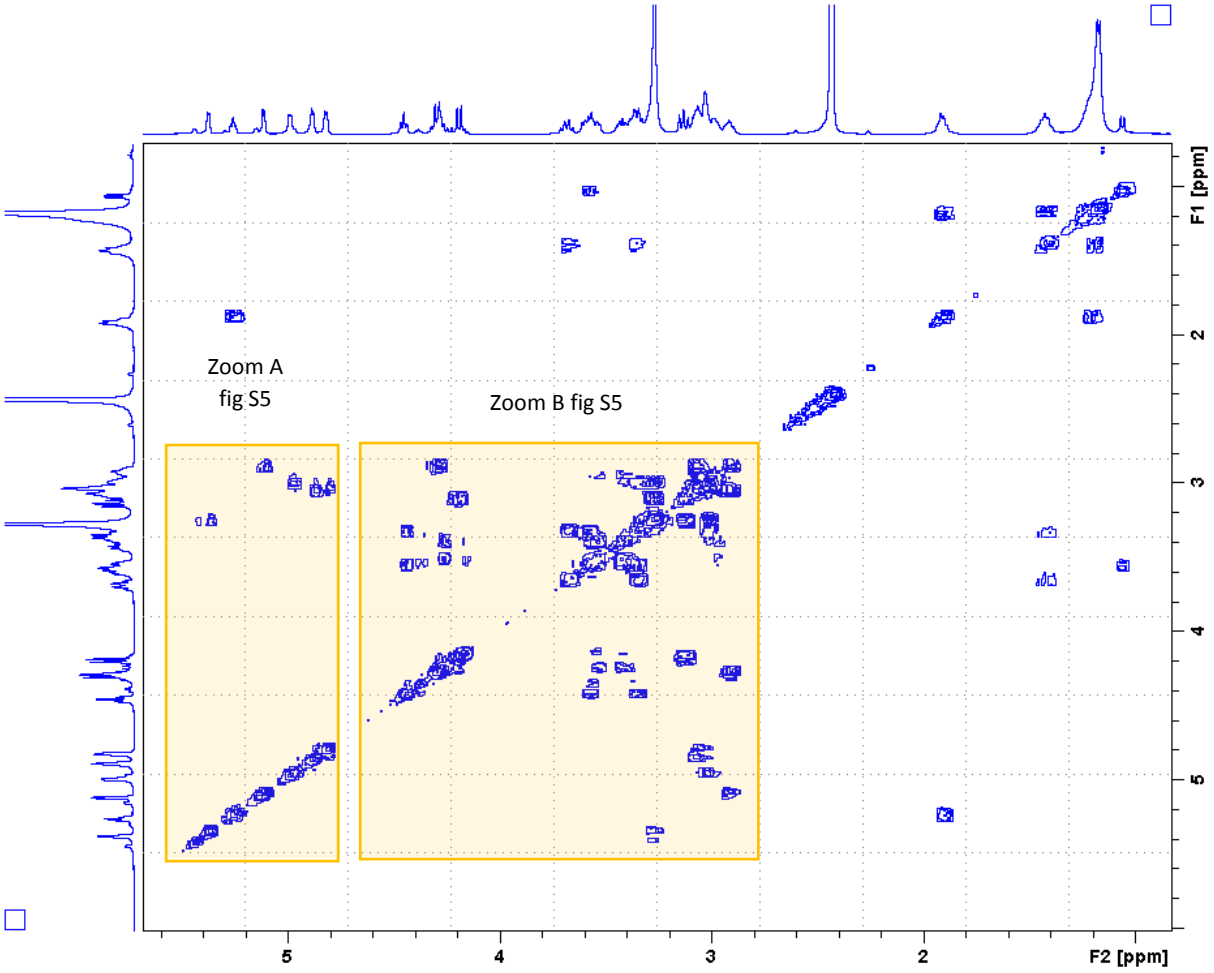



**Fig. S26:** COSY spectrum

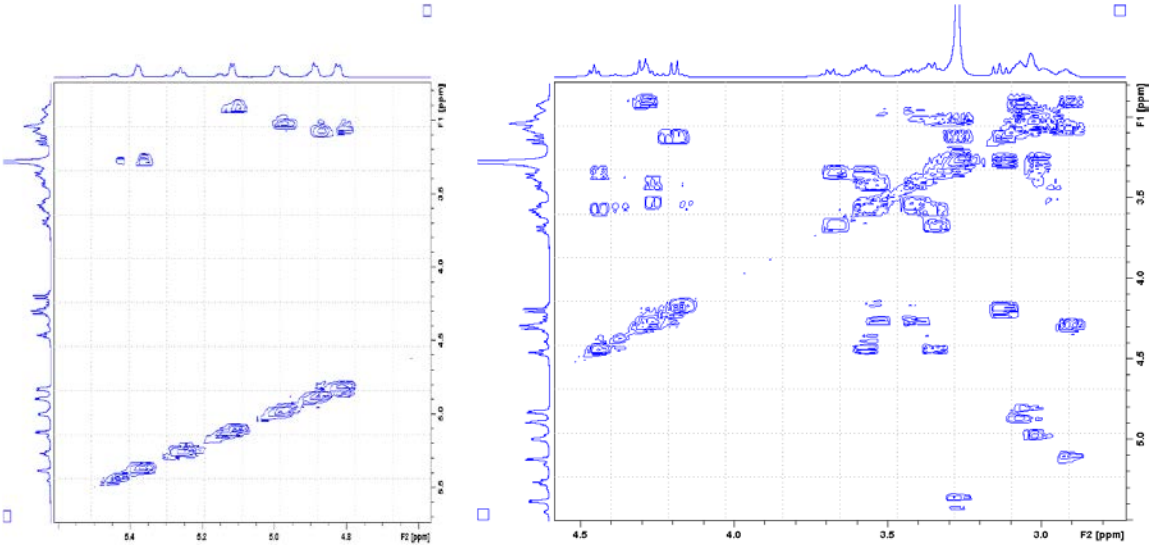


**Fig. S27:** COSY spectrum zoom A (left) and zoom B (right)

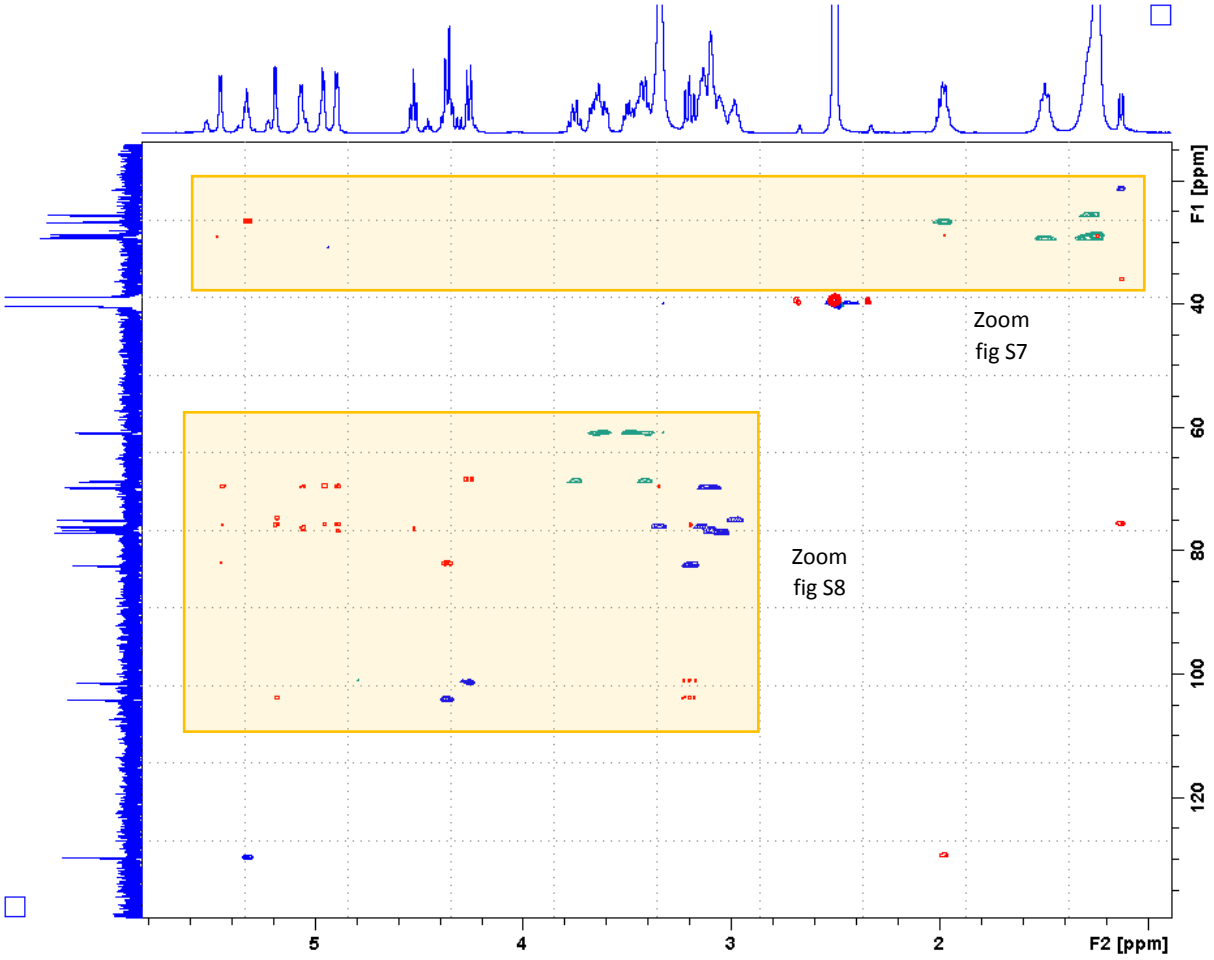



**Fig. S28:** HSQC (blue ($CH_2$)/ green (CH &$CH_3$)) + HMBC (red)

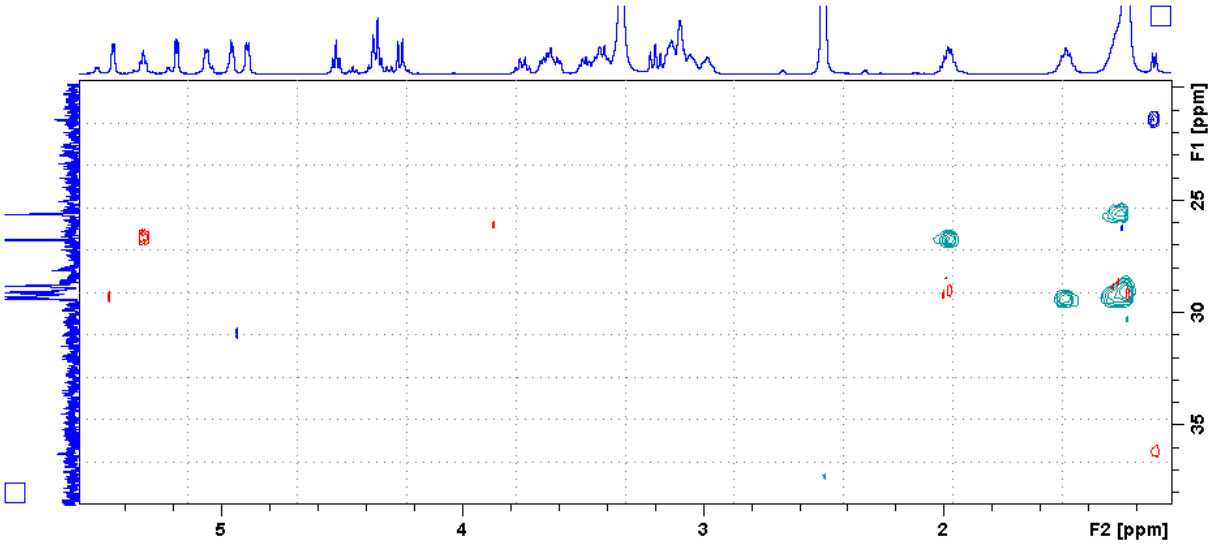



**Fig. S29:** HSQC/HMBC zoom

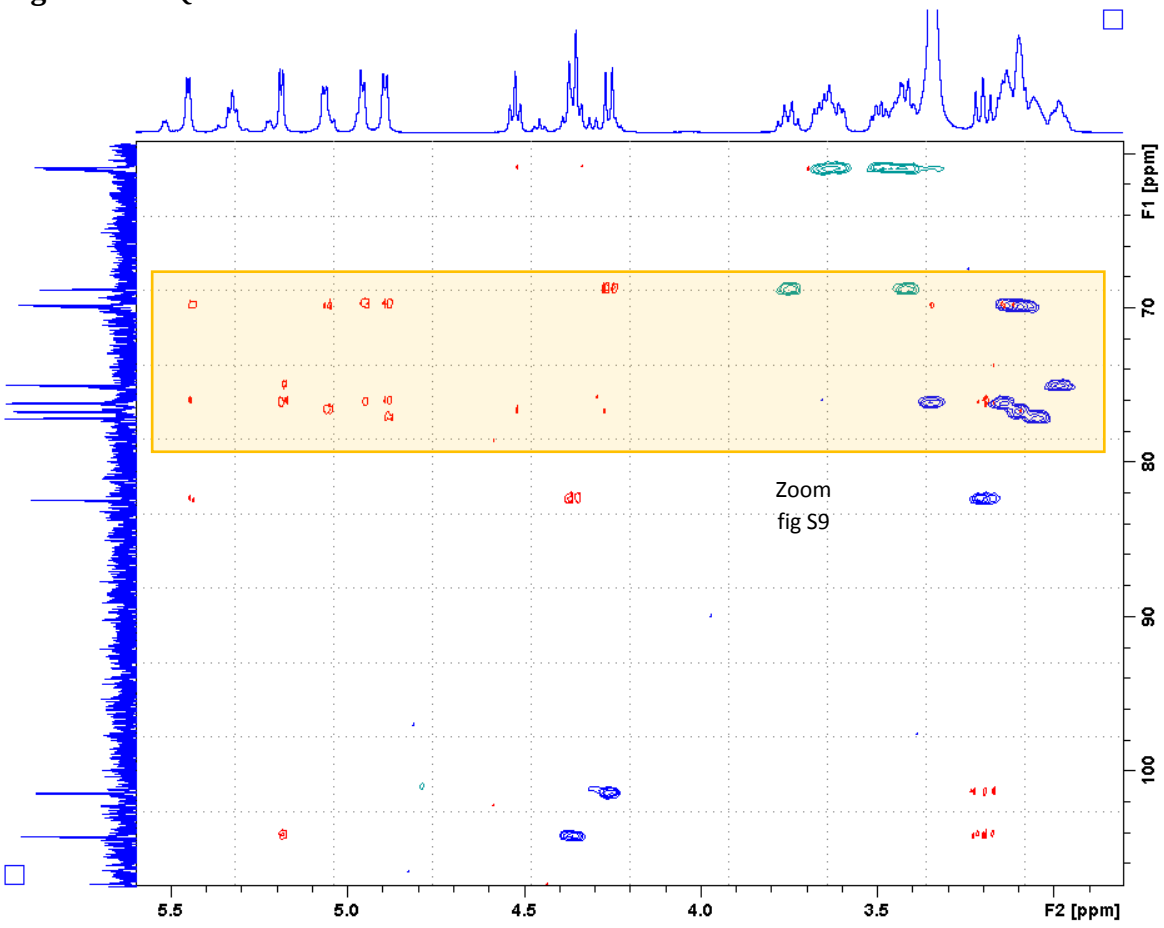



**Fig. S30:** HSQC/HMBC zoom

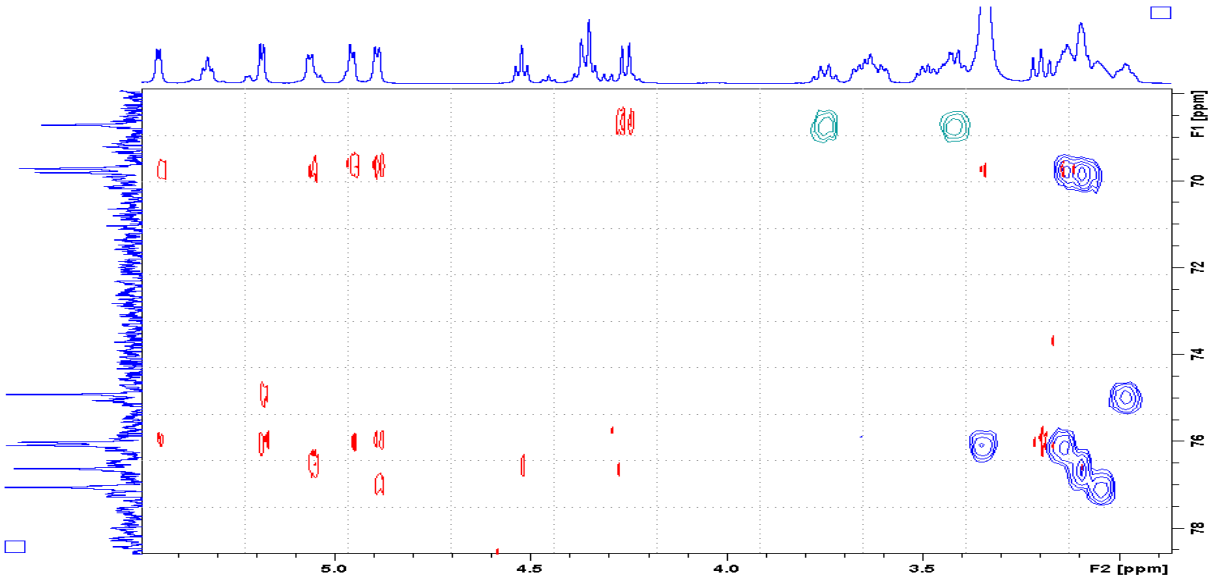



**Fig. S31:** HSQC/HMBC zoom

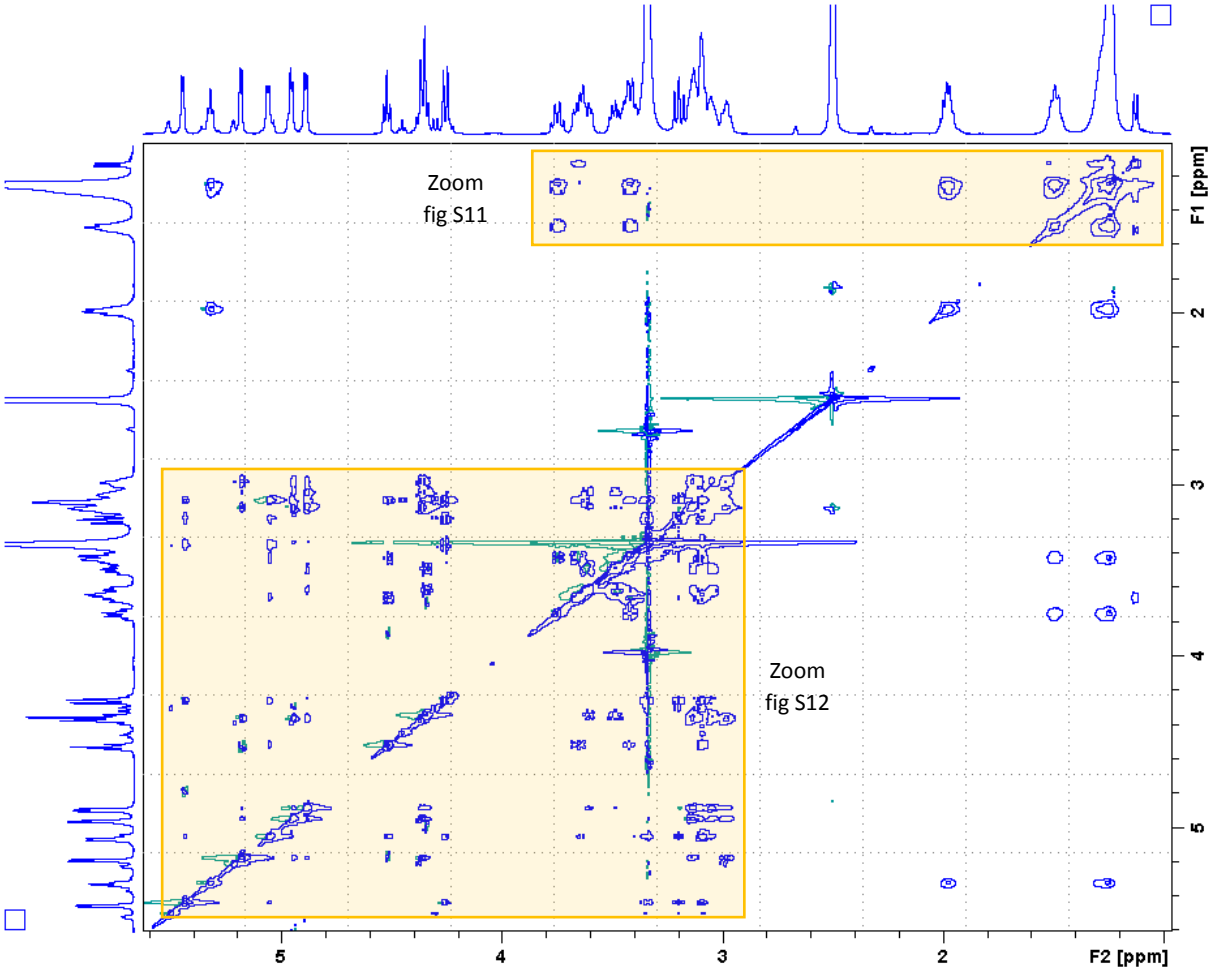



**Fig. S32:** TOCSY spectrum

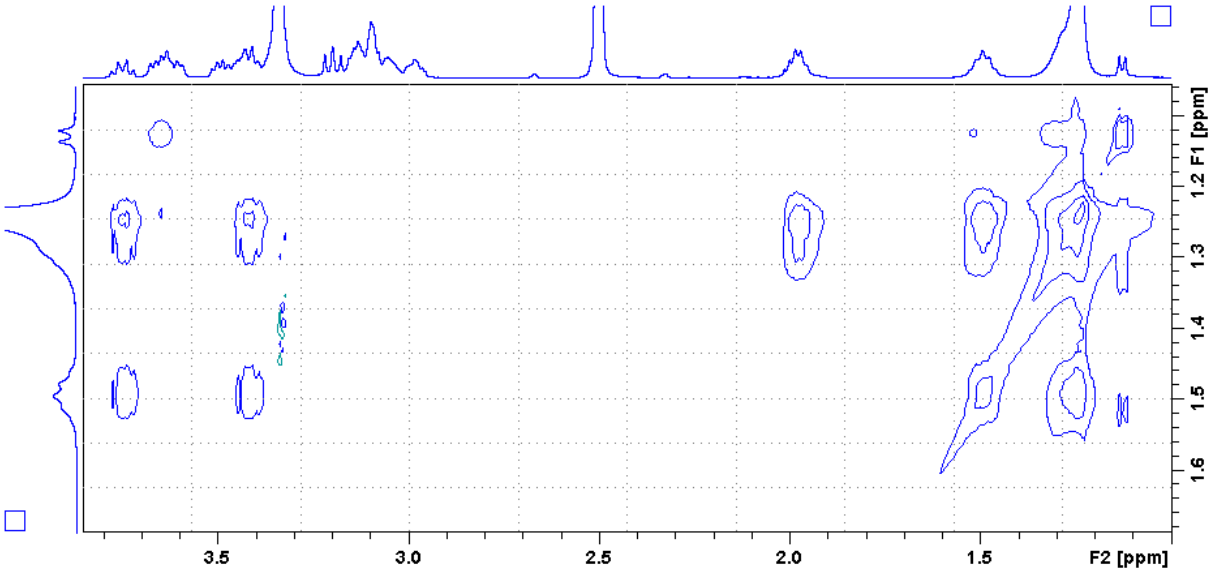



**Fig. S33:** TOCSY spectrum zoom

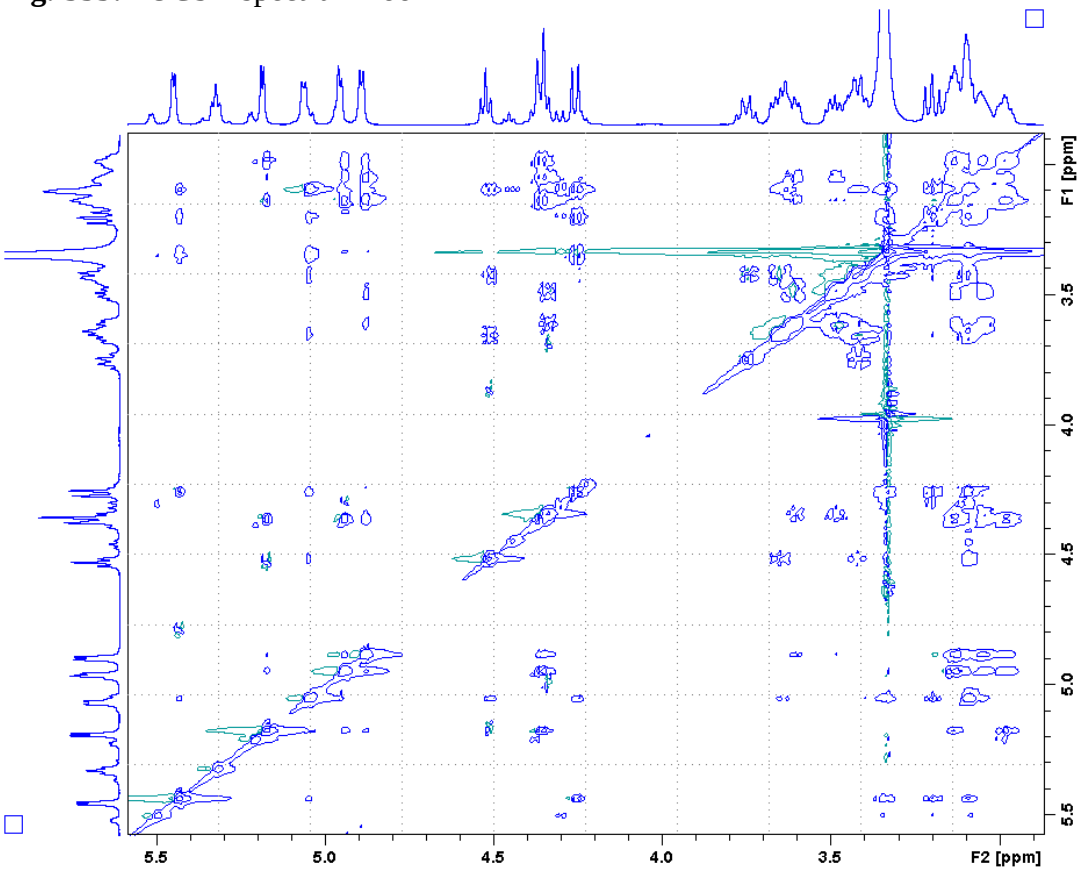



**Fig. S34:** TOCSY spectrum zoom

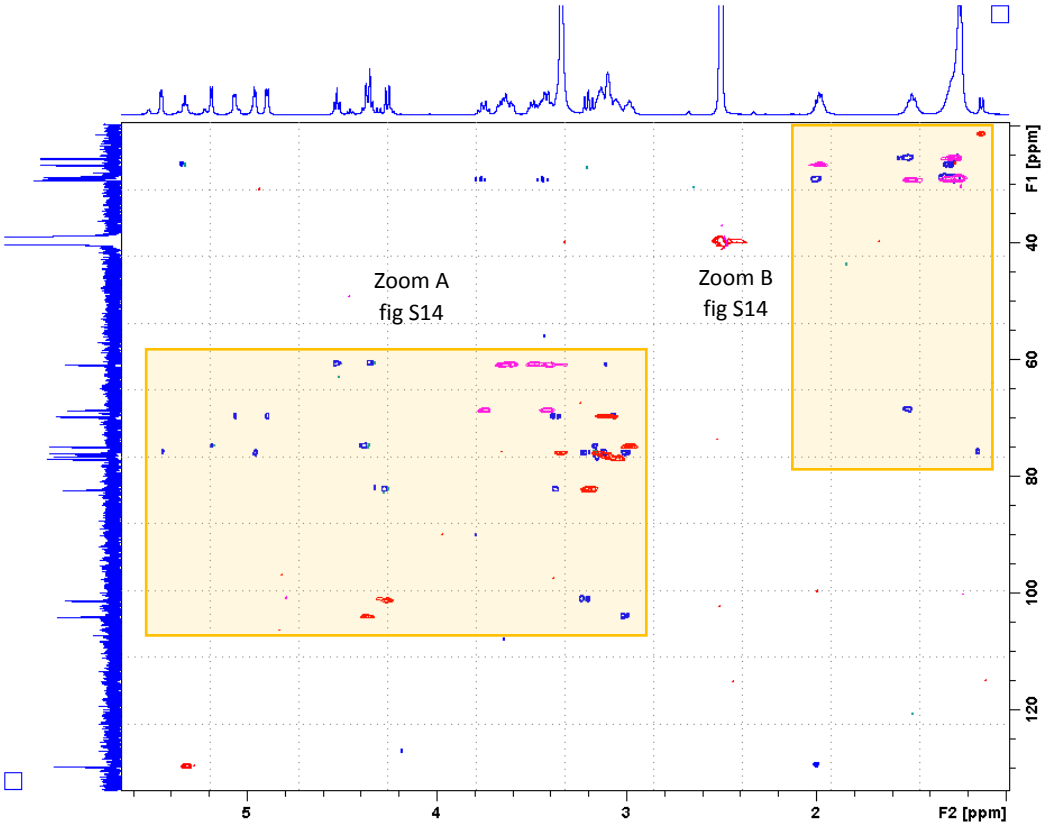



**Fig. S35:** H2BC (blue) + HSQC (red ($CH_2$)/ pink (CH & $CH_3$))

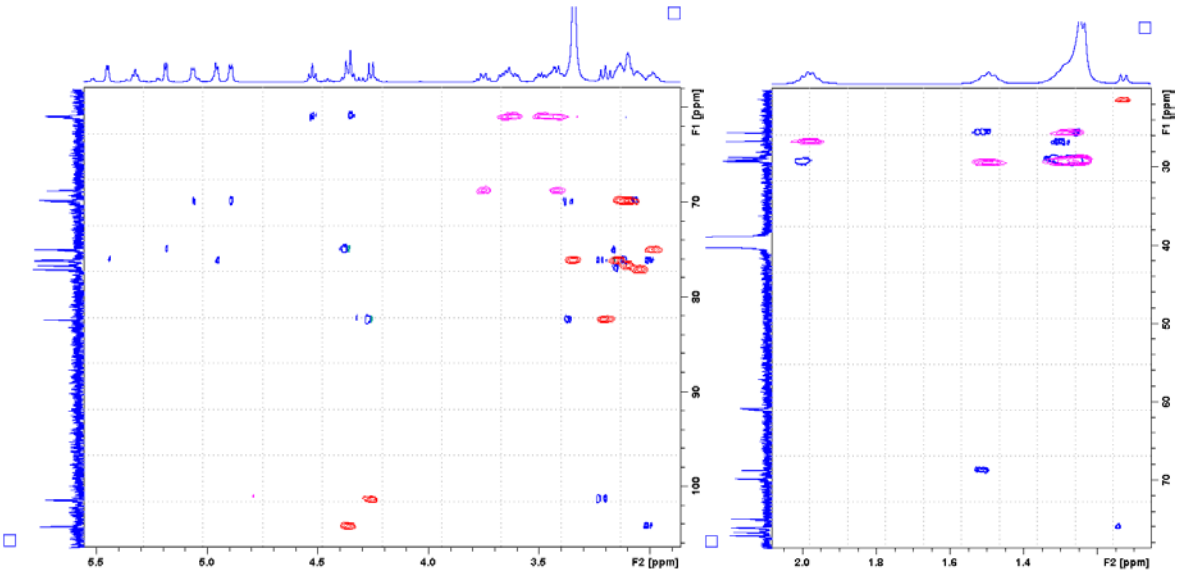



**Fig. S36:** H2BC/HSQC spectrum zoom A (left) and zoom B (right)

## References

Baccile N, Nassif N, Malfatti L, Van Bogaert INA, Soetaert W, Pehau-Arnaudet G, Babonneau F. (2010). Sophorolipids: a yeast-derived glycolipid as greener structure directing agents for self-assembled nanomaterials. Green Chemistry 12:1564-1567.

Green MR, Sambrook J. 2012. Molecular cloning: a laboratory manual (Fourth edition). New York: Cold Spring Harbor Laboratoy Press, 2028p.

Manet S, Cuvier A-S, Valotteau C, Fadda GC, Perez J, Karakas E, Abel S, Baccile N. (2015). Structure of bolaamphiphile sophorolipid micelles characterized with SAXS, SANS, and MD simulations. Physical Chemistry B 119:13113-13133.

Nicolai T, Colombani O, Chassenieux C. (2010). Dynamic polymeric micelles versus frozen nanoparticles formed by block copolymers. Soft Matter 6:3111–3118.

Nagarajan R. (1987). Self-assembly of bola amphiphiles. Chemical Engineering Communications 55: 251-273.

Percus JK, Yevick J. (1958). Analysis of classical statistical mechanics by means of collective coordinates. Physics Review 110: 1-13.

Saerens KMJ. 2012. Synthesis of glycolipids by *Candida bombicola*. PhD, Ghent University. 262p.