Melamine-cored Glucosides for Membrane Protein Solubilization and Stabilization: Importance of Water-mediated Intermolecular Hydrogen Bonding in Detergent Performance

Supplementary Information

Table of Contents

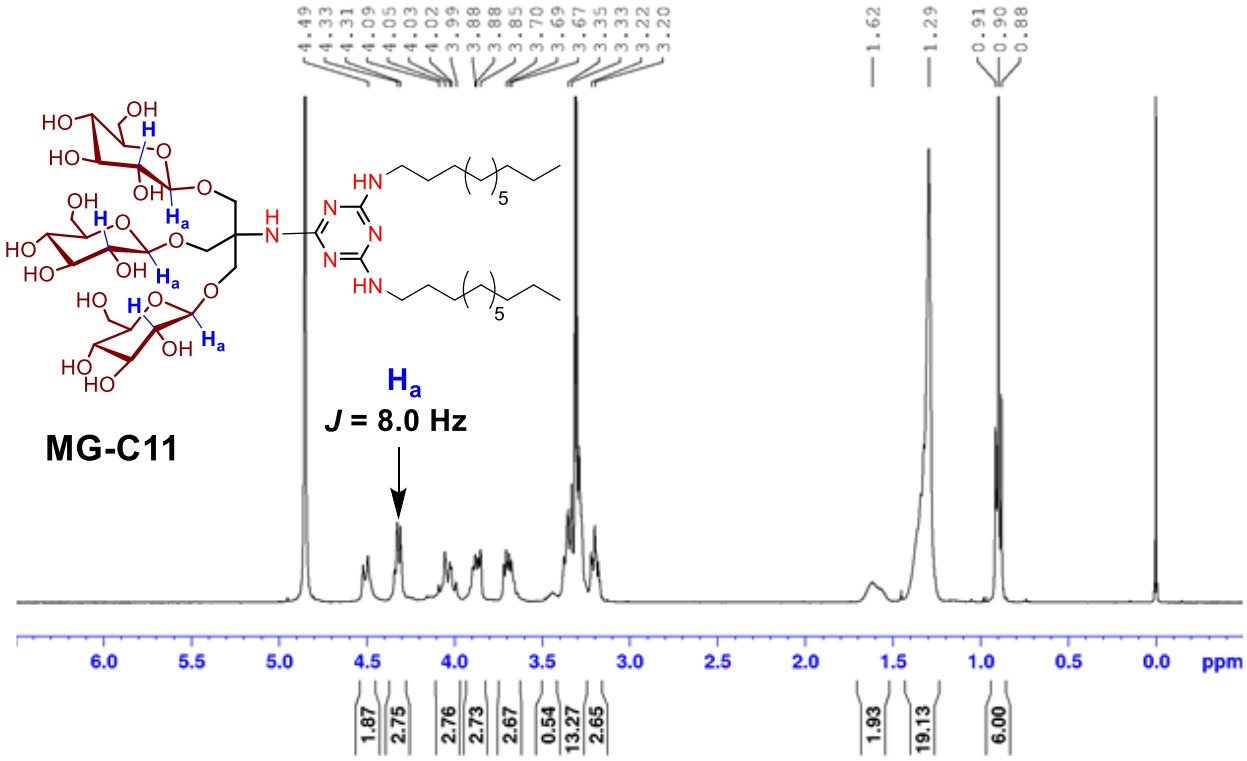


Figure S1. ^{1}H NMR spectra of MG-C11 dissolved in CD_3OD at room temperature. The NMR peaks for the *β*-anomeric protons of MG-C11, indicated by H_a, appeared as a triplet at 4.31 ppm, with a vicinal coupling constant (*J*) of 8.0 Hz.

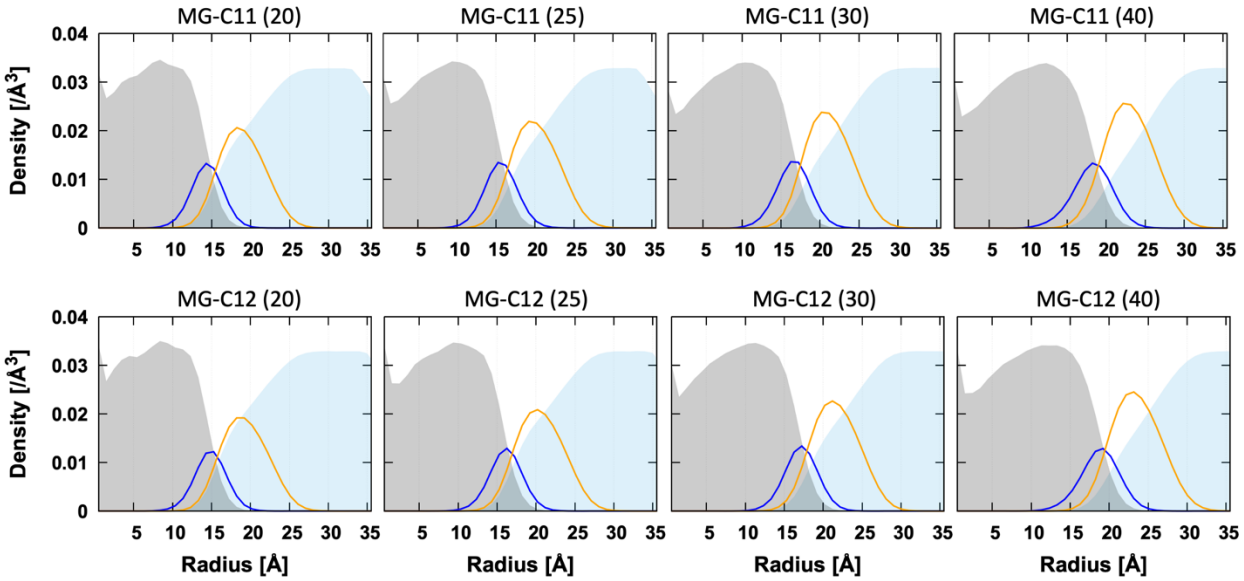


Figure S2. Density profiles for the MG micelle from the micelle center of mass: water (skyblue), the alkyl tail (gray), TRIS-triazine (blue), and glucoside (orange), and. The number in parentheses is the aggregation number of each micelle.

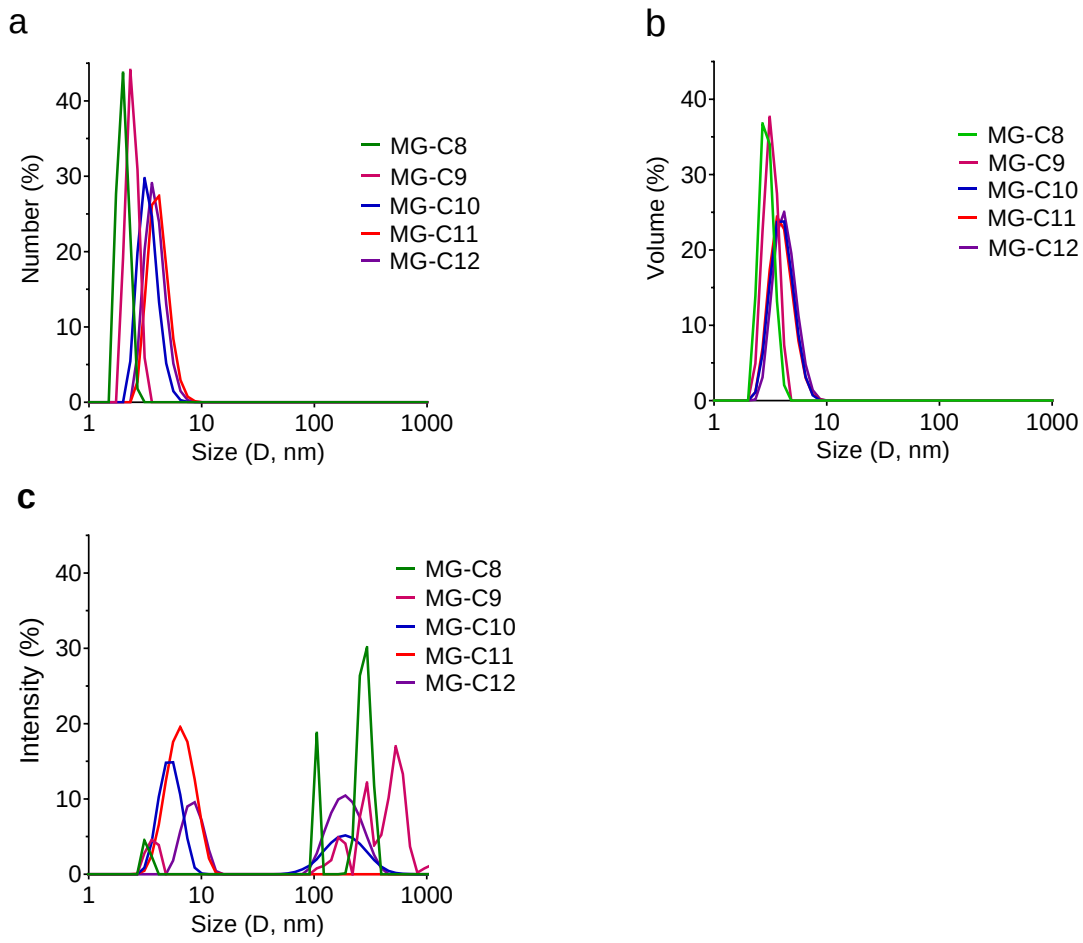


Figure S3. (a) Number- (b) volume- and (c) intensity-weighted DLS profiles of the MGs. Hydrodynamic radii of detergent micelles were measured via DLS at a detergent concentration of 1.0 wt%. Time-dependent fluctuations in the scattered light intensity were analyzed by an autocorrelation function to obtain diffusion coefficients and R_h of detergent micelles.

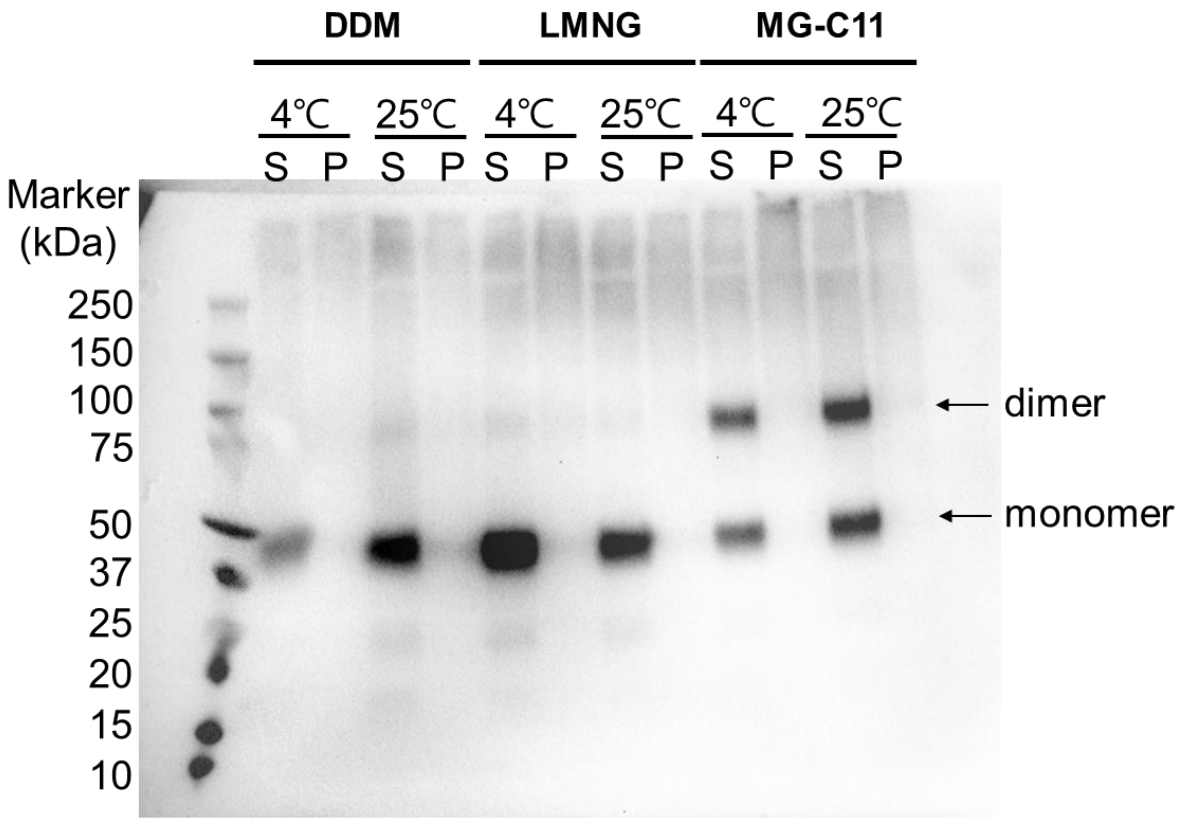


Figure S4. Western blotting of β_2AR extracted by DDM, LMNG and MG-C11. The receptor was extracted using 1.0 wt% individual detergents at both 4 and 25 °C. The amount of solubilized receptor in each sample was estimated by the band density after protein separation based on their molecular weights via SDS-PAGE. The bands corresponding to the monomeric and dimeric receptors were indicated in arrow. The known molecular weights of the monomeric and dimeric β_2AR are ~ 46 and ~92 kDa, respectively. S and P represent the supernatant and pellet, respectively, obtained after ultracentrifugation of the extracted protein samples.

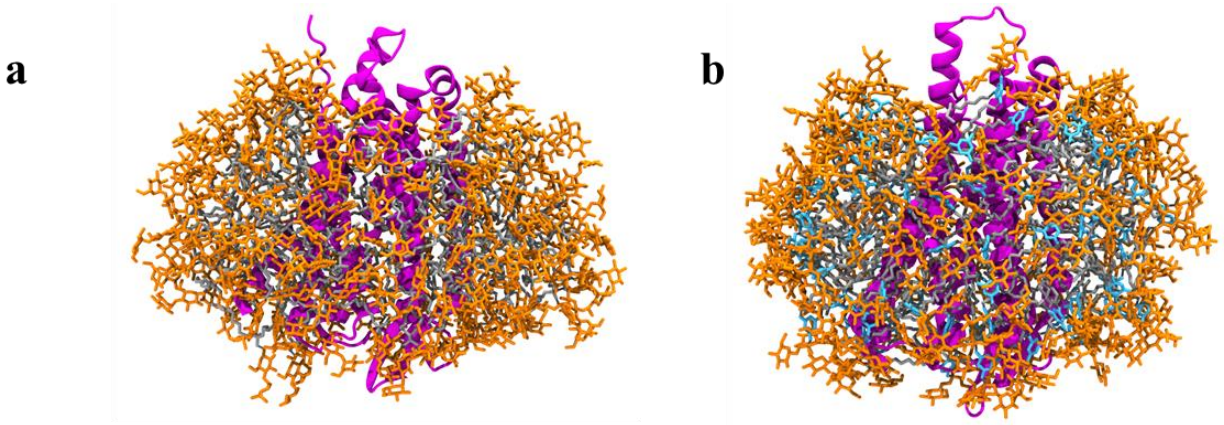


Figure S5. MOR in complex with LMNG (a) or MG-C11 micelles (b) obtained from 500 ns-long MD simulations. 96 detergent molecules were used for these simulations. The receptor helices are represented in purple, while the alkyl chains and carbohydrates of the detergent molecules are indicated in grey and orange, respectively. The melamine core of MG-C11 is indicated in light-blue.

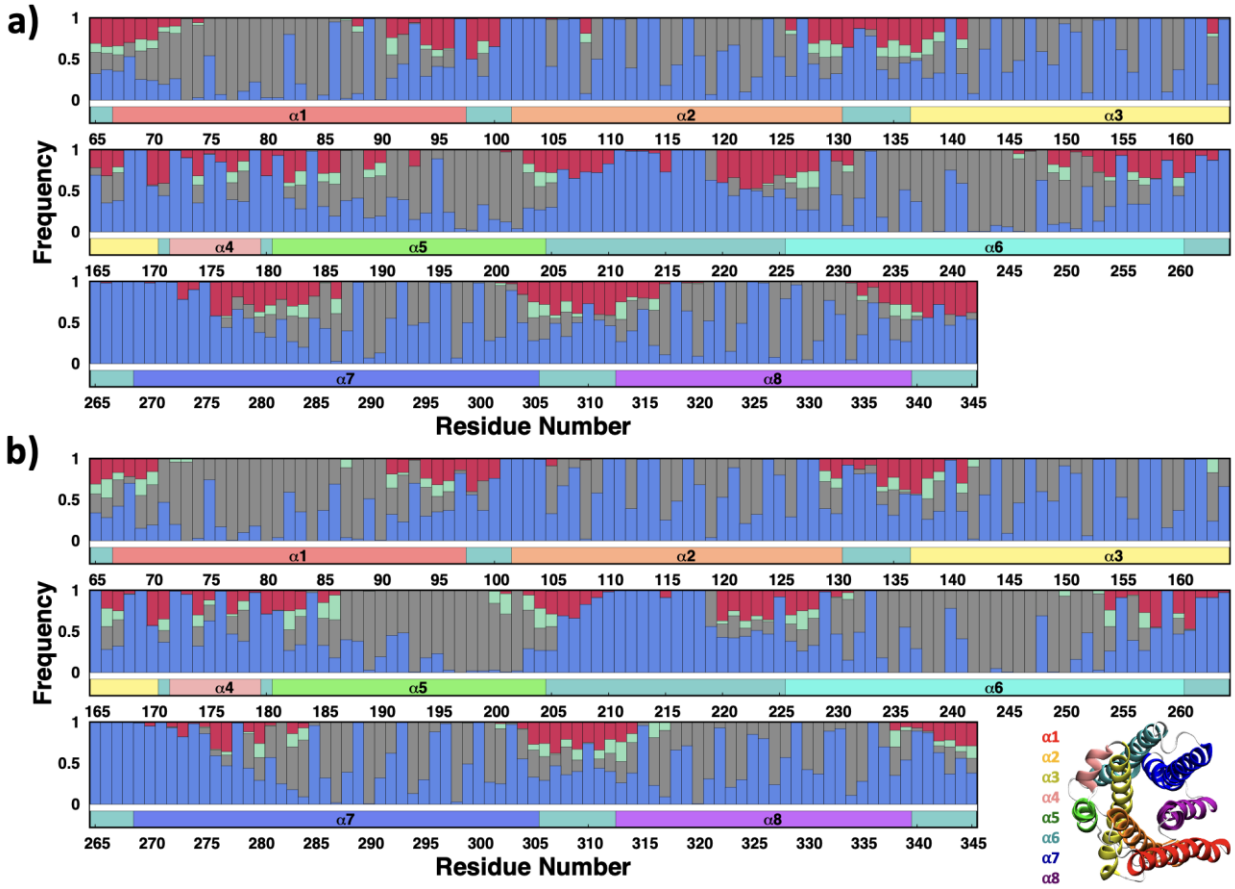


Figure S6. Interaction frequency of the individual parts (alkyl tail, neopentyl glycol (NG)/TRIS-Triazine, and glucoside/maltoside) of (a) LMNG and (b) MG-C11 with MOR. The frequency is displayed according to the amino acid residue numbers of the receptor. The interaction frequencies of the alkyl tail, NG/TRIS-triazine, and glucoside/maltoside with the receptor are represented in grey, cyan, and red respectively. Water interaction frequency is represented in blue.

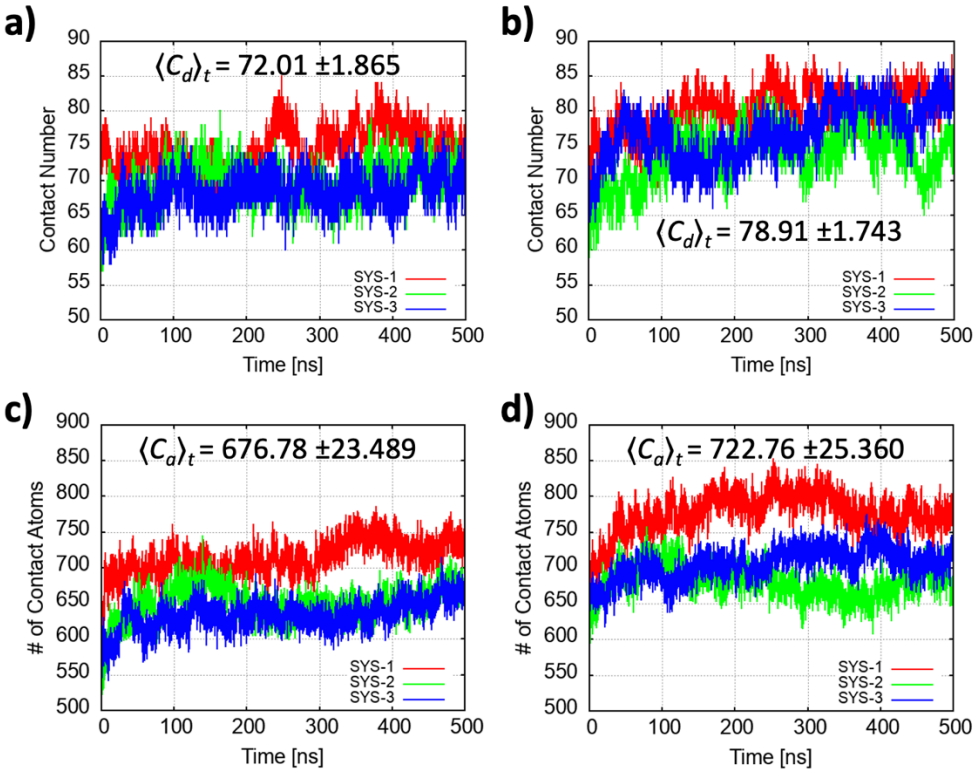


Figure S7. Contact numbers of (a) LMNG and (b) MG-C11 with the protein surfaces of MOR. Average contact number of detergent molecules in micelles ($\langle C_d \rangle_t$) was obtained using the simulation results of the last 200 ns trajectory. MG-C11 gave a higher contact number than LMNG with the receptor (79 *vs* 72). (c,d) The atom numbers of detergent alkyl chains contacting the receptor surface ($\langle C_a \rangle_t$) were also obtained using a similar method. The alkyl chain of MG-C11 gave a higher contact atom number than that of LMNG with the receptor (677 vs 723).

Table S1. Molecular weights (M.W.) and hydrophilic–lipophilic balance (HLB) values of the MGs compared with LMNG and DDM.

Detergent	*M.W.*[a]	HLB[b]
MG-C8	947.1	15.2
MG-C9	969.1	14.7
MG-C10	997.1	14.3
MG-C11	1025.2	13.9
MG-C12	1053.3	13.6
TTG-C11	1059.3	12.7
LMNG	1005.2	13.6
DDM	510.6	13.4

[a] Molecular weight of detergents. [b] Values obtained by Griffin's method.[1]

Table S2. Average distance (Å) of the individual detergent parts (triazine (Tz), TRIS and glucoside (Glu)) from the center of micelles formed by MG-C11 and MG-12. Micelles were generated using aggregation number of 20, 25, 30 or 40 via MD simulations. The values correspond to the maximum density of the individual parts in the detergent density profiles as a function of distance from the micellar center.

	MG-C11 (20)[a]	MG-C12 (20)	MG-C11 (25)	MG-C12 (25)	MG-C11 (30)	MG-C12 (30)	MG-C11 (40)	MG-C12 (40)
Tz	15.27	15.27	16.25	16.25	17.23	17.23	19.20	19.20
TRIS	18.22	18.22	19.20	19.20	20.19	20.19	22.16	22.16
Glu	19.20	20.18	21.17	21.17	22.16	22.16	24.13	24.13

[a] Aggregation number of detergent micelles.

Table S3. Radius of gyration (R_g) and solvent accessible surface area (*SASA*) of micelles formed by MG-C11 and MG-C12, obtained from MD simulations.

	AN[a]	R_g [Å]	*SASA* [Å^2]
MG-C11	20	17.17 ± 0.032	11.57 ± 0.134
	25	18.30 ± 0.004	8.96 ± 0.190
	30	19.24 ± 0.012	6.83 ± 0.110
	40	20.95 ± 0.018	4.54 ± 0.144
MG-C12	20	17.40 ± 0.012	12.95 ± 0.197
	25	18.55 ± 0.020	10.17 ± 0.072
	30	19.49 ± 0.008	7.95 ± 0.116
	40	21.25 ± 0.009	5.36 ± 0.076

[a] Aggregation number of detergent micelles.

Table S4. Averaged root mean square deviations (*RMSD*) of detergent molecules in micelles formed by MG-C11 and MG-C12, depending on aggregation numbers (*AN*s). The calculations were carried out following the trajectory of detergent alkyl chains during the last 100 ns MD simulations.

	AN[a]	<*RMSD*> [Å]
MG-C11	20	10.28 ± 0.36
	25	9.39 ± 0.40
	30	8.19 ± 0.24
	40	7.34 ± 0.35
MG-C12	20	10.08 ± 0.16
	25	10.14 ± 0.22
	30	8.42 ± 0.19
	40	7.98 ± 0.28
DDM	130	14.14 ± 1.59

[a] Aggregation number of detergent micelles.

Molecular dynamics (MD) simulations

For computational study, the melamine-cored glucoside (MG) detergent models were prepared with C11 and C12 alkyl chains, referred to as MG-C11 and MG-C12, respectively. The force field parameters for the melamine core units were generated using the CHARMM General Force Field (CGenFF).[2,3] Subsequently, the melamine core was assembled with three β-D-glucose and alkyl chains by analogy with the CHARMM36 force field.[4,5] To generate the simulation systems for pure micelle and μ-opioid receptor (MOR)/micelle complex structures, both CHARMM-GUI Micelle Builder and Force Field Converter were utilized.[6-8] The initial configuration of the pure micelle models was positioned as spherical micelles with an aggregation number (*AN*) of 20, 25, 30, and 40. In addition, MOR (chain E in PDB:6DDE)[9] was surrounded with 96 of the MG-C11 or lauryl β-maltose neopentyl glycol (LMNG) detergent as a representative model of protein/micelle complex, which was then subjected to 500 ns-long molecular dynamics (MD) simulations using the OpenMM-7.5.0 package.[10] For all simulation models, the hydrogen mass repartitioning (HMR) scheme was applied to enhance the efficiency of the simulation speed.[11-12] To mimic the experimental conditions, we placed each model in a TIP3P water box with 150 mM KCl ions.[13] Prior to the production simulations, we conducted a series of equilibration steps to ensure the stability of the system. Specifically, we carried out a short minimization and applied a 6-step process of gradually decreasing positional and sugar dihedral restraint forces. The 500-ns HMR MD simulations were conducted at 300 K and 1 bar for each system. To improve sampling and verify convergence, we replicated each system three times with different initial velocities.

Detergent CMC determination by diphenylhexatriene (DPH) encapsulation

Stock solutions containing 10.0 mM individual detergents (MGs) were prepared in distilled water. A series of detergent solutions with different concentrations were made from the stock solutions and 200 μL of each detergent sample was then transferred to a 96-well plate in duplicate. A 50 μL DPH stock solution (3.0 mg DPH in 5.0 mL THF) was diluted to 950 μL of distilled water to give a working solution. 2.0 μL DPH working solution was gently added into each well containing a detergent solution. After an incubation for 20 min at room temperature, fluorescence intensities of the solutions were measured at 430 nm upon excitation at 358 nm using a Synergy Mx Monochromator-Based Multi-Mode Microplate reader. Detergent CMCs were determined by identifying an intersection of two trend lines in plots of fluorescence intensity as a function of detergent concentration.

Detergent micelle size measurement by dynamic light scattering (DLS) experiment

DDM and the MGs were individually dissolved in distilled water to give a final concentration of 1.0 wt %. These solutions were filtered by a syringe filter with a pore size of 0.22 μm. The hydrodynamic radii (R_h) of the micelles produced by DDM and the MGs were measured using a Malvern ZetaSizer Nano ZS90 particle analyzer. With a maximum power of 5 mW, a He–Ne laser operating at 633 nm was used as the

light source. The scattered light was collected at an angle of 90°. All experiments were carried out at 25 °C. The hydrodynamic radius (R_h) of detergent micelles were calculated by autocorrelation analysis on time-dependent scattered light intensity. Hydrodynamic radius (R_h) values for micelles formed by the individual detergents were expressed as mean ± S.D. (n = 4).

Protein stability evaluation

LeuT stability assay

The wild type amino acid transporter, LeuT, from *Aquifex aeolicus* was purified as described previously.[14] In short, LeuT was expressed in *E. coli* C41 (DE3) from a pET16b plasmid encoding the 8xHis-tagged transporter and solubilised in 1.0 % DDM before being incubated with Ni^{2+}-NTA resin for 2 hrs. The immobilized LeuT was eluted in buffer containing 20 mM Tris-HCl (pH 8.0), 200 mM KCl, 20 % glycerol, 0.05 % DDM and 300 mM imidazole, concentrated to approx. 1.5 mg/mL (Vivaspin 500 columns, Sigma) and diluted to a 4 µM stock. To test the ability of the MGs to stabilize LeuT, this protein stock was diluted 10-fold in buffer containing 20 mM Tris-HCl (pH 8.0) and 200 mM KCl supplemented with the respective MG detergent. The final MG concentrations were CMCs + 0.04 wt% or 0.2 wt%. DDM was included as a control. Protein samples were stored at room temperature and LeuT activity was determined by measuring the degree of [^{3}H]-leucine (Leu) binding over a 12-day period, using the scintillation proximity assay (SPA).[15] Protein samples were allowed to equilibrate in the new detergent solution for 20 hours before the first measurement (day 0). Activity was assessed using 20 nM LeuT in the respective detergents, 20 nM [^{3}H]-leucine (36.4 Ci/mmol) and 1.25 mg/mL copper chelate (His-tag) Ysi beads (both from Perkin Elmer) in buffer containing 20 mM Tris-HCl (pH 8.0) and 450 mM NaCl. The total [^{3}H]-leucine binding was determined using a MicroBeta^2 liquid scintillation counter. Non-specific binding was determined in the presence of 10 µM leucine.

MelB solubilization and thermal stability assay

Melibiose and other galactoside transporter MelB was produced by *E. coli* DW2 strain (Δ*melB and* Δ*lacZY*) harboring pK95ΔAHB/WT $MelB_{St}$/CH10 plasmid.[16,17] The plasmid contains the gene encoding the wild-type melibiose permease of *Salmonella typhimurium* ($MelB_{St}$) with a 10-His tag at the C-terminus. Cell growth and membrane preparation were carried out as described.[18] Protein assay was carried out with a Micro BCA kit (Thermo Scientific). The membrane samples containing $MelB_{St}$ (10 mg/mL) in a solubilization buffer (20 mM sodium phosphate, pH 7.5, 200 mM NaCl, 10% glycerol and 20 mM melibiose) were mixed with each individual detergent (MG-C8/C9/C10/C11/C12 and DDM) at 1.5% (w/v). Protein extractions were carried out at 0 °C for 90 min. The protein samples were further incubated at three different temperatures (45, 55, and 65 °C) for 90 min. After extraction and incubation, insoluble fractions were removed by ultracentrifugation at 355, 590 g in a Beckman OptimaTM MAX Ultracentrifuge using a TLA-100 rotor for 45 min at 4 °C. 20 mg membrane proteins without ultracentrifugation and detergent extracts

after the ultracentrifugation with the same volume were loaded for analysis by SDS-15% PAGE, and immunoblotting with a HisProbe- HRP antibody (Thermo Scientific) was conducted for visualizing the $MelB_{St}$.

β_2AR solubilization and long-term stability assay

Receptor solubilization: Membranes (10 µL) expressing wild type β_2AR (WT β_2AR) were homogenized in solubilization buffer (50 µL, pH 7.5) containing 20 mM HEPES and 100 mM NaCl, supplemented with 1.0 % individual detergents (DDM, LMNG, and MG-C11). The cell membranes were incubated for 1 hr at 4 ℃ or 25 ℃, with mild agitation. Following solubilization, the supernatant was separated from the pellet in a 1.5 mL microtube via centrifugation at 13000 rpm for 20 min. The supernatant was transferred to another microtube, while the pellet was resuspended using 60 µL solubilization buffer. Both supernatant and resuspended pellet fractions were separated by SDS-PAGE (4 to 20% HEPES-Tris) and then transferred onto polyvinylidene difluoride (PVDF) membranes at 20 V for 30 min. The PVDF membranes were then blocked in TBST buffer (10 mM Tris-HCl, pH 7.5, 150 mM NaCl, 2 mM $CaCl_2$ and 0.1% Tween-20) containing 5% milk for 1 hr at room temperature. After washing the PVDF membranes with TBST buffer 3 times for 5 min, the membranes were incubated with homemade mouse M1 antibody (1:4000 diluted in TBST) for 1 hr at room temperature. The PVDF membranes were further incubated with HRP-conjugated goat-anti-mouse IgG (1:4000 diluted in TBST) for 1 hr at room temperature and treated with chemiluminescent substrate in order to detect the protein bands.

Receptor stability: β_2AR in 0.1% LMNG was purified based on a previously reported protocol[18,9] and concentrated to around 10 mg/mL (approximately 200 mM). The DDM-purified receptor was diluted into buffer solutions containing DDM, LMNG, TTG-C11, or MG-C11 to reach a final detergent concentration of 0.1 wt%. β_2AR in each detergent was incubated for seven days at room temperature and its ligand binding capacity was measured at regular intervals (0, 1, 2, 5, and 7 days) during the incubation. The ligand binding ability of the receptor was measured using 1.0 nM [^{3}H]-dihydroalprenolol (DHA). The [^{3}H]-DHA containing mixture was immobilized to an anti-flag resin and the resin was collected using GF/B filter paper and scintillation fluid was added into the filter paper. Receptor-bound [^{3}H]-DHA was measured with a scintillation counter (Perkin Elmer). Non-specific bindings of [^{3}H]-DHA to the receptor in the individual detergents were measured in the presence of 100 µM alprenolol in the same binding reaction. The specific binding of the protein in the individual detergents was calculated by subtracting respective non-specific binding from three specific binding values (n = 3). Data was analyzed using GraphPad 9.0 software.

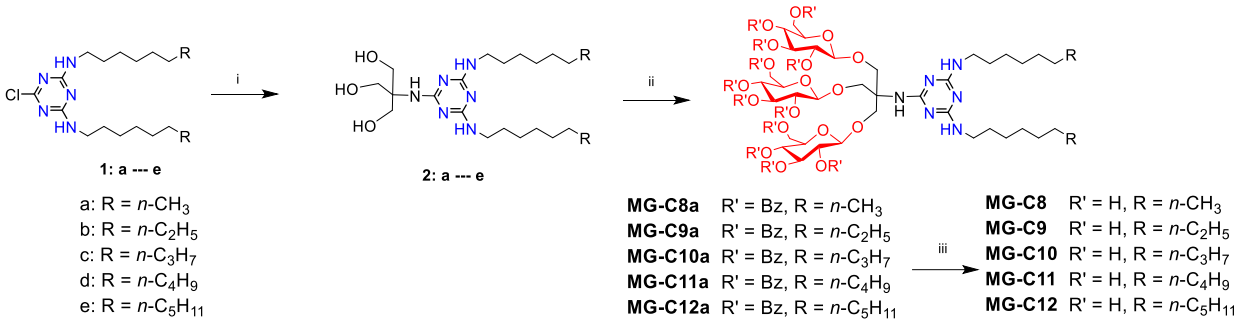
MOR long-term stability assay

MOR in 0.1% LMNG was purified based on a previously reported protocol.[9] The resulting LMNG-purified receptor was diluted into buffer solutions containing DDM, LMNG, TTG-C11, or MG-C11 to reach a final

detergent concentration of 0.1 wt%. MOR in each detergent was incubated for seven days at 4 °C and its ligand binding capacity was measured at regular intervals during the incubation. The ligand binding ability of the receptor was measured by adding ~10 nM [^{3}H]-diprenorphine (DPN). The [^{3}H]-DPN-containing mixture was immobilized to an anti-flag resin and the resin was collected using GF/B filter paper and scintillation fluid was added into the filter paper. Receptor-bound [^{3}H]-DPN was measured with a scintillation counter (Perkin Elmer). Non-specific binding of [^{3}H]-DPN to the receptor in the individual detergents were measured in the presence of excess naloxone in the same binding reaction. The specific binding of the protein in the individual detergents was calculated by subtracting respective non-specific binding from three specific binding values (n = 3). Data was analyzed using GraphPad 9.0 software.

Amphiphile synthesis

Supplementary scheme



i) DIPEA, Tris(hydroxymethyl)aminomethane, THF, 100°C, 34h; **ii)** AgOTf, 2,4,6-collidine, DCM, perbenzoylated glucosylbromide, -45°C to RT; **iii)** NaOMe, MeOH, RT

General synthetic protocols

*General procedure for the synthesis of 2-chloro-4, 6-dialkylated-1,3,5-triazine (**1a** to **1e**)*

A mixture of 2,4,6-trichloro-1,3,5-triazine (3.0 mmol) and diisopropylamine (3.2 mmol) were stirred in THF (10 mL) for 10 min. A respective amine (RNH_2) (6.0 mmol) in THF was added dropwise for 15 min. The resulting reaction mixture was kept at room temperature for 30 min. The oily residue obtained after removal of solvent was subjected to column chromatographic purification to obtain the target product ***(1a-1e)***.

6-chloro-N_2,N_4-dioctyl-1,3,5-triazine-2,4-diamine (1a) was prepared using octylamine in 80% yield. **^{1}H NMR** (400 MHz, $CDCl_3$): δ 5.86 (br, 2H, NH), 3.27-3.42 (m, 4H), 1.50-1.63 (m, 4H), 1.19-1.41 (m, 20H),

0.88 (t, *J* = 7.8 Hz, 6H); **^{13}C NMR** (100 MHz, $CDCl_3$): 167.8, 65.4, 41.5, 32.5, 30.2, 29.8, 27.4, 23.3, 14.4.

6-chloro-*N*$_2$,*N*$_4$-dinonyl-1,3,5-triazine-2,4-diamine (1b) was prepared using nonylamine in 78% yield. **^{1}H NMR** (400 MHz, $CDCl_3$): δ 5.87 (br, 2H, NH), 3.27-3.43 (m, 4H), 1.50-1.63 (m, 4H), 1.19-1.41 (m, 24H), 0.88 (t, *J* = 7.8 Hz); **^{13}C NMR** (100 MHz, $CDCl_3$): 167.9, 65.5, 41.5, 32.5, 30.2, 29.9, 27.4, 23.3, 14.4.

6-chloro-*N*$_2$,*N*$_4$-didecyl-1,3,5-triazine-2,4-diamine (1c) was prepared using decylamine in 80% yield. **^{1}H NMR** (400 MHz, $CDCl_3$): δ 5.86 (br, 2H, NH), 3.28-3.43 (m, 4H), 1.51-1.63 (m, 4H), 1.18-1.41 (m, 28H), 0.88 (t, *J* = 7.8 Hz, 6H); **^{13}C NMR** (100 MHz, $CDCl_3$): 167.8, 165.6, 41.6, 32.5, 30.6, 29.9, 29.8, 27.4, 23.2, 14.3.

6-chloro-*N*$_2$,*N*$_4$-diundecyl-1,3,5-triazine-2,4-diamine (1d) was prepared using undecylamine in 83% yield. **^{1}H NMR** (400 MHz, $CDCl_3$): δ 5.85 (br, 2H, NH), 3.28-3.41 (m, 4H), 1.52-1.63 (m, 4H), 1.18-1.42 (m, 32H), 0.88 (t, *J* = 7.8 Hz, 6H); **^{13}C NMR** (100 MHz, $CDCl_3$): 167.8, 165.7, 41.5, 32.5, 30.6, 29.9, 29.7, 29.5, 27.4, 23.2, 14.3.

6-chloro-*N*$_2$,*N*$_4$-didodecyl-1,3,5-triazine-2,4-diamine (1e) was prepared using dodecylamine in 82% yield. **^{1}H NMR** (400 MHz, $CDCl_3$): δ 5.84 (br, 2H, NH), 3.29-3.41 (m, 4H), 1.53-1.63 (m, 4H), 1.18-1.42 (m, 36H), 0.88 (t, *J* = 7.8 Hz, 6H); **^{13}C NMR** (100 MHz, $CDCl_3$): 167.8, 165.5, 41.5, 32.7, 30.6, 29.9, 29.7, 29.5, 27.8, 23.1, 14.2.

General procedure for the TRIS coupling with the dialkylated triazine (2a-2e; step i)

To a dry flask solution of compound **1** (**1a**/**1b**/**1c**/**1d**/**1e**) (1.00 g, 5.42 mmol) in THF (50 mL) was added tris(hydroxymethyl)aminomethane (1.5 equiv.) and diisopropylethylamine (DIPEA) under nitrogen. The solution temperature was gradually increased to 100 ℃, and stirring was continued for 34 hr. The reaction mixture was diluted with water and then extracted with ethylacetate (EtOAc). The organic layer was washed with 1.0 M HCl, brine and dried over anhydrous Na_2SO_4. After concentration of the ethylacetate solution, the residue was purified by flash column chromatography (EtOAc/hexane), providing a desired product (**2a**/**2b**/**2c**/**2d**/**2e**/**2f**) as a white solid.

2-((4,6-bis(octylamino)-1,3,5-triazin-2-yl)amino)-2-(hydroxymethyl)propane-1,3-diol (2a) was prepared from **1a** in 80% yield according to the general procedure for Tris coupling. **^{1}H NMR** (400 MHz,

$CDCl_3$): δ 3.78-3.72 (m, 6H), 3.41-3.38 (m, 4H), 1.53-1.50 (m, 4H), 1.28-1.24 (m, 20H), 0.88 (t, *J* = 7.2 Hz, 6H); **^{13}C NMR** (100 MHz, $CDCl_3$): δ 180.3, 179.4, 162.2, 62.4, 32.1, 30.1, 29.8, 29.4, 29.3, 29.1, 22.8, 14.2.

2-((4,6-bis(nonylamino)-1,3,5-triazin-2-yl)amino)-2-(hydroxymethyl)propane-1,3-diol (2b) was prepared from **1b** in 82% yield according to the general procedure for Tris coupling. was prepared in 82% yield according to the general procedure for Tris coupling. **^{1}H NMR** (400 MHz, $CDCl_3$): δ 3.77-3.73 (m, 6H), 3.40-3.37 (m, 4H), 1.52-1.48 (m, 4H), 1.26-1.23 (m, 24H), 0.87 (t, *J* = 7.2 Hz, 6H); **^{13}C NMR** (100 MHz, $CDCl_3$): δ 180.4, 179.5, 162.3, 62.5, 54.1, 32.1, 30.0, 29.8, 29.5, 29.4, 29.1, 22.8, 14.2.

2-((4,6-bis(decylamino)-1,3,5-triazin-2-yl)amino)-2-(hydroxymethyl)propane-1,3-diol (2c) was prepared from **1c** in 81% yield according to the general procedure for Tris coupling. **^{1}H NMR** (400 MHz, $CDCl_3$): δ 3.78-3.74 (m, 6H), 3.41-3.38 (m, 4H), 1.53-1.48 (m, 4H), 1.27-1.24 (m, 28H), 0.88 (t, *J* = 7.2 Hz, 6H); **^{13}C NMR** (100 MHz, $CDCl_3$): δ 180.2, 179.2, 162.2, 62.5, 54.2, 32.3, 30.0, 29.7, 29.3, 29.4, 29.0, 22.6, 14.2.

2-((4,6-bis(undecylamino)-1,3,5-triazin-2-yl)amino)-2-(hydroxymethyl)propane-1,3-diol (2d) was prepared from **1d** in 84% yield according to the general procedure for Tris coupling. **^{1}H NMR** (400 MHz, $CDCl_3$): δ 3.77-3.73 (m, 6H), 3.41-3.38 (m, 4H), 1.53-1.549 (m, 4H), 1.25-1.23 (m, 32H), 0.88 (t, *J* = 7.2 Hz, 6H); **^{13}C NMR** (100 MHz, $CDCl_3$): δ 180.3, 179.4, 162.2, 62.4, 54.1, 32.1, 30.1, 29.8, 29.5, 29.4, 29.1, 22.8, 14.2.

2-((4,6-bis(dodecylamino)-1,3,5-triazin-2-yl)amino)-2-(hydroxymethyl)propane-1,3-diol (2e) was prepared from **1e** in 85% yield according to the general procedure for Tris coupling. **^{1}H NMR** (400 MHz, $CDCl_3$): δ 3.79-3.72 (m, 6H), 3.40-3.39 (m, 4H), 1.54-1.50 (m, 4H), 1.27-1.24 (m, 36H), 0.88 (t, *J* = 7.2 Hz, 6H); **^{13}C NMR** (100 MHz, $CDCl_3$): δ 180.2, 179.4, 162.5, 62.5, 54.1, 32.1, 30.0, 29.9, 29.8, 29.5, 29.3, 29.2, 22.7, 14.2.

General procedure for glycosylation reactions *(MG-Ca; step ii)*

This procedure followed a literature method[19] with slight modification. A mixture of a dialkylated diol derivative (**2a**/**2b**/**2c**/**2d**/**2e**; 1 equiv.), AgOTf (3.8 equiv.) and 2,4,6-collidine (1.0 equiv.) in anhydrous CH_2Cl_2 (40 mL) was stirred at - 45 °C. A solution of perbenzoylated glucosylbromide (3.8 equiv.) in CH_2Cl_2 (10 mL) was added dropwise over 0.5 hr to this suspension. Stirring was continued for 0.5 hr at – 45 °C,

and then the reaction mixture was allowed to warm to 0 °C and left stirring for 1 hr. After completion of the reaction, pyridine was added to the reaction mixture, and the reaction mixture was diluted with CH_2Cl_2 (40 mL) before being filtered through celite. The filtrate was washed successively with a 1.0 M aqueous $Na_2S_2O_3$ (40 mL), a 0.1 M aqueous HCl solution (40 mL), and brine (2 x 40 mL). The organic layer was dried with anhydrous Na_2SO_4, and the solvent was removed by rotary evaporation. The residue was purified by silica gel column chromatography (EtOAc/ hexane), which provided the desired glycosylated product (**MG-Ca**) as a glossy white solid.

General procedure for benzoyl group deprotection reactions *(MG-C; step iii)*

This procedure followed the de-*O*-benzoylation or de-O-acetylation under Zemplén's conditions.[19] The O-protected glycosylated compound (**MG-Ca**) was dissolved in MeOH and treated with the required amount of a methanolic solution of 0.5 M NaOMe such that the final concentration of NaOMe was 0.05 M. The reaction mixture was stirred for 14 hr at room temperature, and then neutralized with Amberlite IR-120 resin (H^+ form). The resin was removed by filtration and washed with MeOH, and the solvent was removed from the combined filtrate in vacuo. The residue was purified by silica gel column chromatography (eluting with MeOH/CH_2Cl_2). Recrystallization using CH_2Cl_2/MeOH/diethyl ether afforded fully deprotected product (**MG-C**) as a white solid.

MG-C8a was prepared from **2a** in 85% yield according to the general procedure for glycosylation reactions. **^{1}H NMR** (400 MHz, $CDCl_3$): δ 8.08-8.01 (m, 12H), 7.91 (d, *J* = 7.2 Hz, 6H), 7.83 (d, *J* = 8.0 Hz, 6H), 7.71-7.26 (m, 46H), 5.97 (s, 1H), 5.60 (t, , *J* = 8.0 Hz, 3H), 5.44 (t, *J* = 8.0 Hz, 3H), 5.29-5.2 (m, 4H), 4.50-4.37 (m, 8H), 4.10-3.90 (m, 4H), 3.57-3.50 (m, 5H), 2.94-2.90 (m 4H), 1.59-1.55 (m, 4H), 1.26-1.21 (m, 20H), 0.88 (t, *J* = 7.4 Hz, 6H); **^{13}C NMR** (100 MHz, $CDCl_3$): δ 166.2, 165.3, 165.1, 164.7, 162.6, 133.8, 133.6, 133.4, 133.3, 130.1, 129.9, 129.8, 129.6, 129.5, 129.2, 128.9, 128.8, 128.6, 128.4, 101.5, 72.6, 72.0, 71.8, 69.6, 68.3, 63.2, 59.8, 31.8, 29.2, 28.9, 22.7, 14.2.

MG-C9a was prepared from **2b** in 86% yield according to the general procedure for glycosylation reactions. **^{1}H NMR** (400 MHz, $CDCl_3$): δ 8.08-8.01 (m, 12H), 7.91 (d, *J* = 7.2 Hz, 6H), 7.83 (d, *J* = 8.0 Hz, 6H), 7.71-7.27 (m, 46H), 6.03 (s, 2H), 5.73-5.28 (m, 22H), 4.51-4.47 (m, 18H), 4.03-3.89 (m, 8H), 3.59-3.48 (m, 10H), 3.23-3.12 (m, 4H), 1.60-1.57 (m, 4H), 1.25-1.18 (m, 24H), 0.87 (t, *J* = 7.4 Hz, 6H); **^{13}C NMR** (100 MHz, $CDCl_3$): δ 166.2, 165.1, 164.8, 162.5, 133.7, 133.5, 133.3, 133.2, 130.0, 129.8, 129.7, 129.6, 129.1, 128.8, 128.5, 128.4, 101.5, 101.4, 92.8, 72.5, 72.0, 71.8, 69.6, 68.2, 63.2, 59.8, 48.4, 31.9, 32.0, 29.7, 29.2, 29.0, 22.9, 14.2

MG-C10a was prepared from **2c** in 85% yield according to the general procedure for glycosylation reactions. **^{1}H NMR** (400 MHz, $CDCl_3$): δ 8.08-8.01 (m, 12H), 7.91 (d, *J* = 7.2 Hz 6H), 7.83 (d, *J* = 8.0 Hz, 6H), 7.71-7.27 (m, 46H), 6.03 (s, 2H), 5.73-5.28 (m, 22H), 4.51-4.47 (m, 18H), 4.03-3.89 (m, 8H), 3.59-3.48 (m, 10H), 3.23-3.12 (m, 4H), 1.60-1.57 (m, 4H), 1.25-1.18 (m, 28H), 0.87 (t, *J* = 7.4 Hz, 6H); **^{13}C NMR** (100 MHz, $CDCl_3$): δ 166.2, 165.1, 164.7, 162.5, 133.8, 133.5, 133.3, 133.2, 130.0, 129.8, 129.7, 129.6, 129.1, 128.8, 128.5, 128.4, 101.5, 101.4, 92.7, 72.6, 72.0, 71.9, 69.8, 68.2, 63.1, 59.8, 48.4, 31.8, 32.1, 29.7, 29.2, 29.0, 22.8, 14.2.

MG-C11a was prepared from **2d** in 84% yield according to the general procedure for glycosylation reactions. **^{1}H NMR** (400 MHz, $CDCl_3$): δ 8.08-8.01 (m, 12H), 7.91 (d, *J* = 7.2 Hz 6H), 7.83 (d, *J* = 8.0 Hz, 6H), 7.71-7.27 (m, 46H), 6.03 (s, 2H), 5.73-5.28 (m, 22H), 4.51-4.47 (m, 18H), 4.03-3.89 (m, 8H), 3.59-3.48 (m, 10H), 3.23-3.12 (m, 4H), 1.60-1.57 (m, 4H), 1.25-1.18 (m, 32H), 0.87 (t, *J* = 7.4 Hz, 6H); **^{13}C NMR** (100 MHz, $CDCl_3$): δ 166.2, 165.1, 164.7, 162.5, 133.8, 133.5, 133.3, 133.2, 130.0, 129.8, 129.7, 129.5, 129.1, 128.7, 128.5, 128.4, 101.5, 101.4, 92.8, 72.5, 72.1, 71.8, 69.6, 68.2, 63.1, 59.8, 48.4, 31.9, 32.0, 29.8, 29.2, 29.0, 22.7, 14.2.

MG-C12a was prepared from **2e** in 85% yield according to the general procedure for glycosylation reactions. **^{1}H NMR** (400 MHz, $CDCl_3$): δ 8.08-8.01 (m, 12H), 7.91 (d, *J* = 7.2 Hz 6H), 7.83 (d, *J* = 8.0 Hz, 6H), 7.71-7.27 (m, 46H), 6.04 (s, 2H), 5.71-5.26 (m, 22H), 4.51-4.47 (m, 18H), 4.03-3.89 (m, 8H), 3.59-3.48 (m, 10H), 3.23-3.12 (m, 4H), 1.59-1.57 (m, 4H), 1.26-1.18 (m, 36H), 0.87 (t, *J* = 7.4, Hz 6H); **^{13}C NMR** (100 MHz, $CDCl_3$): δ 166.1, 165.3, 164.5, 162.2, 133.8, 133.5, 133.3, 133.2, 130.0, 129.8, 129.7, 129.6, 129.1, 128.8, 128.5, 128.4, 101.5, 101.4, 92.7, 72.6, 72.0, 71.8, 69.6, 68.2, 63.0, 59.8, 48.4, 31.9, 32.0, 29.7, 29.3, 29.0, 22.8, 14.1.

MG-C8 was prepared from **MG-C8a** in 85% yield according to the general procedure for deprotection reaction. **^{1}H NMR** (400 MHz, CD_3OD): δ 4.48 (d, *J* = 8.0 Hz, 3H), 4.32-4.29 (m, 6H), 4.09-4.05 (m, 6H), 3.92-3.86 (m, 4H), 3.73-3.68 (m, 4H), 3.34-3.20 (m, 24H), 1.62-1.60 (m, 4H), 1.33-1.28 (m, 20H), 0.88 (t, *J* = 7.8 Hz, 6H); **^{13}C NMR** (100 MHz, CD_3OD): δ 164.7, 105.4, 78.2, 78.0, 75.1, 71.6, 70.2, 62.7, 33.2, 30.7, 30.6, 30.2, 23.6, 14.5. **HRMS (FAB$^+$)**: calcd. for $C_{41}H_{76}N_6O_{18}$ $[M]^+$ 941.2412, found 941.2428.

MG-C9 was prepared from **MG-C9a** in 84% yield according to the general procedure for deprotection reaction. **^{1}H NMR** (400 MHz, CD_3OD): δ 4.48 (d, *J* = 8.0 Hz, 3H), 4.31-4.29 (m, 6H), 4.09-4.06 (m, 6H),

3.91-3.86 (m, 4H), 3.73-3.67 (m, 4H), 3.34-3.21 (m, 24H), 1.62-1.60 (m, 4H), 1.33-1.28 (m, 24H), 0.88 (t, *J* = 7.8 Hz, 6H); **^{13}C NMR** (100 MHz, CD_3OD): δ 164.6, 105.3, 78.2, 78.1, 75.1, 71.5, 70.2, 62.7, 33.2, 30.8, 30.5, 30.2, 23.5, 14.5. **HRMS (FAB$^+$)**: calcd. for $C_{43}H_{80}N_6O_{18}$ $[M]^+$ 969.4261, found 969.4274.

MG-C10 was prepared from **MG-C10a** in 86% yield according to the general procedure for deprotection reaction. **^{1}H NMR** (400 MHz, CD_3OD): δ 4.48 (d, *J* = 8 Hz, 3H), 4.32-4.29 (m, 6H), 4.08-4.05 (m, 6H), 3.91-3.87 (m, 4H), 3.72-3.67 (m, 4H), 3.33-3.21 (m, 24H), 1.62-1.60 (m, 4H), 1.33-1.28 (m, 28H), 0.88 (t, *J* = 7.8 Hz, 6H); **^{13}C NMR** (100 MHz, CD_3OD): δ 164.8, 105.3, 78.2, 78.0, 75.1, 71.4, 70.3, 62.8, 33.3, 30.9, 30.2, 28.8, 23.7, 14.5. **HRMS (FAB$^+$)**: calcd. for $C_{47}H_{84}N_6O_{18}$ $[M]^+$ 997.5232, found 997.5245.

MG-C11 was prepared from **MG-C11a** in 85% yield according to the general procedure for deprotection reaction. **^{1}H NMR** (400 MHz, CD_3OD): δ 4.49 (d, *J* = 8.0 Hz, 3H), 4.33-4.30 (m, 6H), 4.08-4.04 (m, 6H), 3.90-3.87 (m, 4H), 3.72-3.68 (m, 4H), 3.36-3.21 (m, 24H), 1.62-1.60 (m, 4H), 1.32-1.28 (m, 32H), 0.88 (t, *J* = 7.8 Hz, 6H); **^{13}C NMR** (100 MHz, CD_3OD): δ 164.7, 105.4, 78.1, 78.0, 75.1, 71.5, 70.3, 62.7, 33.2, 30.9, 30.6, 30.3, 28.8, 23.8, 14.5. **HRMS (FAB$^+$)**: calcd. for $C_{47}H_{84}N_6O_{18}$ $[M]^+$ 1025.3942, found 1025.3961.

MG-C12 was prepared from **MG-C12a** in 84% yield according to the general procedure for deprotection reaction. **^{1}H NMR** (400 MHz, CD_3OD): δ 4.49 (d, *J* = 8.0 Hz, 3H), 4.33-4.29 (m, 6H), 4.08-4.05 (m, 6H), 3.89-3.86 (m, 4H), 3.71-3.68 (m, 4H), 3.36-3.21 (m, 24H), 1.63-1.60 (m, 4H), 1.32-1.28 (m, 36H), 0.89 (t, *J* = 7.8 Hz, 6H); **^{13}C NMR** (100 MHz, CD_3OD): δ 164.8, 105.4, 78.1, 78.0, 75.1, 71.5, 70.3, 62.7, 33.2, 30.9, 30.7, 30.6, 30.3, 23.8, 23.5, 14.5. **HRMS (FAB$^+$)**: calcd. for $C_{49}H_{92}N_6O_{18}$ $[M]^+$ 1053.6543, found 1053.6546.

References

1. W. C. Griffin, *J. Soc. Cosmet. Chem.*, 1954, **5**, 249-256.
2. K. Vanommeslaeghe and A. D. MacKerell Jr, *J. Chem. Inf. Model.*, 2012, **52**, 3144-3154.
3. K. Vanommeslaeghe, E. P. Raman and A. D. MacKerell Jr, *J. Chem. Inf. Model.*, 2012, **52**, 3155-3168.
4. O. Guvench, S. S. Mallajosyula, E. P. Raman, E. Hatcher, K. Vanommeslaeghe, T. J. Foster, F. W. Jamison and A. D. MacKerell Jr, *J. Chem. Theory Comput.*, 2011, **7**, 3162-3180.
5. O. Guvench, E. Hatcher, R. M. Venable, R. W. Pastor and A. D. MacKerell Jr, *J. Chem. Theory Comput*., 2009, **5**, 2353-2370.
6. S. Jo, T. Kim, V. G. Iyer and W. Im, *J. Comput. Chem.*, 2008, **29**, 1859-1865.
7. X. Cheng, S. Jo, H. S. Lee, J. B. Klauda and W. Im, *J. Chem. Inf. Model.* 2013, **53**, 8, 2171–2180.

8. J. Lee, X. Cheng, J. M. Swails, M. S. Yeom, P. K. Eastman, J. A. Lemkul, S. Wei, J. Buckner, J. C. Jeong, Y. Qi, S. Jo, V. S. Pande, D. A. Case, C. L. Brooks 3rd, A. D. Mackerell Jr, J. B. Klauda and W. Im, *J. Chem. Theory Comput.*, 2016, **12**, 405-413.
9. A. Koehl, H. Hu, S. Maeda, Y. Zhang, Q. Qu, J. M. Paggi, N. R. Latorraca, D. Hilger, R. Dawson, H. Matile, G. F. X. Schertler, S. Granier, W. I. Weis, R. O. Dror, A. Manglik, G. Skiniotis, and B. K. Kobilka, *Nature*, 2018, **558**, 547-552.
10. P. Eastman, J. Swails, J. D. Chodera, R. T. McGibbon, Y. Zhao, K. A. Beauchamp, L. P. Wang, A. C. Simmonett, M. P. Harrigan, C. D. Stern, R. P. Wiewiora, B. R. Brooks and V. S. Pande, *PLoS Comput. Biol.,* 2017, **13**, e1005659.
11. C.W. Hopkins, S. Le Grand, R. C. Walker and A. Roitberg, *J. Chem. Theory Comput.*, 2015, **11**, 1864-1874.
12. Y. Gao, J. Lee, I. P. S. Smith, H. Lee, S. Kim, Y. Qi, J. B. Klauda, G. Widmalm, S. Khalid and W. Im, *J. Chem. Inf. Model.*, 2021, **61**, 831-839.
13. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926-935.
14. C. B. Billesbolle, M. B. Kruger, L. Shi, M. Quick, Z. Li, S. Stolzenberg, J. Kniazeff, K. Gotfryd, J. S. Mortensen, J. A. Javitch, H. Weinstein, C. J. Loland and U. Gether, *J. Biol. Chem*. **2015,** *290* (44), 26725-38.
15. M. Quick and J. A. Javitch, *Proc. Natl. Acad. Sci. U. S. A.,* 2007, **104**, 3603-3608.
16. S. A. Short, H. R. Kaback, and L. D. Kohn, *Proc. Natl. Acad. Sci. U. S. A.,* 1974, **71**, 1461-1465.
17. D. M. Rosenbaum, V. Cherezov, M. A. Hanson, S. G. Rasmussen, F. S. Thian, T. S. Kobilka, H.-J. Choi, X.-J. Yao, W. I. Weis and R. C. Stevens, *Science*, 2007, **318**, 1266-1273.
18. B. K. Kobilka, *Ana. biochemistry*, 1995, **231**, 269-271.
19. P. R. Ashton, S. E. Boyd, C. L. Brown, N. Jayaraman, S. A. Nepogodiev and J. F. Stoddart, *Chem.–Eur. J.,* 1996, **2**, 1115-1128.