The Lazy Life of Lipid-Linked Oligosaccharides in All Life Domains

Pablo R. Arantes ,[a*,†,‡] Conrado Pedebos[a,†,¶] Marcelo D. Polêto,[†,§] Laércio Pol-Fachin,[‖] and Hugo Verli[†]

†*Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, 91509-900, Brazil.*

‡*Laboratory of Genetic Toxicology, Federal University of Health Sciences of Porto Alegre – UFCSPA, Sarmento Leite, 245, Lab.714, Porto Alegre, RS, 90050-170, Brazil.*

¶*School of Chemistry, University of Southampton, Southampton, SO17 1BJ, U.K.*

§*Departamento de Biologia Geral, Universidade Federal de Viçosa, Viçosa, MG, 36570-000, Brazil.*

‖*Centro Universitário CESMAC, Maceió, AL, 57081-350, Brazil.*

E-mail: pablitoarantes@gmail.com

Phone: +55 (51) 3303-8861

---

[a]These authors contributed equally to this work

Abstract

Lipid-linked oligosaccharides (LLOs) play an important role in the N-glycosylation pathway as the donor substrate of oligosaccharyltransferases (OSTs), which are responsible for the *en bloc* transfer of glycan chains onto a nascent polypeptide. The lipid component of LLO in both eukarya and archaea consists of a dolichol, and an undecaprenol in prokarya, whereas the number of isoprene units may change between species. Given the potential relevance of LLOs and their related enzymes to diverse biotechnological applications, obtaining reliable LLO models from distinct domains of life could support further studies on complex formation and their processing by OSTs, as well as protein engineering on such systems. In this work, molecular modeling, such as quantum mechanics calculations, molecular dynamics simulations, and metadynamics were employed to study eukaryotic (Glc_3-Man_9-$GlcNAc_2$-PP-Dolichol), bacterial (Glc_1-$GalNAc_5$-Bac_1-PP-Undecaprenol) and archaeal (Glc_1-Man1-Gal_1-Man1-Glc_1-Gal_1-Glc_1-P-Dolichol) LLOs in membrane bilayers. Microsecond molecular dynamics simulations and metadynamics calculations of LLOs revealed that glycan chains are more prone to interact with the membrane lipid head groups, while the PP linkages are positioned at the lipid phosphate head groups level. Dynamics of isoprenoid chains embedded within the bilayer are described and membrane dynamics and its related properties are also investigated. Overall, there are similarities regarding the structural and dynamics of the eukaryotic, the bacterial and the archaeal LLOs in bilayers, which can support the comprehension of their association with OSTs. This data may support future studies on the transferring mechanism of the oligosaccharide chain to an acceptor protein.

Introduction

The glycosylation of asparagine residues is the predominant protein modification throughout all domains of life.[1] This post-translational modification is called N-glycosylation, which is important in many aspects of biology and affect various properties of the proteins, including folding, conformation and solubility.[1,2] N-glycosylation occurs in the consensus motif

2

(referred to as sequon) represented by Asn-X-Ser/Thr, where X can be any residue except proline.[3,4] Lipid-linked oligosaccharides (LLOs) are the substrates of oligosaccharyltransferase (OST), which catalyzes the transfer of the oligosaccharide to the acceptor asparagine residue of nascent proteins.[5]

The LLO molecule can be subdivided in two main parts: the isoprenoid moiety and the oligosaccharide chain. The link between these two portions is made by one or two pyrophosphate groups.[6] The lipid component of LLOs varies for each species, as well as the number of isoprene units. On eukarya and archaea, the isoprenoid moiety is a dolichol, while in bacteria, the isoprenoid is an undecaprenol. In eukarya, the oligosaccharide portion is more conserved, composed by a Glc_3-Man_9-$GlcNAc_2$ chain, while on the other domains of life it is more diverse. Despite of all the differences in their structure, the oligosaccharide transfer mechanism seems to be conserved in all domains of life.[7]

In Bacteria, the N-glycosylation final step is performed by a single-subunit oligosaccharyltransferase (PglB).[8] However, in most eukaryotes, the same is achieved by the action of a large, multi-subunit, membrane-embedded OST complex, usually containing eight subunits in yeast, but possibly showing even more in multicellular organisms. Recently, the structure of the yeast OST was characterized by employing single-particle cryo-EM,[9] revealing a conserved subunit arrangement. In this structure, the catalytic STT3 subunit displays full access to its substrates, since its active site is not located in the inner area of the complex. The dolichol-pyrophosphate portion of the eukaryotic LLO binds to a hydrophobic groove of STT3, similarly to other OSTs,[10,11] while the glycan chain binds to the pocket formed in the interface of both domains of the STT3 unit, also comprised by two noncatalytic subunits and an ordered N-glycan. This suggests that these LLO molecules might populate conformations in which the sugar moieties are parallel to the membrane prior to binding. In this context, there is only one previous computational report[12] that assessed this information at an atomic level (only for bacterial and eukaryotic LLOs) which demonstrated the importance of the oligosaccharide chain behavior and its orientation with respect to the bilayer before

its binding.

In our work, we employed a combination of quantum mechanics (QM) calculations, molecular modelling techniques, molecular dynamics (MD) and enhanced sampling simulations to study conformational properties, such as the orientation and the dynamics of LLOs from all domains of life: i) an eukaryotic LLO[12] (Glc_3-Man_9-$GlcNAc_2$-PP- Dolichol; $G_3M_9Gn_2$-PP-Dol; Fig. 1 A) embedded in 1-Palmitoyl-2-oleoyl-SN-glycero-3-phosphocholine (POPC) membrane; ii) a bacterial LLO[12] (Glc_1-$GalNAc_5$-Bac_1-PP-Undecaprenol; $G_1Gn_5B_1$- PP-Und; Fig. 1B) embedded in 1-palmitoyl-2-oleoyl-phosphatidylethanolamine (POPE) membrane; iii) an archaeal LLO[13,14] (Glc_1-Man_1-Gal_1-Man_1-Glc_1-Gal_1-Glc_1-P-Dolichol; $G_1M_1Gal_1M_1G_1Gal_1G_1$-P-Dol; Fig. 1C) embedded in a POPE membrane. The eukaryotic, bacterial and archaeal LLOs without its glycan chains were also analyzed. Improvements on the parameters of a torsional angle were achieved, an important correction for a better reproduction of some experimental properties of the isoprenoid unit. In addition, novel parameters for the sulfate groups in the archaeal LLO were also generated. Our microsecond molecular dynamics (MD) simulations analysis allowed us to identify clusters of conformations that were submitted to molecular docking calculations with the available crystallographic structures for all domains of life. We generated enzyme-ligand poses that were in agreement with expected crystal structure distances for all cases, achieving the first set of full complexes OSTs for the eukaryal and the archaeal enzymes.

Results and discussion

Simulation results are presented and discussed for the eukaryotic, bacterial and archaeal LLOs embedded in their respective membrane bilayers consisting of POPC, POPE and POPE lipids, respectively. The systems without glycan portion are presented as controls for further comparisons.
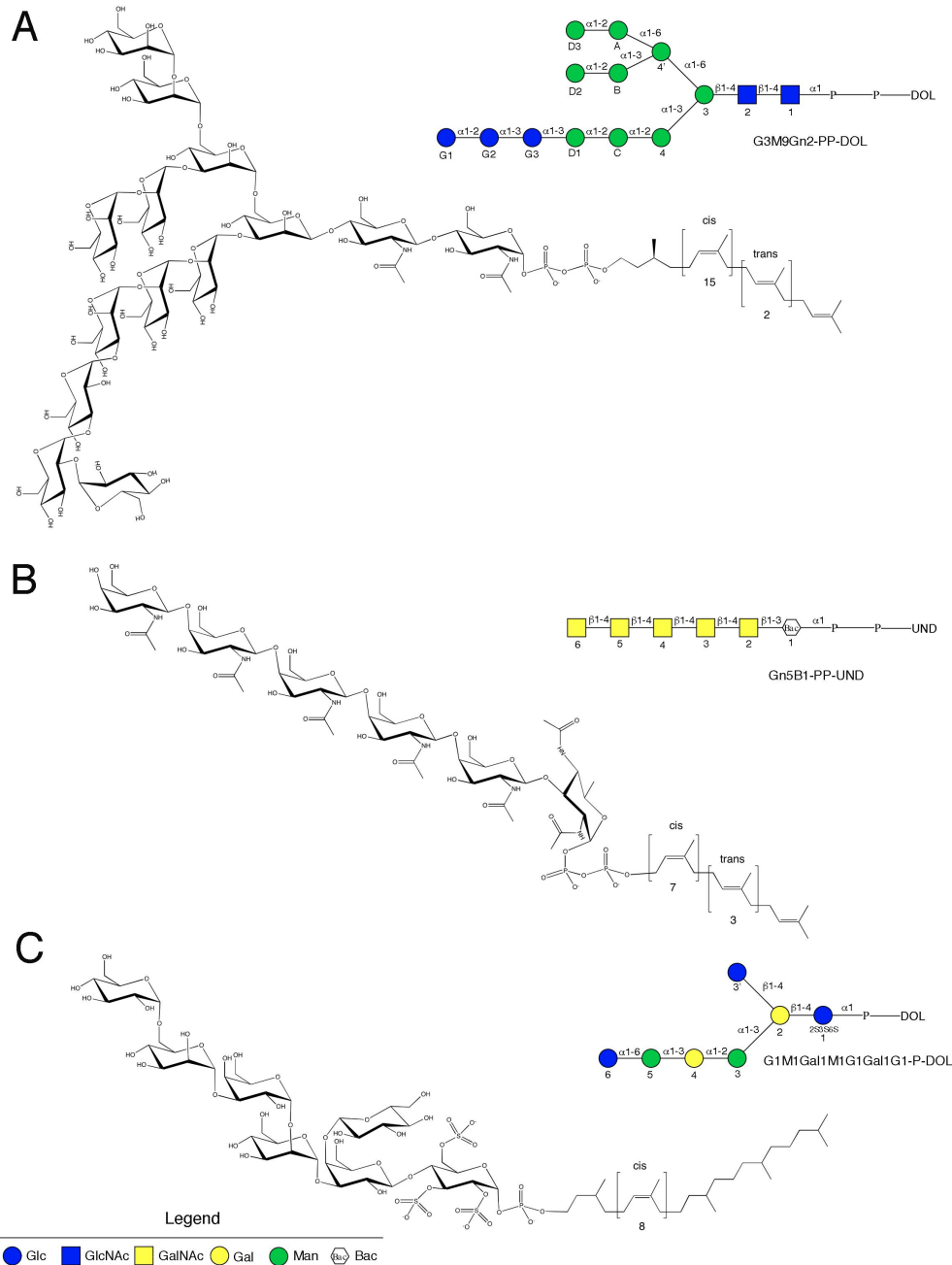
Figure 1: Schematic representations of the (A) eukaryotic, (B) bacterial and (C) Archaea LLOs used in this study. Blue circle = D-Glucose, Blue square = N-Acetyl-D-glucosamine, Yellow square = N-Acetyl-D-galactosamine, Yellow circle = D-Galactose, Green circle = D-Mannose and Bac = Bacillosamine.

Torsional potential and force field calibration

The functional form of the potential energy term, associated with the torsion around the dihedral angle m, is defined in GROMOS by:

$$V_{\phi,m} = k_{\phi,m}[1 + cos\delta_m cos(n_m\phi_m)] \tag{1}$$

where ϕ_m is the dihedral angle value, n_m the multiplicity of the term, δ_m the associated phase shift, and $k_{\phi,m}$, the corresponding force constant, which are applied. It is worth noting that a given dihedral angle may be described by more than one torsional potential energy term with different multiplicities and/or phase shifts.

While some of the torsional potentials were preserved from the GROMOS53A6 original parameter set when building LLOs topology, torsional preferences of the isoprenoid chain were re-evaluated based on QM data fitting. For this, QM calculations at MP2/6-31G(d) theory level was carried for the dihedral rotation of 2-Methylbut-1-ene fragment due to its similarity with insaturation pattern presented in POPE and POPC lipid chains (Figure S1A). A set of three torsional potentials were derived by fitting MM energies into the QM torsional profile generated (Table 1). Due to the presence of *cis* and *trans* configurations observed in the lipid chains that could influence torsional barriers, we have tested the newly obtained parameters for the reproduction of torsional profiles of 3-Methyl-cis-2-pentene and 3-Methyl-trans-2-pentene (Figure 2). Moreover, we have tested additional QM theory levels (HF/6-31G(d), B3LYP/6-31G(d) and MP2/6-31++G(d)) to test whether differences in torsional barrier could be observed (Figure S1).

As shown in Figure 2 (left), the newly obtained dihedral potentials were able to reproduce energy global and local *minima* at 110º/250º and 0º/360º, respectively, of QM torsional profile of 3-Methyl-trans-2-pentene, while also reproducing torsional barriers heights. Moreover, the same set of torsional parameters were able to reproduce QM torsional profiles of 3-Methyl-cis-2-pentene, although with less accuracy. Although energy global *minima* are shifted 20° from QM to MM torsional profile (90°/270° to 110°/250°, respectively), the energy difference in those regions is approximately 5 kJ/mol, which is easily surpassed in common MD fluctuations. The same pattern is observed on the torsional barrier in 0°, which presents a QM to MM difference of 6 kJ/mol. In fact, further investigations of dihedral distribu-
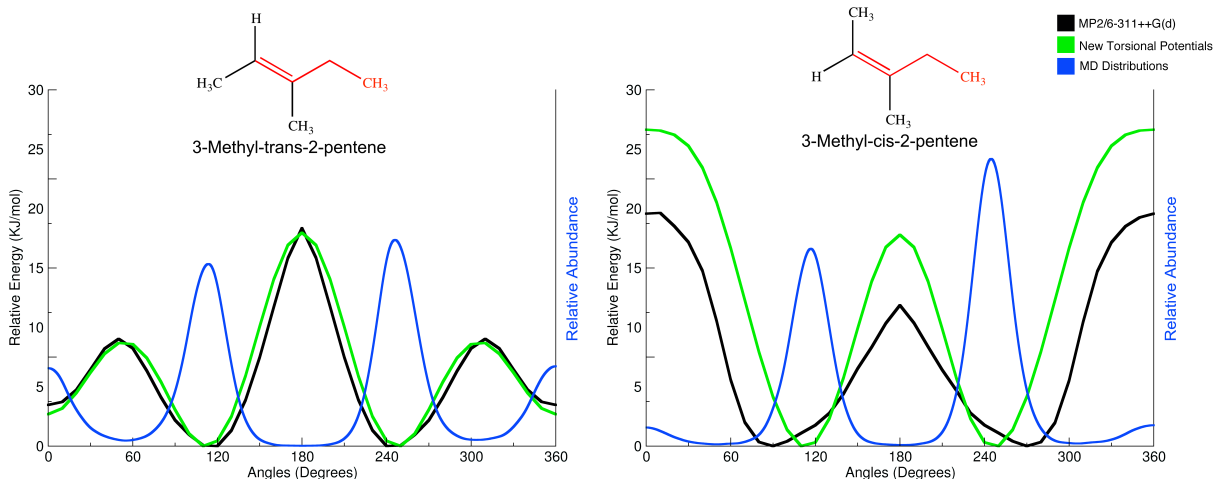
Figure 2: Comparison of energy profiles calculated at classical (green) and quantum mechanics (MP2/6-311++G(d) theory level in black) levels in the gas phase for *cis* and *trans* configurations. Dihedral distribution in further LLO molecular dynamics simulations (blue) reveals a good agreement between QM and MM torsional preferences.

tions throughout LLO simulations revealed a broad distribution covering both QM and MM energy *minima*, suggesting that the single set of torsional potentials used to model both *cis* and *trans* configurations yields energies and torsional preferences in good agreement with their respective QM torsional profiles.

Organic liquid simulations were performed in order to validate the parameters used in topologies for LLOs building blocks in conformity with GROMOS philosophy.[15] This validation strategy of comparing experimental and simulated condensed-phase (ρ and ΔH_{vap}) thermodynamic properties was employed in other works involving molecules parametrization in GROMOS force field[16,17] and for the establishment of force field benchmarks.[18] The simulation of geraniol in liquid phase yielded ρ and ΔH_{vap} values in good agreement with experimental data (Table 2), which indicates that the obtained set of topological parameters, specially nonbonded, is suitable to reproduce condensed phase properties of such hydrophobic fragment and, therefore, can be extrapolated to lipid tails.

Table 1: Torsional parameters obtained based on QM calculations of 2-Methylbut-1-ene dihedral angle and used to model torsions in 3-Methyl-cis-2-pentene and 3-Methyl-trans-2-pentene.

Compound	δ	$k_{\phi,\mathrm{m}}$	n
	0	5.97	0
2-Methylbut-1-ene	0	-3.67	1
	0	-5.17	3

Table 2: Obtained values of thermodynamic properties obtained from geraniol simulation in liquid phase

Compound	Temp. [K]	Exp. ρ [g/cm^3]	Calc. ρ [g/cm^3]	Error	Exp. ΔH_{vap} [kJ/mol]	Calc. ΔH_{vap} [kJ/mol]	Error
Geraniol	298.15[19]	0.89[20]	0.89	0%	58.83[19]	64.24	9.33%

LLOs dynamics and spatial distribution within membrane

Calculated density profiles calculated along the vector normal to the membrane (Z axis) indicated that, in general, the saccharidic residues remains above the lipid head group, while the pyrophosphate (PP) group mostly populate the same area as the lipid head groups. Moreover, LLOs hydrophobic chain (Dolichol-PP, Dolichol-P, and Undecaprenol-PP) fluctuates inside the membrane bilayer (Figure 3). We detected a general asymmetric behavior for the LLO lipid chain, with the highest density peaks being found in the middle of the two leaflets, along with a higher population of the upper leaflet area. The carbon at the end of the LLO tail demonstrates the flexibility of this moiety, as it populates distinct areas of the bilayer, with a higher density between the center area of the membrane and the lower leaflet. This same pattern was observed for the non-glycosilated LLO chains (Figure S2), suggesting that the presence of oligossacharide moieties in LLO structure does not influence the dynamics of its lipid tails within the membrane bilayer in terms of Z axis.

To better understand the conformational space of LLOs lipid tails among the membrane XY axis, we measured the distance distributions between the first carbon (C1A - all simulations) and the middle carbon (C9E - Eukarya, C6E - Bacteria, and C8E - Archaea), and between the first carbon (C1A - all simulations) and the last carbon (C19E - Eukarya, C11E
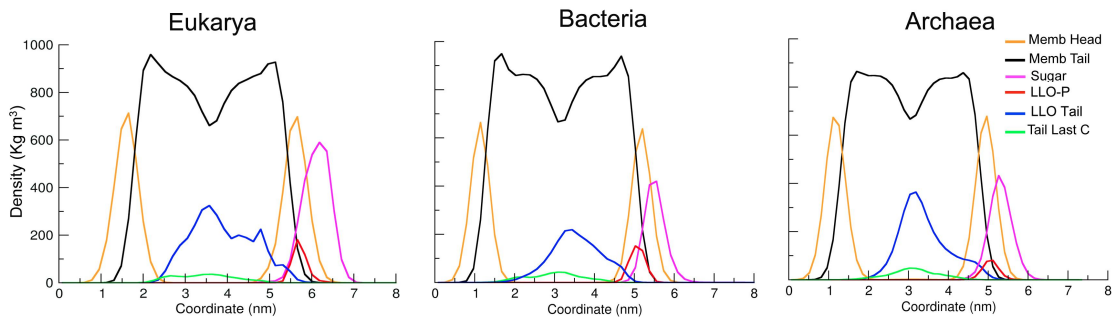
Figure 3: Density profiles for the subdivided parts of the studied LLOs considering the vector normal to the membrane bilayer (Z axis): membrane head groups (yellow), lipid acyl chains (black), isoprenoid chain (blue), oligosaccharide (purple), pyrophosphate linkage (red), and the last carbon atom of isoprenoid chain (green). We enhanced the oligosaccharide distribution scaling it by a factor of five to facilitate the observation of data. The same procedure was performed for the pyrophosphate linkage, dolichol chain, and dolichol's last carbon distributions, but using a scale factor of ten.

- Bacteria, and C12E - Archaea) of the lipid chain (Figure 4). The distributions found for bacterial LLOs revealed a more restricted area of occupation for the first half of the chain (Figure 4A - red line), specially on the bacterial molecule, mainly due to its reduced length, while the full length distance distribution (Figure 4A - blue line) displays a pronounced flexibility, with a 10 Å difference in the stretch of the eukaryotic LLO. The same carbon-carbon distances were calculated as a function of trajectory time blocks of 200ns to confirm convergence (Table S1). Interestingly, the archaeal LLO first half chain presents a distribution pattern comparable to the eukaryotic LLO, despite having a difference of almost half of the isoprenoid units. When comparing the lipid dynamics of the complete LLO structures (Figure 4B) with the dynamics of LLOs without their glycan chains (Figure S3B), eukaryotic LLO presented a greater lipid motion in the absence of its glycan chain, while the the archaea LLO presented a more limited motion in the presence of its glycan chain. Together with the densities of LLO portion along the membrane, it is possible to infer that LLOs lipid tail motion is mostly confined to the XY axis at the region between the membrane leaflets. This behavior demonstrates the influence of the sugar portion on the dynamics of these glycoconjugates.
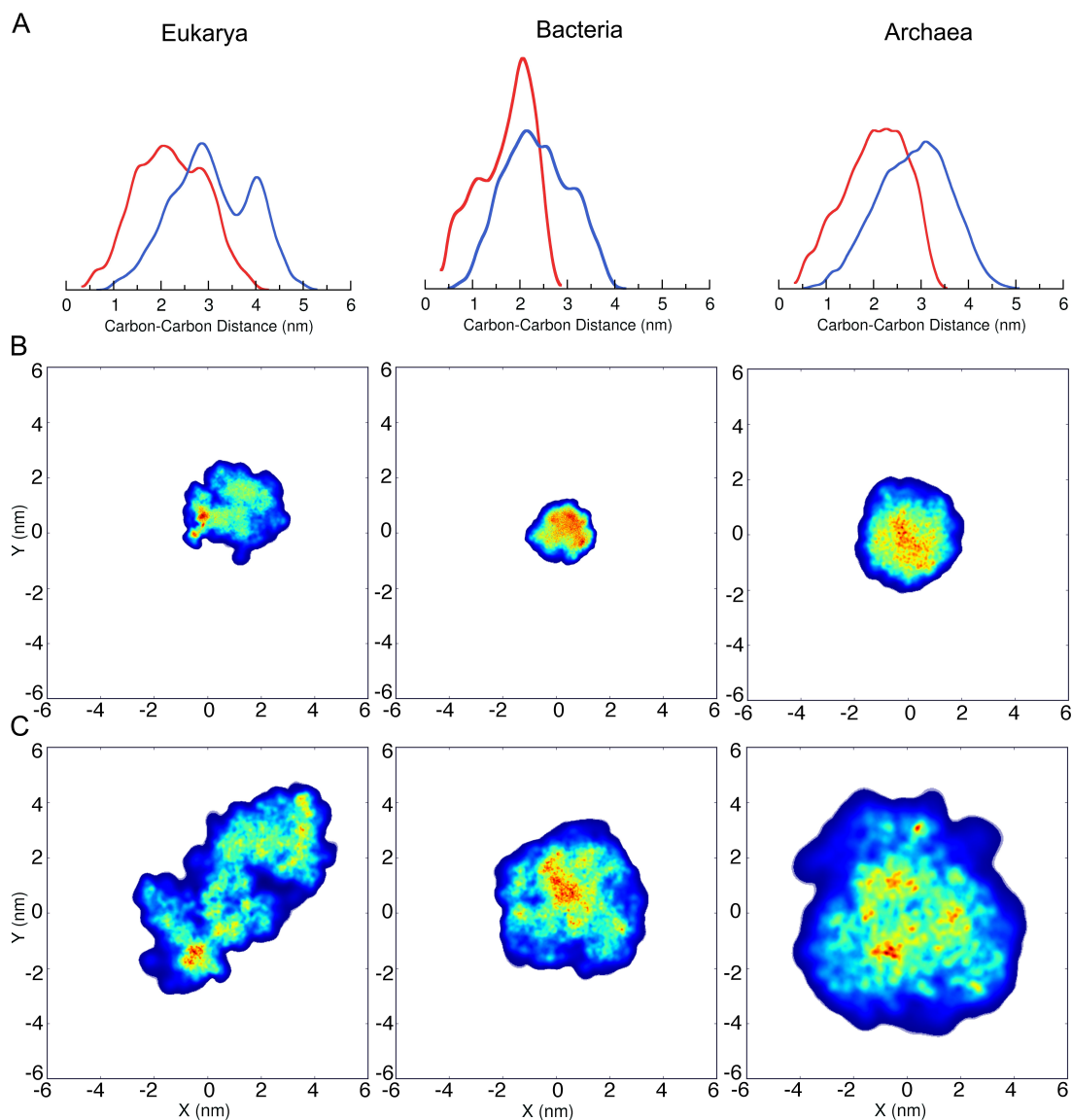
9

Figure 4: Distributions of the LLOs hydrophobic tail in lipid bilayer. A) histograms of the distances between C1A and C9E - Eukarya, C6E - Bacteria, C8E - Archaea (red line) and carbons C1A and C19E - Eukarya, C11E - Bacteria, and C12E - Archaea (blue line) for every LLO. B) 3D heatmap of C9E - Eukarya, C6E - Bacteria, C8E - Archaea occupancy and C) of C19E - Eukarya, C11E - Bacteria, and C12E - Archaea occupancy in membrane, taking C1A as [0,0] cartesian coordinates. In B and C, each value corresponds to the coordinates populated in X (horizontal) and Y (vertical) by the selected carbons. Highly occupied regions are colored in red, while poorly occupied region are colored in blue.

Membrane bilayer dynamics in the presence of LLOs

In order to detect possible effects of LLOs presence on the surrounding biochemical environment, we evaluated membrane properties, such as area per lipid and thickness, throughout

the 1 μs simulations (Table 3). For Eukarya and Bacteria LLOs, the presence of glycan chains did not influenced the area per lipid, even though they present drastic differences on the number of saccharides ramifications or sugar length. However, the presence of glycan chain on Archaea LLO yielded a higher area per lipid than its LLO in absence of the sugar moiety, suggesting that its composing monossacharides are capable to recruit lipids. Moreover, the presence of glycan chains did not altered membrane thickness for any of the simulated systems. However, despite being composed by the same POPE membrane lipids, both Bacteria and Archaea systems presented different membrane thickness values, suggesting that the length of the lipid tail may influence on local lipid structure and dynamics. These findings suggest that, depending on LLO chemical composition, these glycoconjugates may rearrange local membrane structure properties, which might be related to its biological function.

Table 3: Average Area per lipid and thickness of membranes throughout MD trajectory.

LLO	Area per lipid (Å^2)	Thickness (Å)
Eukarya	62.69 ± 1.02	38.07 ± 0.57
Bacteria	59.02 ± 0.83	40.15 ± 0.52
Archaea	61.83 ± 1.28	37.75 ± 0.67
Eukarya (no glycans)	61.41 ± 1.14	38.63 ± 0.62
Bacteria (no glycans)	58.49 ± 0.92	40.16 ± 0.56
Archaea (no glycans)	57.44 ± 1.13	39.22 ± 0.62

LLOs' oligosaccharidic moiety orientation in membranes

Aiming to properly describe the motion and orientation of the oligosaccharidic portion from LLOs, we analyzed the tilt angle of the glycan chain relative to the vector normal to the membrane. For this, we defined a vector connecting distant monosaccharides, as follows: for the bacterial LLO, the monosaccharide bacilosamine (Bac) nearest to the reducing end and the one in the oxidizing end (NAc-Galactosamine) were selected; for the archaeal LLO, two measurements were performed (due to the existing branch ramifications in the glycan), one

using 1 and 4' (red line), and the other one using 1 and 4 (blue line); for the eukaryotic LLO, also two measurements were taken, using 1 and 3' (red line), and 1 and 3 (blue line). As shown in Figure 5A, the main vector indicates that the preferred position for all glycan chains is in parallel with the bilayer (~ 90 degrees). This observations are further confirmed by the average positions of the saccharadic residues relative to the Z axis (Figure 5B), showing that the majority of the monosaccharide residues are close to the membrane.

Correspondingly, the same behavior is depicted by the residence time of all saccharidic residues at the membrane (Figure 5C). Most residues interact with the membrane during the whole simulations time, with few exceptions, such as Archaea residues 4, 5, and 3' with a small decrease and Eukarya residues 3, 4, and D3, which possess lower residence time interacting with the membrane. Possibly, an higher number of ramifications and the presence of flexible glycosidic bonds (β-1,6 bond type) could yield a decrease on the residence time for some monosaccharide residues.

Furthermore, in order to confirm the orientation behavior observed in unbiased MD simulations, we employed metadynamics calculations to describe the free-energy surfaces (FES) of the oligosaccharidic motion relative to the membrane. To obtain this, we employed a collective variable (CV - See Experimental Section) which could differentiate between the possible orientation states of the glycan chain. As shown in Figure 6, there are two low energy states depicted by the FES of the tilt angle: one is at 0 degrees, where the oligosaccharide chain is perpendicular to the membrane, and one comprehending values between 60 and 100 degrees (depending on the LLO type), where the glycan is in parallel with the membrane. For all systems, the lowest energy state detected is the one in parallel with the bilayer, with some of the structures located in this minimum possessing a slightly curved shape (Figure 6). The ΔG between states ranges from 35 to 70 kJ/mol, reinforcing a clear preference for this conformational state.

Looking at the carbohydrate residues coloured by root mean square fluctuation (Figure 5D), it is possible to identify that, in general, the central core of the glycan chains is manly

rigid, while flexibility is higher at the ramifications and at the terminal residues. The bacterial LLO is indeed the most rigid among them, probably due to the absence of branching in its structure. The archaeal and eukaryotic LLOs show increased mobility in their chains, especially the latter with its multiple branched residues, which can sometimes prevent a full interaction in a parallel pattern with the membrane. As seen in the FES, the archaeal oligosaccharide chain has a shifted minimum, but this does not correlate with an increased flexibility in its structure. A possible explanation to this is the presence of the three sulphate groups that can interact firmly with the polar headgroups, recruiting more phospholipids to its surrounding, a feature that might block a further increase in the tilt angle.

Finally, we performed conformational analysis on LLOs glycan chains based on the glycosidic dihedral linkages using the same methodology proposed by Arantes et al.[21] (Table 4). For better understanding, in the Supporting Information, Figure S4 shows each glycosidic linkage on the LLOs structure with their dihedrals angles ϕ, ψ and ω highlighted. As previously demonstrated for bacterial LLO,[22] the average conformational profile of the oligosaccharide chain attached to the lipid remains mostly rigid during the simulations of all LLOs, with minor fluctuations to less frequent conformational states. These results may be related to the interaction of oligosaccharide chains with the phosphate head groups of the membranes during the MD simulations of all LLOs. On the other hand, a higher flexibility is observed for the dihedral angles of monosaccharides linked to phosphate (ϕ_1 and ψ_1), when compared to the remaining glycosidic linkages. This behavior may be related to the lipid chain movement within the membrane. These data were used for identification of the most abundant conformations of LLOs molecules and subsequent docking calculations.

Molecular Docking of LLOs in each OST model

By exploring the conformational space of glycan moieties sampled of each simulation using the same approach as described in Arantes et al., we were able to identify conformational populations for each LLO ensemble and submit them to molecular docking calculations

Table 4: Average dihedral angles for the glycosidic linkages composing LLOs and their relative abundance throughout MD trajectory.

Archea

	Frequency	Average(^{o})
ϕ_1	1.00	99.77 ± 17.80
ψ_1	0.88	172.43 ± 25.67
	0.11	-84.07 ± 14.33
ϕ_2	0.91	-34.96 ± 19.99
	0.09	28.38 ± 9.93
ψ_2	1.00	119.64 ± 12.77
ϕ_3	1.00	-66.91 ± 14.20
ψ_3	1.00	-97.08 ± 13.73
ϕ_4	1.00	79.45 ± 15.26
ψ_4	1.00	-131.40 ± 18.43
ϕ_5	1.00	91.84 ± 14.10
ψ_5	1.00	-98.68 ± 20.19
ϕ_6	0.99	75.68 ± 12.57
ψ_6	1.00	-149.85 ± 19.64
ϕ_7	0.99	79.58 ± 12.76
ψ_7	0.25	81.38 ± 13.72
	0.75	178.34 ± 19.11
ω_7	1.00	88.28 ± 10.36

Bacteria

	Frequency	Average(^{o})
ϕ_1	0.58	109.78 ± 10.98
	0.42	143.13 ± 10.11
ψ_1	0.68	-156.49 ± 11.02
	0.32	-111.59 ± 6.62
ϕ_2	1.00	71.32 ± 7.51
ψ_2	1.00	-141.53 ± 7.95
ϕ_3	0.99	92.46 ± 8.37
ψ_3	0.99	-88.41 ± 10.04
ϕ_4	1.00	91.91 ± 17.20
ψ_4	0.26	-124.25 ± 7.45
	0.74	-90.60 ± 10.10
ϕ_5	0.99	91.98 ± 9.40
ψ_5	1.00	-92.17 ± 14.83
ϕ_6	1.00	97.06 ± 17.30
ψ_6	0.14	-129.59 ± 8.69
	0.85	-87.16 ± 11.31

Eukarya

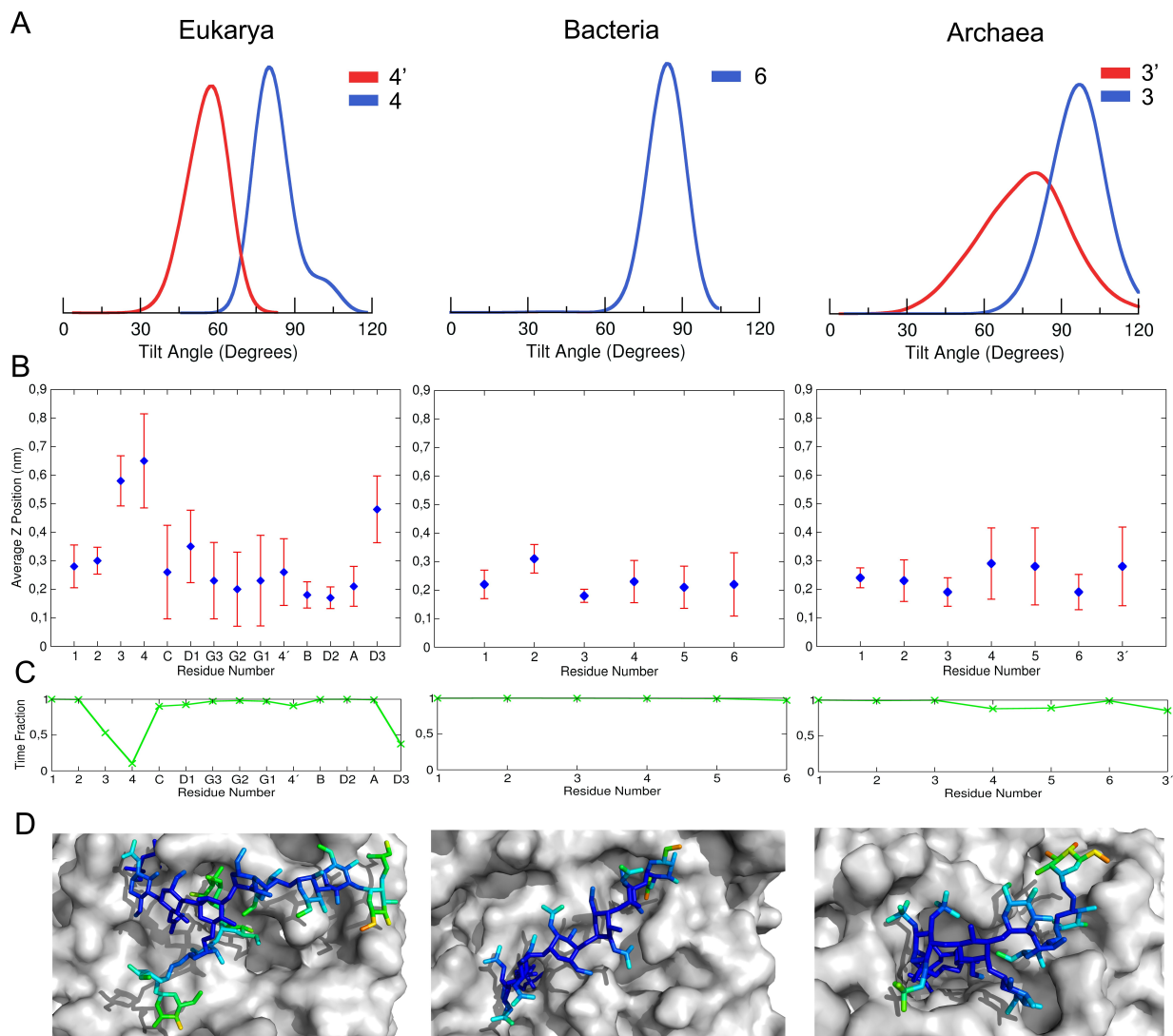|	Frequency	Average(^{o})
ϕ_1	1.00	113.21 ± 13.63
	0.23	90.43 ± 9.85
ψ_1	0.70	139.88 ± 12.41
	0.07	-86.32 ± 13.57
ϕ_2	1.00	-37.80 ± 17.85
ψ_2	1.00	105.13 ± 10.16
ϕ_3	1.00	-85.92 ± 9.93
ψ_3	1.00	94.30 ± 9.57
ϕ_4	1.00	142.94 ± 12.57
ψ_4	1.00	-178.73 ± 11.27
ω_4	1.00	169.45 ± 9.83
ϕ_5	1.00	138.98 ± 18.13
ψ_5	0.92	-171.39 ± 19.19
	0.08	-102.05 ± 9.54
ω_5	0.97	169.78 ± 11.42
ϕ_6	0.99	84.19 ± 18.58
ψ_6	1.00	-102.21 ± 19.96
ϕ_7	1.00	92.11 ± 10.78
ψ_7	1.00	-93.32 ± 10.98
ϕ_8	0.98	71.05 ± 15.37
ψ_8	0.87	-139.40 ± 11.19
	0.13	-86.10 ± 10.83
ϕ_9	1.00	97.31 ± 14.10
ψ_9	1.00	-90.93 ± 17.47
ϕ_{10}	0.52	76.63 ± 12.72
	0.48	123.15 ± 13.66
ψ_{10}	0.40	-125.12 ± 13.73
	0.60	-73.58 ± 13.51
ϕ_{11}	1.00	96.60 ± 14.56
ψ_{11}	1.00	-86.29 ± 16.47
ϕ_{12}	1.00	88.98 ± 12.42
ψ_{12}	0.32	-129.95 ± 9.06
	0.68	-92.60 ± 13.29
ϕ_{13}	1.00	76.13 ± 10.96
ψ_{13}	0.91	-137.17 ± 10.47
	0.09	-95.70 ± 7.42
ϕ_{14}	1.00	77.51 ± 8.88
ψ_{14}	0.99	83.97 ± 11.24

Figure 5: Tilt angles calculated for the vectors describing the motion from LLOs' glycan chains. A) Probability density histograms describing tilt angle values found during unbiased MD simulations. The blue line describes the vector connecting 1 and 4 in Eukarya, 1 and 6 in Bacteria, and 1 and 3 in Archaea. The red line describes the vector connecting residues 1 and 4' in Eukarya, and 1 and 3' in Archaea. B) Positions of each sugar residue relative to the Z-axis are represented by blue dots (average position in blue, error bars in red); C) Residence time of each sugar residue in the membrane interface, calculated using a cutoff of 4.5 Å.D) Root mean square fluctuation B-factor of LLOs. The color scale ranges from blue (hard) to yellow/orange (flexible).

(Experimental Section) in order to test if the obtained models were capable of binding to its corresponding OST counterpart (Figure 7).

For bacterial LLO, the molecule mostly displayed a similar orientation as seen in the
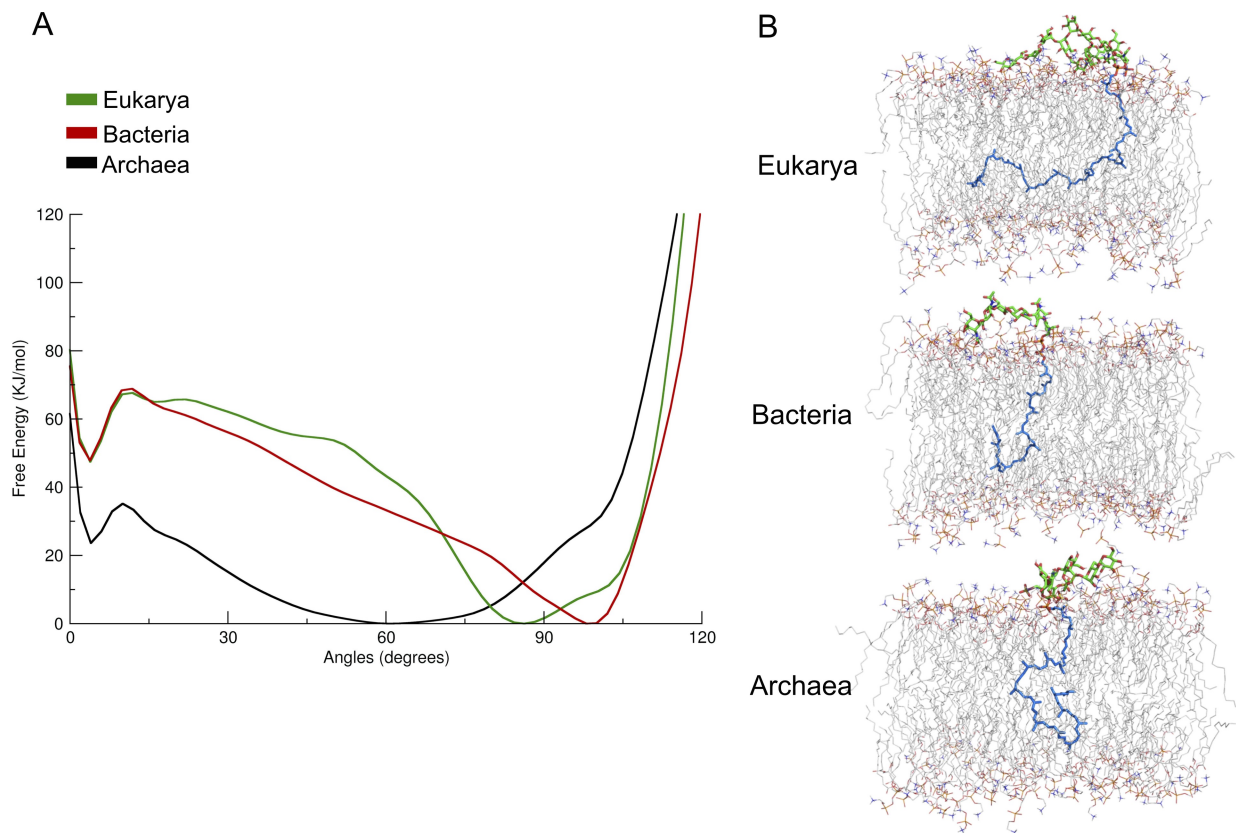
Figure 6: Metadynamics simulations results: A) Free-energy surface (FES) in function of tilt angle for LLOs' glycan chains. B) Representative conformations of the eukaryotic, the bacterial and the archaeal LLOs, in their respective main minimum energy region. Average Error in tilt angles of LLOs: Eukarya = 1.72 ± 2.84 KJ/mol, Bacteria = 2.23 ± 2.73 KJ/mol and Archaea = 2.63 ± 0.88 KJ/mol.

recent crystal structure (PDB ID 6GXC)[23] of the full complex with an acceptor peptide and an inactive LLO, but with closer distances between the peptide acceptor and the LLO donor (3.5 Å vs 4.0 Å in the recent crystallographic structure). Main relevant distances reproduced on our model (apart from the restricted ones) includes: i) atom O7 from the N-acetylation in C2 and the OH group of Tyr468; ii) OH group from Tyr196 and the phosphate groups in LLO; iii) OH group from Ser198 and the phosphate groups. The undecaprenyl portion fitted into the previously identified hydrophobic groove, while the saccharidic moiety interacted along the cavity formed in the interface between transmembrane (TM) domain and PP domain.

The archaeal LLO docking revealed an interesting feature for the WWDXGX motif, which
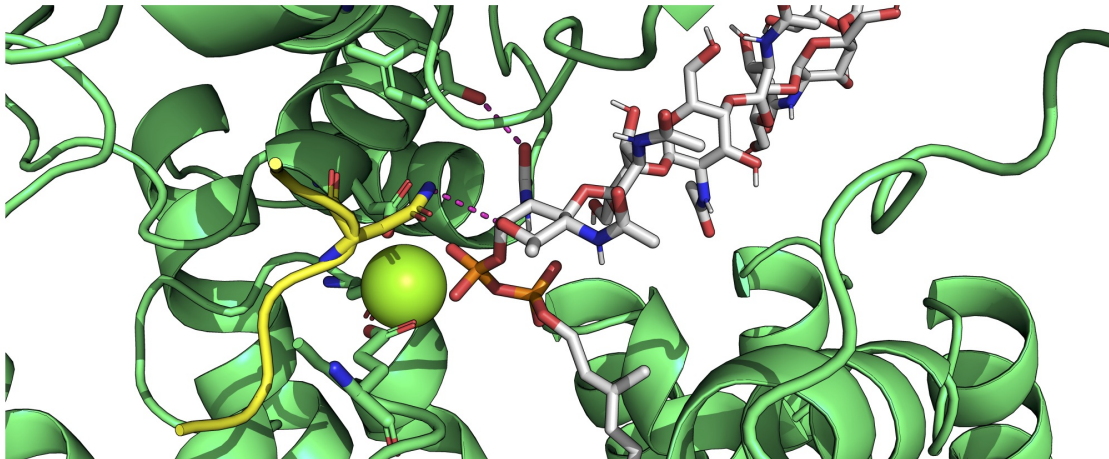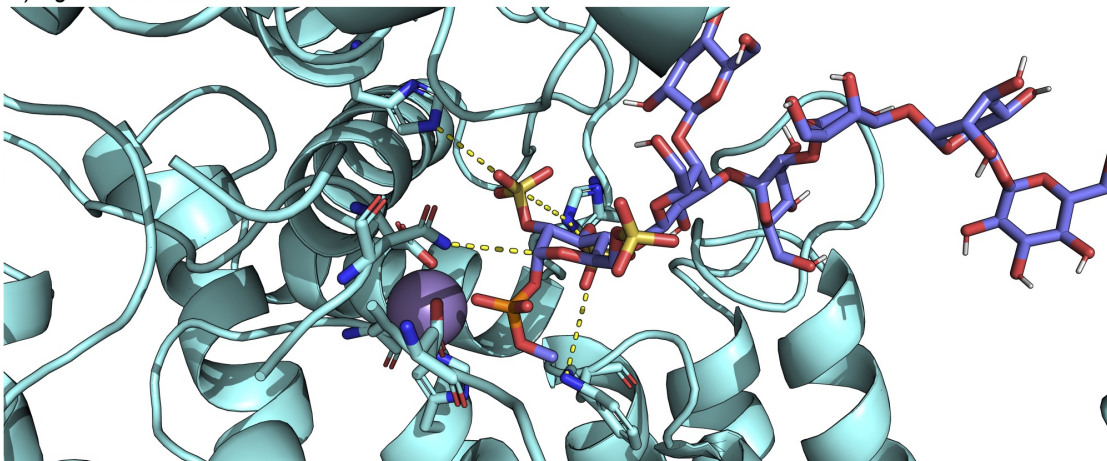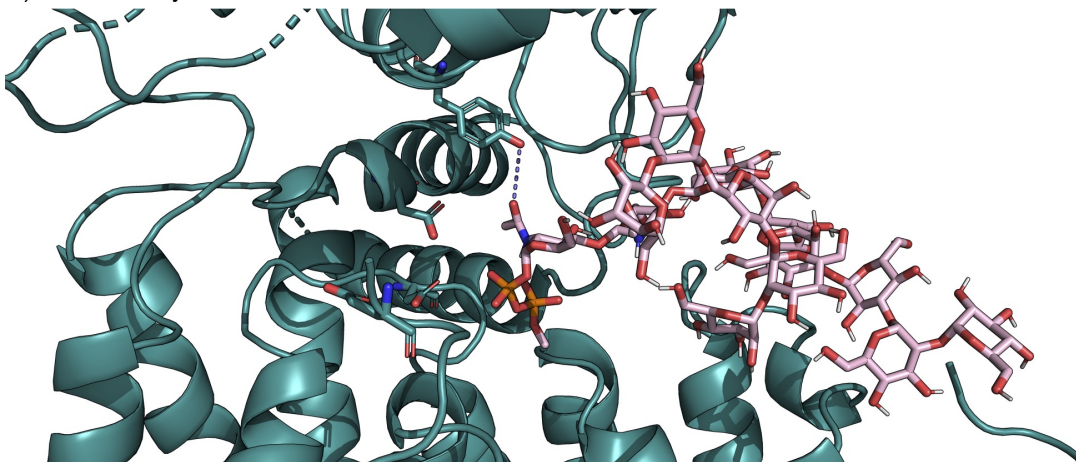
Figure 7: Docking poses obtained for each studied LLO inside the corresponding OSTs. A) Bacterial LLO complexed with PglB depicting interactions Bac-O7 and Tyr468-OH, Asn-ND2 and Bac-C1; B) Archaeal LLO complexed with AglB depicting interactions Glc1-S2 and His555-ND1, Glc1-S3 and Trp215-NE1, Asn-ND2 and Glc1-C1; C) Eukaryotic LLO complexed with STT3 subunit depicting interactions NAcGlc1-O7 and Tyr530-OH, Asn-ND2 and NAcGlc-C1. All phosphate groups interacted with the respective ion, when present in the crystallographic structure.

in *A. fulgidus* AglB (PDB ID 5GMY)[24] is composed of WWDYGH. His555 plays the role of Tyr468 in the WWDYGY motif of *C. jejuni* PglB, performing electrostatic interactions between its imidazole side-chain ring and the sulfate group connected to the C2 of the reducing end monosaccharide from archaeal LLO (Glc1). This observations indicate that the WWDXGX motif is directly involved in the selectivity of the donor substrate of OSTs and the enzyme possibly evolved to better accommodate distinct glycan ligands. Another residue substitution occurs in the archaeal enzyme, where residue Tyr79 in *C. lari* becomes His81 in *A. fulgidus*. The position occupied by this amino acid, in close proximity to the sulfate group connected to C2 of the first monosaccharide, suggests that His81 (and Tyr79 in *C. lari*) acts as an additional interaction that ensures correct positioning of the LLO inside the enzyme cavity. Besides that, another residue was observed to interact with LLO phosphate groups, namely Trp215, performing a similar role seen in the enzymes from the other domains of life (Trp208 in eukarya and Tyr196 in Bacteria),[25,26] explaining the decrease in the enzyme activity when it was mutated to alanine.[11] The remaining carbohydrate residues of the molecule are in contact with AglB domains interface, where the last two monosaccharide are located near the N-terminal region of external loop 5 (EL5).

The eukaryotic LLO docking was performed in the single catalytic subunit (PDB ID 6FTI),[9] achieving a pose that fitted adequately into the donor cavity, despite this structure being obtained without the coordinating ion at the catalytic site, which organizes the structure around this area. When comparing our final docking complex to the small LLO found in the cryo-EM structure of the mammalian OST complex, it is possible to observe that the full LLO obtained from our data was not able to reproduce the exact conformation found in the experimental data. This is probably due to the larger number of monosaccharide units and the distinct conformations sampled in our simulations. Despite that, alignment of the currently formed complex to the full OST complex did not generate any clashes with the other subunits. Based on our obtained model and on the three reports of the recently released structures of the full OST complex, we verified that the glycan portion of the LLO molecule

must interact with EL5, while inducing the ordering of transmembrane helice 9 (TMH9), and that EL5 must be in a flexible state, not defined by a specific secondary structure. We could not observe a predicted entry route, since the high fluctuations of the hydrophobic tail generated conformations that were not compatible for fitting of its whole structure within the OST transmembrane domain.

Conclusion

Understanding conformation and dynamics of distinct LLOs could support a better comprehension of their complex formation with OSTs, and, consequently, provide insights to enhance the efficiency of OSTs glycosylation turnover, as well as allow the engineering of these enzymes aiming biotechnological applications. In this work, we characterized three LLOs, representatives for each domain of life, employing a series of molecular modelling techniques.

By properly parameterizing the isoprenoid unit (one of the building blocks of the LLOs hydrophobic tail) and achieving good agreement with experimental data (error less than 10%), we were able to describe the highly flexible behavior of this portion of the molecule, demonstrating its distribution inside the membranes. Additionally, glycan chains were also evaluated in respect of their motions and interactions, demonstrating the clear preference for a parallel position relative to the membrane bilayer, mostly in close contact to the phospholipids head groups, similarly as previous reports.[12]

Molecular docking trials accessed the complexes for each LLO inside the respective OSTs cavities, which respected cutoff distances expected from previously determined OSTs structures. By correlating recent crystallographic data from the bacterial OST (PglB)[26] with our results from the archaeal LLO docking model, we were able to elaborate an explanation for the change in amino acids in the WWDXGX motif. This observations highlights the substrate selectivity role of this motif, allowing the prediction of interactions between glycans

19

and this region. Furthermore, we described a rationale for the amino acid substitutions seen in Archaea, and the sulfate group in C2 observed on this LLO. Besides that, all of our models sampled conformations that could properly fit inside the donor substrate binding site, further validating our findings. We hope this work contributes for future investigations regarding the glycosylation activity and the distinct carbohydrate moieties distributed among other species.

Experimental

Nomenclature and software

The IUPAC proposed recommendations for nomenclature and symbols were used. Regarding MD simulations, the GROMACS 5.0.7 simulation suite[27] was employed, along with the GROMOS53A6 force field[15] for the systems without the glycan chain and the GROMOS53A6GLYC force field[28,29] for glycan chain, in which X is "2", "3", "4" or "6" and its ϕ and ψ angles are defined as shown below:

$$\phi_{(1\to X)} = O5 - C1 - OX - CX \tag{2}$$

$$\psi_{(1\to X)} = C1 - OX - CX - C(X-1) \tag{3}$$

Finally, for (1→6) linkage, the ω is defined as shown below:

$$\omega = O6 - C6 - C5 - C4 \tag{4}$$

Generation of new torsional parameter for isoprenoid chain

The QM torsional profile for the dihedrals within the isoprenoid structure were obtained using Gaussian09.[30] The 2-Methylbut-1-ene molecule was used due to their similarity to the

isoprenoid portion. QM calculations were carried out using the relaxed *scan* routine (with geometry optimization at each scan point) combined with a *tight* convergence criterion, at HF/6-31G(d), MP2/6-31G(d), B3LYP/6-31G(d) and MP2/6-311++G(d) theory levels, obtaining the relative energy associated with the rotation of dihedral by increments of 10°. Moreover, similar calculations were carried out for the torsion in 3-Methyl-trans-2-pentene and 3-Methyl-cis-2-pentene molecules in order to evaluate possible differences in transition barriers. Analogue MM calculations were performed in GROMACS 5.0.7, using the force field parameter set GROMOS53A6, as described in Pol-Fachin et al. and Arantes et al..[17,21,29] The MM torsional profile was fitted to the QM profile of 2-Methylbut-1-ene obtained with MP2/6-31G(d) theory level in the Rotational Profiler server,[31] providing proper torsional parameters for MM calculations to yield a torsional profile similar to the QM. These new parameters were then implemented in the LLOs topologies for MD simulations.

Parametrization strategy and topology construction

In order to describe LLOs through molecular mechanics simulations, a set of lipids commonly found in membranes was selected to act as building blocks in this work. Our parametrization strategy was based on accurately reproducing experimental values for physical-chemical properties of liquids. Topologies were constructed for LLOs using the potentials for bond stretching, bond-angle bending, improper dihedral deformation and proper dihedral, as well as van der Waals interactions terms retrieved directly from GROMOS53A6[15] set, while potentials for diphosphate linkages were retrieved from previous works.[29] Moreover, in order to maintain compatibility with the general GROMOS force field, partial charges for the sulfate group were selected from those already distributed within GROMOS53A6 for the phosphate group (Table S2). Glycan chains topologies were built using the potentials for bond stretching, bond-angle bending, improper dihedral deformation and van der Waals interactions terms were retrieved directly from GROMOS53A6GLYC force field.[28,29] Such models had their glycosidic linkage geometries adjusted to the main conformational states

for each linkage, based on their relative abundance in the isolated disaccharides in water, as previously described.[22,32,33]

Liquid and gas-phase simulations for assessment of thermodynamic properties

MD simulations were carried out by using GROMACS 5.0.7 package and physical-chemical properties of organic liquids (density and enthalpy of vaporization) were used as target to validate topology parameters, as previous works of parametrization of small biomolecules[16,17,21] and benchmark of force fields.[18] The geraniol fragment was chosen due to the availability of experimental values of density and enthalpy of vaporization. In order to calculate such thermodynamic properties, liquid phase was induced by simulating 125 randomly placed molecules in a cubic box under 100 bar and, after, scaling the simulation box in X, Y and Z axis in order to obtain a simulation box of 1000 molecules in liquid phase. All simulations were carried out with the Berendsen pressure and temperature coupling algorithms[34], using $\tau_T = 0.2$ ps and $\tau_P = 0.5$ ps, following previous works of GROMOS parametrization strategies[15]. Non-bonded interactions settings followed the reccomendations of Gonçalves et al.[35] by using particle mesh Ewald (PME) method[36] and a cutoff of 1.2 Å to treat electrostatic interactions, while using a cutoff scheme to treat van der Waals interactions within 1.2 Å. Additionally to the liquid-phase simulation of 10ns, a gas-phase simulation was carried out for 100 ns using *leap-frog* and langevin algorithm[37], respectively. The potential energies associated with these systems ($E_{pot}(g)$ for gas-phase and $E_{pot}(l)$ for liquid-phase) were extracted and used to calculate (Eq. 5) the enthalpy of vaporization (ΔH_{vap}) of the fragments.

$$\Delta H_{vap} = (E_{pot}(g) + k_B T) - E_{pot}(l) \tag{5}$$

Organic liquid densities (ρ) were calculated from liquid-phase simulations using block averages of 5 blocks, as for ΔH_{vap}.

Membrane insertion of LLOs

Initial LLOs structures were oriented along the Z axis, perpendicular to the membrane patch. A bilayer model constituted of 120 POPE lipids for bacterial and archaea systems and a bilayer constituted of 120 POPC lipids for eukaryotic, were obtained at the end of the protocol (Figure 8). In order to insert LLOs into their respective membrane patch, The InflateGRO methodology[38] was employed. Briefly, this consists in inserting LLOs into a pre-equilibrated bilayer patch, with lipids overlapping the LLOs, and then expanding the dimensions of the box, as well as, translating all lipids laterally, so that no more overlap is found. After that, a series of minimization steps are applied, along with the box dimensions being compressed and the lipids translated back to the center of the system, until the system reaches the desired density. During all this process, the LLOs are under strong position restraints, so that the structures are not affected.
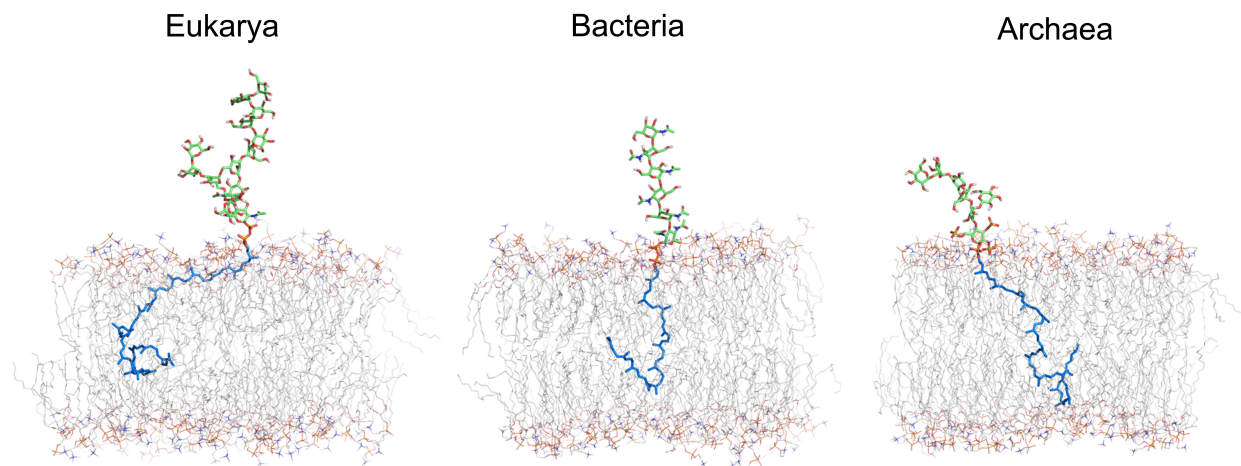

Figure 8: Initial models of the eukaryotic, bacterial, and archaeal LLOs embedded in membrane bilayers. Water molecules are omitted in this representation.

Molecular Dynamics Simulations

Following the membrane insertion steps, the rectangular box was then solvated with SPC water model[39] and periodic boundary conditions. Before this process, we did the common procedure of raising the van der Waals radius of the C atom from 0.15 to 0.375, aiming to

avoid water molecules filling any left space between the lipids. After solvation, this parameter was returned to its original value. Counter ions were added to neutralize the systems, when needed. Steepest Descent algorithm was used in the energy minimizations performed. First, two simulations of equilibration were performed with position restraints: an NVT and an NPT ensembles of 2 ns and 50 ns, respectively. Subsequently, one microsecond unrestrained NPT MD simulations were performed for each of the six systems, generating the production run where data were collected for the systems analysis. The LINCS algorithm[40] was chosen to constrain covalent bond lengths and an integration step of 2 fs was applied. Calculations of electrostatic interactions were performed by using the particle mesh Ewald (PME) method[36] and a cutoff of 1.2 Å, while a cutoff scheme and a cutoff of 1.2 Å was used for Lennard-Jones contributions[35]. The pressure barostat chosen was Parrinello–Rahman,[41,42] with a 2.0 ps coupling constant, while the temperature thermostats chosen were V-rescale (NVT step)[43] and Nosé–Hoover (NPT equilibration and production MD),[44,45] with a coupling constant of $\tau = 0.5$. Additionally, semiisotropic pressure coupling was applied, accounting for the presence of the membrane. Constant temperature of 310 K (Eukarya and Bacterial), 353 K (Archaea) and constant pressure of 1 atm were also implemented.

Well-tempered metadynamics simulations

For the structural assessment of glycan chains tilt angle in respect to the membrane, well-tempered metadynamics calculations were performed. Following the membrane insertion steps, a rectangular box was used and periodic boundary conditions applied. The systems were submitted to energy minimization by Steepest Descent algorithm, followed by an equilibration phase of 20 ns and subsequently to well-tempered (WT) metadynamics simulations. A CV was chosen to obtain discrimination between the orientation states of the glycan chain. This CV computed the angle between two fixed virtual atoms, positioned at the three-dimensional space of the box with XYZ coordinates corresponding to [3.0,3.0,8.57] and [3.0,3.0,5.55], and the center of geometry from the last monosaccharide from each LLO.

These coordinates for the fixed virtual atoms were selected so that when the glycan chain is fully perpendicular this value would be equal to 0 and when the glycan chain is fully parallel it would equals to around 90 degrees. A bias factor of 15 was employed to rescale gaussian functions. The initial height of the Gaussian hills was set as 1.0 kcal·mol^{-1}, while the hill width was chosen to be 0.1 degrees. An upper limit was defined for the CV around 115 degrees to avoid unwanted membrane insertion. The Parrinello–Rahman barostat[41,42] was employed, with a 2.0 ps coupling constant, to keep pressure constant at 1 atm. Temperature values were coupled by a V-rescale thermostat (NVT step),[43] using a constant of $\tau = 0.1$. Covalent bond lengths were constrained by the Lincs method,[40,46] allowing an integration step of 2 fs. PME method[36] was applied to treat electrostatic interactions. GROMACS 2018[47] interfaced with the PLUMED plugin package 2.5b[48] was chosen to perform the enhanced sampling simulations. To generate the free energy surfaces of the tilt angle, the sum hills tool from the PLUMED package was employed. Block-analysis technique was utilized for error calculation and the reweighting procedures were applied similarly to a previous report.[49] The eukaryotic and the archael LLOs were simulated during 500 ns, at 310 K and 353K respectively. The bacterial LLO was simulated during 250ns, at 310K. Systems were simulated until the FES were considered converged (hence, the different simulation run times): this was verified by evaluating the average deviation between free-energy surfaces generated at every 10ns. FES profiles were considered converged when reaching an average deviation below 5 kJ/mol.

Identification of LLOs Conformations

A conformational characterization tool initially described in Arantes et al.(2019) was used in order to characterize conformational populations of LLOs embedded in membranes. Briefly, dihedral angles are calculated throughout MD simulations for each ligand torsion and the code identifies dihedral populations based on dihedral distributions sampled throughout the trajectory. After, the algorithm evaluates each time step in search for different tuples of

dihedral populations arrays. A unique tuple of dihedral populations is defined as a ligand conformer and, therefore, all conformations sampled in MD trajectory and identified by the same tuple belong to the same conformational population. The algorithm outputs the time percentage in which a simulation visited each dihedral population and the relative abundance of each conformational population throughout MD trajectory. For the characterization of LLOs conformers, all ϕ, ψ (and ω, when applicable) torsions of each LLO were used. For further details of torsions used, see Table 4.

Molecular Dockings

After identifying the most abundant LLO conformations embedded in membrane, we submitted the most prevalent structures to PatchDock molecular docking[50] server against their corresponding OST counterparts, PglB (PDB ID 6GXC),[23] AglB (PDB ID 5GMY)[24] and STT3 (PDB ID 6FTI).[9] Patchdock molecular docking, which uses a rigid docking algorithm, uses shape complementarity principles with two main steps: i) determining the molecular shape representation to find possible binding geometric patches, and ii) matching of the surface patches between the two molecules. Additionally, we employed distance restraints to reduce the search space, similarly to a previous report.[12] For the eukaryotic LLO docking, restraints were employed between OST residue Tyr530 at atom OH to monosaccharide Gn1 at atom O7, with boundaries limited between 1.0 to 5.0 Å. For the archaeal LLO docking, pairs of restraints were added for more atoms: i) acceptor peptide residue Asn-ND2 and atom C1 of the LLO G1 (min. 1.0 Å and max. 4.0 Å); ii) Aglb residue Arg426-NH1 and NH2 and atom P76 of LLO (min. 1.0 Å and max. 6.0 Å); iii) Zn^{2+} catalytic ion to atom P76 of LLO (min. 1.0 Å and max. 5.0 Å). The bacterial LLO docking was performed by setting restraints in the distances between: i) acceptor peptide residue Asn-ND2 and C1 of the LLO B1 (min. 1.0 Å and max. 4.0 Å); ii) Mg^{2+} catalytic ion and atom P2 of LLO (min. 1.0 Å and max. 6.0 Å); iii) Both NH1 and NH2 of Arg375 and atom P2 of LLO (min. 1.0 Å and max. 6.0 Å). From this methodology, we have obtained at least one complex that

respected the expected distances and was located around the expected donor cavity. For the manipulation and visualization of structures, the softwares VMD[51] and PyMOL[52] were employed.

Acknowledgement

Supporting Information Available

The following files are available free of charge.

- Supplementary material: Density profiles for the subdivided parts of the studied LLOs, without glycan chains. Distributions of the LLOs, without glycan chains, hydrophobic tail in lipid bilayer and schematic representations of the oligosaccharide chains of LLOs.

References

(1) Varki, A. Biological Roles of Oligosaccharides: All of the Theories are Correct. *Glycobiology* **1993**, *3*, 97–130.

(2) Larkin, A.; Imperiali, B. The Expanding Horizons of Asparagine-Linked Glycosylation. *Biochemistry* **2011**, *50*, 4411–4426.

(3) Kowarik, M.; Numao, S.; Feldman, M. F.; Schulz, B. L.; Callewaert, N.; Kiermaier, E.; Catrein, I.; Aebi, M. N-Linked Glycosylation of Folded Proteins by the Bacterial Oligosaccharyltransferase. *Science (New York, N.Y.)* **2006**, *314*, 1148–50.

(4) Kowarik, M.; Young, N. M.; Numao, S.; Schulz, B. L.; Hug, I.; Callewaert, N.; Mills, D. C.; Watson, D. C.; Hernandez, M.; Kelly, J. F.; Wacker, M.; Aebi, M. Definition of the Bacterial N-Glycosylation Site Consensus Sequence. *The EMBO journal* **2006**, *25*, 1957–1966.

(5) Weerapana, E.; Imperiali, B. Asparagine-Linked Protein Glycosylation: from Eukaryotic to Prokaryotic Systems. *Glycobiology* **2006**, *16*, 91–101.

(6) Schwarz, F.; Lizak, C.; Fan, Y. Y.; Fleurkens, S.; Kowarik, M.; Aebi, M. Relaxed Acceptor Site Specificity of Bacterial Oligosaccharyltransferase in Vivo. *Glycobiology* **2011**, *21*, 45–54.

(7) Wacker, M.; Feldman, M. F.; Callewaert, N.; Kowarik, M.; Clarke, B. R.; Pohl, N. L.; Hernandez, M.; Vines, E. D.; Valvano, M. A.; Whitfield, C.; Aebi, M. Substrate Specificity of Bacterial Oligosaccharyltransferase Suggests a Common Transfer Mechanism for the Bacterial and Eukaryotic Systems. *Proceedings of the National Academy of Sciences* **2006**,

(8) Wacker, M.; Linton, D.; Hitchen, P. G.; Nita-Lazar, M.; Haslam, S. M.; North, S. J.; Panico, M.; Morris, H. R.; Dell, A.; Wren, B. W.; Aebi, M. N-Linked Glycosylation in Campylobacter Jejuni and its Functional Transfer into E. Coli. *Science* **2002**, *298*, 1790–1793.

(9) Braunger, K.; Pfeffer, S.; Shrimal, S.; Gilmore, R.; Berninghausen, O.; Mandon, E. C.; Becker, T.; Förster, F.; Beckmann, R. Structural Basis for Coupling Protein Transport and N-glycosylation at the Mammalian Endoplasmic Reticulum. *Science* **2018**, *360*, 215–219.

(10) Lizak, C.; Gerber, S.; Numao, S.; Aebi, M.; Locher, K. P. X-Ray Structure of a Bacterial Oligosaccharyltransferase. *Nature* **2011**, *474*, 350–355.

(11) Matsumoto, S.; Shimada, A.; Nyirenda, J.; Igura, M.; Kawano, Y.; Kohda, D. Crystal Structures of an Archaeal Oligosaccharyltransferase Provide Insights into the Catalytic Cycle of N-Linked Protein Glycosylation. *Proceedings of the National Academy of Sciences of the United States of America* **2013**, *110*, 17868–73.

(12) Kern, N. R.; Lee, H. S.; Wu, E. L.; Park, S.; Vanommeslaeghe, K.; Mackerell, A. D.; Klauda, J. B.; Jo, S.; Im, W. Lipid-Linked Oligosaccharides in Membranes Sample Conformations That Facilitate Binding to Oligosaccharyltransferase. *Biophysical Journal* **2014**, *107*, 1885–1895.

(13) Fujinami, D.; Nyirenda, J.; Matsumoto, S.; Kohda, D. Structural Elucidation of an Asparagine-Linked Oligosaccharide from the Hyperthermophilic Archaeon, Archaeoglobus Fulgidus. *Carbohydrate Research* **2015**, *413*, 55–62.

(14) Taguchi, Y.; Fujinami, D.; Kohda, D. Comparative Analysis of Archaeal Lipid-Linked Oligosaccharides That Serve as Oligosaccharide Donors for Asn Glycosylation. *Journal of Biological Chemistry* **2016**, *291*, 11042–11054.

(15) Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. A Biomolecular Force Field Based on the Free Enthalpy of Hydration and Solvation: the Gromos Force-Field Parameter Sets 53A5 and 53A6. *Journal of computational chemistry* **2004**, *25*, 1656–76.

(16) Pedebos, C.; Pol-Fachin, L.; Verli, H. Unrestrained Conformational Characterization of Stenocereus Eruca Saponins in Aqueous and Nonaqueous Solvents. *Journal of Natural Products* **2012**, *75*, 1196–1200.

(17) Polêto, M.; Rusu, V.; Grisci, B.; Dorn, M.; Lins, R.; Verli, H. Aromatic Rings Commonly Used in Medicinal Chemistry: Force Fields Comparison and Interactions with

Water Toward the Design of New Chemical Entities. *Frontiers in Pharmacology* **2018**, *9*, 395.

(18) Caleman, C.; Van Maaren, P. J.; Hong, M.; Hub, J. S.; Costa, L. T.; Van Der Spoel, D. Force Field Benchmark of Organic Liquids: Density, Enthalpy of Vaporization, Heat Capacities, Surface Tension, Isothermal Compressibility, Volumetric Expansion Coefficient, and Dielectric Constant. *Journal of Chemical Theory and Computation* **2012**, *8*, 61–74.

(19) Chickos, J. S.; Jr., W. E. A. Enthalpies of Vaporization of Organic and Organometallic Compounds, 1880–2002. *Journal of Physical and Chemical Reference Data* **2003**, *32*, 519.

(20) Haynes, W. M. *CRC handbook of chemistry and physics*; CRC press, 2014.

(21) Arantes, P. R.; Polêto, M. D.; John, E. B. O.; Pedebos, C.; Grisci, B. I.; Dorn, M.; Verli, H. Development of Gromos-Compatible Parameter Set for Simulations of Chalcones and Flavonoids. *The Journal of Physical Chemistry B* **2019**, *123*, 994–1008.

(22) Pedebos, C.; Arantes, P. R.; Giesel, G. M.; Verli, H. In Silico Investigation of the Pglb Active Site Reveals Transient Catalytic States and Octahedral Metal Ion Coordination. *Glycobiology* **2015**, *25*, 1183–1195.

(23) Napiórkowska, M.; Boilevin, J.; Darbre, T.; Reymond, J.-L.; Locher, K. P. Structure of Bacterial Oligosaccharyltransferase Pglb Bound to a Reactive LLO and an Inhibitory Peptide. *Scientific Reports* **2018**, *8*, 16297.

(24) Matsumoto, S.; Taguchi, Y.; Shimada, A.; Igura, M.; Kohda, D. Tethering an N-Glycosylation Sequon-Containing Peptide Creates a Catalytically Competent Oligosaccharyltransferase Complex. *Biochemistry* **2017**, *56*, 602–611.

(25) Li, H.; Chavan, M.; Schindelin, H.; Lennarz, W. J.; Li, H. Structure of the Oligosaccharyl Transferase Complex at 12 Å Resolution. *Structure* **2008**, *16*, 432–440.

(26) Napiórkowska, M.; Boilevin, J.; Sovdat, T.; Darbre, T.; Reymond, J. L.; Aebi, M.; Locher, K. P. Molecular Basis of Lipid-Linked Oligosaccharide Recognition and Processing by Bacterial Oligosaccharyltransferase. *Nature Structural and Molecular Biology* **2017**, *24*, 1100–1106.

(27) Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; Van Der Spoel, D.; Hess, B.; Lindahl, E. Gromacs 4.5: a High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *29*, 845–854.

(28) Pol-Fachin, L.; Rusu, V. H.; Verli, H.; Lins, R. D. Gromos 53A6 Glyc , an Improved Gromos Force Field for Hexopyranose-Based Carbohydrates. *Journal of Chemical Theory and Computation* **2012**, *8*, 4681–4690.

(29) Pol-Fachin, L.; Verli, H.; Lins, R. D. Extension and Validation of the Gromos 53A6(Glyc) Parameter Set for Glycoproteins. *Journal of computational chemistry* **2014**, *35*, 2087–95.

(30) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.;

Knox, J. E.; Cross, J. B. et al. Gaussian09 Revision E.01. Gaussian Inc. Wallingford CT 2009.

(31) Rusu, V. H.; Baron, R.; Lins, R. D. PITOMBA: Parameter Interface for Oligosaccharide Molecules Based on Atoms. *Journal of chemical theory and computation* **2014**, *10*, 5068–5080.

(32) Pol-Fachin, L.; Serrato, R. V.; Verli, H. Solution Conformation and Dynamics of Exopolysaccharides from Burkholderia Species. *Carbohydrate Research* **2010**, *345*, 1922–1931.

(33) Pol-Fachin, L.; Fernandes, C. L.; Verli, H. Gromos96 43A1 Performance on the Characterization of Glycoprotein Conformational Ensembles through Molecular Dynamics Simulations. *Carbohydrate Research* **2009**, *344*, 491–500.

(34) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *The Journal of Chemical Physics* **1984**, *81*, 3684.

(35) Gonçalves, Y. M. H.; Senac, C.; Fuchs, P. F. J.; Hünenberger, P. H.; Horta, B. A. C. Influence of the Treatment of Nonbonded Interactions on the Thermodynamic and Transport Properties of Pure Liquids Calculated Using the 2016H66 Force Field. *Journal of Chemical Theory and Computation* **2019**, *15*, 1806–1826.

(36) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: an N-Log(N) Method for Ewald Sums in Large Systems. *The Journal of Chemical Physics* **1993**, *98*, 10089.

(37) Van Gunsteren, W. F.; Berendsen, H. J. C. A Leap-Frog Algorithm for Stochastic Dynamics. *Mol. Simulat.* **1988**, *1*, 173–185.

(38) Kandt, C.; Ash, W. L.; Peter Tieleman, D. Setting Up and Running Molecular Dynamics Simulations of Membrane Proteins. *Methods* **2007**, *41*, 475–488.

(39) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *Journal of Physical Chemistry* **1987**, *91*, 6269–6271.

(40) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. Lincs: a Linear Constraint Solver for Molecular Simulations. *Journal of Computational Chemistry* **1997**, *18*, 1463–1472.

(41) Parrinello, M. Polymorphic Transitions in Single Crystals: a New Molecular Dynamics Method. *Journal of Applied Physics* **1981**, *52*, 7182.

(42) Nosé, S.; Klein, M. L. Constant Pressure Molecular Dynamics for Molecular Systems. *Molecular Physics* **1983**, *50*, 1055–1076.

(43) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *The Journal of chemical physics* **2007**, *126*, 014101.

(44) Nosé, S. A Molecular Dynamics Method for Simulations in the Canonical Ensemble. *Molecular Physics* **1984**, *52*, 255–268.

(45) Hoover, W. G. Canonical Dynamics: Equilibrium Phase-Space Distributions. *Physical Review A* **1985**, *31*, 1695–1697.

(46) Hess, B. P-Lincs: a Parallel Linear Constraint Solver for Molecular Simulation. *Journal of Chemical Theory and Computation* **2008**, *4*, 116–122.

(47) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations Through Multi-level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1*, 19–25.

(48) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Computer Physics Communications* **2014**, *185*, 604 – 613.

(49) Branduardi, D.; Bussi, G.; Parrinello, M. Metadynamics with Adaptive Gaussians. *Journal of Chemical Theory and Computation* **2012**, *8*, 2247–2254.

(50) Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H. J. Patchdock and Symmdock: Servers for Rigid and Symmetric Docking. *Nucleic Acids Research* **2005**, *33*, 363–367.

(51) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *Journal of Molecular Graphics* **1996**, *14*, 33–38.

(52) Schrödinger, LLC, The PyMOL Molecular Graphics System, Version 1.8. **2015**,

Graphical TOC Entry