



**HAL**  
open science

## Martini 3 Coarse-Grained Force Field for Carbohydrates

Fabian Grünewald, Mats Punt, Elizabeth Jefferys, Petteri Vainikka, Melanie König, Valtteri Virtanen, Travis Meyer, Weria Pezeshkian, Adam Gormley, Maarit Karonen, et al.

► **To cite this version:**

Fabian Grünewald, Mats Punt, Elizabeth Jefferys, Petteri Vainikka, Melanie König, et al.. Martini 3 Coarse-Grained Force Field for Carbohydrates. *Journal of Chemical Theory and Computation*, 2022, 10.1021/acs.jctc.2c00757 . hal-03864651

**HAL Id: hal-03864651**

**<https://cnrs.hal.science/hal-03864651>**

Submitted on 25 Nov 2022

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

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Martini 3 Coarse-Grained Force Field for Carbohydrates

Fabian Grünewald<sup>1\*</sup>, Mats H. Punt<sup>1\*</sup>, Elizabeth E. Jefferys<sup>2</sup>, Petteri A. Vainikka<sup>1</sup>, Valtteri Virtanen<sup>3</sup>, Melanie König<sup>1</sup>, Weria Pezeshkian<sup>1,5</sup>, Maarit Karonen<sup>3</sup>, Mark S. P. Sansom<sup>2</sup>, Paulo C. T Souza<sup>4†</sup>, Siewert J. Marrink<sup>1†</sup>

<sup>1</sup> Groningen Biomolecular Sciences and Biotechnology Institute and Zernike Institute for Advanced Materials, University of Groningen, Groningen, The Netherlands

<sup>2</sup> Department of Biochemistry, University of Oxford, South Parks Road, Oxford, United Kingdom, OX1 3QU

<sup>3</sup> Natural Chemistry Research Group, Department of Chemistry, University of Turku, FI-20014 Turku, Finland

<sup>4</sup> Molecular Microbiology and Structural Biochemistry, UMR 5086 CNRS and University of Lyon, Lyon, France

<sup>5</sup> The Niels Bohr International Academy, Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark

\* shared first authors

† corresponding author

*Supporting Information Placeholder*

---

**ABSTRACT:** The Martini 3 force field is a full re-parametrization of the Martini coarse-grained model for biomolecular simulations. Due to the improved interaction balance it allows for more accurate description of condensed phase systems. In the present work we develop a consistent strategy to parametrize carbohydrate molecules accurately within the framework of Martini 3. In particular, we develop a canonical mapping scheme that decomposes arbitrarily large carbohydrates into a limited number of fragments. Bead types for these fragments have been assigned by matching physicochemical properties of mono- and disaccharides. In addition, guidelines for assigning bonds, angles, and dihedrals are developed. These guidelines enable a more accurate description of carbohydrate conformations than in the Martini 2 force field. We show that models obtained with this approach are able to accurately reproduce osmotic pressures of carbohydrate water solutions. Furthermore, we provide evidence that the model differentiates correctly the solubility of the poly-glucoses dextran (water soluble) and cellulose (water insoluble, but soluble in ionic-liquids). Finally, we demonstrate that the new building blocks can be applied to glycolipids, being able to reproduce membrane properties and to induce binding of peripheral membrane proteins. These test cases demonstrate the validity and transferability of our approach.

---

## 1 – Introduction

Carbohydrates (sugars) are an important class of biomolecules. They play an active role in cell biology as they are, for example, part of the cell metabolism<sup>1</sup>, or signaling pathways<sup>2</sup>. In addition, they are structural building blocks for many bio-macromolecules such as polysaccharides, glycosylated proteins and lipids, and nucleotides. Furthermore, in research for sustainable materials carbohydrates are a key factor as they can be obtained from renewable stock.<sup>3</sup> Therefore, simulating these molecules in complex systems by molecular dynamics (MD) is of high interest to a wide audience of researchers. MD studies can give near atomistic resolution of processes not possible to capture with experimental techniques and therefore often complement experimental studies.

Due to the limits in special temporal resolution of models representing all atoms explicitly, so-called coarse-grained (CG) models are often used in MD simulations. In CG models, several atoms are grouped into one effective interaction site. This greatly increases the simulation speed and reduces computational costs. Among the most popular CG models for (bio-) molecular dynamics is the Martini model.<sup>4,5</sup> The Martini model has been widely applied across many fields ranging from biomolecular science to material science.<sup>6-9</sup> In the Martini model,<sup>5</sup> about four heavy atoms are grouped into one interaction center, called bead. The interactions between beads represent the nature of the underlying chemical groups; the strength of the interaction is selected from a discrete set of LJ interactions by

reproducing thermodynamic data mostly the free energies of transfer between water and different organic solvents. In addition to the regular Martini beads, smaller bead sizes (S- and T-beads) are used for groups that are represented at higher resolution such as aliphatic or aromatic ring fragments.<sup>5,10</sup>

Within the framework of the previous version of the Martini model (i.e. version 2), several parameters for carbohydrates have been developed and successfully applied.<sup>11–26</sup> However, Martini 2 has several pitfalls when it comes to parameterization of molecules, which lead to unphysical behavior.<sup>27</sup> This was especially apparent for the carbohydrate model. As pointed out by several authors carbohydrates in Martini 2 tend to largely overestimate the self-aggregation propensity.<sup>12,13</sup> Although some of these problems could be alleviated by either increasing the interaction strength of carbohydrates with water<sup>13</sup>, or by replacing regular bead types with small beads<sup>18</sup>, these solutions were *ad-hoc* and did not resolve the underlying imbalance of the bead interactions. To overcome these deficiencies the third edition of the Martini force field comprises a complete reparameterization of the original force field. Rebalancing of the non-bonded interaction as well as extended verification against physicochemical reference data, leads to an improved description of previously problematic molecular interactions.<sup>5,10,28–31</sup>

In the present work we develop a consistent strategy to parametrize arbitrary carbohydrate molecules accurately within the framework of Martini 3. In particular, we develop a canonical mapping scheme that decomposes large carbohydrates into mono- and disaccharides, which are parameterized based on matching physicochemical reference properties and atomistic reference simulations. To facilitate application of this scheme, automatic mapping from all-atom simulations is implemented in the fast-forward program<sup>32</sup> for all carbohydrate fragments considered. In addition, we propose guidelines for assigning bonds, angles, and dihedrals to allow for a more accurate description of carbohydrate conformations than in the Martini 2 force field. At the moment bonded interactions for specific complex carbohydrates need to be mapped from an atomistic reference simulation unless the specific compounds are presented in this paper. Generic bonded parameters are subject to a forthcoming publication.

The remainder of this paper is organized as follows: First we present the parametrization strategy of carbohydrates starting with monosaccharides (Section 2.1) and subsequently extending to disaccharides and more complex carbohydrates (Section 2.2). Afterwards, we validate the transferability of our approach by demonstrating that we can accurately model four example systems, previously impossible to consistently model with Martini 2. In particular, reproduction of osmotic pressures of

monosaccharides (Section 3.1), solution and solubility behavior of two polysaccharides namely dextran and cellulose (Section 3.2 & 3.3), and glycolipid mediated binding of peripheral membrane proteins (Section 3.4). Finally, we discuss limitations of our new carbohydrate modeling strategy and conclude.

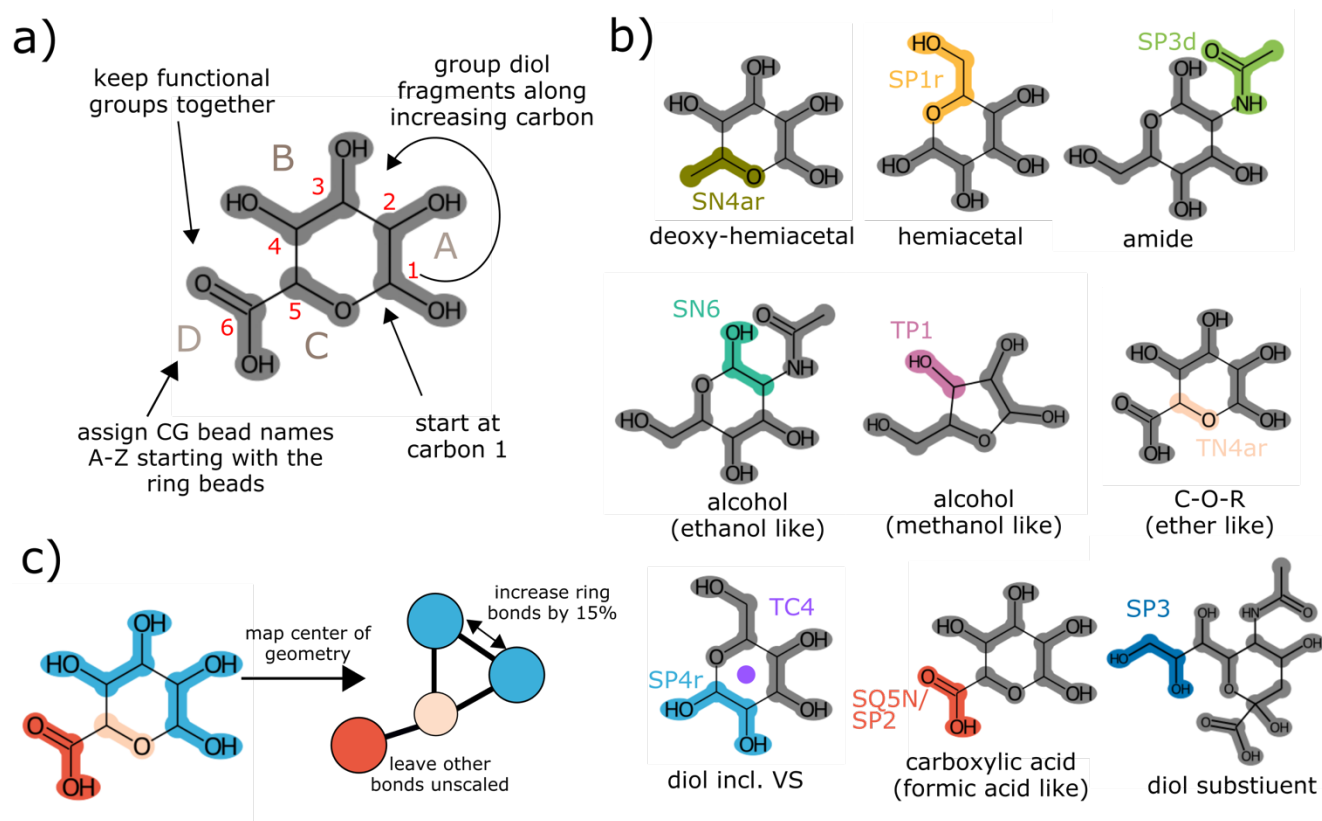
## 2 – Parametrization of carbohydrates with Martini 3

We derived parameters for Martini 3 carbohydrates following the general rules for creating Martini models as outlined in the main parameterization paper.<sup>5</sup> However, we aimed at not only deriving the optimal parameters for the specific carbohydrates considered, but also casting them in a consistent framework as much as possible such that we obtain a generalized strategy for modeling arbitrary carbohydrates.

### 2.1 – Monosaccharides

**Scope.** Carbohydrates as a molecule class display a large heterogeneity in size, structural connectivity and isomerization states. Most biologically and technologically important carbohydrates are either hemiacetal monosaccharides or formed by condensation reactions of hemiacetal sugars. The hemiacetal monosaccharides exist mostly as six membered carbon rings (pyranoses) or five membered carbon rings (furanoses). These monosaccharides can undergo enantiomerization reactions switching between an open form and a ring-closed form in solution. However, the total fraction of ring-open structures is typically very low. Therefore, in this paper we only consider ring-closed monosaccharides. Furthermore, the C1 carbon, which is called anomeric carbon, is chiral. This chirality of hemiacetal monosaccharides also causes a specific type of isomerization called anomerization. Depending on the position of the alcohol group connected to the anomeric carbon a carbohydrate either has a  $\alpha$ - (axial alcohol) or  $\beta$ - (equatorial alcohol) conformation. Within the Martini 3 carbohydrate model we do not distinguish between the anomers in the case of monosaccharides. As shown in the Table S1 the geometry (i.e. bond lengths) are so similar that we can treat these molecules as one class as was done in the previous Martini carbohydrate models.

**Mapping.** The mapping describes which atoms at the all-atom level are represented by a single bead in the CG model. The mapping choice determines all subsequent model choices and therefore requires careful consideration. In general, in the context of Martini 3 one maps 2-5 heavy atoms into one bead, where the number of mapped atoms and their connectivity determines the bead-sizes. All carbohydrate mappings are derived obeying the following three base rules, which are aimed at making the mappings transferable and consistent across also complex carbohydrates (Figure 1a).



**Figure 1. Parametrization strategy for monosaccharides. a) Systematic mapping scheme; b) Bead assignment for all fragments found in monosaccharides; c) Design principle for bonded interactions.**

A mapping: 1) maximizes the number of diols assigned to a single bead therefore maximizing the number of 4:1 mapped atoms; 2) keeps functional groups together as much as possible; 3) starts at the anomeric carbon and proceeds counterclockwise for grouping fragments. The first rule recognizes that the most commonly found fragments in sugars are diols and the hemiacetal group. Having as many similar fragments as possible simplifies the model and allows to find a good bead type for that fragment across the many test cases. The second rule is needed in cases where substituent groups are present and supersedes rule 1, if needed. For example, in D-glucuronic acid (Figure 1a), the acid substituent group is kept together making the ring-fragment smaller and in the case of Neu5Ac the three side chains are also kept together (Figure S1). Thus, the ring fragment becomes a 3-1 mapping. The third rule ensures that equivalent fragments are generated for the different sugars and makes a canonical naming scheme possible. To simplify handling and analysis of our model we have developed a canonical naming scheme. The ring-beads in carbohydrates are named A, B, C, where the A bead is the first bead and always includes the anomeric carbon according to the previously defined mapping direction and the C bead always includes the ether oxygen. Substituents

are named after letters in the alphabet in order in which they are attached to the main ring beads. For example, in the case of glucuronic acid we have one substituent named D.

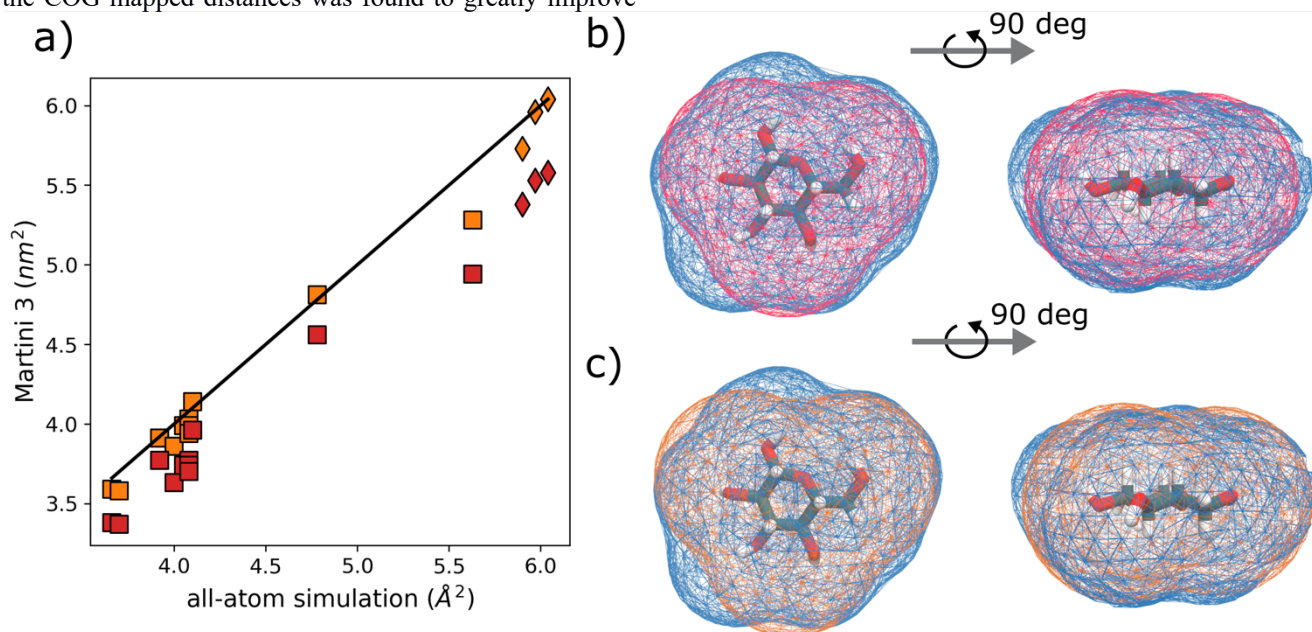
**Bonded interactions.** As the monosaccharides are rigid triangles at the CG level, we decided to model them using constraints. The constraint length was derived by mapping the center of geometry (COG) of the atomistic reference structure to the CG representation following the previously derived mapping scheme. One of the key problems of sugars in the old Martini 2 model was the overestimated aggregation propensity.<sup>12,13,27</sup> As analyzed by Alessandri and coworkers, this is partially caused by too short bond lengths and a bad representation of the molecular volume at the CG level.<sup>27</sup> Thus in Martini 3 it is recommended to match the molecular volume of the atomistic structure as closely as possible.<sup>27</sup> To assess the molecular volume, we conducted atomistic simulations of 11 monosaccharides, and 3 disaccharides at the all-atom level using the most popular force fields for sugars (i.e., Glycam06h<sup>33</sup>, Charmm36<sup>34</sup>, and GROMOS<sup>35</sup>). Based on these simulations the SASA was computed as outlined in the method section and compared to CG SASA values. As the SASA depends only on the bead size we assigned bead-types of an appropriate size, based on the mapping guidelines of the

Martini 3 model. As most of the fragments are 4:1 mapped that display branched moieties or 3:1 mapped linear groups, we considered mostly S-beads in accordance with the Martini 3 mapping guidelines. For some of the remaining 2:1 mapped fragments, the appropriate class of T-beads was used instead. Figure 2a shows the comparison of the Martini 3 SASA values against atomistic simulation data. The models, which are obtained by simply mapping the center of geometry (red symbols), underestimate the molecular volume significantly. Overall deviations for the unscaled model are of the order of 8%. Thus, it stands to reason that this approach leads to similar problems as observed in Martini 2.<sup>27</sup> To further elucidate the problem, we computed the Connolly surfaces<sup>36</sup> of the molecules involved. An example is shown for glucose in Figure 2 b-e. One can clearly see that the unscaled coarse-grained surface (blue) does not match the atomistic reference surface (red). To improve the agreement with the molecular volume we followed the approach suggested by Alessandri *et al.*<sup>27</sup> and increased the bond lengths of the beads forming the sugar rings. The scaling approach has also been used previously to improve interactions of PIP lipids, which contain a carbohydrate head group.<sup>30</sup> To keep our model transferable to carbohydrates not considered, we explored a compound independent scaling factor. A uniform scaling of 15% over the COG mapped distances was found to greatly improve

the agreement of the SASA (orange data-points Fig. 2a) and at the same time be applicable to all monosaccharides. Also, the Connolly surfaces show a better agreement (Figure 2 d,e). Now the red and the blue surface align well for most parts of the molecule.

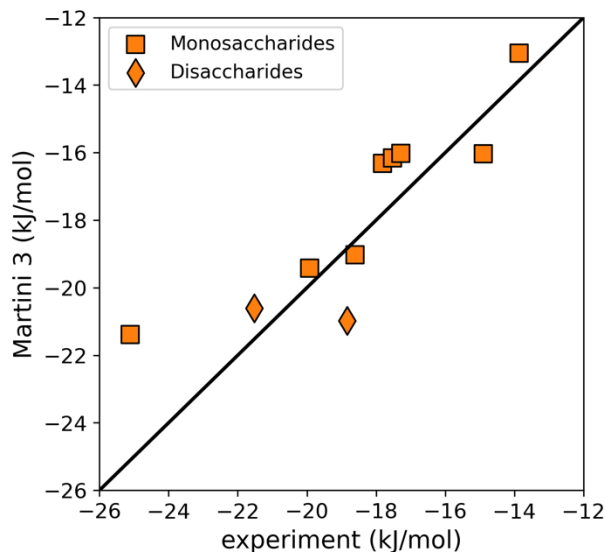
**Bead choices.** Non-bonded interactions are assigned from a discrete set of interaction levels referred to as bead-type by selecting those types that reproduce best the available physicochemical reference data. In this particular study we selected the bead types by matching the free energies of transfer from octanol to water for 11 monosaccharides. We note that only experimental values for glucose are available<sup>37-39</sup> ( $-17.52 \pm 1$ ). Thus, we set out to measure the remaining partition coefficients experimentally ourselves. The value obtained for glucose ( $-17.81 \pm 0.5$ ) matches the previously published values well, giving confidence in the choice of experimental method. Table S2 summarizes the experimentally determined partition coefficients.

The bead assignments were then optimized under the constraint that the same fragments must have the same bead type to be consistent with the building block approach of Martini. For example, inositol consists of three diol units. Thus, all beads in inositol must have the same type and it fixes the choice for the diol bead fragment.



**Figure 2. Molecular shape optimization.** a) Solvent accessible surface area (SASA) compared between atomistic reference simulations and Martini 3 with unscaled bonds (red) and scaled bonds (orange), for monosaccharides (squares) and disaccharides (diamonds); b) Connolly surfaces for glucose comparing atomistic (blue) to Martini 3 (red) before scaling the bonds c) Connolly surface for glucose after scaling the bonds comparing atomistic (blue) to Martini3 (orange)

Starting with that assignment the choices for the other fragments could be optimized. Figure 1b shows the final bead assignments for the monosaccharides considered. Figure 3 shows the correlation of the experimental versus coarse-grained free energies of transfer. We note that the mean-absolute error across all monosaccharides is only 1.5 kJ/mol which we consider excellent. For comparison, it is about the same as the average error in transfer free energy for the small molecules considered in the Martini 3 parametrization, which is 2.0 kJ/mol.<sup>5</sup>



**Figure 3. Free energies of transfer. Octanol - water free energies of transfer were computed using the Martini 3 model developed here and compared to either newly experimentally measured or existing literature values for monosaccharides (squares) and disaccharides (diamonds). See Table S2 for actual data. The error for all points was less than 0.4 kJ/mol.**

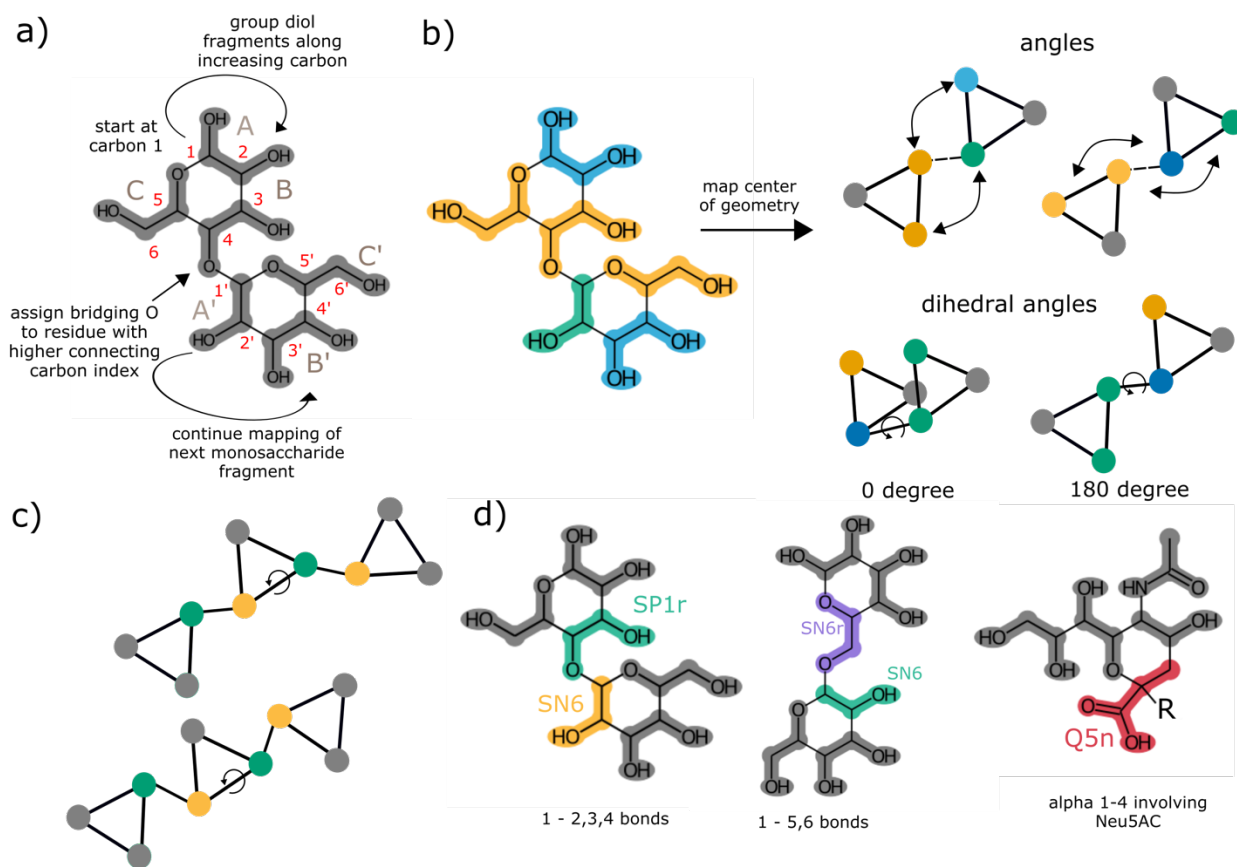
In addition to assigning a bead type for each fragment in the three membered ring our model also contains a virtual interaction site (VS), which is placed at the center of geometry of the ring. The VS is a mass-less interaction site that has the bead type TC4 across all monosaccharides. Note that all bead assignments were done with the TC4 interaction site present. This VS helps to reproduce interactions with aromatic groups — through so-called ring stacking<sup>40-42</sup> — which in Martini needs to be captured through a hydrophobicity component to the interaction. In addition, the extra TC4 bead avoids excessive mismatch of number of non-hydrogen by using S-beads representing 4 branched non-hydrogen atoms. As described in the Martini 3 guidelines, the maximum mismatch should be 1 non

hydrogen atom for each 10 non hydrogen atoms mapped by CG beads.<sup>5</sup> Such an approach has already been used successfully in parametrization of phosphatidylinositide lipids with Martini 3.<sup>30</sup> In the Supporting Information we assess the effectiveness of the virtual site by computing the potential of mean force (PMF) profiles between indol and glucose (Figure S2).

## 2.2 Disaccharides and more complex carbohydrates

**Mapping.** The mapping of disaccharides directly follows from the mapping of the monosaccharides, that is each constituting monosaccharide is mapped following the rules outlined above. The only thing to consider is the problem of where to separate the disaccharide into its constituting monomers. We adopt the same approach as previously used by the CHARMM-GUI glycan reader<sup>43</sup>, considering the connecting atoms to belong to the monosaccharide with the higher bead number at the CG level. For example, in the case of lactose ( $C_{12}H_{22}O_{11}$ ), which is an alpha-1,4 linkage of glucose and galactose (see Figure 4a), the connection is between the B and A bead of the CG model and between carbon 5 and 1 of the atomistic model. We consider the ether fragment to belong to the B bead fragment. This mapping allows to obtain transferable mappings between all disaccharides and since it is consistent with the CharmmGUI convention it allows automatic forward and backward mapping using already existing tools such as fast-forward<sup>32</sup> or backwards<sup>44</sup>. The complex carbohydrates and polysaccharides are mapped following the same principle, which also holds for branched carbohydrates such as the GM1 lipid. This scheme also allows Martini 3 carbohydrates to keep a building block approach, which was a concern in the previous Martini 2 model, where the CG geometry of the sugar rings needed to change from monosaccharides (triangular topology) to oligosaccharides models (linear).<sup>45</sup>

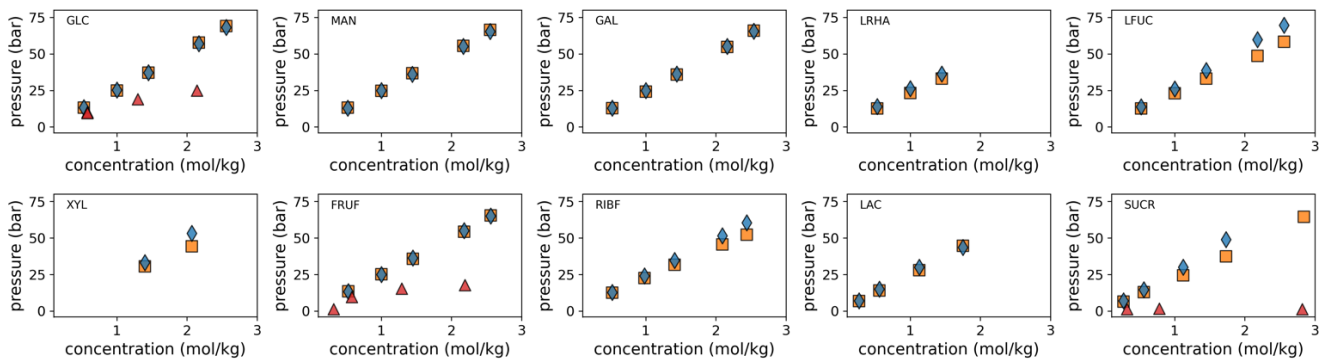
**Bonded interactions.** Bonded interactions are derived following the same strategy as for the monosaccharides that is mapping and matching the underlying atomistic reference distributions. To best represent the conformational space underlying carbohydrate oligomers and polymers we define all angles spanning the bond between two monosaccharide repeat units as well as one dihedral controlling the rotation around the glycosidic bond (Figure 4b). In case more than two monosaccharide repeat units are connected, an additional dihedral angle is introduced that spans three repeat units (Figure 4c). This dihedral angle defines the relative orientation of the n and n+2 residue with respect to the plane spanned by the n+1 residue.



**Figure 4. Parametrization strategy for oligo- and polysaccharides. a) Systematic mapping strategy for complex carbohydrates; b) Angles and dihedrals introduced between two linked monosaccharide fragments; c) Dihedral angle introduced between three consecutive monosaccharide fragments; d) Bead assignment for all fragments not found in monosaccharides.**

We notice that this dihedral is important especially for longer carbohydrates as it relates to the stiffness of the polysaccharides. Polysaccharides formed by condensation can either have a  $\alpha$  or  $\beta$  based bonds. We noticed that the difference in bond length at the CG level is significant between an  $\alpha$  or  $\beta$  bond due to the relative positioning of the two rings with respect to each other. Therefore, in our model we distinguish explicitly between the two anomers when found in a poly- or oligo-saccharide. In the case of dextran polymer to allow for better matching of the underlying AA distributions and improved stability, we use 3-bonded neighbor exclusions. However, for all other models only the 1-bonded neighbors are excluded as is standard in Martini lipids. Furthermore, for angles that are covered by any dihedral potential, the restricted bending potential introduced by Bulacu *et al.*<sup>46</sup> is used to reduce instabilities from angles becoming co-linear. As for the monosaccharides, we have assessed how well our model represents the molecular volume by computing SASA

values for three disaccharides lactose, sucrose, and trehalose (Figure 2a). The correlation with the atomistic SASA values is similarly good as for the monosaccharide case given the bond scaling is retained. However, it should be noted the connecting bonds between two monosaccharide repeat units are left unscaled. This minimal number of bonded interactions provides a good representation of the all-atom conformations and also leads to a numerically stable model. Moreover, it improves in selectivity and functionality compared to the Martini 2 carbohydrates, which treated glycosidic bonds indiscriminately and even had problems modelling flexible  $\alpha$ -1,6 linkages<sup>11</sup>. Note that bonded parameters for other complex carbohydrates (especially dihedrals) should be mapped from an atomistic reference simulation. Using generic bonded interaction to combine arbitrary carbohydrates is beyond the scope of the current paper and will be discussed in a forthcoming publication.



**Figure 5. Osmotic pressure of carbohydrate solutions.** Osmotic pressure for ten different carbohydrates measured from simulations using the presented Martini 3 model (orange) and the regular Martini 2 model (red) in comparison to experimental data (blue diamonds) collected from various sources<sup>47-50</sup>. Mono- and disaccharide codes are indicated with each panel: GLC=D-glucose, MAN=D-mannose, GAL=D-galactose, LRHA=L-rhamnose, LFUC=L-fucose, XYL=D-xylose, FRUF=D-fructose, RIBF=D-ribose, LAC=lactose, SUCR=sucrose. The error on CG osmotic pressures was less than 1 bar and thus smaller than the symbols shown.

**Bead choices.** Bead types of the fragments which are equivalent in both mono- and disaccharides are retained following the building block spirit of Martini. Therefore, the only bead types, which need to be defined are those involved in the glycosidic bond. As shown in Figure 4d, hexose bonds can be collected into two groups based on the newly generated fragments. One group contains the 1-1, 1-2, 1-3, and 1-4 bonds. The appropriate bead type of this fragment, SP1r, is directly taken from the monosaccharides (cf. Fig 1b). However, we further validated this choice by reproducing the free energies of transfer between octanol and water of the disaccharide’s trehalose and sucrose. Deviations for both compounds were acceptable with errors of 0.5 kJ/mol for trehalose and about 3.0 kJ/mol for sucrose. The other group contains 1-5 and 1-6 glycosidic bonds, in which case another bead type needs to be assigned.

The new bead is similar to the hemiacetal fragment but for the change of one OH group to an ether group. Such a change will likely result into less strong self-interactions, which is captured by the SN6r bead type having one level less strong self-interaction. Finally, in case of N-acetylated neuraminic acid attached via a 1-4 bond (Figure 4d last panel), we group the carboxylic acid together with the remaining carbon fragment in order to avoid a 2:1 fragment being generated with a short bond length. As a consequence, a new large bead is used instead of the two smaller beads. The bead type was determined to be the standard carboxylic acid bead from the original Martini 3 publication. We note that as a result of the generalized mapping scheme as well as the fact that biologically relevant carbohydrates are confined to certain linkages, with this small number of new

bead-types almost all biologically relevant sugars can be constructed.

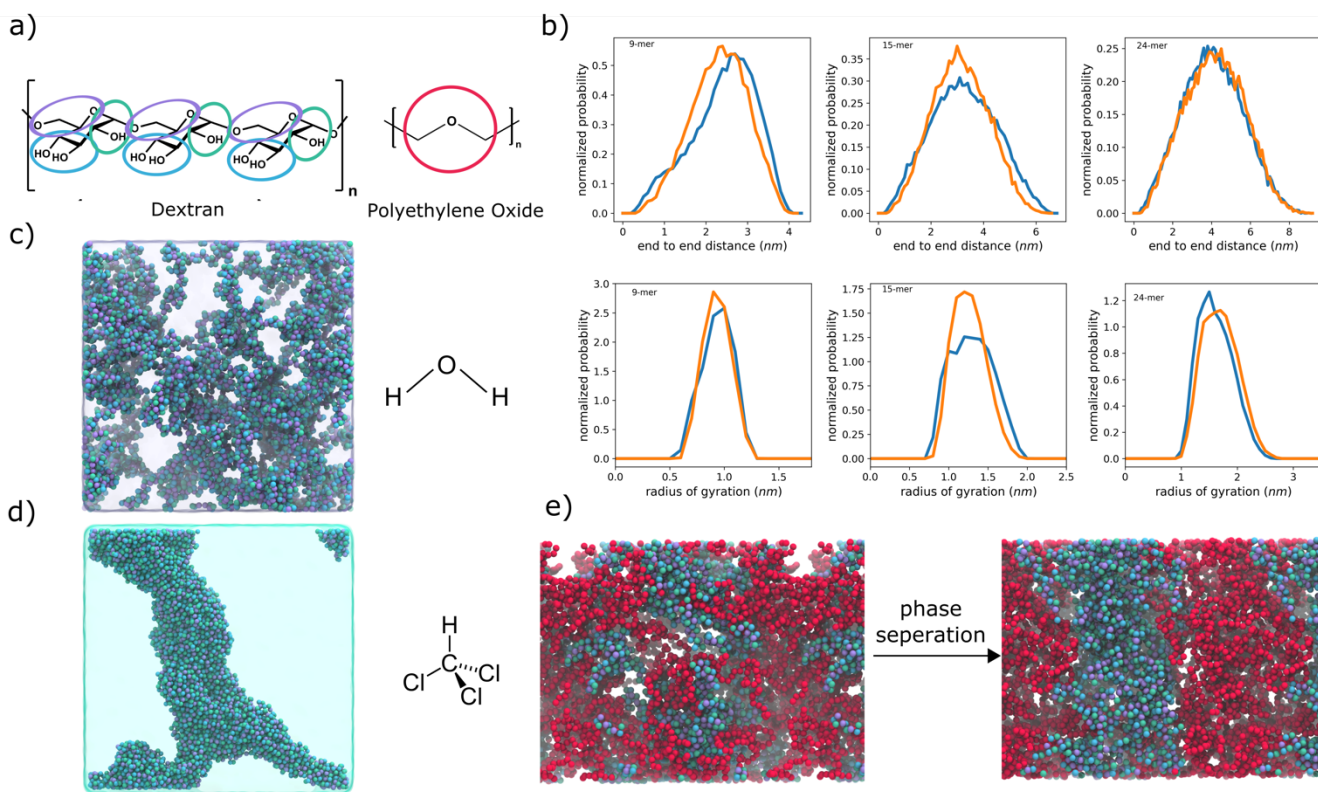
### 3.0 – Validation

In order to demonstrate the transferability of the model and assess if the model improves on the multiple issues of the Martini 2 model, we analyzed a number of test cases, considering four different target systems.

#### 3.1 Osmotic pressures

The major drawback of the Martini 2 carbohydrate model is the overestimation of the aggregation propensity.<sup>12,13,27</sup> To quantitatively assess the aggregation propensity of solute molecules in solution the osmotic pressure is frequently computed as function of the concentration. An osmotic pressure lower than experiment is indicative of too strong aggregation propensity. This procedure has been applied to reparametrize both CG and AA force fields for carbohydrates and other molecules.<sup>13,34,51,52</sup> To compute the osmotic pressure for our carbohydrates we have adopted the procedure originally proposed by Luo and Roux.<sup>53</sup> The molal concentrations were determined from the box density after preparing it at a certain molar concentration. Since experimental measurements are generally reported in molal units, we considered this approach to be more accurate at higher concentrations. Figure 5 shows the osmotic pressures for eight monosaccharides and two disaccharides in the concentration range from 0-2.5 molal. The scaled bond model (orange squares) shows an excellent agreement with the experimental data (blue diamonds) in the lower concentration range (<1.5 molal) across all carbohydrates. This already presents a significant improvement over the Martini 2 model for which data was available for only three carbohydrates (red triangles).





**Figure 6. Solution properties of dextran. a) Mapping of dextran and polyethylene oxide (PEO) at the Martini level. The colors of dextran correspond to the bead-types as found in Figure 1 and 4. b) Radius of gyration and end-to-end distance of dextran oligomers with 9, 15 and 24 repeat units from all-atom CHARMM36m<sup>43,54</sup> simulations (blue) and the new Martini 3 model (orange). c) Snapshot of aqueous solution of dextran (50 repeat units). d) Dextran globule formed in chloroform, a non-solvent, starting from dispersed polymers. e) Aqueous solution of dextran (65 repeat units) and PEO (180 repeat units) at the beginning of the simulation in the fully mixed initial state (left) and after 3 microseconds of simulation (right). Solvent is omitted for clarity.**

For the higher concentration range ( $>1.5$  molal), we see that our model follows the overall trends well but shows some deviations, in particular, for ribose, sucrose, fucose, and xylose. The lower pressures observed for these carbohydrates suggests some remnants of the stickiness problem to be still present. However, overestimated solute-solute interactions in concentrated solutions are not unique to CG force fields like Martini. Even popular atomistic force fields such as CHARMM36 or GLYCAM06 have been shown to significantly underestimate osmotic pressures and therefore exaggerate aggregation in simulation of carbohydrates in water. Agreement with experimental data is especially bad at higher concentrations ( $>2.5$  molal).<sup>34,51,55</sup> The reported deviations for these atomistic force fields are similar or even much worse than the deviations observed for our CG carbohydrate model.

Keeping this in mind, we conclude that our model reproduces the osmotic pressure very well overall and constitute a significant improvement over the Martini 2 carbohydrate models. In addition, we note that the accuracy

of our model is comparable to default atomistic force fields. Thus, we consider the match to be good enough.

### 3.2 Solution properties of dextran

Dextran is a branched polysaccharide, consisting of  $\alpha_{1,6}$  glucose units with  $\alpha_{1,3}$  connected glucose units branching off it.<sup>56,57</sup> Unlike other poly-glucoses such as cellulose or amylose, dextran is fully water soluble even at high molecular weight fractions.<sup>58</sup> Here we investigate the solution properties of dextran in water to demonstrate that our model is not only capable of reproducing properties of mono- and disaccharides but also properties of complex polysaccharide solutions. Whereas high molecular weight dextran is usually highly branched, lower molecular weights (with degrees of polymerization below 100) typically display only in the order of 5% or less branching.<sup>56,57</sup> Since we mostly utilize such lower molecular weights, all dextran used here is linear and has no branches. Dextran bead-types (Figure 6a) were assigned based on the previously presented concepts and the bonded parameters were obtained by

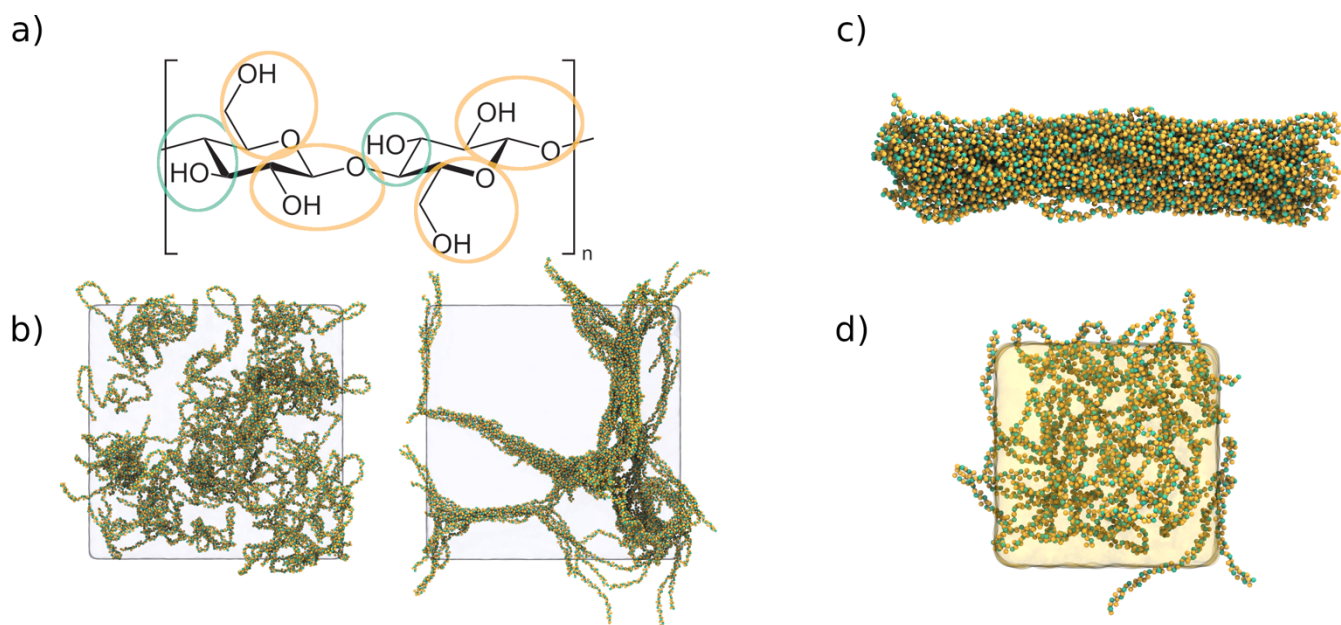
matching atomistic CHARMM36m<sup>43,54</sup> simulations. Subsequently we investigated the dilute solution conformations of three oligomers differing in the degree of polymerization (DoP 9, 15, 24). Figure 6b shows distributions of the radius of gyration as well as the end-to-end distance for the all-atom model (blue) and the Martini 3 model (orange). We note that the agreement is excellent between both sets of simulations for both metrics. The end-to-end distance is typically related to polymer stiffness via the persistence length. Seeing that our end-to-end distance distributions agree well with the AA model, this indicates that our persistence length at least for dextran oligomers is very close to what is predicted by the AA model. In contrast, the scaling of the radius of gyration is related to the dilute solution thermodynamics via for example Flory theory<sup>59</sup>. Being able to match the AA distributions closely indicates that we capture the solution thermodynamics well, suggesting that our model has a good balance between self and water interactions. We continued to investigate further, if dextran is still soluble at higher weight fractions. Figure 6c shows a snapshot of a dextran (DoP 50) solution at 10w/w%, after 5 microseconds of simulation. Clearly no phase separation or strong aggregation is visible. In the Supporting Information (Figure S3), we further show the radial distribution functions (RDFs) for the polymer-polymer and polymer water interactions. Both indicate that the dextran is fully solubilized in water. To set these results into perspective, we also simulated dextran in chloroform, which is a known non-solvent<sup>58</sup>, for 5 microseconds. Already within a few hundred nanoseconds the polymers all aggregate into a periodic cluster (Figure 6d). Note that the concentration is the same in both simulations. The RDF for the polymer and polymer chloroform interactions (Figure S3) also shows the increased aggregation and depleted solvent interaction. We conclude our dextran model is fully water soluble and insoluble in chloroform, as expected for these concentrations and molecular weights.

Our final test-case involves an aqueous system of dextran and polyethylene oxide (PEO). Dextran is known to phase separate from PEO in a ternary mixture within water forming an aqueous two-phase system (ATPS) via liquid-liquid phase separation (LLPS).<sup>60,61</sup> ATPSs are important both in biomedical applications<sup>62</sup>, for instance, microfluidic separation and in biological research<sup>60</sup>, where they are used as compartmentalizing cytosol mimetic. Martini 3 has previously been shown to be able to capture LLPS of biomimetic compounds<sup>63</sup>. Thus, our model should also be capable of simulating those systems accurately. To this end, we generated a mixed PEO (DoP 180)-dextran (DoP 65) system using polyply<sup>64</sup>. We note that the used molecular

weights and concentrations have previously been reported to form an ATPS.<sup>60</sup> Figure 6e shows the system at the start and end of a 2  $\mu$ s simulation. Clearly the system has phase separated from the initial mixed state into a dextran rich phase and a phase enriched in PEO. Analysis of density profiles along the cartesian z-axis supports this conclusion (Figure S4). Overall, we showed that our dextran model matches solution conformations of atomistic simulations well, dissolves up to high weight fractions in water, does not dissolve in chloroform and forms an ATPS with PEO. All these observations are consistent with experimental data and demonstrate the validity of this carbohydrate polymer.

### 3.3 Solution properties of cellulose

In the previous two sections, we have shown that our Martini model produces carbohydrate molecules, which are water soluble and do not suffer from the same aggregation effects as seen for Martini2. However, not all carbohydrates are water soluble. For example, cellulose is famously known to be insoluble in water. In order to verify that our balance of interactions is reasonable and does not favor water soluble systems too much, we assessed solution properties of short cellulose analogs. In particular we simulated poly-( $\beta$ -1,4) glucoses with a DoP of 50 (Figure 7a). At these lengths it is known to form stable crystals in water that do not solubilize.<sup>65</sup> We started by building a system in a perfectly mixed state and random chain conformations and simulated it for 3  $\mu$ s. Figure 7b shows the starting configurations as well as the last frame of the simulation. Clearly the cellulose analog starts aggregating even forming small fiber like structures. We further investigated if pre-build crystals of cellulose are stable. To this end, a cellulose crystal (1 $\beta$ ) with was build using cellulose-builder<sup>66</sup>, solvated and run for 2 microseconds. Reassuringly, the fibril remains stable and insoluble in water, although we do observe the structure deviating from the original forward mapped crystal structure. Whereas cellulose is insoluble in water it does solve in ionic liquids (IL). As Martini 3 is also capable of simulating ILs<sup>67</sup> we proceeded to investigate what happens, if the solvent of the above systems is changed from water to [BMIM][Cl], which cellulose is known to dissolve in.<sup>68</sup> Figure 7d shows the system after 2  $\mu$ s of simulation. In contrast to the simulation with water we do not observe a fiber like structure being formed. In addition, RDFs (Figure S5) show that the cellulose remains fully solvated. This test-case demonstrates that our Martini 3 model is in principle capable of investigating cellulose solubility.



**Figure 7. Solution properties of cellulose. a) Mapping of cellulose at the Martini level. The colors correspond to the bead-types as found in Figure 1. b) Cellulose solvated in water. Figure on the left shows 100 glucose chains with a degree of polymerization of 50, placed randomly in water. Figure on the right shows the state of the system after 2  $\mu$ s of simulation. c) A cellulose I $\beta$  fibril (36 chains with a degree of polymerization of 50) after a 2  $\mu$ s simulation, solvated in water. d) Cellulose chains after 2  $\mu$ s simulation, being solvated in [BMIM][Cl] ionic liquid.**

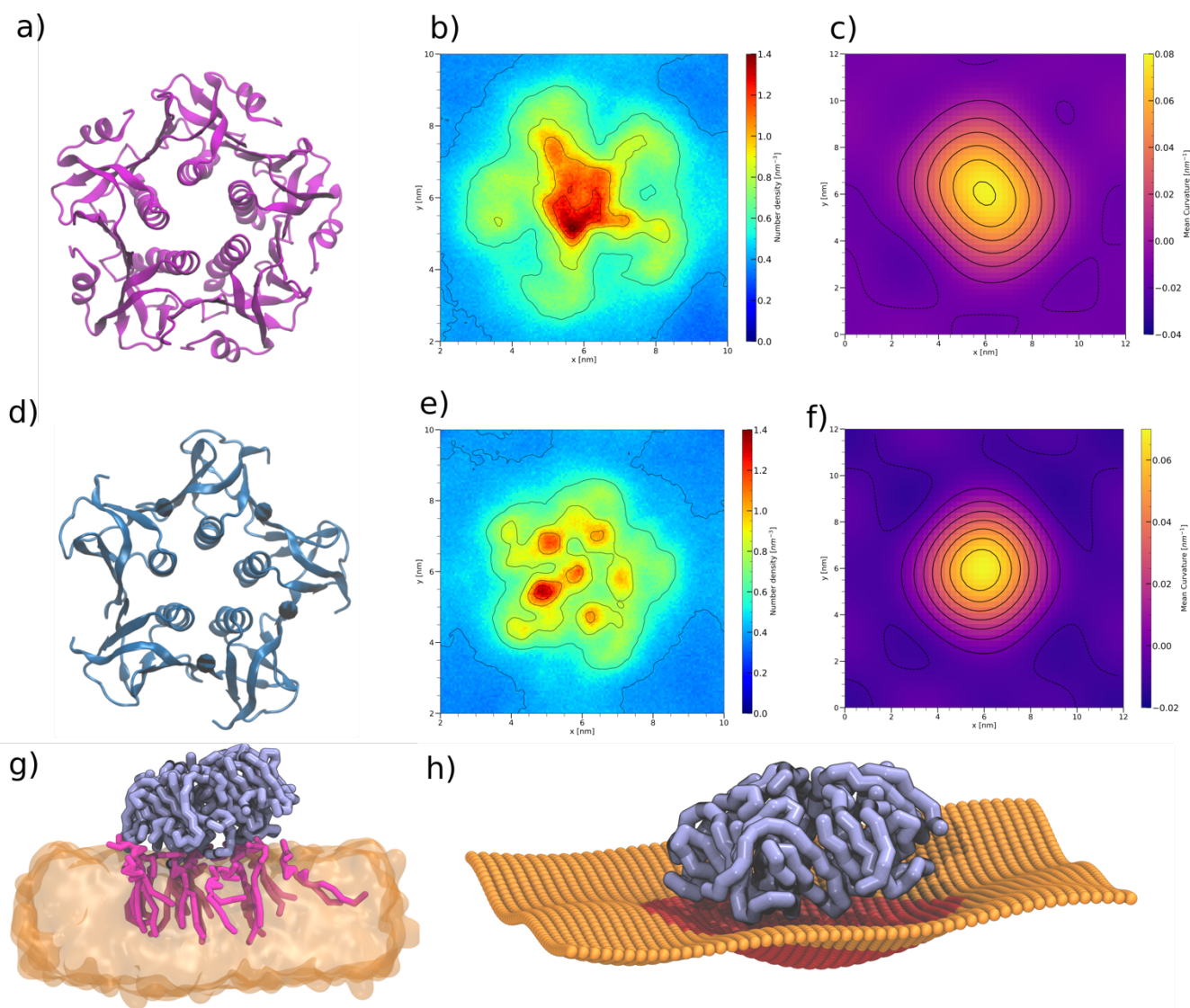
### 3.4 Binding of peripheral membrane proteins to glycolipids

Lipids and lipid protein interactions in complex membranes are one of the main application areas of Martini.<sup>69–71</sup> However, glycolipids, which consists of a carbohydrate headgroup and various lipid tails, suffered from the same problem of excess aggregation as other sugar molecules.<sup>18</sup> To show that our model is transferable to lipids and proteins, we study the interaction of peripheral membrane proteins with glycolipids. In particular, we focus on bacterial Shiga and Cholera toxins, secreted by *Shigella dysenteriae* and *Vibrio cholerae*, respectively. Both toxins are associated with several human diseases e.g., diarrhea.<sup>34</sup> In addition, these toxins are of special interest for their applications in biophysical experiments, targeted drug delivery and cancer therapy.<sup>72–74</sup> Both toxins are AB<sub>5</sub> proteins that are composed of an enzymatic active A and membrane binding homopentameric B subunit. The B subunit of Cholera toxin (CTxB; Figure 8a) and Shiga toxin (STxB; Figure 8d) initiate the toxin internalization through binding to their natural receptors on the targeted cell plasma membrane: the glycolipid globotriaosylceramide (Gb3) for STxB and ganglioside (GM) for CTxB. Parameters for Gb3 and GM3 (monosialodihexosyl-ganglioside) have been designed using the presented strategy for parametrizing the carbohydrate headgroup. Lipid tail parameters were the same as in the default Martini 3 force field with adjusted

linker mapping as explained in the Supporting Information, where the complete mapping for both lipids is shown (Figure S6).

First, we studied cluster formation of the GM3 lipids in a 10% POPC bilayer. Figure S7 shows the cluster-size distribution from the all-atom CHARMM36m simulation, an improved Martini 2 model proposed by Gu and coworkers<sup>18</sup>, as well as the present model. In the AA simulation GM3 mostly exists as monomers with dimers being much less likely. Higher order clustering is almost non-existent. The same trends are captured by the fixed Martini 2 model as well as our newly proposed Martini 3 model. We conclude that both Martini models perform equivalently well, but slightly overestimate aggregation in relation to the AA simulations. However, we consider this a satisfactory result for a generic coarse-grained model. We note that the lipid tail parameters in Martini 3 are currently subject to further optimization.

Whereas realistic lipid mixtures in complex membranes were already possible to capture with the optimized Martini 2 models, binding of peripheral membrane proteins to their natural glycolipid receptor remained problematic. The Martini 2 carbohydrate model shows no specific binding sites of Gb3 to STxB. In previous studies using Martini 2, Gb3 lipids were therefore tethered to the protein via a covalent bond based on the atomistic reference structure.<sup>75</sup>



**Figure 8. Binding of peripheral membrane proteins to glycolipids. a) Rendering of Cholera toxin subunit B (CTxB) protein structure (PDB 3CHB); b) 2D lipid density map of GM3 around CTxB; c) 2D curvature plot of membrane around CTxB; d) Rendering of Shiga toxin subunit B (STxB) protein structure (PDB 2C5C); e) 2D lipid density map of Gb3 around STxB with equivalent binding sites indicated by 1-3; f) 2D curvature plot of membrane around STxB; g) CTxB (violet) bound to membrane with Gb3 lipids shown in pink and the rest of the membrane shown in transparent; h) projected membrane surface with CTxB protein structure in the center.**

To study the process of STxB binding to Gb3 lipids with our new Martini 3 model, we simulated the system under the same conditions as the previous Martini 2 study.<sup>75</sup> To this end, a single STxB was placed above but not in contact with a POPC membrane, containing 10% mol fraction of Gb3. During the simulation, we observe that the protein stably binds to the membrane, not leaving during the 10  $\mu$ s long simulation. A control simulation without Gb3 lipids showed no binding of STxB to the membrane over the entire course of 10  $\mu$ s. Similar to Shiga toxin, Cholera toxin also binds to glycolipid receptors, but in this case GM lipids including

GM3. Hence, we have simulated the binding of CTxB to a POPC membrane with 10% GM3 lipids, analogously to the STxB case. Also, CTxB spontaneously binds to the membrane within a few hundreds of nanoseconds, and stays bound for the remaining 10  $\mu$ s. We further analyzed binding of the glycolipids to the proteins by computing a 2D lipid density map around the centered proteins. The density maps (Figure 8b, Figure 8e) show an enrichment of glycolipids under the proteins in specific spots. For CTxB we see the most dominant binding site to be in the center of the protein and weaker binding sites in the peripheries. To our

knowledge binding sites of GM3 have not been resolved for CTxB. In contrast, for Gb3 binding to STxB three distinct binding sites per monomer have been resolved experimentally by X-ray diffraction.<sup>76</sup> Two binding sites are found on the peripheries close to each other and a third binding site is found at the bottom of the alpha helix. In order to investigate to what extent our lipids bind in similar spots, we have computed the site-specific RDFs (see Figure S8) between the sugar parts of the glycolipids and those residues experimentally identified in binding. We see an increased probability to find a carbohydrate around each binding site indicating that binding locations appear reasonable.

Finally, we have assessed the membrane curvature induced by binding of the two proteins. Both are known to induce curvature, which is an essential step in their endocytosis.<sup>77–79</sup> Figure 8c and Figure 8f show the 2D mean curvature under each protein clearly demonstrating that our model can capture this behavior qualitatively. Previously the curvature for STxB bound to Gb3 has been computed from atomistic simulations ( $0.034 \pm 0.004 \text{ nm}^{-1}$ )<sup>78</sup>. The value obtained from the CG simulations is in the same ballpark ( $0.0260 \pm 0.0001 \text{ nm}^{-1}$ ). The level of membrane curvature induced by binding of CTxB is further illustrated in Figure 8h.

#### 4.0 – Discussion & Conclusion

In the present paper we developed a consistent strategy to parametrize arbitrary carbohydrates with the Martini 3 force field. In particular, we presented a canonical mapping scheme that decomposes arbitrarily large carbohydrates into a limited number of fragments. Bead types for these fragments have been assigned by matching atomistic volumes and free energies of transfer from water to octanol. The best bead assignment yields a mean-absolute error of about 1.3 kJ/mol compared to the experimental reference partitioning data. In addition to the bead type assignment for fragments, guidelines for assigning bonds, angles, and dihedrals have been presented. These guidelines allow for a more accurate description of carbohydrate conformations than in the Martini 2 force field, and can easily be expanded to more complex carbohydrates. We showed that models obtained with this parametrization strategy are able to reproduce osmotic pressures of carbohydrate water solutions to very good accuracy. Furthermore, we demonstrated that the model differentiates correctly the solubility of the poly-glucoses dextran (water soluble) and cellulose (water insoluble). Given that the difference between both models is only a single bead-type and different bonded interactions, it speaks for the accuracy of our model being able to capture their differences. In the final test-case we illustrate that the model is applicable to glycolipids by showing that the clustering of GM3 is in good agreement with all-atom reference simulations. As last validation we analyzed the binding of peripheral membrane

proteins Shiga and Cholera toxin to two glycolipid receptors. Here we found that both proteins bind to the glycolipids and induce membrane curvature as expected. Taken together, these test cases demonstrate the validity and transferability of our approach.

However, some limitations apply as well. The osmotic pressure for certain monosaccharides indicated a too high self-interaction in the high concentration regime (conc. > 1.5 molal). Therefore, simulations concerning highly concentrated solutions need to be verified carefully. Furthermore, we note that while inclusion of the TC4 virtual site greatly helps in interactions with aromatic moieties, they remain lower than observed in all-atom models. This is especially true for conformations where a stacked interaction is enforced. For example, protein binding could be influenced by this effect. Whereas part of it is an intrinsic limitation of a CG model with less degrees of freedom, further improvement can be obtained by future improvements in the protein model or even carefully revision of the Martini 3 interaction matrix. These are ongoing processes. We conclude that the rules for making carbohydrate models within Martini 3 lead to CG models that greatly improve in accuracy over Martini 2, and, at least in some aspects, are comparable to standard atomistic force fields employed in the field.

#### 5.0 – Methods

##### 5.1 Experimental measurements of the partition coefficients

Measurements were performed following a similar methodology as outlined in Virtanen *et al*<sup>80</sup>. Measurements were done with an UPLC-DAD-HESI-Orbitrap-MS instrument. The column in the UPLC was an Aquity BEH Phenyl ( $100 \times 2.1 \text{ mm i.d.}, 1.7 \mu\text{m}$ ) and the mobile phase consisted of acetonitrile (A) and 0.1% aqueous formic acid (B). The elution gradient was carried out with a constant flow rate of 0.65 mL/min as follows: 0–0.1 min: 3% A; 0.1–3.0 min: 3.0–45.0% A (linear gradient); 3.0–3.1 min: 45.0–90.0% A (linear gradient); 3.1–4.0 min: 90% A; 4.0–4.1 min: 90.0–3.0% A (linear gradient); 4.1–4.2 min: 3.0% A. The ionization mode (negative/positive) of the mass spectrometer that was used for each compound depended on their ionization efficiency in either negative or positive mode; the one where each compound ionized more effectively in the test samples was then used for quantitative measurements. All measurements were done in triplicate and quantitation for each compound was done from extracted ion chromatograms (EICs) from full scan MS analysis with a specific *m/z*-range for each compound. Integrated EIC areas were converted to concentrations before partition coefficient calculations with a calibration series done with a dilution series of each compound. Both the calibration series samples and the actual  $K_{ow}$  samples

mass responses (integrated EIC areas) were normalized with an external standards mass response so that the possible variation in the mass spectrometers performance during the measurements and on different days could be taken into account.

## 5.2 All-atom MD simulations

All atomistic simulations were performed with GROMACS (2018.8 or 2021.5)<sup>81</sup>.

### Simulations of mono- and disaccharides

Carbohydrates were simulated using the GLYCAM06<sup>33</sup> force field, the CHARMM36m<sup>43,82</sup> force field, and the GROMOS54a7<sup>35,83</sup> force field as outlined in Table S3. For each simulation a single carbohydrate was solvated in a box of water with size 2.4 x 2.4 x 2.4 nm<sup>3</sup> and simulated after short equilibration in the isobaric-isochoric ensemble at 1bar. Temperatures were fixed at 310K, 303.15K or 298.15K. Each simulation was run for at least 200 ns using the default leap-frog integrator. All bonds were restrained with the LINCS algorithm.<sup>84</sup> The Glycam06 and CHARMM36 simulations used TIP3P<sup>85</sup> as water model, whereas the GROMOS ones used the SPC model<sup>86</sup>. For all force fields the GROMACS specific recommended run settings were used. Itp files for the carbohydrates were obtained from the Glycam-Web and converted to GROMACS with acpype<sup>87,88</sup>, or the CharmmGUI<sup>43</sup>, or the automated topology builder (ATB)<sup>35,83</sup>.

### Simulations of dextran oligomers

Single chains of dextran oligomers in water were simulated using the CHARMM36m force field<sup>54</sup>. Parameters and coordinates for three different degrees of polymerization (9, 14, 24) were obtained from the CHARMM-GUI<sup>43,89</sup>. After equilibration each simulation was run under constant temperature at 298.15K using the v-rescale<sup>90</sup> temperature coupling ( $\tau=1$ ps) with a coupling group for solvent and polymer. Pressure was kept constant at 1 bar using the Parrinello-Rahman pressure coupling algorithm ( $\tau=5$ ps,  $\beta=4.5 \times 10^{-5}$  bar<sup>-1</sup>). The simulations for the three oligomers were run for 6 $\mu$ s, 3 $\mu$ s, and 3 $\mu$ s respectively. Radius of gyration and end-to-end distance were obtained from the simulation using the 'gmxdist' tool. Distributions were subsequently computed after discarding an equilibration time.

### Simulations of lipid bilayers

All atomistic resolution lipid bilayer simulations used the CHARMM36m force field<sup>54</sup>. The bilayers consisted of POPC as major component and 10% glycolipids either GM3 or Gb3. Parameters and coordinates for were obtained from the CHARMM-GUI<sup>43,89</sup>. After equilibration each simulation was run under constant temperature at 310K using the Nose-Hoover temperature coupling ( $\tau=1$ ps) with a coupling group for solvent and membrane. Pressure was

kept constant at 1 bar using the semiisotropic Parrinello-Rahman pressure coupling algorithm in xy and z direction ( $\tau=5$ ps,  $\beta=4.5 \times 10^{-5}$  bar<sup>-1</sup>). The simulations were run for 2 $\mu$ s. Clustering of GM3 lipids was analyzed after mapping the trajectories to CG resolution with fast\_forward<sup>32</sup>. Subsequently using the 'gmxcsize' tool lipids were counted as being in the same cluster, if the distance between the linker beads was less than 1.4 nm.

### SASA calculations

The solvent accessible surface area was computed using the double cubic lattice method by Eisenhaber *et al.* as implemented in the GROMACS software suite (i.e. gmcsasa).<sup>91</sup> Instead of using the default VdW-radii for this calculation, the more recent VdW-radii proposed by Rowland and Taylor were used for the atomistic simulations.<sup>92</sup> For the Martini simulations the VdW-radii were taken to be the minimum of the LJ self-interaction, which leads to three radii for the regular (0.264 nm), small (0.230 nm), tiny (0.191 nm) beads. The probe size for both atomistic and CG simulations was 0.191 nm and the SASA was averaged over at least 200 ns both for atomistic simulations and CG Martini simulations.

## 5.3 Coarse-grained MD simulations

CG simulations were performed with the Martini 3 force field<sup>5</sup>, or Martini 2 force field<sup>4</sup>. Each Martini 3 simulation followed the standard simulation settings as outlined in the main publication<sup>5</sup> and the Martini 2 simulations followed the parameters as outlined by de Jong *et al.*<sup>93</sup>, unless specified otherwise. The velocity rescaling thermostat<sup>90</sup> ( $\tau=1$ ps) and Parrinello-Rahman barostat<sup>94</sup> ( $\tau=12$ ps,  $\beta=4.5 \times 10^{-5}$  bar<sup>-1</sup>) were used to maintain temperature and pressure in production simulations. GROMACS version 2021.5 was used for all simulations unless otherwise stated. Bonds within monosaccharides were constrained with the LINCS algorithm.<sup>84</sup>

### Free energies of transfer

All simulations pertaining free energies of transfer were carried out with the GROMACS software (version 2021.5)<sup>81</sup>, using the stochastic dynamics integrator<sup>95</sup> (with inverse friction constant 1.0 ps<sup>-1</sup>) and a time step of 20 fs. Free energies of transfer of the carbohydrates were calculated as differences between free energies of solvation in water and octanol. Solvation free energies were computed by alchemical free energy transformations as implemented in the GROMACS package. All systems for the solvation free energy in water consisted of a single carbohydrate solute molecule and 1023 Martini water beads. The systems for the octanol solvation free energies consisted of 920 octanol molecules and 80 water beads representing a saturated octanol composition.

The calculations used in total 19 non-equally spaced windows, switching only the LJ interactions as the Martini

molecules considered have no partial charges. Soft-core LJ potentials were applied following the recommended values<sup>96</sup>. Each window was run under NpT conditions for 12 ns at 1 bar pressure maintained ( $\tau=4$  ps). Temperature was maintained at 298.15K. The derivative of the potential energy was recorded every 10 steps. All the free energies of the transformation were estimated using the Bennetts-Acceptance-Ratio (BAR) method as implemented in the 'gmx bar' tool. The error reported with the calculations is the statistical error estimate. The intramolecular interactions were not switched off for both sets of simulations.

#### Osmotic pressure calculations

The osmotic pressure was computed from simulations adopting the protocol originally proposed by Luo and Roux.<sup>53</sup> A rectangular box was created in which the solute molecules were confined in z-direction by a flat-bottomed potential to the center of the box. At a distance of 2.52078 nm from the center of the box, a harmonic force with a force constant 1000 kJ/nm<sup>2</sup> was applied to the solute molecules. The box dimensions were taken to be 10.08312 nm in z dimensions and 5.04156 nm in x and y. Previous to each run, a random configuration of solute and solvent molecules was created with polyply<sup>64</sup> placing solute molecules only in the center of the box and the solvent in the entire box. After energy minimization, this setup was subjected to a 10 ns equilibration using Berendsen barostat<sup>97</sup>. Production simulations were run for 500 ns as previously used for atomistic simulation<sup>52</sup> at a pressure of 1 bar. The temperature matched the temperatures reported with the experimental data-sets. Following Sauter and Grafmuller<sup>51,55</sup>, the pressure was coupled only in z-dimensions. The osmotic pressure was computed from the trajectory by recalculating the total force exerted by the solute particles onto the flat-bottomed potential averaged over the two potentials. Subsequently that force is divided by the xy area of the box. The ensemble average as well as an error were computed from the time-series of the osmotic pressure.

#### Simulation of dextran systems

Initial structures were built using polyply<sup>64</sup> and subsequently subjected to an energy minimization. For the scaling simulations of the oligomers first a short relaxation using the Berendsen barostat was run. Afterwards they were sampled for 3  $\mu$ s using the v-rescale barostat ( $\tau=6$  ps,  $\beta=4.5 \times 10^{-5}$  bar<sup>-1</sup>).<sup>98</sup> Mixing of PEO and dextran was studied in the same fashion; however, simulations were run for 5  $\mu$ s. RDFs were computed using the 'gmx rdf' tool. PEO parameters were taken from the polyply<sup>64</sup> library (v1.3.0). The polymer-polymer RDF was computed as an average of RDFs for each polymer separately with the other polymers as to remove the correlation induced simply by the fact that neighboring repeat units are covalently bound to each other.

#### Simulation of cellulose systems

Starting configurations for the mixed state simulations in water and [BMIM][Cl] ionic liquid were generated by placing glucose chains with a DoP of 50 randomly to a large simulation volume with the 'gmx insert-molecules' tool, and then solvating with the appropriate solvent. The Martini 3 IL parameters as published earlier were used.<sup>5,67</sup> The starting structure of the cellulose fibril was created with the Cellulose Builder<sup>66</sup> after which it was solvated in the same way as the previous systems. All systems were equilibrated for 50 ns at a temperature of 310K with the system pressure controlled using the Berendsen barostat. Production simulations were run for 2  $\mu$ s in the same temperature, using the Parrinello-Rahman barostat with isotropic coupling ( $\tau = 12$  ps,  $\beta = 4.5 \times 10^{-5}$  bar<sup>-1</sup>).

#### Simulation of peripheral membrane protein binding

Initial structures of the lipid bilayers were built using TS2CG<sup>75</sup> or obtained from the atomistic simulation by mapping the bilayer and re-solvating it. All simulations were subjected to an energy minimization and equilibration. Subsequently all simulations were run under semiisotropic pressure coupling at 1bar at 310K temperature for 10 $\mu$ s. Protein itp files were obtained using the martinize2 code as available on GitHub. The clustering of the GM3 lipids was analyzed following the same protocol as used previously.<sup>18</sup> Membrane curvature was analyzed as described previously.<sup>78</sup> 2D density maps were computed with 'gmx densmap'. Both properties were computed as time-average over the last 7 $\mu$ s. Specific binding sites were analyzed using gmx rdf.

## 6.0 – Acknowledgments

We would like to thank the Center for Information Technology of the University of Groningen for their support and for providing access to the Peregrine high performance computing cluster. S.J.M. acknowledges funding from the ERC via an Advanced grant "COMP-MICR-CROW-MEM". W.P. acknowledges funding from the Novo Nordisk Foundation (grant No. NNF18SA0035142) and INTERACTIONS, Marie Skłodowska-Curie grant agreement No 847523. P.C.T.S acknowledges the support of the French National Center for Scientific Research (CNRS) and the research collaboration with PharmCADD. M.S.P.S and E.E. J. acknowledge funding by the Wellcome Trust (grant No. 203815/Z/16/Z). We would like to thank Mateusz Sikora and Philipp Schmalhorst for the early tests with the preliminary models of Martini 3 carbohydrates.

## 7.0 – Associated Content

Mono- and Disaccharide Parameters developed in this paper were deposited in a GitHub repository (<https://github.com/marrink-lab/martini-forcefields>)

Dextran and Cellulose parameters are available from the polyply force field library starting from polyply version >1.3.2. See the following link for more details [https://github.com/marrink-lab/polyply\\_1.0](https://github.com/marrink-lab/polyply_1.0).

## 8.0 – References

- (1) Maughan, R. Carbohydrate Metabolism. *Surgery (Oxford)* **2009**, *27* (1), 6–10. <https://doi.org/10.1016/j.mpsur.2008.12.002>.
- (2) Rini, J. M.; Leffler, H. Carbohydrate Recognition and Signaling. In *Handbook of Cell Signaling*; Elsevier, 2010; pp 85–91. <https://doi.org/10.1016/B978-0-12-374145-5.00013-9>.
- (3) Ragauskas, A. J.; Williams, C. K.; Davison, B. H.; Britovsek, G.; Cairney, J.; Eckert, C. A.; Frederick, W. J.; Hallett, J. P.; Leak, D. J.; Liotta, C. L.; Mielenz, J. R.; Murphy, R.; Templer, R.; Tschaplinski, T. The Path Forward for Biofuels and Biomaterials. *Science (1979)* **2006**, *311* (5760), 484–489. <https://doi.org/10.1126/science.1114736>.
- (4) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *Journal of Physical Chemistry B* **2007**, *111* (27), 7812–7824. <https://doi.org/10.1021/jp071097f>.
- (5) Souza, P. C. T.; Alessandri, R.; Barnoud, J.; Thallmair, S.; Faustino, I.; Grünwald, F.; Patmanidis, I.; Abdizadeh, H.; Bruininks, B. M. H.; Wassenaar, T. A.; Kroon, P. C.; Melcr, J.; Nieto, V.; Corradi, V.; Khan, H. M.; Domański, J.; Javanainen, M.; Martinez-Seara, H.; Reuter, N.; Best, R. B.; Vattulainen, I.; Monticelli, L.; Periolo, X.; Tieleman, D. P.; de Vries, A. H.; Marrink, S. J. Martini 3: A General Purpose Force Field for Coarse-Grained Molecular Dynamics. *Nature Methods* **2021**, *18* (4), 382–388. <https://doi.org/10.1038/s41592-021-01098-3>.
- (6) Marrink, S. J.; Tieleman, D. P. Perspective on the Martini Model. *Chemical Society Reviews* **2013**, *42* (16), 6801. <https://doi.org/10.1039/c3cs60093a>.
- (7) Marrink, S. J.; Monticelli, L.; Melo, M. N.; Alessandri, R.; Tieleman, D. P.; Souza, P. C. T. Two Decades of Martini: Better Beads, Broader Scope. *WIREs Computational Molecular Science* **2022**. <https://doi.org/10.1002/wcms.1620>.
- (8) Alessandri, R.; Grünwald, F.; Marrink, S. J. The Martini Model in Materials Science. *Advanced Materials* **2021**, *33* (24), 2008635. <https://doi.org/10.1002/adma.202008635>.
- (9) Frederix, P. W. J. M.; Marrink, S. J. Molecular Simulations of Self-Assembling Bio-Inspired Supramolecular Systems and Their Connection to Experiments. **2018**, *47* (10). <https://doi.org/10.1039/c8cs00040a>.
- (10) Alessandri, R.; Barnoud, J.; Gertsen, A. S.; Patmanidis, I.; de Vries, A. H.; Souza, P. C. T.; Marrink, S. J. Martini 3 Coarse-Grained Force Field: Small Molecules. *Advanced Theory and Simulations* **2022**, *5* (1), 2100391. <https://doi.org/10.1002/adts.202100391>.
- (11) López, C. A.; Rzepiela, A. J.; de Vries, A. H.; Dijkhuizen, L.; Hünenberger, P. H.; Marrink, S. J. Martini Coarse-Grained Force Field: Extension to Carbohydrates. *Journal of Chemical Theory and Computation* **2009**, *5* (12), 3195–3210. <https://doi.org/10.1021/ct900313w>.
- (12) Shivgan, A. T.; Marzinek, J. K.; Huber, R. G.; Krah, A.; Henchman, R. H.; Matsudaira, P.; Verma, C. S.; Bond, P. J. Extending the Martini Coarse-Grained Force Field to N-Glycans. *Journal of Chemical Information and Modeling* **2020**, *60* (8), 3864–3883. <https://doi.org/10.1021/acs.jcim.0c00495>.
- (13) Schmalhorst, P. S.; Deluweit, F.; Scherrers, R.; Heisenberg, C. P.; Sikora, M. Overcoming the Limitations of the MARTINI Force Field in Simulations of Polysaccharides. *Journal of Chemical Theory and Computation* **2017**, *13* (10), 5039–5053. <https://doi.org/10.1021/acs.jctc.7b00374>.
- (14) Hsu, P.-C.; Bruininks, B. M. H.; Jefferies, D.; de Souza, P.; Lee, J.; Patel, D. S.; Marrink, S. J.; Qi, Y.; Khalid, S.; Im, W. CHARMM-GUI Martini Maker for Modeling and Simulation of Complex Bacterial Membranes with Lipopolysaccharides. *J. Comput. Chem.* **2017**, *38* (27), 2354–2363.
- (15) Vaiwala, R.; Sharma, P.; Puranik, M.; Ayappa, K. G. Developing a Coarse-Grained Model for Bacterial Cell Walls: Evaluating Mechanical Properties and Free Energy Barriers. *Journal of Chemical Theory and Computation* **2020**, *16* (8), 5369–5384. <https://doi.org/10.1021/acs.jctc.0c00539>.
- (16) López, C. A.; de Vries, A. H.; Marrink, S. J. Computational Microscopy of Cyclodextrin Mediated Cholesterol Extraction from Lipid Model Membranes. *Scientific Reports* **2013**, *3* (1), 2071. <https://doi.org/10.1038/srep02071>.
- (17) Xu, H.; Matysiak, S. Effect of PH on Chitosan Hydrogel Polymer Network Structure. *Chem. Commun.* **2017**, *53* (53), 7373–7376. <https://doi.org/10.1039/C7CC01826F>.
- (18) Gu, R.-X.; Ingólfsson, H. I.; de Vries, A. H.; Marrink, S. J.; Tieleman, D. P. Ganglioside-Lipid and Ganglioside-Protein Interactions Revealed by Coarse-Grained and Atomistic Molecular Dynamics Simulations. *The Journal of Physical Chemistry B* **2017**, *121* (15), 3262–3275. <https://doi.org/10.1021/acs.jpcc.6b07142>.
- (19) Thallmair, S.; Vainikka, P. A.; Marrink, S. J. Lipid Fingerprints and Cofactor Dynamics of Light-Harvesting Complex II in Different Membranes.



- Biophysical Journal* **2019**, *116* (8), 1446–1455.  
<https://doi.org/10.1016/j.bpj.2019.03.009>.
- (20) López, C. A.; Bellesia, G.; Redondo, A.; Langan, P.; Chundawat, S. P. S.; Dale, B. E.; Marrink, S. J.; Gnanakaran, S. MARTINI Coarse-Grained Model for Crystalline Cellulose Microfibers. *The Journal of Physical Chemistry B* **2015**, *119* (2), 465–473.  
<https://doi.org/10.1021/jp5105938>.
- (21) López, C. A.; Sovova, Z.; van Eerden, F. J.; de Vries, A. H.; Marrink, S. J. Martini Force Field Parameters for Glycolipids. *Journal of Chemical Theory and Computation* **2013**, *9* (3), 1694–1708.  
<https://doi.org/10.1021/ct3009655>.
- (22) de Jong, D. H.; Lopez, C. A.; Marrink, S. J. Molecular View on Protein Sorting into Liquid-Ordered Membrane Domains Mediated by Gangliosides and Lipid Anchors. *Faraday Discuss.* **2013**, *161*, 347–363.  
<https://doi.org/10.1039/C2FD20086D>.
- (23) Hsu, P.-C.; Jefferies, D.; Khalid, S. Molecular Dynamics Simulations Predict the Pathways via Which Pristine Fullerenes Penetrate Bacterial Membranes. *The Journal of Physical Chemistry B* **2016**, *120* (43), 11170–11179.  
<https://doi.org/10.1021/acs.jpcc.6b06615>.
- (24) Boags, A.; Hsu, P.-C.; Samsudin, F.; Bond, P. J.; Khalid, S. Progress in Molecular Dynamics Simulations of Gram-Negative Bacterial Cell Envelopes. *The Journal of Physical Chemistry Letters* **2017**, *8* (11), 2513–2518. <https://doi.org/10.1021/acs.jpcl.7b00473>.
- (25) Shearer, J.; Khalid, S. Communication between the Leaflets of Asymmetric Membranes Revealed from Coarse-Grain Molecular Dynamics Simulations. *Scientific Reports* **2018**, *8* (1), 1805.  
<https://doi.org/10.1038/s41598-018-20227-1>.
- (26) Moiset, G.; López, C. A.; Bartelds, R.; Syga, L.; Rijpkema, E.; Cukkemane, A.; Baldus, M.; Poolman, B.; Marrink, S. J. Disaccharides Impact the Lateral Organization of Lipid Membranes. *J Am Chem Soc* **2014**, *136* (46), 16167–16175.  
<https://doi.org/10.1021/ja505476c>.
- (27) Alessandri, R.; Souza, P. C. T.; Thallmair, S.; Melo, M. N.; de Vries, A. H.; Marrink, S. J. Pitfalls of the Martini Model. *Journal of Chemical Theory and Computation* **2019**, *15* (10), 5448–5460.  
<https://doi.org/10.1021/acs.jctc.9b00473>.
- (28) Souza, P. C. T.; Thallmair, S.; Marrink, S. J.; Mera-Adasme, R. An Allosteric Pathway in Copper, Zinc Superoxide Dismutase Unravels the Molecular Mechanism of the G93A Amyotrophic Lateral Sclerosis-Linked Mutation. *The Journal of Physical Chemistry Letters* **2019**, *10* (24), 7740–7744.  
<https://doi.org/10.1021/acs.jpcl.9b02868>.
- (29) Vainikka, P.; Thallmair, S.; Souza, P. C. T.; Marrink, S. J. Martini 3 Coarse-Grained Model for Type III Deep Eutectic Solvents: Thermodynamic, Structural, and Extraction Properties. *ACS Sustainable Chemistry & Engineering* **2021**, *9* (51), 17338–17350.  
<https://doi.org/10.1021/acssuschemeng.1c06521>.
- (30) Borges-Araújo, L.; Souza, P. C. T.; Fernandes, F.; Melo, M. N. Improved Parameterization of Phosphatidylinositol Lipid Headgroups for the Martini 3 Coarse-Grain Force Field. *Journal of Chemical Theory and Computation* **2022**, *18* (1), 357–373. <https://doi.org/10.1021/acs.jctc.1c00615>.
- (31) Souza, P. C. T.; Thallmair, S.; Conflitti, P.; Ramírez-Palacios, C.; Alessandri, R.; Raniolo, S.; Limongelli, V.; Marrink, S. J. Protein–Ligand Binding with the Coarse-Grained {M}artini Model. *Nat. Commun.* **2020**, *11*, 3714.
- (32) Grünewald, F.; Marrink, S. J. Fast\_forward. 2022.
- (33) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. {GLYCAM}06: A Generalizable Biomolecular Force Field. *Carbohydrates. Journal of Computational Chemistry* **2007**, *29* (4), 622–655.  
<https://doi.org/10.1002/jcc.20820>.
- (34) Lay, W. K.; Miller, M. S.; Elcock, A. H. Optimizing Solute–Solute Interactions in the GLYCAM06 and CHARMM36 Carbohydrate Force Fields Using Osmotic Pressure Measurements. *Journal of Chemical Theory and Computation* **2016**, *12* (4), 1401–1407. <https://doi.org/10.1021/acs.jctc.5b01136>.
- (35) Malde, A. K.; Zuo, L.; Breeze, M.; Stroet, M.; Poger, D.; Nair, P. C.; Oostenbrink, C.; Mark, A. E. An Automated Force Field Topology Builder (ATB) and Repository: Version 1.0. *Journal of Chemical Theory and Computation* **2011**, *7* (12), 4026–4037.  
<https://doi.org/10.1021/ct200196m>.
- (36) Connolly, M. L. Analytical Molecular Surface Calculation. *Journal of Applied Crystallography* **1983**, *16* (5), 548–558.  
<https://doi.org/10.1107/s0021889883010985>.
- (37) Bas, D.; Dorison-Duval, D.; Moreau, S.; Bruneau, P.; Chipot, C. Rational Determination of Transfer Free Energies of Small Drugs across the Water–Oil Interface. *Journal of Medicinal Chemistry* **2002**, *45* (1), 151–159.  
<https://doi.org/10.1021/jm010289a>.
- (38) Mazzobre, M. F.; Román, M. v.; Mourelle, A. F.; Corti, H. R. Octanol–Water Partition Coefficient of Glucose, Sucrose, and Trehalose. *Carbohydrate Research* **2005**, *340* (6), 1207–1211.  
<https://doi.org/10.1016/j.carres.2004.12.038>.
- (39) Lennernäs, H.; Nylander, S.; Ungell, A.-L. Jejunal Permeability A Comparison between Ussing Chambers and in Humana. *Pharmaceutical Research*. 1997, pp 667–671.
- (40) Fernández-Alonso, M. del C.; Cañada, F. J.; Jiménez-Barbero, J.; Cuevas, G. Molecular Recognition of Saccharides by Proteins. Insights on the Origin of the

- Carbohydrate–Aromatic Interactions. *J Am Chem Soc* **2005**, *127* (20), 7379–7386. <https://doi.org/10.1021/ja051020+>.
- (41) Houser, J.; Kozmon, S.; Mishra, D.; Hammerová, Z.; Wimmerová, M.; Koča, J. The CH– $\pi$  Interaction in Protein–Carbohydrate Binding: Bioinformatics and In Vitro Quantification. *Chemistry – A European Journal* **2020**, *26* (47), 10769–10780. <https://doi.org/10.1002/chem.202000593>.
- (42) Asensio, J. L.; Ardá, A.; Cañada, F. J.; Jiménez-Barbero, J. Carbohydrate–Aromatic Interactions. *Accounts of Chemical Research* **2013**, *46* (4), 946–954. <https://doi.org/10.1021/ar300024d>.
- (43) Jo, S.; Song, K. C.; Desaire, H.; MacKerell, A. D.; Im, W. Glycan Reader: Automated Sugar Identification and Simulation Preparation for Carbohydrates and Glycoproteins. *Journal of Computational Chemistry* **2011**, *32* (14), 3135–3141. <https://doi.org/10.1002/jcc.21886>.
- (44) Wassenaar, T. A.; Pluhackova, K.; Böckmann, R. A.; Marrink, S. J.; Tieleman, D. P. Going Backward: A Flexible Geometric Approach to Reverse Transformation from Coarse Grained to Atomistic Models. *Journal of Chemical Theory and Computation* **2014**, *10* (2), 676–690. <https://doi.org/10.1021/ct400617g>.
- (45) Lo, C. A.; Rzepiela, A. J.; Vries, A. H. de; Dijkhuizen, L.; Hu, P. H.; Marrink, S. J.; Lopez, C. A.; de Vries, A. H.; Hunenberger, P. H.; López, C. A.; Hünenberger, P. H.; Rzepiela, A. J.; de Vries, A. H.; Dijkhuizen, L.; Hunenberger, P. H.; Marrink, S. J.; Lo, C. A.; Rzepiela, A. J.; Vries, A. H. de; Dijkhuizen, L.; Hu, P. H.; Marrink, S. J. Martini Coarse-Grained Force Field: Extension to Carbohydrates. *Journal of Chemical Theory and Computation* **2009**, *5* (12), 3195–3210. <https://doi.org/10.1021/ct900313w>.
- (46) Bulacu, M.; Goga, N.; Zhao, W.; Rossi, G.; Monticelli, L.; Periolo, X.; Tieleman, D. P.; Marrink, S. J. Improved Angle Potentials for Coarse-Grained Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2013**, *9* (8), 3282–3292.
- (47) Miyajima, K.; Kyoto, J.; Sawada, M.; Nakagaki, M. STUDIES ON AQUEOUS SOLUTIONS OF SACCHARIDES - I. ACTIVITY COEFFICIENTS OF MONOSACCHARIDES IN AQUEOUS SOLUTIONS AT 25 Degree C. *Bulletin of the Chemical Society of Japan*. 1983, pp 1620–1623. <https://doi.org/10.1246/bcsj.56.1620>.
- (48) Barone, G.; Cacace, P.; Castronuovo, G.; Elia, V.; Iappelli, F. Excess Thermodynamic Properties of Aqueous Solutions of L-Fucose and l-Rhamnose at 25°. *Carbohydrate Research* **1981**, *93* (1), 11–18. [https://doi.org/10.1016/S0008-6215\(00\)80748-X](https://doi.org/10.1016/S0008-6215(00)80748-X).
- (49) Jamehbozorg, B.; Sadeghi, R. Evaluation of the Effect of Carbohydrates as Renewable, Non-Charged and Non-Toxic Soluting-out Agents on the Ionic-Liquid-Based ABS Implementation. *Journal of Molecular Liquids* **2018**, *255*, 476–491. <https://doi.org/10.1016/j.molliq.2018.01.093>.
- (50) Ebrahimi, N.; Sadeghi, R. Osmotic Properties of Carbohydrate Aqueous Solutions. *Fluid Phase Equilibria* **2016**, *417*, 171–180. <https://doi.org/10.1016/j.fluid.2016.02.030>.
- (51) Sauter, J.; Grafmüller, A. Predicting the Chemical Potential and Osmotic Pressure of Polysaccharide Solutions by Molecular Simulations. *Journal of Chemical Theory and Computation* **2016**, *12* (9), 4375–4384. <https://doi.org/10.1021/acs.jctc.6b00295>.
- (52) Piana, S.; Robustelli, P.; Tan, D.; Chen, S.; Shaw, D. E. Development of a Force Field for the Simulation of Single-Chain Proteins and Protein–Protein Complexes. *Journal of Chemical Theory and Computation* **2020**, *16* (4), 2494–2507. <https://doi.org/10.1021/acs.jctc.9b00251>.
- (53) Luo, Y.; Roux, B. Simulation of Osmotic Pressure in Concentrated Aqueous Salt Solutions. *The Journal of Physical Chemistry Letters* **2010**, *1* (1), 183–189. <https://doi.org/10.1021/jz900079w>.
- (54) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nature Methods* **2017**, *14* (1), 71–73. <https://doi.org/10.1038/nmeth.4067>.
- (55) Sauter, J.; Grafmüller, A. Solution Properties of Hemicellulose Polysaccharides with Four Common Carbohydrate Force Fields. *Journal of Chemical Theory and Computation* **2015**, *11* (4), 1765–1774. <https://doi.org/10.1021/ct500924f>.
- (56) Gekko, K.; Noguchi, H. Physicochemical Studies of Oligodextran. I. Molecular Weight Dependence of Intrinsic Viscosity, Partial Specific Compressibility and Hydrated Water. *Biopolymers* **1971**, *10* (9), 1513–1524. <https://doi.org/10.1002/bip.360100907>.
- (57) Gascioli, V.; Choplin, L.; Paul, F.; Monsan, P. Viscous Properties and Molecular Characterization of Enzymatically Size-Controlled Oligodextrans in Aqueous Solutions. *Journal of Biotechnology* **1991**, *19* (2–3), 193–202. [https://doi.org/10.1016/0168-1656\(91\)90058-4](https://doi.org/10.1016/0168-1656(91)90058-4).
- (58) Meyer, T. A.; Gormley, A. J. Dextran Solubility; Personal Communication. 2022.
- (59) Colby, R. H.; Rubinstein, M. *Polymer Physics*; Oxford University Press: New-York, 2003.
- (60) Andes-Koback, M.; Keating, C. D. Complete Budding and Asymmetric Division of Primitive Model Cells To Produce Daughter Vesicles with Different Interior and Membrane Compositions. *J Am Chem Soc*

- 2011**, *133* (24), 9545–9555.  
<https://doi.org/10.1021/ja202406v>.
- (61) MacE, C. R.; Akbulut, O.; Kumar, A. A.; Shapiro, N. D.; Derda, R.; Patton, M. R.; Whitesides, G. M. Aqueous Multiphase Systems of Polymers and Surfactants Provide Self-Assembling Step-Gradients in Density. *J Am Chem Soc* **2012**, *134* (22), 9094–9097.  
<https://doi.org/10.1021/ja303183z>.
- (62) Chao, Y.; Shum, H. C. Emerging Aqueous Two-Phase Systems: From Fundamentals of Interfaces to Biomedical Applications. *Chemical Society Reviews* **2020**, *49* (1), 114–142.  
<https://doi.org/10.1039/C9CS00466A>.
- (63) Tsanai, M.; Frederix, P. W. J. M.; Schroer, C. F. E.; Souza, P. C. T.; Marrink, S. J. Coacervate Formation Studied by Explicit Solvent Coarse-Grain Molecular Dynamics with the Martini Model. *Chemical Science* **2021**, *12* (24), 8521–8530.  
<https://doi.org/10.1039/D1SC00374G>.
- (64) Grünewald, F.; Alessandri, R.; Kroon, P. C.; Monticelli, L.; Souza, P. C. T.; Marrink, S. J. PolyPy: A Python Suite for Facilitating Simulations of (Bio-)Macromolecules and Nanomaterials. **2021**.
- (65) Medronho, B.; Romano, A.; Miguel, M. G.; Stigsson, L.; Lindman, B. Rationalizing Cellulose (in)Solubility: Reviewing Basic Physicochemical Aspects and Role of Hydrophobic Interactions. *Cellulose* **2012**, *19* (3), 581–587. <https://doi.org/10.1007/s10570-011-9644-6>.
- (66) Gomes, T. C. F.; Skaf, M. S. Cellulose-Builder: A Toolkit for Building Crystalline Structures of Cellulose. *Journal of Computational Chemistry* **2012**, *33* (14), 1338–1346. <https://doi.org/10.1002/jcc.22959>.
- (67) Vazquez-Salazar, L. I.; Selle, M.; de Vries, A. H.; Marrink, S. J.; Souza, P. C. T. Martini Coarse-Grained Models of Imidazolium-Based Ionic Liquids: From Nanostructural Organization to Liquid–Liquid Extraction. *Green Chemistry* **2020**, *22* (21), 7376–7386.  
<https://doi.org/10.1039/D0GC01823F>.
- (68) Zhu, S.; Wu, Y.; Chen, Q.; Yu, Z.; Wang, C.; Jin, S.; Ding, Y.; Wu, G. Dissolution of Cellulose with Ionic Liquids and Its Application: A Mini-Review. *Green Chem.* **2006**, *8* (4), 325–327.  
<https://doi.org/10.1039/B601395C>.
- (69) Corradi, V.; Sejdiu, B. I.; Mesa-Galoso, H.; Abdizadeh, H.; Noskov, S. Yu.; Marrink, S. J.; Tieleman, D. P. Emerging Diversity in Lipid–Protein Interactions. *Chemical Reviews* **2019**, *119* (9), 5775–5848. <https://doi.org/10.1021/acs.chemrev.8b00451>.
- (70) Khan, H. M.; Souza, P. C. T.; Thallmair, S.; Barnoud, J.; de Vries, A. H.; Marrink, S. J.; Reuter, N. Capturing Choline–Aromatics Cation– $\pi$  Interactions in the MARTINI Force Field. *Journal of Chemical Theory and Computation* **2020**, *16* (4), 2550–2560.  
<https://doi.org/10.1021/acs.jctc.9b01194>.
- (71) Marrink, S. J.; Corradi, V.; Souza, P. C. T.; Ingólfsson, H. I.; Tieleman, D. P.; Sansom, M. S. P. Computational Modeling of Realistic Cell Membranes. *Chemical Reviews* **2019**, *119* (9), 6184–6226.  
<https://doi.org/10.1021/acs.chemrev.8b00460>.
- (72) Tartour, E.; Johannes, L. STxB as an Antigen Delivery Tool for Mucosal Vaccination. *Toxins (Basel)* **2022**, *14* (3), 202.  
<https://doi.org/10.3390/toxins14030202>.
- (73) Johannes, L. Application of Protein Toxins as Cell Biological and Pharmacological Tools. *Toxins (Basel)* **2022**, *14* (4), 242.  
<https://doi.org/10.3390/toxins14040242>.
- (74) Pezeshkian, W.; Shillcock, J. C.; Ipsen, J. H. Computational Approaches to Explore Bacterial Toxin Entry into the Host Cell. *Toxins (Basel)* **2021**, *13* (7), 449. <https://doi.org/10.3390/toxins13070449>.
- (75) Pezeshkian, W.; König, M.; Wassenaar, T. A.; Marrink, S. J. Backmapping Triangulated Surfaces to Coarse-Grained Membrane Models. *Nat Commun* **2020**, *11* (1), 1–9.
- (76) Ling, H.; Boodhoo, A.; Hazes, B.; Cummings, M. D.; Armstrong, G. D.; Brunton, J. L.; Read, R. J. Structure of the Shiga-like Toxin I B-Pentamer Complexed with an Analogue of Its Receptor Gb<sub>3</sub>. *Biochemistry* **1998**, *37* (7), 1777–1788.  
<https://doi.org/10.1021/bi971806n>.
- (77) Pezeshkian, W.; Nåbo, L. J.; Ipsen, J. H. Cholera Toxin B Subunit Induces Local Curvature on Lipid Bilayers. *FEBS Open Bio* **2017**, *7* (11), 1638–1645. <https://doi.org/10.1002/2211-5463.12321>.
- (78) Pezeshkian, W.; Hansen, A. G.; Johannes, L.; Khandelia, H.; Shillcock, J. C.; Kumar, P. B. S.; Ipsen, J. H. Membrane Invagination Induced by Shiga Toxin B-Subunit: From Molecular Structure to Tube Formation. *Soft Matter* **2016**, *12* (23), 5164–5171.  
<https://doi.org/10.1039/C6SM00464D>.
- (79) Pezeshkian, W.; Gao, H.; Arumugam, S.; Becken, U.; Bassereau, P.; Florent, J.-C.; Ipsen, J. H.; Johannes, L.; Shillcock, J. C. Mechanism of Shiga Toxin Clustering on Membranes. *ACS Nano* **2017**, *11* (1), 314–324. <https://doi.org/10.1021/acs.nano.6b05706>.
- (80) Karonen, M.; Virtanen, V. Partition Coefficients (LogP) of Hydrolysable Tannins. *Molecules* **2020**, *25*, 3691.
- (81) 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**, *1–2*, 19–25.  
<https://doi.org/10.1016/j.softx.2015.06.001>.
- (82) Guvench, O.; Mallajosyula, S. S.; Raman, E. P.; Hatcher, E.; Vanommeslaeghe, K.; Foster, T. J.; Jamison, F. W.; MacKerell, A. D. {CHARMM} {A}dditive {A}ll-{A}tom {F}orce {F}ield for

- {C}arbohydrate {D}erivatives and {I}ts {U}tility in {P}olysaccharide and {C}arbohydrate {P}rotein {M}odeling. *Journal of Chemical Theory and Computation* **2011**, *7* (10), 3162–3180. <https://doi.org/10.1021/ct200328p>.
- (83) Stroet, M.; Caron, B.; Visscher, K. M.; Geerke, D. P.; Malde, A. K.; Mark, A. E. Automated Topology Builder Version 3.0: Prediction of Solvation Free Enthalpies in Water and Hexane. *Journal of Chemical Theory and Computation* **2018**, *14* (11), 5834–5845. <https://doi.org/10.1021/acs.jctc.8b00768>.
- (84) 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* (12), 1463–1472. [https://doi.org/10.1002/\(SICI\)1096-987X\(199709\)18:12<1463::AID-JCC4>3.0.CO;2-H](https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H).
- (85) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *The Journal of Chemical Physics* **1983**, *79* (2), 926–935. <https://doi.org/10.1063/1.445869>.
- (86) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *The Journal of Physical Chemistry* **1987**, *91* (24), 6269–6271. <https://doi.org/10.1021/j100308a038>.
- (87) Batista, P. R.; Wilter, A.; Durham, E. H. A. B.; Pascutti, P. G. Molecular Dynamics Simulations Applied to the Study of Subtypes of HIV-1 Protease Common to Brazil, Africa, and Asia. *Cell Biochemistry and Biophysics* **2006**, *44* (3), 395–404. <https://doi.org/10.1385/CBB:44:3:395>.
- (88) Sousa da Silva, A. W.; Vranken, W. F. ACPYPE - AnteChamber PYthon Parser InterfacE. *BMC Research Notes* **2012**, *5* (1), 367. <https://doi.org/10.1186/1756-0500-5-367>.
- (89) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. {CHARMM-GUI}: A Web-Based Graphical User Interface for CHARMM. *J. Comput. Chem.* **2008**, *29* (11), 1859–1865.
- (90) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *The Journal of Chemical Physics* **2007**, *126* (1), 14101. <https://doi.org/10.1063/1.2408420>.
- (91) Eisenhaber, F.; Lijnzaad, P.; Argos, P.; Sander, C.; Scharf, M. The Double Cubic Lattice Method: Efficient Approaches to Numerical Integration of Surface Area and Volume and to Dot Surface Contouring of Molecular Assemblies. *Journal of Computational Chemistry* **1995**, *16* (3), 273–284. <https://doi.org/10.1002/jcc.540160303>.
- (92) Rowland, R. S.; Taylor, R. Intermolecular Nonbonded Contact Distances in Organic Crystal Structures: Comparison with Distances Expected from van Der Waals Radii. *The Journal of Physical Chemistry* **1996**, *100* (18), 7384–7391. <https://doi.org/10.1021/jp953141+>.
- (93) de Jong, D. H.; Baoukina, S.; Ingólfsson, H. I.; Marrink, S. J. Martini Straight: Boosting Performance Using a Shorter Cutoff and GPUs. *Computer Physics Communications* **2016**, *199*, 1–7. <https://doi.org/10.1016/j.cpc.2015.09.014>.
- (94) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *Journal of Applied Physics* **1981**, *52* (12), 7182–7190. <https://doi.org/10.1063/1.328693>.
- (95) Goga, N.; Rzeplia, A. J.; de Vries, A. H.; Marrink, S. J.; Berendsen, H. J. C. Efficient Algorithms for Langevin and DPD Dynamics. *Journal of Chemical Theory and Computation* **2012**, *8* (10), 3637–3649. <https://doi.org/10.1021/ct3000876>.
- (96) Shirts, M.R.; Mobley, D. L. An Introduction to Best Practices in Free Energy Calculations. In *Biomolecular Simulations. Methods in Molecular Biology (Methods and Protocols)*; Monticelli, L.; Salonen, E., Ed.; Humana Press: Totowa, 2013. [https://doi.org/https://doi.org/10.1007/978-1-62703-017-5\\_11](https://doi.org/https://doi.org/10.1007/978-1-62703-017-5_11).
- (97) 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* (8), 3684–3690. <https://doi.org/10.1063/1.448118>.
- (98) del Tatto, V.; Raiteri, P.; Bernetti, M.; Bussi, G. Molecular Dynamics of Solids at Constant Pressure and Stress Using Anisotropic Stochastic Cell Rescaling. *Applied Sciences* **2022**, *12* (3), 1139. <https://doi.org/10.3390/app12031139>.

# Supporting Information

## Martini 3 Coarse-Grained Force Field for Carbohydrates

Fabian Grünewald<sup>1\*</sup>, Mats H. Punt<sup>1\*</sup>, Elizabeth E. Jefferys<sup>2</sup>, Petteri A. Vainikka<sup>1</sup>, Valtteri Virtanen<sup>3</sup>, Melanie König<sup>1</sup>, Weria Pezeshkian<sup>1,5</sup>, Maarit Karonen<sup>3</sup>, Mark S. P. Sansom<sup>2</sup>, Paulo C. T. Souza<sup>4</sup>, Siewert J. Marrink<sup>1</sup>

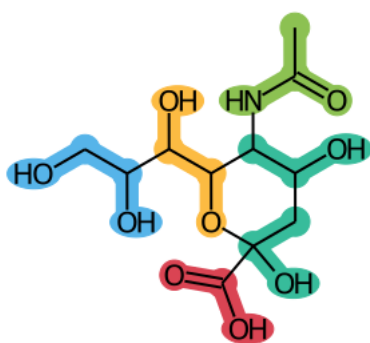
<sup>1</sup> Groningen Biomolecular Sciences and Biotechnology Institute and Zernike Institute for Advanced Materials, University of Groningen, Groningen, The Netherlands

<sup>2</sup> Department of Biochemistry, University of Oxford, South Parks Road, Oxford, United Kingdom, OX1 3QU

<sup>3</sup> Natural Chemistry Research Group, Department of Chemistry, University of Turku, FI-20014 Turku, Finland

<sup>4</sup> Molecular Microbiology and Structural Biochemistry, UMR 5086 CNRS and University of Lyon, Lyon, France

<sup>5</sup> The Niels Bohr International Academy, Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark



**Figure S1. Mapping of Neu5Ac carbohydrate.** Colors correspond to bead types in Figure 1 and Figure 4 of the main paper.

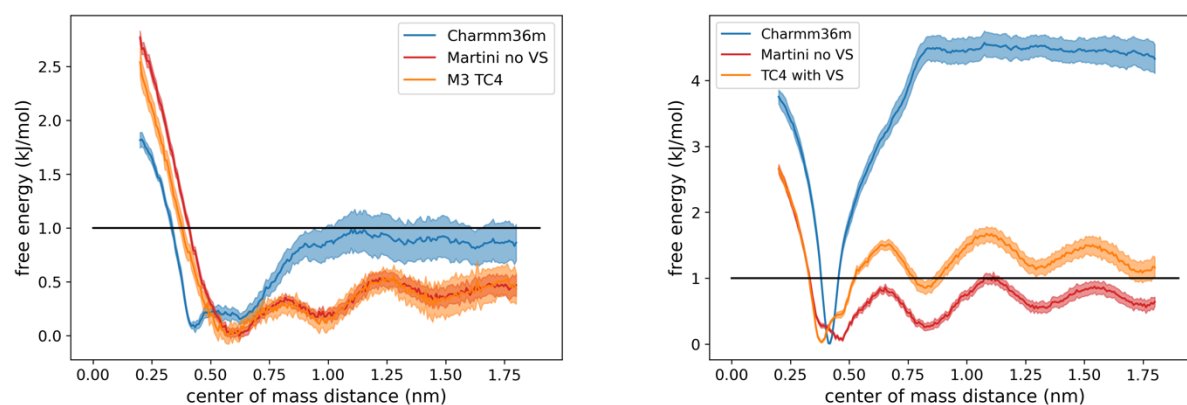
**Table S1. Bond lengths (nm) of  $\alpha$ - and  $\beta$ -glucose.**

Bond	$\alpha$ -D-GLC	$\beta$ -D-GLC	%-diff
A - B	0.322	0.330	2.45
A - C	0.389	0.409	5.0
B - C	0.349	0.344	1.4

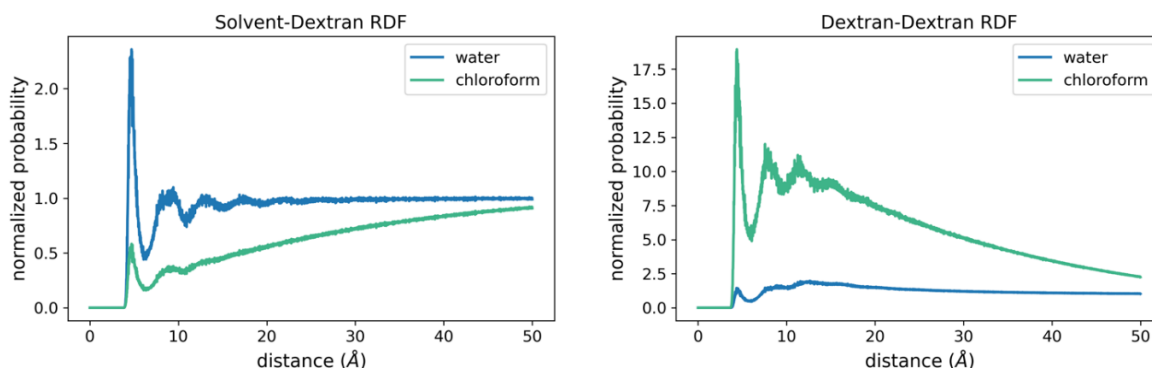
**Table S2. Experimental partitioning coefficients.**

Sugar	CODE	Log P	Exp	SD (n=3)	CG_final	CG_SE
D-glucose	GLC	-3.12	-17.81	0.57	-16.32	0.23
D-mannose	MAN	-2.61	-14.9	0.17	-16.03	0.22
D-galactose	GAL	-3.07	-17.52	0.17	-16.16	0.28
N-acetylglucosamine	GlcNAc	-3.03	-17.29	0.34	-16.02	0.33
N-acetylneuraminic acid	NMC	-4.4	-25.11	0.46	-21.39	0.29
D-glucuronic acid	GLA	-3.26	-18.61	0.11	-18.17	0.31
D-fucose	LFUC	-2.26	-12.9	0.34	-11.09	0.23
L-rhamnose	LRHA	-2.26	-12.9	0.23	-11.11	0.24
D-xylose	XYL	-2.43	-13.87	0.11	-13.06	0.19
Inositol	INO	-3.49	-19.92	0.23	-19.42	0.26
Trehalose*	TREH*	-3.77	-21.52	0.29	-20.61	0.42

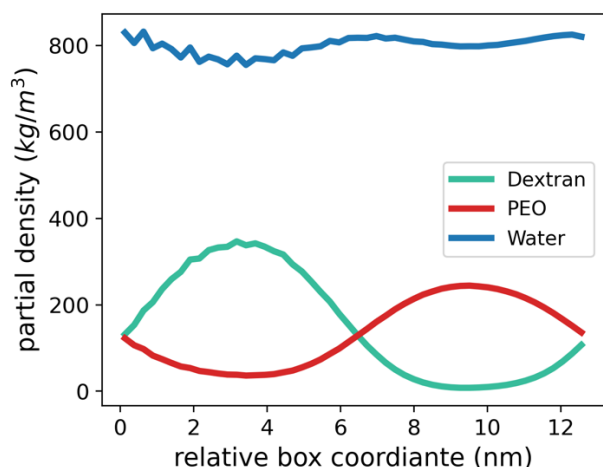
\*values taken from 10.1016/j.carres.2004.12.038



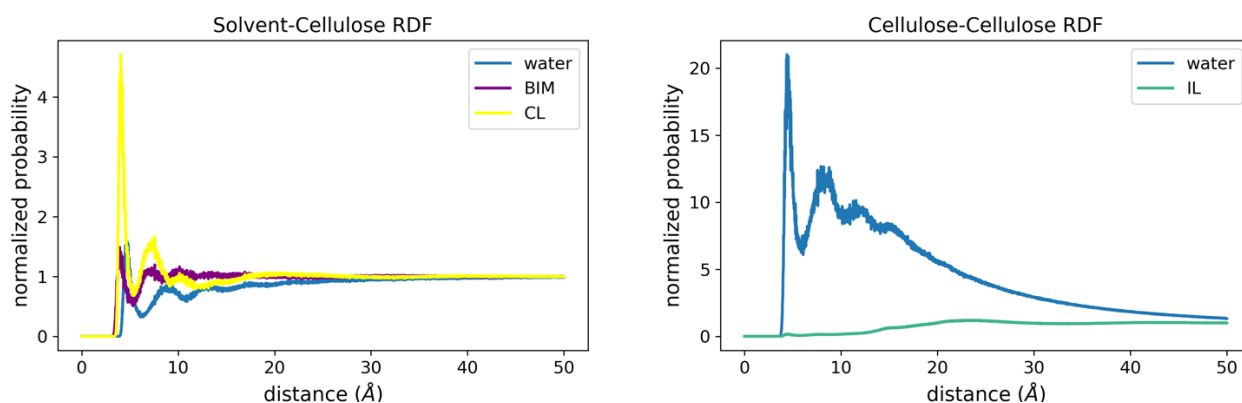
**Figure S2. Potential of mean force (PMF) of glucose - indole interaction.** We assessed the effectiveness of the virtual site (VS) by computing the potential of mean force (PMF) profiles between indole and glucose in solution without (left panel) and with orientation restraints (right panel). The orientation restraints enforce a planar interaction. We observe a reasonable agreement between Martini 3 and CHARMM36m, in case of the orientation free PMF. When planarity is enforced only the Martini model with VS shows an increased binding free energy in agreement with the atomistic simulation. However, the atomistic binding is stronger than what is recovered in Martini. Therefore, on the one hand inclusion of the VS helps in capturing aromatic interactions in general. On the other hand, in situations where a special orientation is enforced as, for example, in a tight binding pocket interactions in Martini could be underestimated.



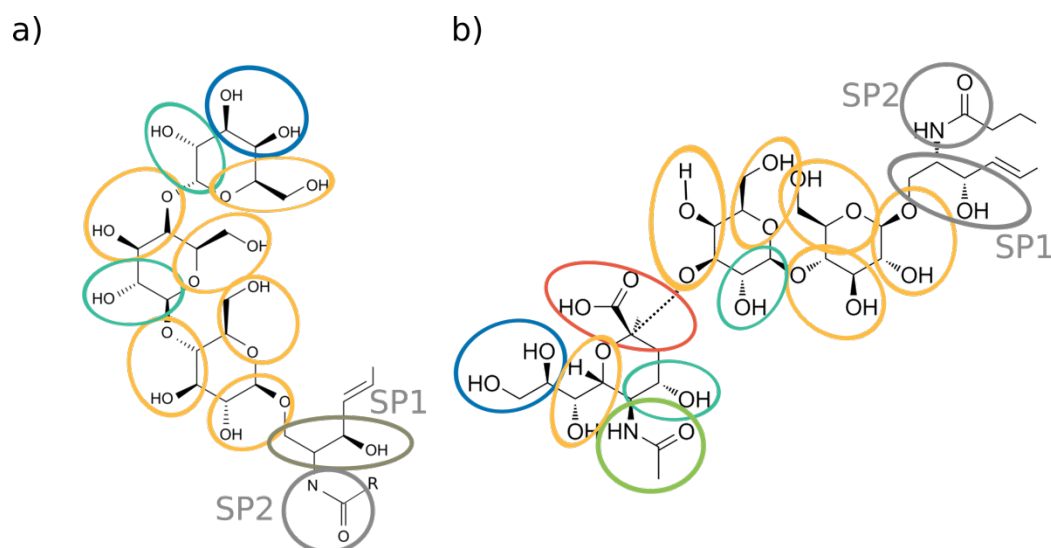
**Figure S3. Radial distribution functions of dextran with solvents and itself.** RDFs of the dextran beads with two different solvents, water and chloroform (left panel) were computed from 500 frames of the solution simulations. The RDF between dextran and water shows a pronounced peak at around 0.45nm corresponding to the first solvation shell. In contrast the green curve corresponding to the RDF between dextran and chloroform shows a peak which is smaller than 1 indicating that the average interactions are unfavorable as expected in a non-solvent. The self RDF between dextran, computed in the same fashion, (right panel) confirms this conclusion. Taken together these RDFs clearly demonstrate that dextran at this concentration is fully solvated in water whereas it is collapsed in chloroform.



**Figure S4. Partial densities of PEO, dextran and water.** Density profiles are shown of the ATPS of dextran and PEO in water along the z-axis of the simulation box. Clearly Dextran is depleted from the PEO phase and vice versa indicating a phase separation between the two polymers. However, the water has an almost constant partial density across the simulation box, which shows that both polymer phases are still hydrated. This corresponds to the expected density profiles for an ATPS system. We further note that water density in the dextran phase is ever so slightly less than in the PEO phase. This behavior is in qualitative agreement with experimental measurements of partitioning in higher molecular weight dextran PEO systems.<sup>1-3</sup>

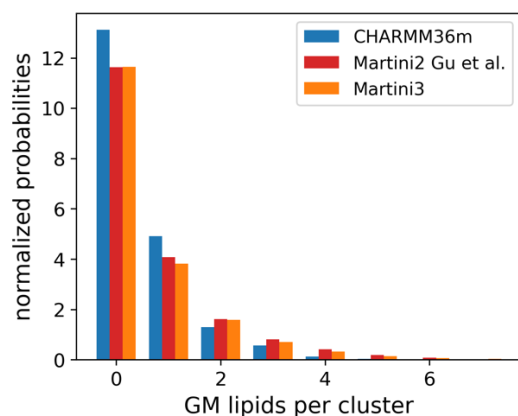


**Figure 5. Radial distribution functions of cellulose with solvents and itself.** RDFs of the cellulose beads with two different solvents (left panel) namely water and the ionic liquid [BMIM][CL] were computed from 500 frames of the solution simulations. The RDF between the CL of the ionic-liquid and cellulose (yellow) shows a pronounced peak at around  $4.5\text{\AA}$  corresponding to the first solvation shell. In contrast the blue curve corresponding to the RDF between cellulose and water shows a peak in that region, which is much smaller indicating that the average interactions are less favorable as expected in a non-solvent. The self RDF between cellulose, computed in the same fashion, (right panel) confirms this conclusion. Clearly the self-interaction in water (blue) is much higher than in the ionic-liquid (green). Taken together these RDFs clearly demonstrate that cellulose at this concentration is fully solvated in the ionic liquid whereas it is aggregated in water.

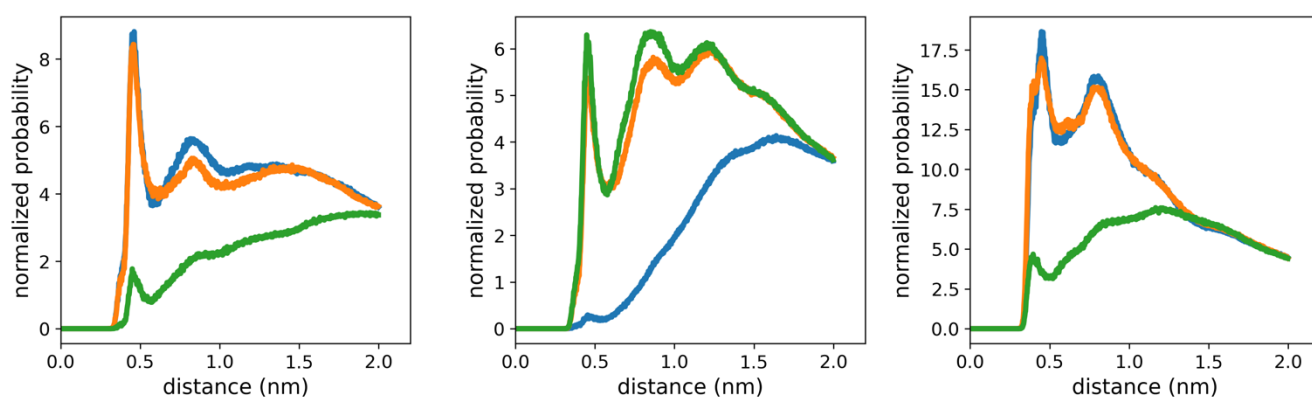


**Figure S6. Mapping of glycolipids.** Gb3 (a) and GM3 (b) with colors of the beads corresponding to Figure 1 and Figure 4 in the main manuscript. Linker beads are mapped as indicated with bead-types SP2 for the amide moiety and SP1 for the alcohol moiety. The rest of the tails is taken from the default Martini 3 paper. Note the TC4 central site has been omitted from the mapping but is part of the parameter set.





**Figure S7. Cluster size distribution for GM3 lipids in CHARMM36m, Martini 2<sup>4</sup> and Martini 3.**



**Figure S8. Site specific RDFs of Gb3 with Shiga Toxin binding sites 1-3. GAL1 (orange) GAL2 (blue) and GLC3 (orange) are sequentially numbered with decreasing distance to the linker. Binding sites were taken as the whole residue at the Martini level as reported from the analysis of the X-ray crystal structure, that is, Table 3 and Table 4 in reference <sup>5</sup>.**

**Table S3. Force-fields used for all-atom simulations of carbohydrates**

AA SUGAR	CODE	FF
$\beta$ -D-glucose	GLC	GLYCAM06h
$\beta$ -D-mannose	MAN	GLYCAM06h
$\beta$ -D-galactose	GAL	CHARMM36
$\beta$ -L-fucose	LFUC	CHARMM36
$\beta$ -L-rhamnose	LRHA	CHARMM36
$\beta$ -D-ribofuranose	RIBF	CHARMM36
$\beta$ -D-xylopyranose	XYL	CHARMM36
$\beta$ -D-fructofuranose	FRUF	CHARMM36
Inosito	INO	GROMOS54a7 AA
$\beta$ -D-glucuronic acid	GLA	CHARMM36
$\beta$ -D-N-acetylglucosamine	GYN	GLYCAM06h
$\beta$ -D-N-acetylneuraminic acid	NMC	CHARMM36
$\beta$ -D-glucosamine	GCN	CHARMM36
Lactose ( $\beta$ 1,4)	LAC	CHARMM36
Sucrose ( $\alpha$ 1,2)	SUCR	GROMOS54a7 AA
Trehalose ( $\alpha$ 1,1)	TREH	CHARMM36

## References

- (1) Großmann, C.; Tintinger, R.; Zhu, J.; Maurer, G. Partitioning of Low Molecular Combination Peptides in Aqueous Two-Phase Systems of Poly(Ethylene Glycol) and Dextran in the Presence of Small Amounts of K<sub>2</sub>HPO<sub>4</sub>/KH<sub>2</sub>PO<sub>4</sub> Buffer at 293 K: Experimental Results and Predictions. *Biotechnology and Bioengineering* **1998**, *60* (6), 699–711. [https://doi.org/10.1002/\(sici\)1097-0290\(19981220\)60:6<699::aid-bit7>3.0.co;2-i](https://doi.org/10.1002/(sici)1097-0290(19981220)60:6<699::aid-bit7>3.0.co;2-i).
- (2) Tintinger, R.; Zhu, J.; Grossmann, C.; Maurer, G. Partitioning of Some Amino Acids and Low Molecular Mass Peptides in Aqueous Two-Phase Systems of Poly(Ethylene Glycol) and Dextran in the Presence of Small Amounts of K<sub>2</sub>HPO<sub>4</sub>/KH<sub>2</sub>PO<sub>4</sub>-Buffer at 293 K: Experimental Results and Correlation. *Journal of Chemical and Engineering Data* **1997**, *42* (5), 975–984. <https://doi.org/10.1021/je970080v>.
- (3) Tintinger, R.; Zhu, J.; Grossmann, C.; Maurer, G. Partitioning of Some Amino Acids and Low Molecular Mass Peptides in Aqueous Two-Phase Systems of Poly(Ethylene Glycol) and Dextran in the Presence of Small Amounts of K<sub>2</sub>HPO<sub>4</sub>/KH<sub>2</sub>PO<sub>4</sub>-Buffer at 293 K: Experimental Results and Correlation. *Journal of Chemical & Engineering Data* **1997**, *42* (5), 975–984. <https://doi.org/10.1021/je970080v>.
- (4) Gu, R.-X.; Ingólfsson, H. I.; de Vries, A. H.; Marrink, S. J.; Tieleman, D. P. Ganglioside-Lipid and Ganglioside-Protein Interactions Revealed by Coarse-Grained and Atomistic Molecular Dynamics Simulations. *The Journal of Physical Chemistry B* **2017**, *121* (15), 3262–3275. <https://doi.org/10.1021/acs.jpcc.6b07142>.

- (5) Ling, H.; Boodhoo, A.; Hazes, B.; Cummings, M. D.; Armstrong, G. D.; Brunton, J. L.; Read, R. J. Structure of the Shiga-like Toxin I B-Pentamer Complexed with an Analogue of Its Receptor Gb<sub>3</sub>. *Biochemistry* **1998**, *37* (7), 1777–1788.  
<https://doi.org/10.1021/bi971806n>.