Supporting information for: TopoGromacs: Automated Topology Conversion from CHARMM to GROMACS within VMD

Josh V Vermaas,[†,‡,¶] David J Hardy,[¶] John E Stone,[¶] Emad Tajkhorshid,[*,†,‡,¶] and Axel Kohlmeyer[*,§]

†*Center for Biophysics and Quantitative Biology*

‡*Department of Biochemistry*

¶*Beckman Insitute, University of Illinois at Urbana-Champaign, Urbana, IL 61801*

§*Insitute for Computational Molecular Science, Temple University, Philadelphia, PA 19122*

E-mail: emad@life.illinois.edu; a.kohlmeyer@temple.edu

TopoTools Overview

The TopoTools package* acts as a middleware in the structure building tools within VMD. While VMD proper has the capability of adding bonds, defining improper torsions, and replicating molecules, the interface can be cumbersome for simple tasks. TopoTools simplifies the process by providing an interface layer that is easier to use for structure preparation tasks, but also extends what is possible by providing an interface for the user to automatically infer properties from the structure. For instance, given a set of bonded atoms, TopoTools can guess the required angle and dihedral terms from connectivity, and can additionally guess which atoms will require an improper torsion to maintain the correct geometry. It also enables

*The full documentation is available for the package is available at `http://www.ks.uiuc.edu/Research/vmd/plugins/topotools/`

the user to focus on building their system one component at a time, by providing utilities to combine either entire of subsets of structures with one another into a single simulation system. Other powerful utilities include molecular replication and the ability to read and write LAMMPS topology files. The conversion from CHARMM to GROMACS was thus a natural extension of these existing features.

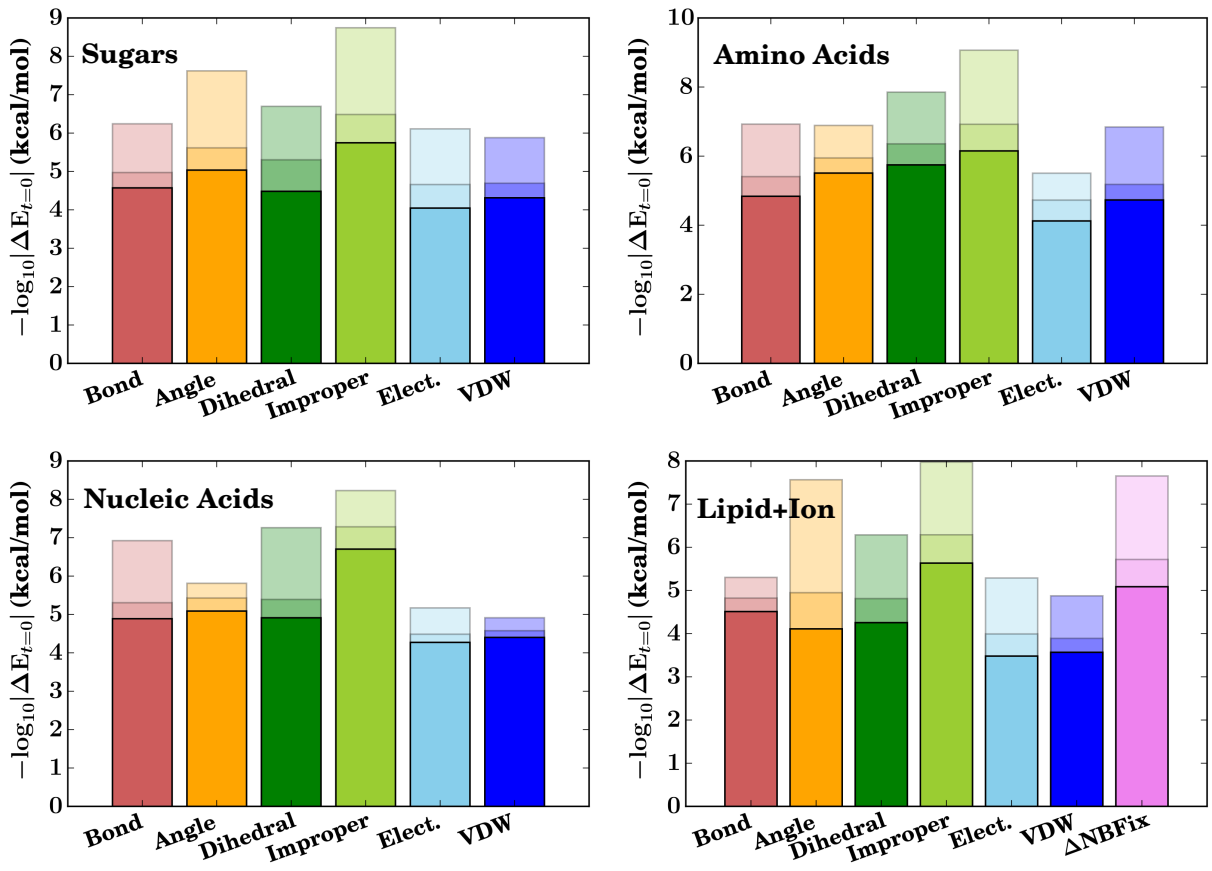

Figure S1: Energy difference of each potential energy term for monomeric systems computed by NAMD and GROMACS. The lowest line in each bar indicates the maximum energy difference observed, the highest line the minimum energy difference observed, and the middle line the mean energy difference. The ΔNBFIX term is calculated by computing the difference between the VDW term both with and without the sodium-carbonyl correction[S1] and comparing the difference between NAMD and GROMACS. These results encompass 73 distinct sugar residues, 5 nucleic acids, 20 amino acids, and 35 lipid types commonly found in the CHARMM36 force field.[S2–S7]

Implementation of the Conversion Process

While the information content of a GROMACS `.top` file is equivalent to the pairing of a structure file (`.psf`) with a parameter file in CHARMM, the structure is quite different, and involves reshuffling the data. In a sense, the conversion process is very mechanical, and not relevant to the use of the conversion tool. However, for interested parties, we detail the algorithmic details of how each section of a `.top` file is written.

The first three sections of the `.top` file are related to how non-bonded forces are treated and enumerate all the atom types in the system. In the defaults section, we specify the same combination rules for the Lennard-Jones parameters ε ($\epsilon_{ij} = (\epsilon_i \epsilon_j)^{\frac{1}{2}}$) and σ ($\mathrm{r}_{ij} = 2^{\frac{1}{6}}\sigma_{ij}$, $\mathrm{r}_{ij} = \frac{1}{2}(\mathrm{r}_i + \mathrm{r}_j)$) as are used in a typical CHARMM force field. When listing the atom types, the respective columns are populated based on the information available to VMD when reading in the `.psf` (or other structure files) and parameter files. The atom types themselves are assumed to be alphanumeric, as in XPLOR-formatted `.psf` files, rather than strictly numeric, as in CHARMM-formatted `.psf` files. (Conversion between the two `.psf` formats can be easily performed using CHARMM or the psfgen plugin within VMD.) The use of alphanumeric atom types reduces the possibility for errors when combining parameters from different sources and additionally improves the readability of the `.top` file when checking correctness. GROMACS-required fields, such as atomic number, are determined from the atomic masses present in the `.psf` file, and the generic σ and ε are computed from the CHARMM parameters, which are combined using the combination rules to provide parameters for the non-bonded interactions between distant pairs. Certain atom types in CHARMM have different parameters if they are involved in a 1-4 interaction, and these cases are handled by their inclusion in the pair-types list, where σ and ε for 1-4 interactions are explicitly precomputed if they differ from the generic case. No attempt is made to reduce the size of the list based on the topology of the system in question, as the file size savings are outweighed by the increased runtime for the conversion process required for large systems to check if a particular pair is involved in a 1-4 interaction. The next five sections of

the resultant `.top` file are related to the calculation of the bonded forces, and are a direct translation of the applicable CHARMM parameters (kcal/mol and Å) to GROMACS units (kJ/mol and nm). Unlike in the case of 1-4 interactions, the check for whether a certain set of parameters are required within the simulation simplifies to testing whether each atom type present in the interaction is also within the simulation system, reducing the file size at low computational cost. Due to the differences in the parsing logic for wildcards in dihedral terms between CHARMM/NAMD and GROMACS, dihedral and improper parameters are reordered such that wildcard terms are picked up last by the grompp `.top` parser, to be used only when no explicit dihedral term exists in the system.

After specifying the parameters, the sections of the `.top` file that follow describe each molecular species present in the simulation system, enumerating the details of each atom (name, type, residue identification, charge, charge group, and mass). Most of these data elements can be copied directly from the input data, however, the charge group and the 1-4 pair list needs to be calculated separately. Charge groups are identified algorithmically by including sequential atoms until the sum of the charges in the group is an integer. For many molecular species (lipids, amino acids, and nucleic acids), this procedure results in compact charge groups. In species where atoms belonging to a charge group are not sequential in the CHARMM topology file (such as in cyclic sugars), the charge groups expand significantly, and a single charge group may contain most of the residue. Since charge groups are not used in the most common electrostatic treatments in GROMACS (PME[S8,S9] and RFZ[S10]), incorrect charge groups in these cases are not an impediment to running simulations.[S11] CHARMM and NAMD automatically calculate the 1-4 pair list for each molecule; however, in GROMACS the 1-4 interactions must be explicitly specified so that they properly take into account different parameters that may be used for 1-4 terms. The pair list is built for each molecule by iterating over the bonds, thereby building up a complete list of pair interactions. Duplicate pairs and pairs in the 1-2 or 1-3 exclusion lists arising from cyclic systems are removed from the pair list as appropriate. For each molecule, the remaining

bonded terms are copied directly from the input structure.

Unlike CHARMM structure files, which list each atom and bonded term explicitly, repeated molecules in GROMACS topology files can be represented as a single copy with a multiplicity. We detect this multiplicity by iterating through the fragments (molecules) within the selected structure, and comparing successive fragments to previous fragments to determine if the names and atom types of the atoms are identical. Only if a new fragment is detected is a new molecule written to the output topology file. As a direct consequence, the system information is significantly more compact than in alternative workflows,[S12] where only repeated water molecules are recognized. This last step completes the topology translation to a GROMACS-formatted topology file.

Observed Energy Drift Distribution

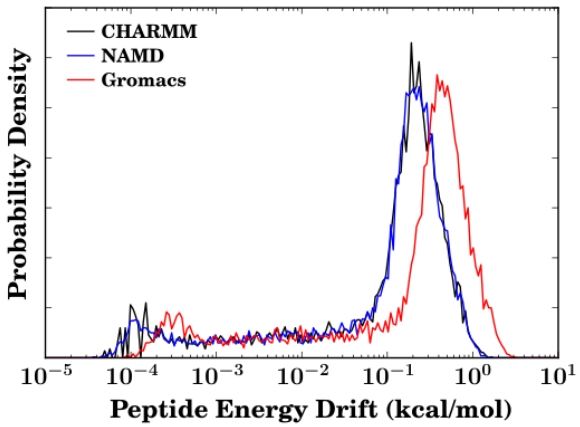

Figure S2: Energy drift distribution of all 8,000 10 ps tripeptide simulations, representing the distribution of the difference between the starting energy value and the value after 10 ps. CHARMM results are in black, NAMD results are in blue, and GROMACS results are in red.

As a general note, for constant energy simulations, such as those carried out here, there is a natural tendency for energy drift to occur due to numerical imprecision propagating errors forward in time. Even in these short simulations, there is a noticeable drift from the beginning

of the simulation to 10 ps later (Fig. S2). The effect of mixed precision in GROMACS here manifests itself as an energy drift that is typically twice as large as the fully-double precision CHARMM or NAMD. For typical simulation ensembles with a thermostat, the modest energy drift is a small effect relative to a thermostat, and should not be considered an impediment to the reliability of the simulation.

Construction of the Application Benchmarks

The 23,558-atom DHFR and 92,224-atom ApoA1 benchmarks were updated to use CHARMM36,[S5,S13] using their customary periodic box sizes for simulation with NAMD[S14] and GROMACS.[S15] A short xyloglucan fibril, similar to that found in plant cell walls,[S16] was built using CarbBuilder,[S17] solvated within a water-box of side length 70 Å using VMD[S18] yielding a solvated system size of 32,271 atoms, and simulated for 10 ns at 300 K using 2 fs timesteps. The initial crystal structure of the vitamin B12 transporter BtuCD-F in its nucleotide bound state (PDBID: 4FI3[S19]) was aligned using the OPM (Orientations of Proteins in Membranes) database,[S20] and inserted into a POPE (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine) membrane using CHARMM-GUI,[S21–S23] resulting in a 191,708-atom system. Since construction, CHARMM-GUI has added the capability of generating input for a variety of MD packages,[S12] however, this functionality was not used here. With the exception of DHFR, the simulations were performed only in NAMD and GROMACS using PME for electrostatics and a 12 Å spherical cutoff distance for the van der Waals interactions with a switching function applied after 10 Å. The DHFR system was, for comparison, also simulated with CHARMM version c39b1,[S24] also using PME for electrostatics but with a shorter 9 Å spherical cutoff for the van der Waals interactions with a switching function applied after 7.5 Å.

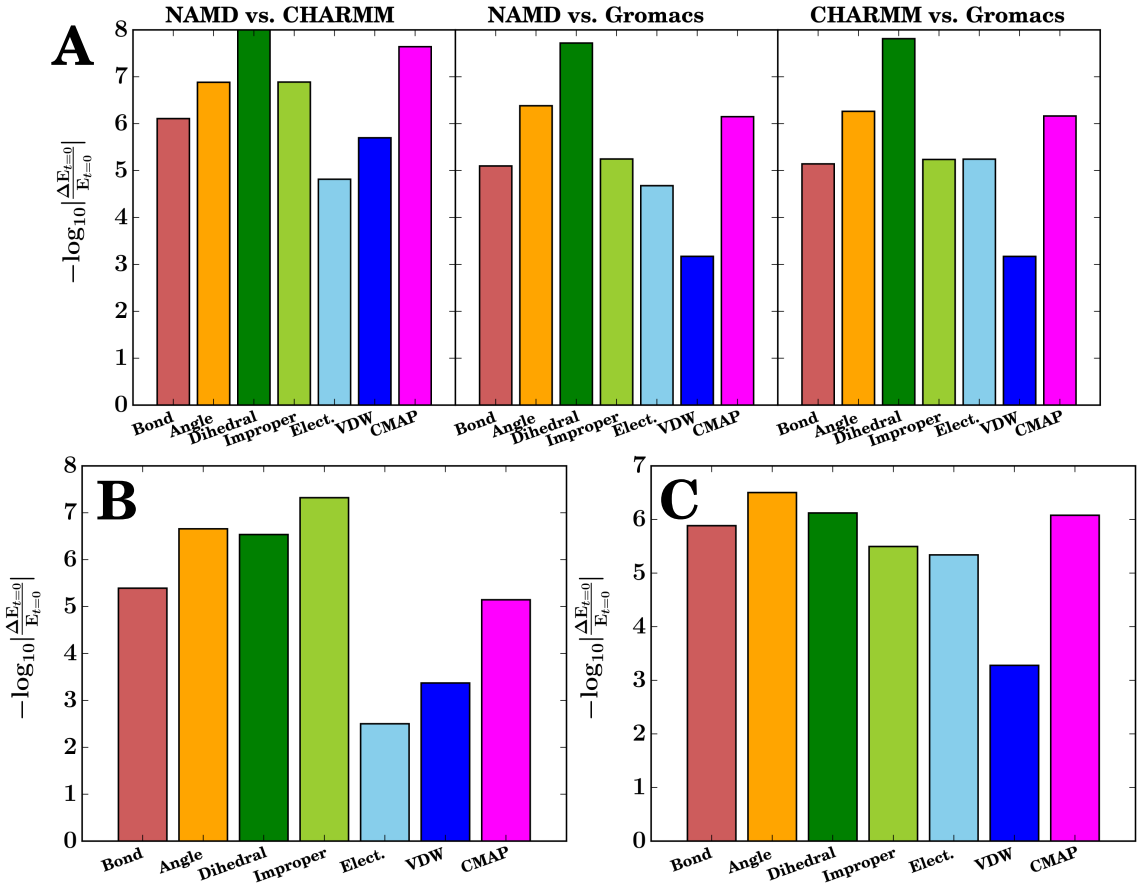

Figure S3: Relative energy difference within each term for periodic benchmark systems between NAMD and GROMACS simulations at t=0. (A) DHFR comparisons between NAMD, GROMACS, and CHARMM. (B) ApoA1 comparison between NAMD and GROMACS. (C) BtuCD-F in membrane comparison between NAMD and GROMACS.

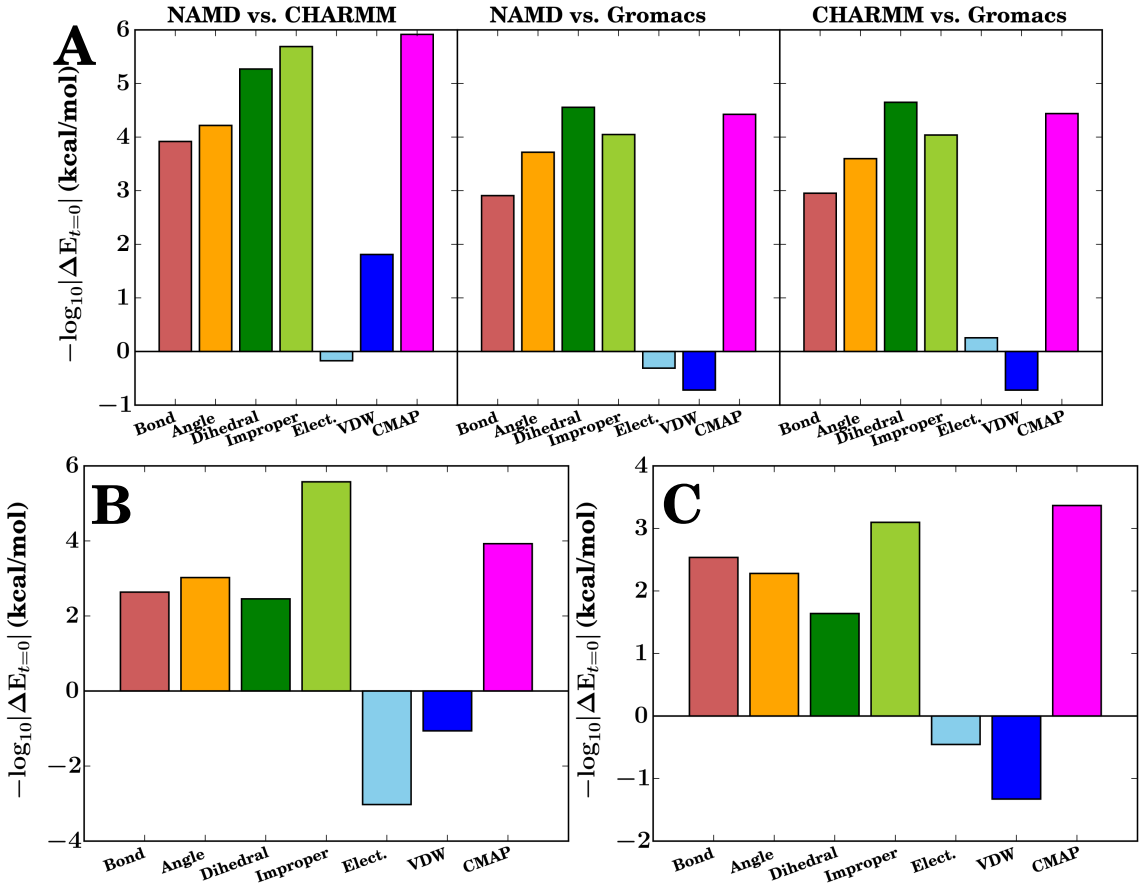

Figure S4: Absolute energy difference within each term for periodic benchmark systems between NAMD and GROMACS simulations. (A) DHFR comparisons between NAMD, GROMACS, and CHARMM. (B) ApoA1 comparison between NAMD and GROMACS. (C) BtuCD-F in membrane comparison between NAMD and GROMACS.

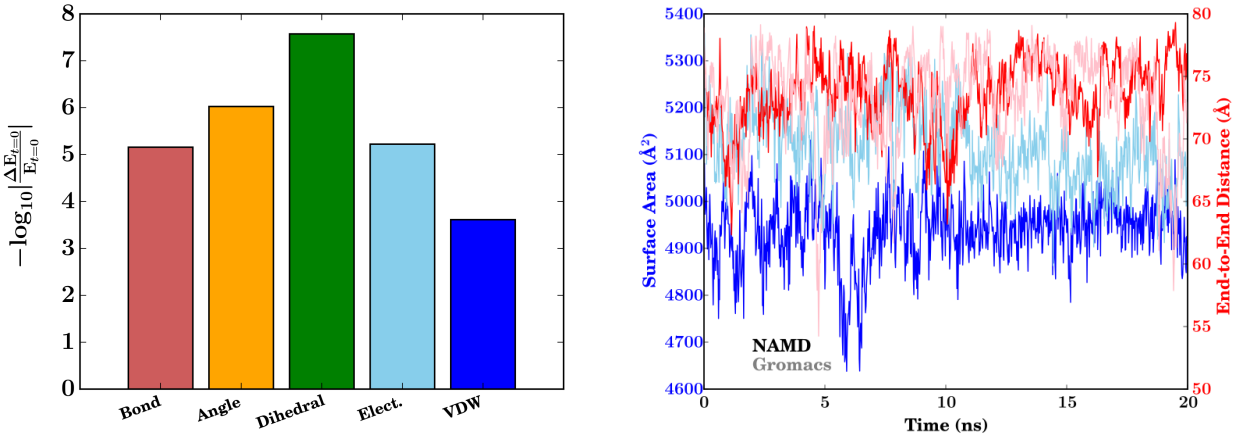

Figure S5: (Left) Relative energy difference for a xyloglucan structure at t=0 between NAMD and GROMACS. (Right) Xyloglucan surface area (blue) and end-to-end distance (red) for a short Xyloglucan over 20 ns of simulation using NAMD (darker colors) and GROMACS (lighter colors).

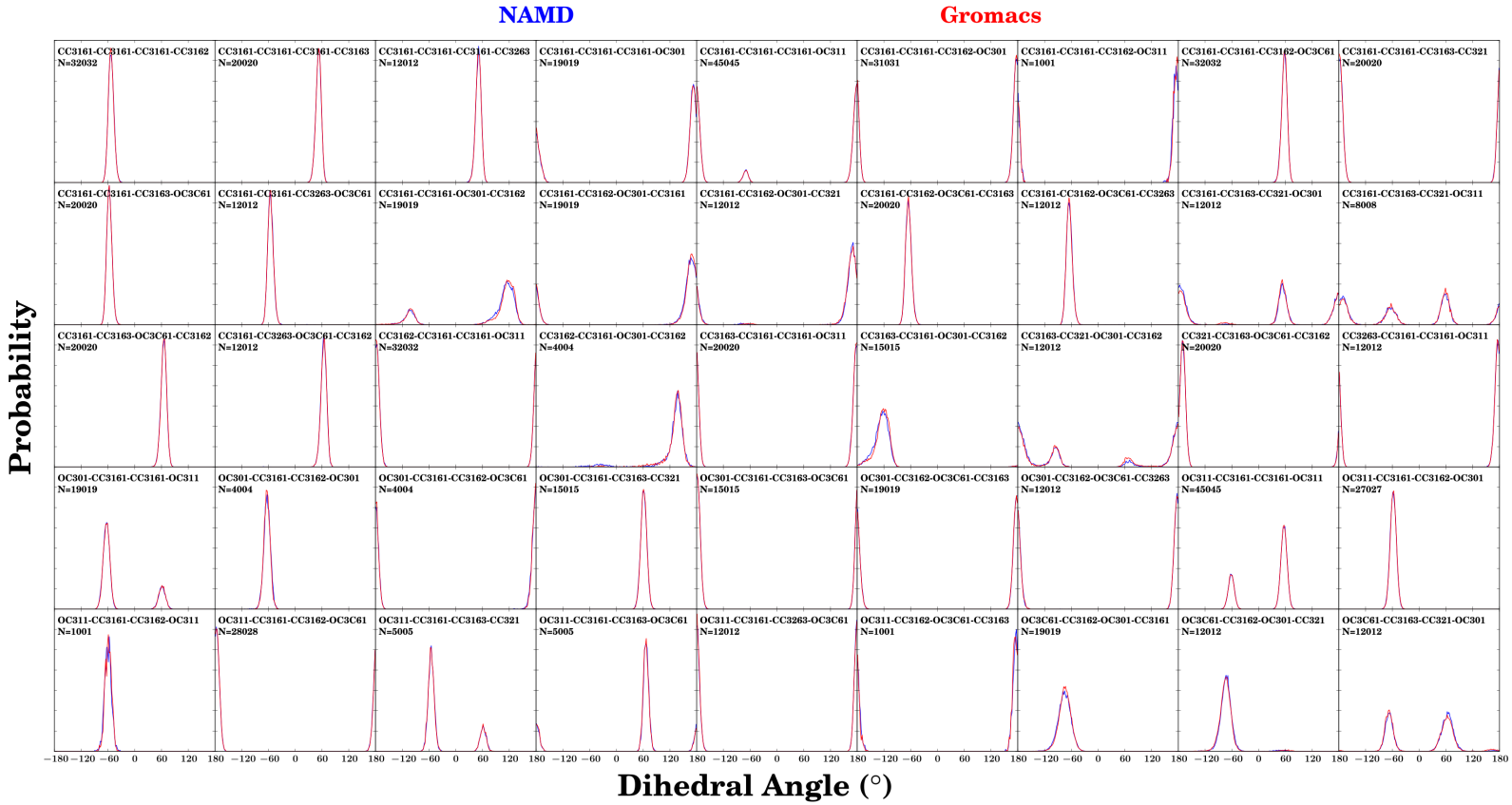

Figure S6: Heavy-atom dihedral distributions for Xyloglucan over 20 ns of simulation. Each subplot is labeled according to the atom types comprising the dihedral, along with the number of distinct samples present, taken every 5 ps. Distributions are colored according to MD engine: blue for NAMD, red for GROMACS.

References

(S1) Venable, R. M.; Luo, Y.; Gawrisch, K.; Roux, B.; Pastor, R. W. Simulations of anionic lipid membranes: Development of interaction-specific ion parameters and validation using NMR data. *J. Phys. Chem. B* **2013**, *117*, 10183–10192.

(S2) Guvench, O.; Greene, S. N.; Kamath, G.; Brady, J. W.; Venable, R. M.; Pastor, R. W.; Mackerell, A. D. Additive empirical force field for hexopyranose monosaccharides. *J. Comput. Chem.* **2008**, *29*, 2543–2564.

(S3) Hatcher, E. R.; Guvench, O.; MacKerell, A. D. CHARMM Additive All-Atom Force Field for Aldopentofuranoses, Methyl-aldopentofuranosides, and Fructofuranose. *J. Phys. Chem. B* **2009**, *113*, 12466–12476.

(S4) Hatcher, E. R.; Guvench, O.; MacKerell, A. D. CHARMM Additive All-Atom Force Field for Acyclic Polyalcohols, Acyclic Carbohydrates, and Inositol. *J. Chem. Theory Comput.* **2009**, *5*, 1315–1327.

(S5) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain χ_1 and χ_2 Dihedral Angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257–3273.

(S6) MacKerell, A. D.; Feig, M.; Brooks, C. L. Improved treatment of the protein backbone in empirical force fields. *J. Am. Chem. Soc.* **2004**, *126*, 698–9.

(S7) Denning, E. J.; Priyakumar, U. D.; Nilsson, L.; MacKerell, A. D. Impact of 2'-hydroxyl sampling on the conformational properties of RNA: Update of the CHARMM all-atom additive force field for RNA. *J. Comput. Chem.* **2011**, *32*, 1929–1943.

(S8) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(S9) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577.

(S10) Sperb, R.; Tironi, I. G.; Tironi, I. G.; Sperb, R.; Smith, P. E.; Smith, P. E.; Gunsteren, W. F. V.; Gunsteren, W. F. V. A generalized reaction field method for molecular dynamics simulations. *J. Chem. Phys.* **1995**, *102*, 5451–5459.

(S11) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.

(S12) Lee, J.; Cheng, X.; Swails, J. M.; Yeom, M. S.; Eastman, P. K.; Lemkul, J. A.; Wei, S.; Buckner, J.; Jeong, J. C.; Qi, Y.; Jo, S.; Pande, V. S.; Case, D. A.; Brooks, C. L.; MacKerell, A. D.; Klauda, J. B.; Im, W. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J. Chem. Theory Comput.* **2016**, *12*, 405–413.

(S13) Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell, A. D.; Pastor, R. W. Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *J. Phys. Chem. B* **2010**, *114*, 7830–7843.

(S14) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802.

(S15) Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 845–854.

(S16) Schultink, A.; Liu, L.; Zhu, L.; Pauly, M. Structural Diversity and Function of Xyloglucan Sidechain Substituents. *Plants* **2014**, *3*, 526–542.

(S17) Kuttel, M.; Mao, Y.; Widmalm, G.; Lundborg, M. CarbBuilder: An Adjustable Tool for Building 3D Molecular Structures of Carbohydrates for Molecular Simulation. 2011 IEEE Seventh Int. Conf. eScience. 2011; pp 395–402.

(S18) Humphrey, W.; Dalke, A.; Schulten, K. VMD – Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.

(S19) Korkhov, V. M.; Mireku, S. A.; Locher, K. P. Structure of AMP-PNP-bound vitamin B12 transporter BtuCD-F. *Nature* **2012**, *490*, 367–72.

(S20) Lomize, M. A.; Lomize, A. L.; Pogozheva, I. D.; Mosberg, H. I. OPM: Orientations of proteins in membranes database. *Bioinformatics* **2006**, *22*, 623–625.

(S21) Wu, Z.; Schulten, K. Synaptotagmin's Role in Neurotransmitter Release Likely Involves Ca^{2+}-induced Conformational Transition. *Biophys. J.* **2014**, *107*, 1156–1166.

(S22) Jo, S.; Lim, J. B.; Klauda, J. B.; Im, W. CHARMM-GUI Membrane Builder for Mixed Bilayers and Its Application to Yeast Membranes. *Biophys. J.* **2009**, *97*, 50–58.

(S23) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *J. Comput. Chem.* **2008**, *29*, 1859–1865.

(S24) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545–1614.