Supporting information:

Conformational Diversity of O-antigen Polysaccharides of the Gram-negative Bacterium *Shigella flexneri* Serotype Y

Yu Kang[a)], Stefanie Barbirz[b)], Reinhard Lipowsky[a)] and Mark Santer[a)]

[a)]Max Planck Institute of Colloids and Interfaces, 14424 Potsdam, Germany

[b)] Physikalische Biochemie, Universität Potsdam, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany

A. Force fields and simulation details

Conventional MD simulations with explicit solvent were conducted in GROMACS according to the following protocol: after initial solvation, the system was first optimized with 500 steps of steepest descent followed by another 500 steps of conjugate gradient minimization. After minimization, the system was heated from 0 K up to 310 K at a pressure of 1 atm with a short 100 ps MD simulation, followed by an equilibration period for 500 ps at a temperature of 310 K and a pressure of 1 atm by using Nosé-Hoover temperature- [1,2] and Parrinello-Rahman [3] pressure coupling. After the equilibration phase, production runs of 200 ns up to 800 ns duration at constant volume were conducted using a velocity-Verlet intergrator [4] with half-step averaged kinetic energies, and a temperature of 310 K was maintained with a multi-chain Nose-Hoover thermostat, with a chain depth of 10. The particle mesh Ewald (PME) method [5] was used to calculate electrostatic interactions, with a cut-off of 1.0 nm for the separation of the direct and reciprocal space summation. The cut-off distance for van der Waals interaction was 1.0 nm, and the parameters of the Lennard-Jones potential for the cross interactions between non-bonded atoms were obtained from the well-known Lorentz-Berthelot combination rule.

MD simulations with implicit solvent were conducted in Amber using the generalized Born solvent model HCT [6,7] and/or OBC [8] with the dielectric constant set to 80. The simulation protocol is similar to the one for explicit solvation, including minimization, heating, equilibration and production. The length

of the production runs varies between 1-4μs. The temperature was controlled by Langevin thermostat with a collision frequency of 1.0 ps^{-1}. The periodicity was disabled and the cut-off was set to be infinite.

In all MD simulations, 1-4 interactions were scaled with the factors 1.0/1.0 (corresponding to the requirements of GLYCAM/CHARMM). Hydrogen bonds within water were treated with the SETTLE algorithm [9], and other constraints were treated with LINCS [10] in Gromacs or SHAKE [11] in Amber. A time step of 2 fs was used and the coordinates were saved every 2 ps. Visualizations were made using VMD [12]. The existence of hydrogen bonds as discussed in the main text has been determined by using a geometrical criterion of donor-acceptor distance no greater than 0.35 nm and hydrogen-donor-acceptor angle no greater than 30º.

The modeling of oligosaccharides consisting of two, four, up to sixteen monosaccharide units was carried out with the software tool tLEaP, as part of the Amber 11 package.

B. Adaptively Biased MD (ABMD) simulations.

For free energy calculation with adaptively biased molecular dynamics simulations disaccharides corresponding to all linkages I-IV were considered (explicit solvent). The corresponding profiles are rendered stationary within 200 ns of simulation time (Figure S1), where for linkages II-IV beyond 100ns simulation time the shape of the free energy curves do not change appreciably any more. This means that from a very early stage onwards, the biased trajectory has uniform access to all relevant parts of glycosidic angular space. A continuous overall shift simply reflects the ongoing flooding within a basin defined by steric hindrances. A slight deviation from this characteristic is only observed for linkage I, where it is evident that some part of conformational space is visited only at a later stage of the flooding simulation, compare the free energy profiles (time slices) between 170ns and 190ns. In Figure S2, we show the probability distribution projected onto single glycosidic angles to the distributions as calculated from a typical 200ns MD trajectory, as a supplement to Table 1. The comparison suggests a nearly quantitative agreement.

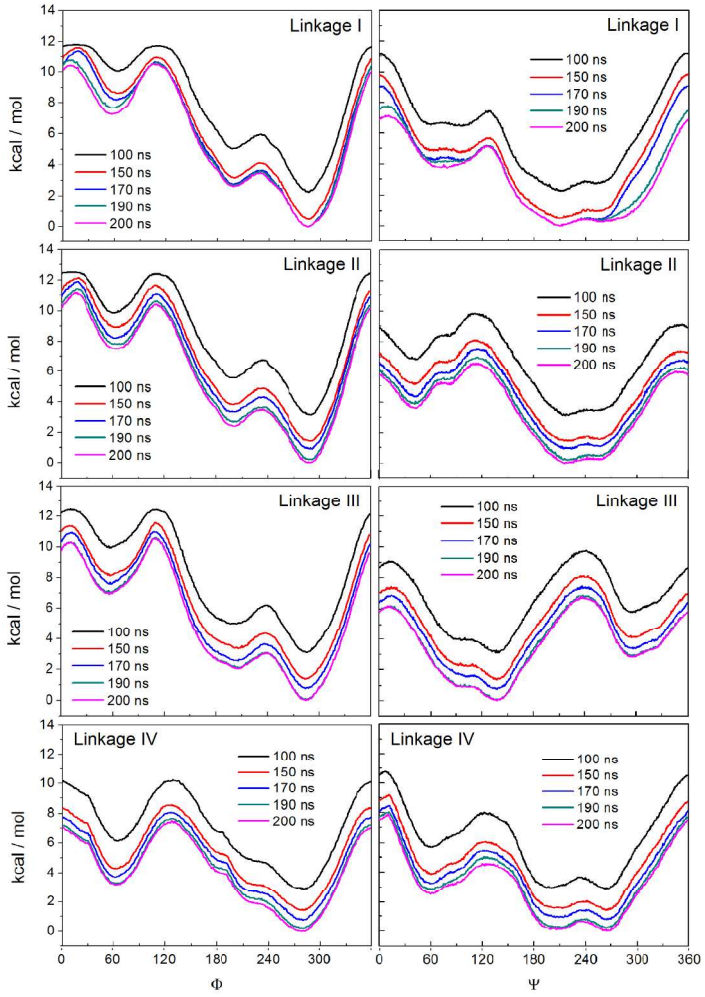

Figure S1. Free energies taken from an ABMD simulation of disaccharides corresponding to linkages I-IV as a function of Φ or Ψ. The different colors correspond to free energy curves obtained at different times of the ABMD run with explicit solvent (a single curve is constructed from a time slice of the ABMD trajectory around the time indicated).

Although the ABMD method directly yields smooth and rather extensive free energy landscapes, from which populations can be inferred by Boltzmann inversion and subsequent integration, we have to recognize that it is actually a non-equilibrium method, and its precision depends most sensitively on the flooding time scale τ_F, governing the growth rate of the biasing potential.

However, in practice the method appears very robust, values for τ_F of about 1ns and even far below usually give residual RMS errors on the whole free energy landscape of around and less than 1kcal/mole after 150ns to 300ns simulation time [13], although the precision also depends on the particular system

investigated. For our case, on account of previous studies with oligosaccharides in explicit solvent [14] we could confirm that there is hardly any difference between results with τ_F in the range of 20ps to 50ps, and the error should be at most a k_BT (~0.6kcal/mole) up to values of 7kcal/mole above the global minimum (and in the majority of cases less than that). In order to rigorously converge occupation probabilities, a systematic and hierarchical procedure employing, for instance, parallel tempering would be needed [13]. This is, however, beyond the scope of the present study. The agreement between occupation probabilities inferred from ABMD simulations and from plain MD trajectory data displayed in Table 1 is, nevertheless, rather convincing. Just to illustrate the effect a residual error function on the free energy landscape would have: if we added penalty functions to the states defined in this work as to emulate a non-converged profile, adding/subtracting a value of ±0.1 kcal/mol would roughly correspond to a variation in population of around ±5% for A states and ±1% for B states.

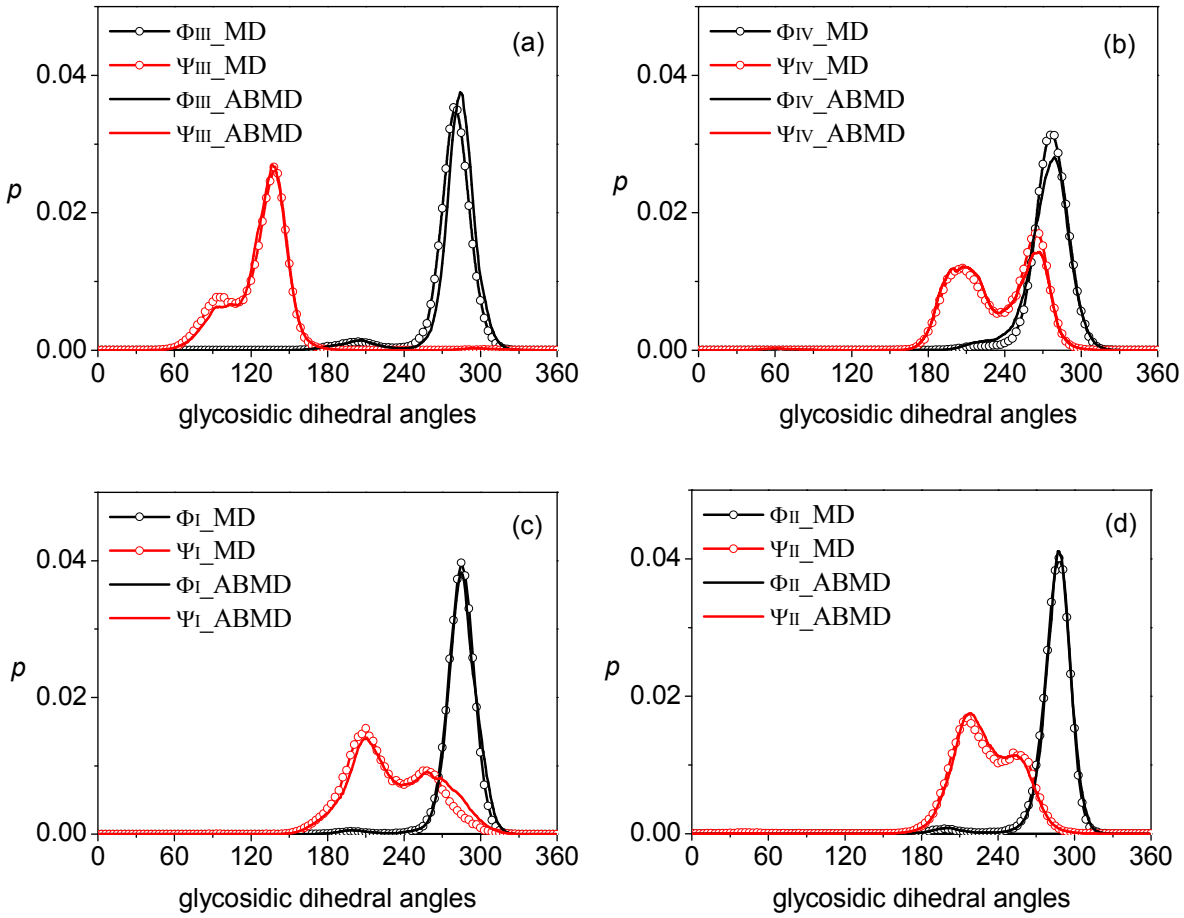

Figure S2. Populations of the Φ-Ψ angles of (a) Linkage III, (b) Linkage IV, (c) Linkage I and (d) Linkage II in disaccharides calculated from MD trajectories and ABMD free energy profiles with explicit solvent.

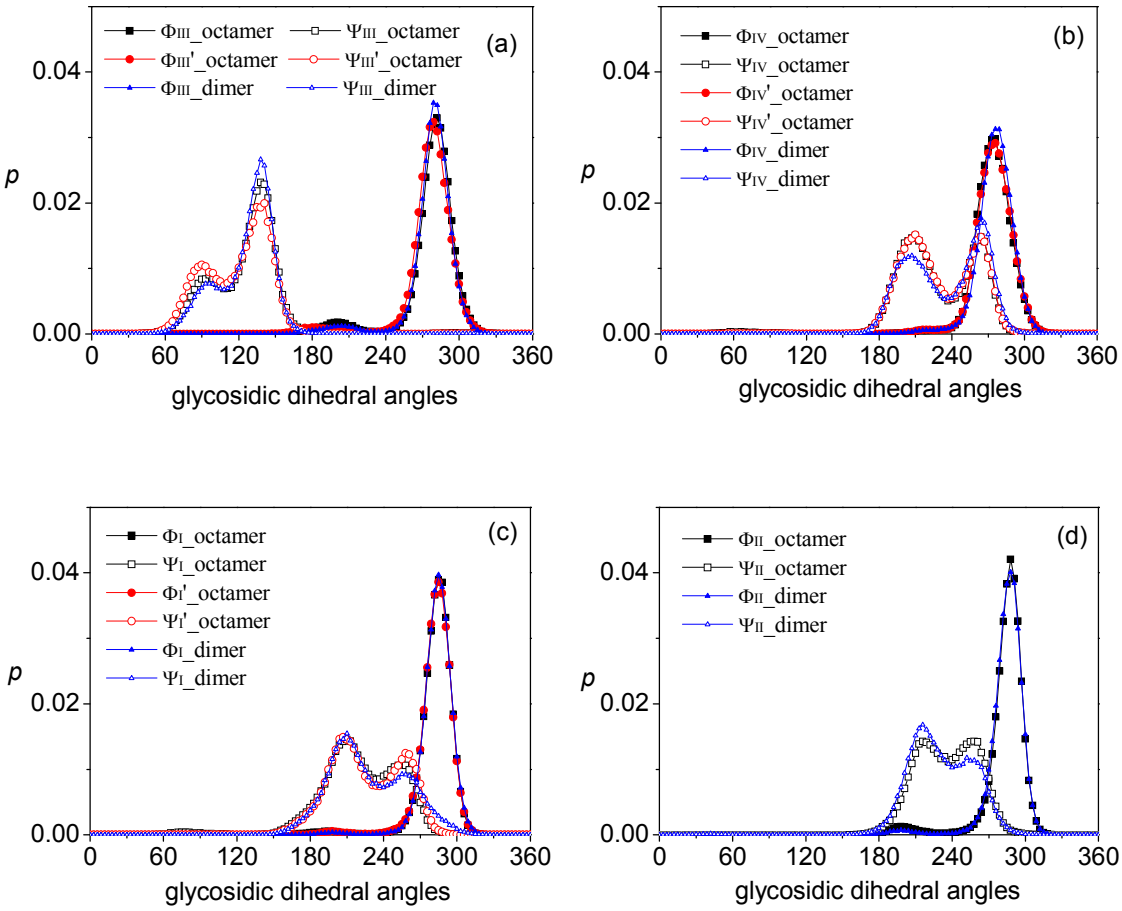

Figure S3. Populations of Φ-Ψ glycosidic angles of (a) Linkage III, (b) Linkage IV, (c) Linkage I and (d) Linkage II in an octasaccharide fragment CDABC'D'A'B' in explicit water compared with those in corresponding disaccharides (blue).

C. Simulations with implicit solvent.

To comprehensively compare the quality of implicit solvent models, we have acquired another series of free energy landscapes using generalized Born-models. Figure S4 displays the results of ABMD simulation of all relevant linkages in the gas phase, explicit and implicit solvent, projected onto single glycosidic angles. The discrepancy between the behavior in vacuum and that in a solvent (explicit/implicit) is obvious and expected. The possibility to form hydrogen bridges not screened

dramatically affects the population of the conformational energy minima and barrier heights. The differences between explicit and implicit solvation are small, 1kcal/mole at maximum, for most of the angular ranges much smaller than $k_B T$ (about 0.6kcal/mole).

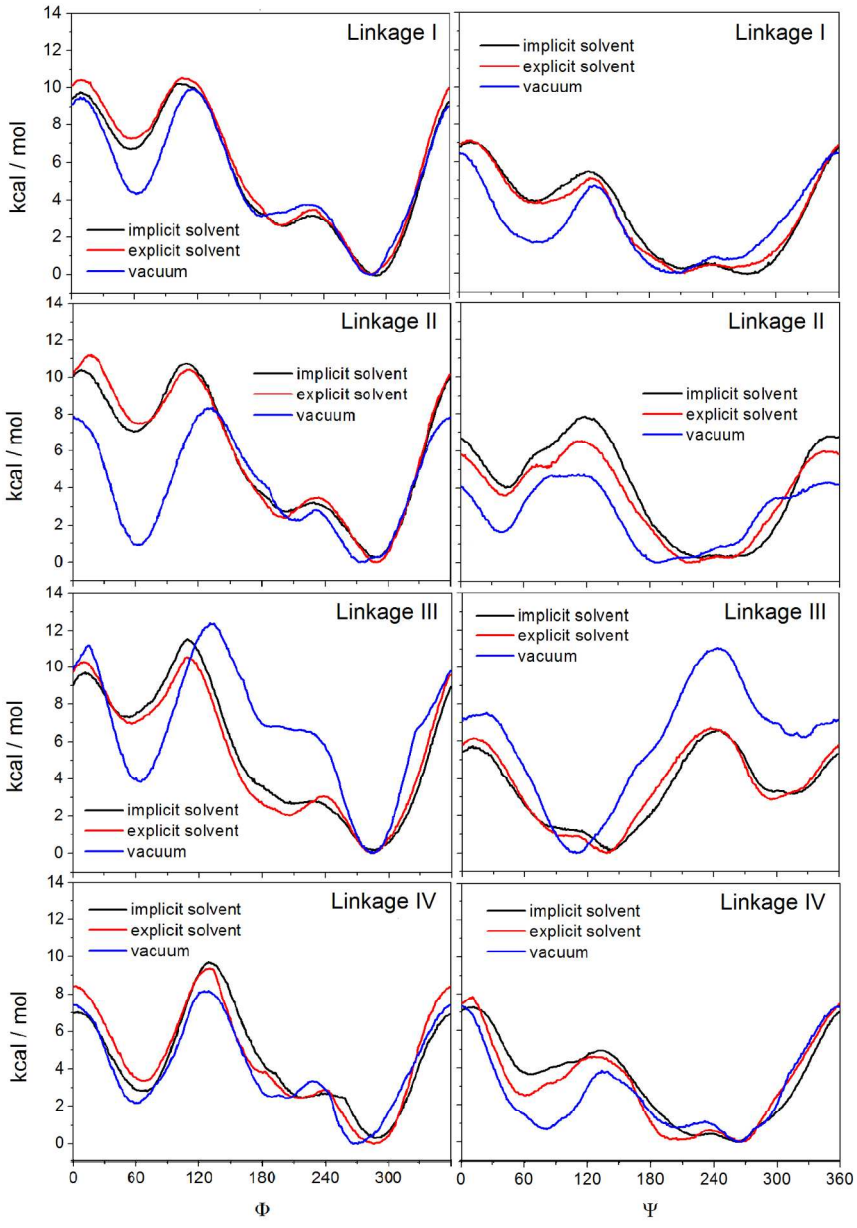

Figure S4. Free energies taken from ABMD simulations of disaccharides corresponding to four linkages as a function of Φ or Ψ with implicit (black) and explicit solvent (red) and in vacuum (blue).

A subtlety arises with respect to the stability of ring conformations in longer O-antigen fragments. It should be expected that ring flips (from $^4C_1$ to $^1C_4$ for GlcNAc and from $^1C_4$ to $^4C_1$ for Rha) are extremely rare [15], with an overwhelming preference for the most stable conformation $^4C_1$ for GlcNAc and $^1C_4$ for Rha. In the simulations with explicit solvent, this situation is sufficiently well represented: in a typical 200ns simulation, we may observe a flip on one saccharide ring once, with a duration of ~10 ps. Thus all simulation results can be safely interpreted in terms of the stable ring conformation. In all simulation runs using the CHARMM36 force field, we have never observed any ring flips. In simulations with the HCT or OBC implicit solvent model, ring flips can occasionally be observed during a simulation of 800 ns duration (GLYCAM), but in combination with the tendency of longer chains to assume compact conformations this leads to crumpled, trapped configurations. These conformations erroneously change the distribution of the end-to-end distance as compared to the results with explicit solvent. It is thus necessary to stabilize the expected dominant ring conformation found in solution. To achieve this, an external torsion restraint on each carbohydrate ring was applied to prevent ring flips, as shown in Table 4 (main text). In this way, good agreement with the explicit solvent results in (Φ, Ψ) and also the end-to-end distance distribution can be obtained.

In addition to free energy landscapes we have also investigated rather fine details such as the formation and occupancies of intra-molecular hydrogen bridges. The trends in occupancy for all hydrogen bridges observed across the series of di-, tetra- and octasaccharide fragments is fully reproduced in implicit solvent.

D. Reduced backbone model and simulation of 4RU O-antigen polysaccharides.

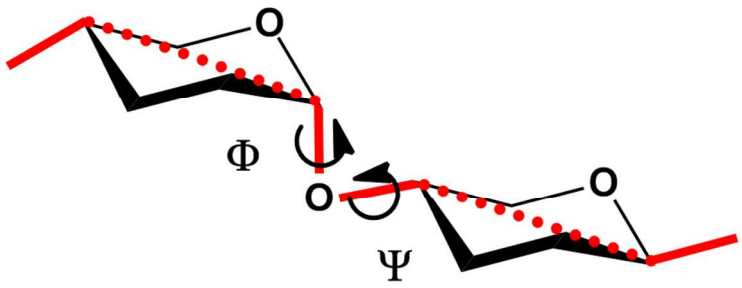

Figure S5: schematic drawing of the reduced backbone model studied with Monte Carlo simulations in this work.

A reduced model is obtained from the fully atomistic one in the way sketched in Figure **S5**. A linear polysaccharide is represented as a sequence of bonds, carbohydrate rings are assumed rigid and are replaced with bonds (dashed lines) around which no rotation is allowed. In comparison, the torsions around glycosidic linkages are soft, and are represented by rotatable bonds (solid lines). If we assume that the total free energy F_{tot} of the polysaccharide with N monomers projected onto the collection of glycosidic angles can be decomposed in additive contributions of its disaccharide fragments, we may write

$$F_{\mathbf{tot}}\left(\{\Phi,\Psi\}\right) \cong F_{\mathbf{add}}\left(\{\Phi,\Psi\}\right) = \sum_{i=1}^{N-1} F_{\mathbf{i,i+1}}\left(\Phi,\Psi\right), \tag{1}$$

where $F_{\mathbf{add}}$ is the approximation to $F_{\mathbf{tot}}$ and $F_{\mathbf{i,i+1}}$ is the free energy landscape of a disaccharide fragment composed of the consecutive rings **i** and **i+1**. For $F_{\mathbf{i,i+1}}$ we can directly use the landscapes shown in Figure 2 (main text), and the whole conformational space can be sampled by independent, consecutive Monte Carlo moves involving trial rotations around the glycosidic linkages, with an acceptance probability of $\min\{1, \exp(-\beta\Delta F)\}$, ΔF being the difference in free energy after single move. With a maximum step size of 30°, we typically arrive at an acceptance ratio between 25-35%. In the reduced backbone topology, we additionally keep a set of atoms so as to keep track of the ring orientations. We usually allow for 10^6-10^7 MC steps in order to arrive at stationary distribution functions in the relevant quantities.

The reduced models always produce smoothly decaying, unstructured tails in the end-to-end distributions for the molecules considered in this work, and in this way give strong support for intra-molecular interactions in case a distribution qualitatively differs from this behavior. Figure S6 again shows r_{ete} for the case of four repeat units in implicit solvent (Figure 9b, main text), where the deviations were most

pronounced, but now as function of time (same dataset as used for Figure 9). Figure S6 shows that frequent transitions between essentially one class of compact and one class of extended conformations take place.

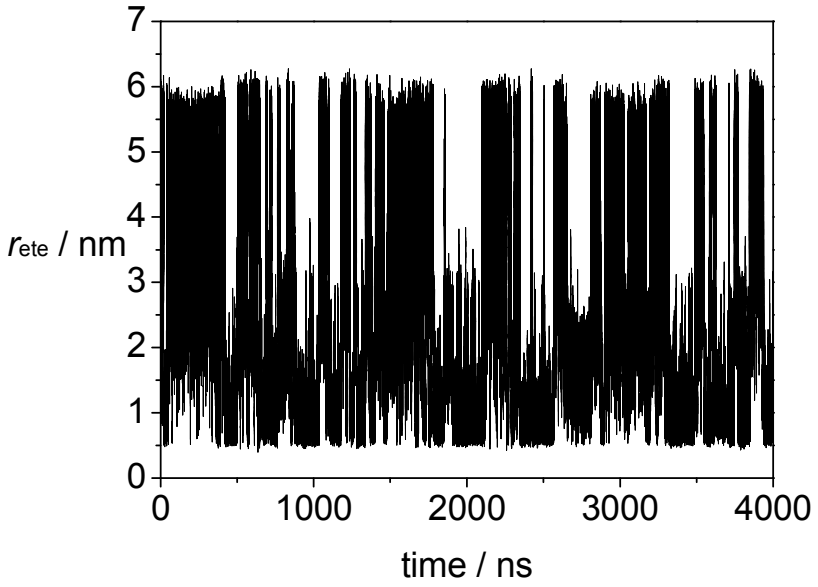

Figure S6. The end-to-end distance (r_{ete}) of the 4RU sequenced by CDAB as a function of simulation time in an implicit solvent MD trajectory.

1. Nosé, S. A Molecular Dynamics Method for Simulations in the Canonical Ensemble. *Mol. Phys.* **1984**, *52*, 255–268.
2. Hoover, W. G. Canonical Dynamics: Equilibrium Phase-Space Distributions. *Phys. Rev. A* **1985**, *31*, 1695–1697.
3. Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.
4. Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. A Computer Simulation Method for the Calculation of Equilibrium Constants for the Formation of Physical Clusters of Molecules: Application to Small Water Clusters. *J. Chem. Phys.* **1982**, *76*, 637–649.
5. 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.

6. Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Parametrized Models of Aqueous Free Energies of Solvation Based on Pairwise Descreening of Solute Atomic Charges from a Dielectric Medium. *J. Phys. Chem.* **1996**, *100*, 19824–19839.
7. Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Pairwise Solute Descreening of Solute Charges from a Dielectric Medium. *Chem. Phys. Lett.*, **1995**, *246*, 122–129.
8. Onufriev, A.; Bashford, D.; Case, D.A. Exploring Protein Native States and Large-Scale Conformational Changes with a Modified Generalized Born Model. *Proteins*, **2004**, *55*, 383–394.
9. Miyamoto, S.; Kollman, P. A. SETTLE: An Analytical Version of the SHAKE and RATTLE Algorithms for Rigid Water Models. *J. Comp. Chem.* **1992**, *13*, 952–962.
10. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem*. **1997**, *18*, 1463–1472.
11. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.
12. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33-38.
13. Babin, V.; Sagui, C. Conformational Free Energies of Methyl-α-L-Iduronic and Methyl-β-D-Glucuronic Acids in Water. *J. Chem. Phys.* **2010**, *132*, 104108.
14. Wehle, M.; Vilotijevic, I.; Lipowsky, R.; Seeberger, P. H.; Silva, D. V.; Santer, M. Mechanical Compressibility of the Glycosylphosphatidylinositol (GPI) Anchor Backbone by Independent Glycosidic Linkages, *J. Am. Chem. Soc.* **2012**, *134*, 18964-18972.
15. Widmalm, G., personal communication. **2013**