Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Supporting information:

A Molecular Dynamics Model for Glycosylphosphatidyl-Inositol Anchors: "flop down" or "lollipop"?

Pallavi Banerjee, † Marko Wehle,† Reinhard Lipowsky,† and Mark Santer†,*

†Max Planck Institute of Colloids and Interfaces, 14424 Potsdam, Germany

S1. Parametrization of the hybrid building block

In Figure S1 (a) we reproduce the link of Figure 3(a) (main text) 6OMe-Ino-PGL(0) (Ino-PGL(0)). Its parametrization is carried out in two steps that mainly involve the reduced fragment Figure S1(b), Ino-POMe and 6OMe-Ino-OMe. (c) shows these molecules with the variables for bonded interactions that involve a mixed set of atom types to be determined in section S1.1. In a second step partial charges suitable for simulations in solution will be computed (section S1.2). Different protocols are tested and it is shown that especially partial charges on the phosphate group vary only mildly; this observation is then exploited in section S1.3 in order to define a hybrid set of charges for 6OMe-Ino-PGL(0) and Ino-PGL(0). The "work horse" compounds 6OMe-Ino-POMe and Ino-POMe have been chosen as a compromise between the somewhat large and sloppy compound S1(a) (which is inconvenient to use with ensemble averaging,see S1.2) and much smaller molecular fragments (see Figure S4 (a)-(c)) that do not capture the presence of the inositol ring. On the other hand, the phosphate group needs to be included because of the

influence on electron density via the connecting O32 oxygen.

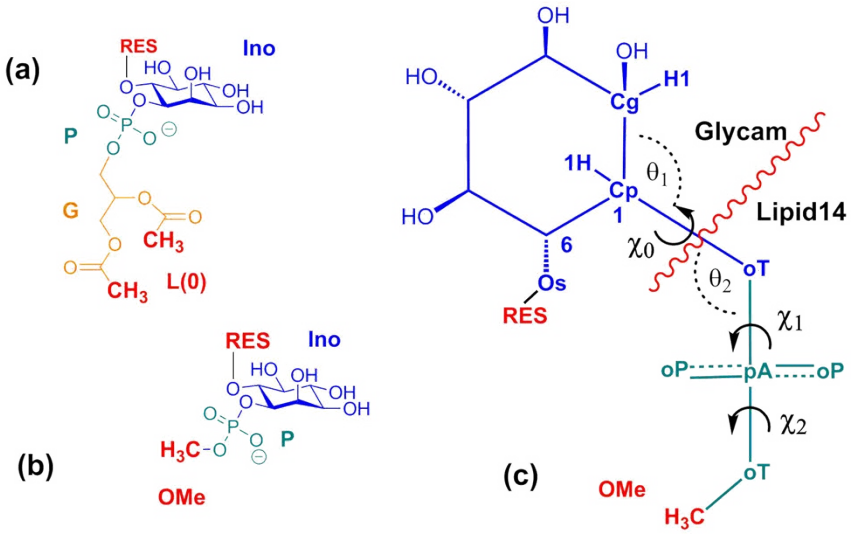


Figure S1. (a) Hybrid link reproduced from Fig. 3(a) of the main text, with RES=CH3 (6OMe-Ino-PGL(0) or RES=H (Ino-PGL(0)), and (b) reproduced from Figure3(b) with RES=CH3 (6OMe-Ino-POMe) or RES=H (Ino-POMe). (c) schematic version of (b) with atom names and types.

S1.1 Adapting bonded interactions.

Figure S1(c) shows the hybrid topology in a schematic way. The dividing line between the domains of atom types runs through the bond Cp-oT. The highlighted atoms Cg, Cp, H1 and Os belong to the GLYCAM definitions, and oT, pA and oP to Lipid14. The two torsions angles, χ_0 and χ_1 involve a mixed set of atom types, the subsequent χ_2 can be encoded fully within Lipid14, but will be considered in the analysis for comparison. In principle, each set of parameters that describes bonded interactions and involves a mixed set of atom types would require either a redefinition/adaption or a choice for one of the force fields (formally, the relevant GLYCAM06 atom types are included in the force field definition of Lipid14). In the latter case the sequence Cg-Cp-oT-pA, for instance, contributing to the torsion potential in χ_0, could be translated into either Cg-Cp-Os-P or

cA-cA-oT-pA with the corresponding set of parameters and 1-4 scaling. cA represents a general sp3-bonded carbon atom in Lipid14 and is used in the models of phosphocholine (PC) and phosphoethanolamine (PE). Its counterpart in GLYCAM06, Cp, is a sp3 carbon linked to an oxygen linked to a phosphorous atom. In section S2 we show with a series of smaller fragments up to and including Ino-POMe, parametrized with *either* GLYCAM06 *or* Lipid14 parameters (Figures S4, S5) that the distinction of the carbon type actually matters, see also Figure S6.

For this reason, we shall define the torsion parameters independently, directly with respect to mixed sequences. Furthermore, while types Cg, Cp and H1 and their Lipid14 counterparts share the same LJ parameters, those of Os and O2, corresponding to oT and oP in Figure S1(c) are slightly different, compare also Table S1. The torsion parameters for χ_0 (H1-Cp-oT-pA, Cg-Cp-oT-pA) and χ_1 (Cp-oT-pA-oP, Cp-oT-pA-oT) are then determined by quantum mechanical (QM) calculations on Ino-POMe, in order to provide an explicit molecular context. We chose to derive the torsion parameters with respect to GLYCAM06 conventions, using a 1.0/1.0 scaling. Note that we exclude sequences such as H1-Cg-Cp-oT that affect only the inositol ring, which is known to assume a rather rigid and stable chair conformation and ring torsions do not require adaption.

The fitting of the torsion parameters involves matching the Energy $E^{MM}(\chi)$ ($\chi=\chi_0,\chi_1$) of the molecular mechanics (MM) or force field model to the corresponding QM-calculated energies $E^{QM}(\chi)$. The QM scan defining the torsion potentials was carried out as follows. In Ino-POMe χ_0 and χ_1 were varied in independent scans (the respective other angle fixed (in its trans conformation) from 0° to 360° in steps of 10°. After each torsion increment (carried out rigidly), the structure was relaxed constraining the current value of the

dihedral angle; possible hydrogen bridges towards the oxygens of the phosphate group were allowed to form. In this way, a "minimum energy" like pathway $E^{QM}(\chi)$ in the gas phase is generated. The QM theory level used in the optimization and final single point (SP) energy calculation on each set of 36 geometries was B3LYP//6-31++(2p,2d), all QM calculations in this work are performed with Gaussian03.[1] For the evaluation of $E^{MM}(\chi)$, gas phase charges were computed separately for each set of 36 geometries according to GLYCAM06 conventions (B3LYP/cc-pVTZ, and a single stage RESP fit), and averaged to provide a single charge set. The resulting energy difference $\Delta E\ (\chi) = E^{QM}(\chi) - E^{MM}(\chi)$ is then subject to minimization by fitting the MM parameters defining the torsions χ_0 and χ_1 using ParamFit.[2] Although less important for the present purposes, also the bending angles θ_1, θ_2 (which contain a mixed set of atom types as well) were included in the fit.

Table S1: Atom types involved in the transition from GLYCAM06 to Lipid14. The numerical values in the GLYCAM06 domain mostly involve parameters from the PARM99 set, Lipid14 parameters partly originate from the general amber force field (GAFF).[3]

Force field		**Atom Name**	**Atom Type**	**σ[Å]**	**ε[kca/mole]**
GLYCAM		**Cg**	sp3 C aliphatic	1.9080	0.1094
		Cp	sp3 C aliphatic - carbon atom bonded to an oxygen atom bonded to a phosphorus atom	1.9080	0.1094
		H1	H aliph. bond. to C with 1 electrwd. Groups	1.3870	0.0157
		Os	Ether oxygen	1.6837	0.1700
		O2	Carboxy oxygen	1.6612	0.2100
		P	Phosphorous	2.1000	0.2000
LIPID14		**oT/oS**	sp3 oxygen bonded to carbon in phosphate group (GAFF os-) / sp3 oxygen in ethers and esters	1.6500	0.1200
		oP	sp2 oxygen with one connected atom (e.g P-O) in phosphate group (GAFF o -)	1.6500	0.1400
		pA	phosphorus with four connected atoms, such as O=P(OH)3 (GAFF p5-)	2.1000	0.2000
		cP	Parameters of type Cp in GLYCAM		
		cA	sp3 carbon (GAFF c3 head, glycerol)	1.9080	0.1094
		hE	H bonded to aliphatic carbon with 1 electron-withdrawing group	1.3870	0.0157

To provide a starting set for the numerical values of all MM parameters to be adjusted, we borrowed from GLYCAM06. The final values were obtained iteratively, first fitting only the torsion parameters to capture the overall variation of the torsion potential profile, whereby χ_0 and χ_1 were varied in independent scans. In a second iteration, the equilibrium values of the bond length Cg-pA as well as those of θ_1 and θ_2 were allowed to vary. In a third step, the latter were kept fixed again and a subsequent fit of torsion parameters did not yield notable changes of their values. The final parameter set is summarized in Table S2.

Table S2. Parameters for bonded interaction within the transition region. The 1-4 interactions w.r.t. torsions receive a 1.0/1.0 scaling (SCNB=SCEE=1.0). PN is the periodicity of the torsion potential and PK the corresponding barrier height, and IDIVF the torsional barrier coefficient divided by 2.

	GLYCAM06		LIPID14				
Bond	**Atoms**		**Atoms**	**RK kcal/mol/(Å²)**	**REQ [Å]**		
	Cp	**—**	**oT**	285,00	1,3848		
Angle	**Atoms**		**Atoms**	**TK kcal/mol/ (rad²)**	**TEQ [°]**		
	H1 — Cp	**—**	**oT**	60,00	113,9240		
	Cg — Cp	**—**	**oT**	70,00	110,6108		
	Cp	**—**	**oT — pA**	50,00	121,4630		
Dihedral	**Atoms**		**Atoms**	**IDIVF**	**PK**	**Phase [°]**	**PN**
	H1 — Cp	**—**	**oT — pA**	1,00	0,0350	0,0000	3,0000
	Cg — Cp	**—**	**oT — pA**	1,00	0,2382	0,0000	3,0000
				1,00	0,5494	180,0000	2,0000
				1,00	-0,8808	0,0000	1,0000
	Cp	**—**	**oT — pA — oT**	1,00	1,0533	180,0000	1,0000
				1,00	1,0879	0,0000	2,0000
				1,00	-1,2340	0,0000	3,0000
	Cp	**—**	**oT— pA — oP**	1,00	0,9006	180,0000	1,0000
				1,00	-0,1392	0,0000	2,0000
				1,00	-0,8218	180,0000	3,0000

S1.2 Partial charges on 6OMe-Ino-POMe and InoP-OMe.

Now we shall determine partial charges for the molecular topology in Figure S1(c), suitable for simulations in solution. Since Lipid14 and GLYCAM06 make use of different charge schemes, we shall briefly review the general protocol how partial atomic charges are determined for most force fields in the Amber family.

In general charges q_i on atoms i =1...N are chosen such that the electrostatic potential (ESP) they produce around a molecule matches the molecular electrostatic potential (MEP) of the continuous charge distribution as closely as possible. The MEP is inferred from a QM calculation and evaluated on a dense grid of points contained in a shell a few Å thick off the Van-der-Waals (VdW) surface of a molecule. In the present work, the grid is created with the CHELPG scheme.[4] The theory level for computing the charge density in order to determine partial charges for solution molecular dynamics with the TIP3P water model is HF/6-31G*.

Bayly et al.[5] have established the widely used "restrained electrostatic potential" (RESP) fitting procedure involving the minimization of a figure-of-merit function of the form[6,7]

$$f(q_1,\ldots,q_N) = \chi^2_{esp} + \chi^2_{rstr} + \sum_k \lambda_k g_k(q_1,\ldots,q_N) \quad (1),$$

with

$$\chi^2_{esp} \equiv \sum_i (V_i - \dot{V}_i)^2 \quad \text{and} \quad \dot{V}_i \equiv \sum_j \frac{q_j}{r_{ij}} \quad (2),$$

where V_i is the electrostatic potential in atomic units at positions $\vec{r}_i$ produced by a set of point charges q_j at positions $\vec{r}_j$, which in the simplest case are just the position of the atomic nuclei of a given molecule, and $r_{ij} = |\vec{r}_i - \vec{r}_j|$. The $\vec{r}_i$ are points of the MEP grid with values V_i that are inferred from the continuous charge distribution obtained from the QM calculation. χ^2_{rstr} is a collection of hyperbolic functions of the form

$$\chi^2_{rstr} = a_{wt} \sum_{j=1}^{natoms} \left(\left(q_j^2 - b^2 \right)^{1/2} - b \right) \quad (3),$$

for restraining each charge to a specified target value (usually 0), avoiding an uncontrolled growth during the fit. The strength of the restraints scales with the common weight a_{wt}, the tightness of the hyperbola is controlled by the parameter b.

The deviation of the potential $\dot{V}_i$ from the quantum mechanically derived V_i can, for instance, be characterized by the relative root mean square error

$$\mathrm{RRMS} \left\{ \chi^2_{esp} \Big/ \sum_i V_i^2 \right\}^{1/2} \quad (4).$$

The λ_k in (1) are Lagrangian multipliers and the functions $g_k(q_1,\ldots,q_N)$ describe additional constraints that can be used to equalize certain charges or to specify a fixed net charge of a group of atoms. In Lipid14, use is made of the conventional two-stage Amber protocol for charges in solvent, with a restraint weight of a_{wt}=0.0005 for stage one without restraints other than (3); and a_{wt}=0.001 in stage two during which charges on equivalent aliphatic hydrogens are forced to be equal.

The corresponding GLYCAM06 protocol involves only one stage, with a_{wt}=0.01, and aliphatic hydrogens constrained to zero. To obtain charges suitable for simulations in

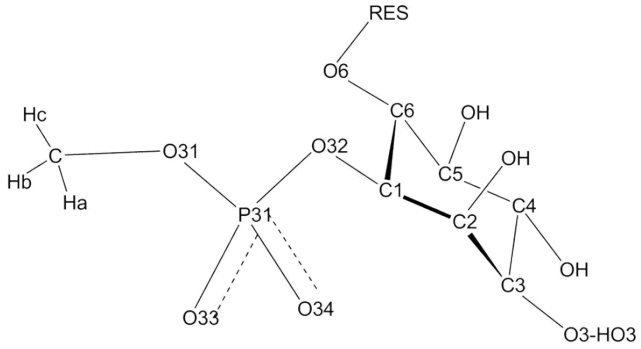
ambient conditions including solvent an average over a conformational ensemble is performed. Especially for molecules that can assume a number of conformations (such as the relative orientations of exocyclic torsions in carbohydrate rings), not only a single reference geometry (e.g. a minimum energy configuration) should enter the charge fit but deviations of the reference configuration of equal or slightly elevated energy should be considered, to take polarization effects due to conformational changes into account or to avoid the accidental overrepresentation of some inconvenient structure.[8] In the present work, we perform ensemble averaging (EA) as described by Woods and co-workers.[9] Given an initial charge set, an ensemble set of 200 molecular configurations is selected, equidistant in time, from a 50ns long MD run in TIP3P water at T=300K. For each snapshot, the QM-MEP is computed after minimizing the configuration inferred from the MD trajectory with respect to angles and bonds (torsions are kept fixed). The final charge set is then chosen as the arithmetic average of the 200 single sets.

In the present work, we were using a stand-alone version of the RESP-program v. 2.4.[10] In Table S3 we show the partial charges resulting from EA for 6OMe-Ino-POMe and Ino-POMe in column A and B, respectively, in comparison to a semi-empirical AM1-BCC calculation[11] for Ino-POMe (column C). The latter has been performed for a single, all-*trans* relaxed configuration of the molecule. The AM1-BCC scheme (facilitated by the tLeaP program shipped with Amber 15) here has been chosen for convenience, to quickly produce charge sets similar to the explicit Amber 2-stage procedure and distinct from ensemble averaging (the magnitudes of charges in AM1-BCC is somewhat lower).

Table S3: Partial charges on 6OMe-Ino-POMe/Ino-POMe according to the ensemble averaging (EA) and the AM1-BCC protocol, all charges reported in units of the elementary charge e. For atom labeling see scheme S1. In the charge set **A** the charge of the methyl group connected to O6 has been constrained to +0.194e.

Atom name	**Residue**	**A** 6-OMe-Ino-POMe **(EA)**	**B** Ino-POMe **(EA)**	**C** Ino-POMe **(AM1-BCC)**
C		0.20744	0.21277	0.1597
Ha		0.00000	0.00000	0.0174
Hb	Me	0.00000	0.00000	0.0174
Hc		0.00000	0.00000	0.0174
O33		-0.82500	-0.82558	-0.8245
O34		-0.82500	-0.82558	-0.8245
O32	Phosphate (P)	-0.45505	-0.54725	-0.5682
O31		-0.53539	-0.54091	-0.5632
P31		1.35977	1.39702	1.4198
C1		0.04740	0.12678	0.1941
C2		0.39906	0.34628	0.1151
C3		0.18631	0.19240	0.1081
C4		0.39383	0.32997	0.1101
C5		0.16520	0.19240	0.1086
C6		0.37216	0.34628	0.1151
O2		-0.76881	-0.70321	-0.5933
O3		-0.72413	-0.69691	-0.6083
O4		-0.74790	-0.69683	-0.6218
O5		-0.72379	-0.69691	-0.6083
O6	Inositol (Ino)	-0.46519	-0.70321	-0.5933
H1		0.00000	0.00000	0.0637
H2		0.00000	0.00000	0.0667
H3		0.00000	0.00000	0.0597
H4		0.00000	0.00000	0.0657
H5		0.00000	0.00000	0.0597
H6		0.00000	0.00000	0.0667
HO2		0.44977	0.42535	0.4090
HO3		0.43484	0.42242	0.4035
HO4		0.42712	0.39692	0.4130
HO5		0.43333	0.42242	0.4035
HO6			0.42535	0.4090

C	RES	0.19400		
H1		0.00000		
H2		0.00000		
H3		0.00000		



Scheme S1: Atom names in Ino-POMe (RES=H) and 6-OMe-Ino-POMe (RES=CH_3), for reporting the partial charges in Table S3. At the ring carbon C3 the labeling of hydroxy atoms is indicated (at C3, we have O3 and HO3 etc.).

S1.3 Joining GLYCAM06 and Lipid14 charges.

In order to construct the desired hub 6OMe-Ino-PGL(0) (Figure 3(a) or S1(a)), to which both, GLYCAM06 (glycan moieties) and Lipid14 components (fatty acid chains) can be attached, we consider the molecular species displayed in Figure S2. Scheme (c) depicts 6OMe-Ino-PGL(0). The carboxyl groups are terminated by methyl cappings carrying zero charge, the O6 oxygen has been amended by a methyl cap (replacing HO6) the net charge of which is constrained to +0.194e in order to comply with the GLYCAM building block principle (see also Table S3). Ideally, we would like to assign GLYCAM06-type charges to the inositol head, while retaining Lipid14-type charges on the glycero-lipid part (to make it compatible with the other lipid species). We now demonstrate that this is feasible. The structure in (a) is 6OMe-Ino-POMe, with net charges on phosphate and methyl caps annotated (inferred from Table S3 (A)). In (b), the

phosophcholine (PC) and phosphoethanolamine (PE) lipid head group building blocks[12] are shown, including glycerol, carboxy- and terminating methyl groups. Net charges on corresponding groups are indicated in the same way.

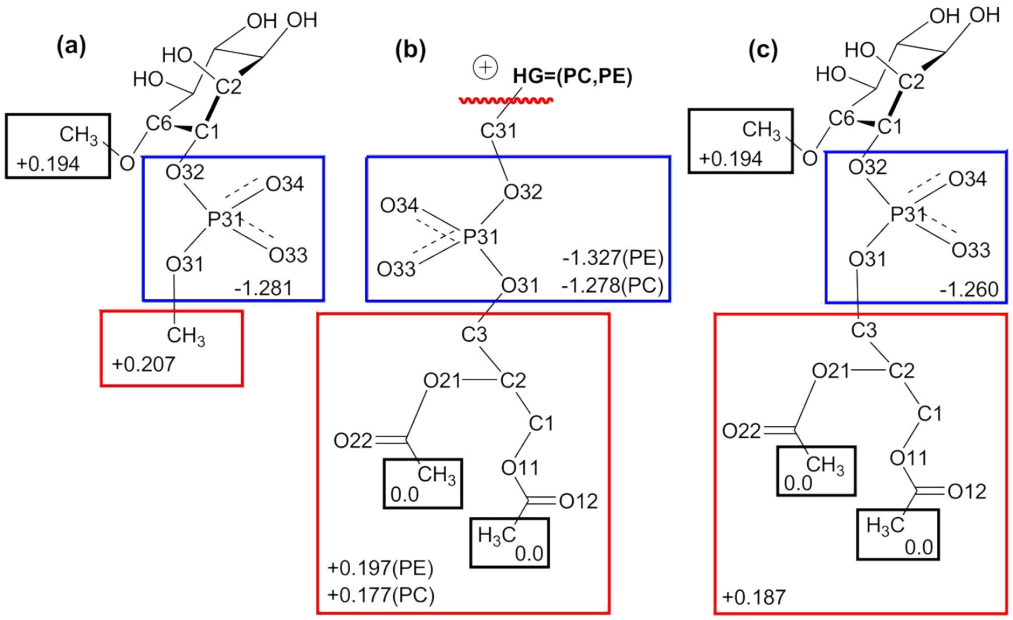


Figure S2. Chemical compounds used for allocation of partial charges on the hybrid link 6OMe-Ino-PGL(0), depicted in (c). (a) 6-OMe-Ino-POMe (b) Phosphocholine (PC-) and Phospho-Ethanolamine (PE)-glycerol. Black solid frames indicate capping groups: the CH_3 moiety at the O6 oxygen represents the link to the GPI-anchor backbone. The methyl groups within the glycerol part of (b) are capping groups in Lipid14 and carry zero net charge. (a) and (c) carry a net charge of -1, for (b) this value is 0.

To arrive at an adequate and plausible charges for (c), we first set those for the glycerol atoms (red frame) as the average of those of the Lipid14 PC and PE head groups in (b); the numerical values of the two a rather close, compare Table S4. We then note the close similarity of net charges on the respective phosphate groups in (a) and (b) (blue frames) and also between those of CH_3 and the trailing glycerol (red frames). In (c) we therefore tentatively equip the glycerol part with the average charges of the PE and PC heads (column A and B in Table 3), and inositol with those from column D (from Table

S3 col. A). The charges on the phosphate group on (c) are chosen as the arithmetic mean of the ensemble average in column D, and the average from PE and PC (Lipid14) in column C; this mean is listed in column E, and with all other charges would result in a net charge of -0.9687e on the whole building block. If we spread the (small) difference ΔQ = -0.0313e evenly across all 5 atoms of the phosphate group, we arrive at the final hybrid charge set in column F.

If the glycan head is just inositol, we repeat the procedure above but with Ino-POMe (charges listed in Table S3 col. B) instead of 6OMe-Ino-POMe, thereby creating Ino-PGL(0) as a separate building block. Here, the residual ΔQ was as low as +0.00579e. The procedure described above is of course not unique, but the results suggests that the partial charges on the head group and those on the glycerol moiety can be decoupled to some extent, if we allow the small residual charge difference ΔQ to be absorbed by the phosphate group.

Table S4. Composition of of the set of partial charges for the hybrid link 6OMe-Ino-PGL(0) (charges in units of e) (a) in Figure S1. Column A and B contain the charges on the glycerol and phosphate part of the PE and PC head groups as provided with Lipid14; C contains their average. Column E is the average of C and D; ΔQ is the residual charge difference when the values of the phosphate group (P) are replaced by those in E.

		A	B	C	D	E	F
	Name	**PE**	**PC**	**Av (A,B)**	**6OMe-Ino-POMe**	**Av(D,C)**	**6OMe -Ino-PGL(0)**
Glycerol	C11	0.79180	0.78360	0.78770			0.78770
	O12	-0.60080	-0.59970	-0.60025			-0.60025
	O11	-0.45030	-0.45500	-0.45265			-0.45265
	O21	-0.43160	-0.41950	-0.42555			-0.42555
	C21	0.77890	0.77040	0.77465			0.77465
	O22	-0.58860	-0.59690	-0.59275			-0.59275
	C1	-0.01030	0.01590	0.00280			0.00280
	HR	0.11800	0.11980	0.11890			0.11890
	HS	0.11800	0.11980	0.11890			0.11890
	C2	0.09410	0.09140	0.09275			0.09275
	HX	0.14590	0.14260	0.14425			0.14425
	C3	0.06460	0.01050	0.03755			0.03755
	HA	0.08370	0.09710	0.09040			0.09040
	HB	0.08370	0.09710	0.09040			0.09040
Phosphate	O31	-0.42260	-0.42240	-0.42250	-0.53539	-0.47894	-0.47269
	P31	1.11540	1.14990	1.13265	1.35977	1.24621	1.25247
	O32	-0.48700	-0.44420	-0.46560	-0.45505	-0.46033	-0.45407
	O33	-0.76640	-0.78070	-0.77355	-0.82500	-0.79927	-0.79302
	O34	-0.76640	-0.78070	-0.77355	-0.82500	-0.79927	-0.79302
CH_3 (O31)					0.20744		---
CH_3 (O6)					0.19400		0.194
ΔQ						0.03129	
Inositol	C1				0.04740		0.04740
	C2				0.39906		0.39906
	C3				0.18631		0.18631
	C4				0.39383		0.39383
	C5				0.16520		0.16520
	C6				0.37216		0.37216
	O2				-0.76881		-0.76881
	O3				-0.72413		-0.72413
	O4				-0.74790		-0.74790
	O5				-0.72370		-0.72370
	O6				-0.46519		-0.46519
	HO2				0.44977		0.44977
	HO3				0.43484		0.43484
	HO4				0.42712		0.42712
	HO5				0.43333		0.43333

S2. Torsions χ_0-χ_2 influenced by intramolecular interactions.

In this section, we shall elucidate the origin and the validitiy of the preferences for the torsion angles χ_0, χ_1 and χ_2 that have been used as an argument for selecting conformations of GPI-core-PGL that should occur with high probability. We will accomplish this in several steps employing Ino-POMe as reference compound. In Figure S3 we compare distributions functions $P(\chi_0)$ to $P(\chi_2)$ inferred from 100ns MD runs in TIP3P water at 303K for two cases: in addition to the re-derived parameters obtained directly with Ino-POMe, we also consider its version with all-GLYCAM06 parameters taken from Tessier et al.[13] Furthermore, two different charge sets are employed listed in column B and C of Table S3 (GLYCAM06, ensemble averaged charges vs. AM1-BCC). We make the observation that whereas the variations in partial charges does not have a notable impact, the bonded interaction parameters from the explicit molecular context of Ino-POMe lead to slight shifts and broadening of the profiles for χ_0 and χ_1. In both cases some general asymmetry is to be noticed, the emphasis for χ_0 with respect to the peak close to the *trans*-, and for χ_1 the peak in +*syn* configuration. The same holds for χ_2; note that in this case the appearance of the torsion profile is different because the torsion is defined entirely in terms of Lipid14 atom types. In what follows we shall argue that the above mentioned asymmetry originates from intramolecular interactions and is thus a generic, robust phenomenon.

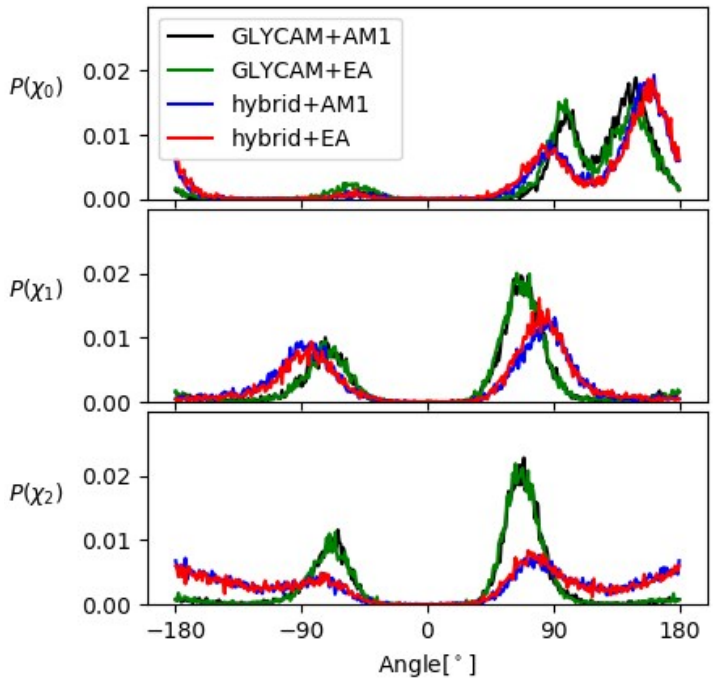


Figure S3: Density distributions $P(\chi_0)$ to $P(\chi_2)$ for Ino-POMe (structure (e) in Figure S1). Black and green graphs correspond to bonded interactions taken entirely from the GLYCAM06 force field, blue and red employ the rederived parameters of this work (Table S2). Simulation time is 100ns at 303K, in TIP3P water.

To do this we track how torsion profiles evolve in a sequence of smaller molecular fragments, see Figure S4, starting di-methyl-phosphatidyl (a) in which one methyl group is successively substituted by an ethyl (b), propyl (c) and cyclohexyl (d) group, (e) corresponds to Ino-POMe. For convenience, AM1-BCC partial charges are chosen throughout, reported in Table S5. The emphasis of χ_0 on positive values emerges with the propyl group (c), partially because of the steric interaction of the two methyl groups with the phosphate oxygens. The transition from cyclohexyl (d) towards inositol (e) is accompanied by a splitting of the broad peak, because now intramolecular interactions with the hydroxygroups are possible, and the OH group at C6 is equatorially, but the one at C2 axially oriented.

Table S5: AM1-BCC partial charges q for fragments (a)-(e) in Figure S4,S5. Labeling of atom names is as in Scheme S1, the carbon towards the head group bonded to the phosphate group is always designated as C1, the trailing methyl carbon as C. Charges are quoted in units of the elementary charge e. The charge set in column (e) and that in col. C of Table S3 have been computed w.r.t. the same molecular configuration, and are thus identical.

Methyl (a)		Ethyl (b)		Propyl (c)		Cyclohexyl (d)		Inositol (e)	
Atom name	q	Atom name	q	Atom name	q	Atom name	q	Atom name	q
C	0.1837	C	0.1827	C	0.1837	C	0.1827	C	0.1597
Ha	0.0037	Ha	0.0037	Ha	0.0034	Ha	0.0040	Ha	0.0174
Hb	0.0037	Hb	0.0037	Hb	0.0034	Hb	0.0040	Hb	0.0174
Hc	0.0037	Hc	0.0037	Hc	0.0034	Hc	0.0040	Hc	0.0174
O31	-0.5802	O31	-0.5822	O31	-0.5812	O31	-0.5812	O31	-0.5632
P31	1.4228	P31	1.4188	P31	1.4208	P31	1.4208	P31	1.4198
O32	-0.5802	O32	-0.5732	O32	-0.5752	O32	-0.5732	O32	-0.5682
O33	-0.8245	O33	-0.8260	O33	-0.8235	O33	-0.8225	O33	-0.8245
O34	-0.8245	O34	-0.8260	O34	-0.8235	O34	-0.8225	O34	-0.8245
C1	0.1837	C1	0.1704	C1	0.1931	C1	0.2021	C1	0.1941
H1	0.0037	H1	0.0272	H1	0.0597	H1	0.0617	H1	0.0637
H1a	0.0037	H1a	0.0272	C2	-0.1036	C2	-0.0954	C2	0.1151
H1b	0.0037	C2	-0.0911	H2	0.0239	H2	0.0442	H2	0.0667
		H2	0.0204	H2a	0.0239	H2a	0.0442	O2	-0.5933
		H2a	0.0204	H2b	0.0239	C3	-0.0719	HO2	0.4090
		H2b	0.0204	C3	-0.1036	H3	0.0270	C3	0.1086
				H3	0.0239	H3a	0.0270	H3	0.0597
				H3a	0.0239	C4	-0.0724	O3	-0.6083
				H3b	0.0239	H4	0.0217	HO3	0.4035
						H4a	0.0217	C4	0.1101
						C5	-0.0719	H4	0.0657
						H5	0.0270	O4	-0.6218
						H5a	0.0270	HO4	0.4130
						C6	-0.0954	C5	0.1086
						H6	0.0442	H5	0.0597
						H6a	0.0442	O5	-0.6083
								HO5	0.4035
								C6	0.1151
								H6	0.0667
								O6	-0.5933
								HO6	0.4090

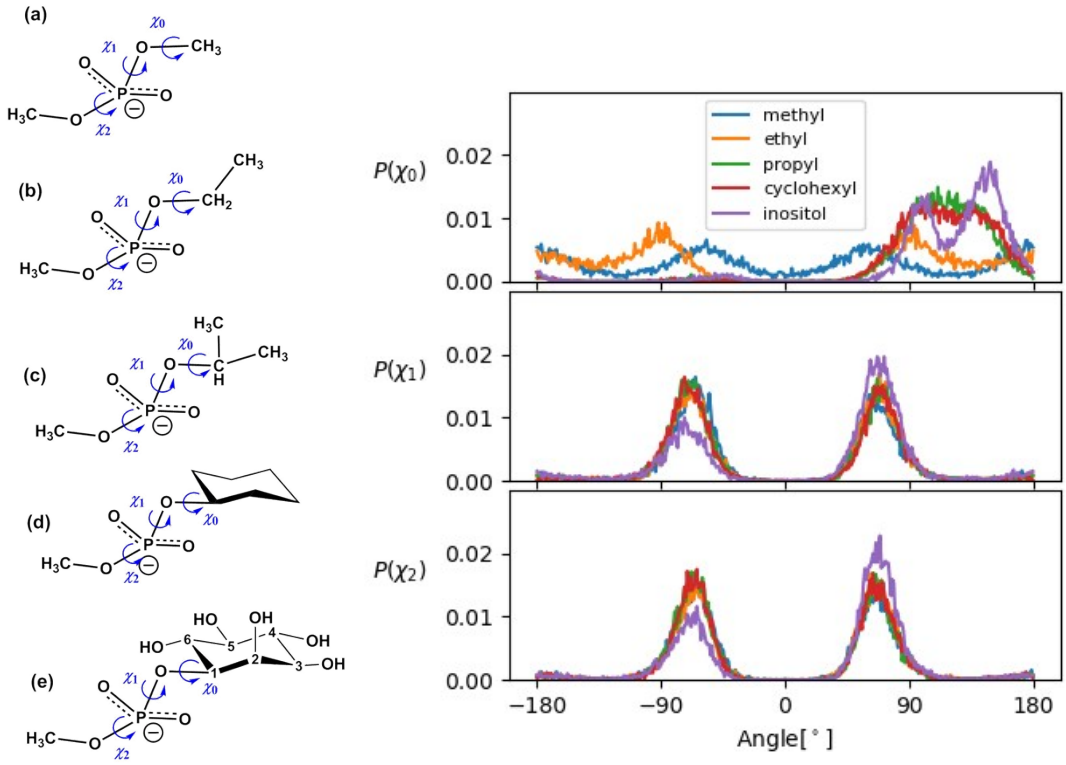


Figure S4: Left: phosphatidyl moieties with methyl capping and (a) methyl, (b) ethyl, (c) propyl, (d) cyclohexyl and (e) inositol head. The graphs to the right show the corresponding distribution of the torsion angles χ_0-χ_2 obtained from 100ns simulations at 303K in TIP3P water using the GLYCAM06 parameter set.

This steric imbalance also leads to the asymmetry in χ_1 in going from cyclohexyl to inositol. In Figure S5, we show the analogous results employing Lipid14 parameters for bonded interactions, using the same charge sets. In comparison to the GLYCAM06 parameters, we note a few things. The preferences for χ_0 are quite similar although the splitting is already emerges with substituting cyclohexyl (d). The behaviour for χ_1 and χ_2 is at variance to the GLYCAM06 case in that a lot of weight is given to *trans*-like (*anti-periplanar*) configurations around ±180°. As indicated above, this is explained by the different atom types used for the carbons that connect to the phosphate oxygens: in Lipid14, type cA is chosen (generic sp3 bonded carbon) and in GLYCAM06, this more specific Cp is assigned to account for the proximity of the phosphate group.[13]

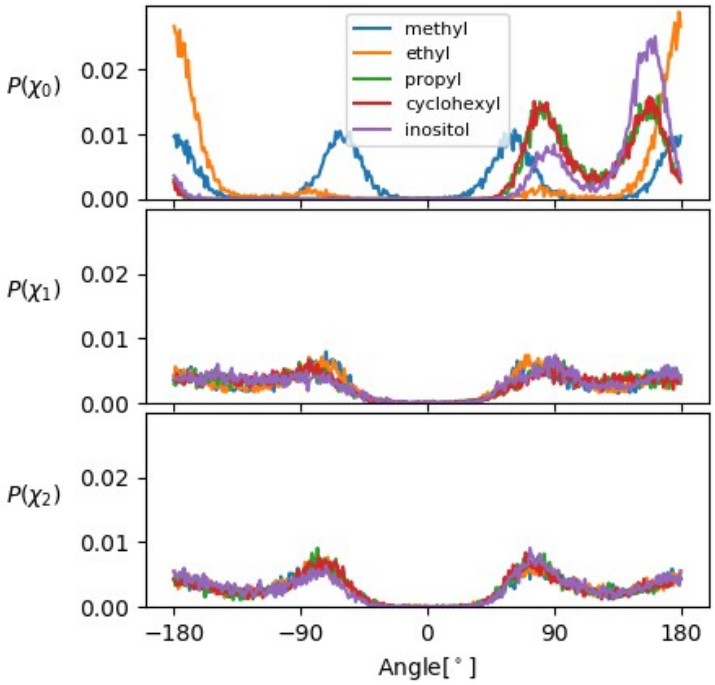


Figure S5: Torsion angle distribution of species (a) to (e) (see Figure S4) employing Lipid14 force field parameters.

We can make the distinction between GLYCAM06 and Lipid14 parameters more obvious by exchanging the type for C1 in the Lipid14 topology, see Figure S6.

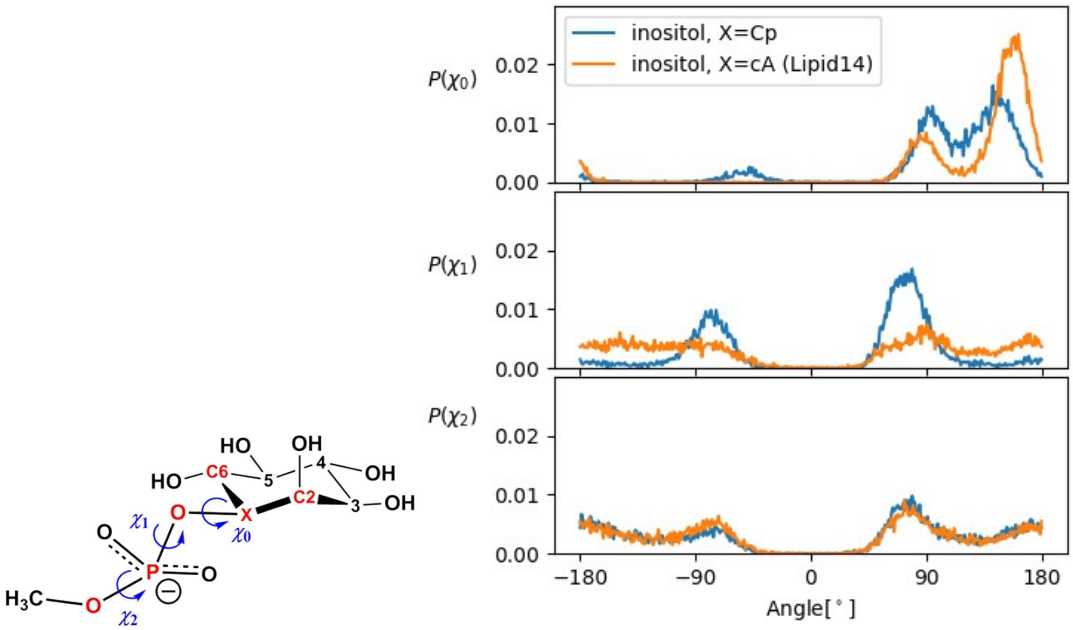
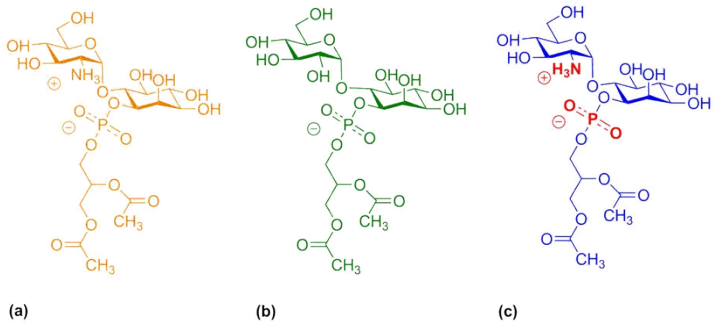


Figure S6: Torsion angle distribution in Ino-POMe, using the force field parameters for Lipid14 and two different atom types for the C1 carbon marked with X (see chemical schematic to the left). Orange solid lines: all-Lipid14 parameter set, X=cA. Blue solid lines: in all sequences containing X, (such as C2-X-O-P) X is set to Cp (cP), and the corresponding torsion parameters are taken from the GLYCAM06-h database, including the change in 1-4 scaling to (1.0/1.0). All other parameters (angle bending, bond stretching, torsions not containing X) are left unchanged. Partial charges are taken from Table S5 (Inositol, (e)).

Cp is formally available in Lipid14 under the name cP, along with other GLYCAM06 atom types. If we replace all torsion parameters in the all-Lipid14 parametrisation pertaining to sequences of four atoms containing C1 (as cP), we arrive at torsion profiles as shown in Figure S6. Effectively, we have now changed the profiles for χ_0 and χ_1 into those displayed in Figure S4. Figure S6 shows more clearly that the asymmetry in χ_1 (and χ_2) exists in the all-Lipid14 case as well, reemphasising the impact of intramolecular interactions. Although at the level of Ino-POMe they might still be considered rather small, the substitution of glucosamine or glucose renders them dominant, see scheme S2 and Figure S7.



Scheme S2: Molecules considered in Figure S7 (same color coding), created from 6OMe-Ino-PGL(0): (a) GlcN-Ino-PGL(0), and (b) Glc-Ino-PGL(0); (c) is the same as (a), but with mutual electrostatic interactions between group NH_3 and atoms P31, O33 and O34 (all highlighted in red) excluded to test the sensitivity on electrostatic interactions.

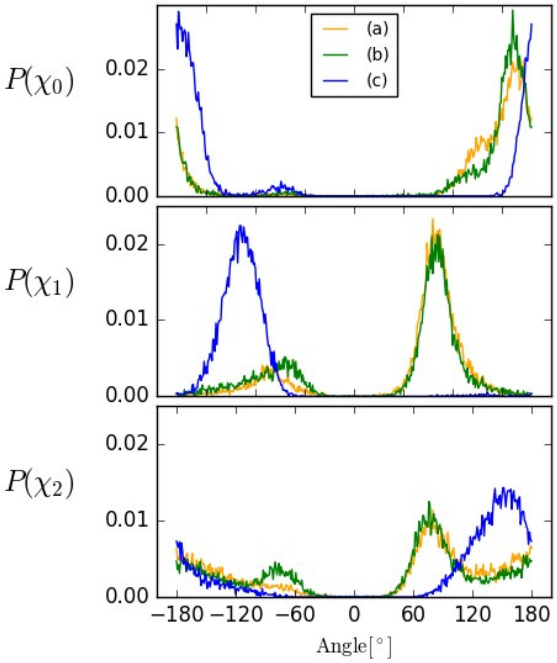


Figure S7: Density distributions $P(\chi_0)$ to $P(\chi_2)$ for the three molecules depicted in scheme S2 above. Simulation time: 100ns each, at 303K in TIP3P water.

In this figure, $P(\chi_0)$ to $P(\chi_2)$ are displayed for the molecular species (a) to (c) Depicted in Scheme 2. The first one (orange) is just 6OMe-Ino-PGL(0), the basic hybrid building block for creating arbitrary GPI anchors. The asymmetry in χ_1 and χ_2 is enhanced compared to Figure S6, the double peak in χ_0 is merged and shifted towards *trans*. This

behavior for χ_0-χ_2 closely parallels the one shown in Figure 7 of the main text, for the various GPI fragments inserted in DMPC and POPC bilayers, and strongly suggests that the preferences for these torsions can largely be attributed to the intramolecular interactions between GlcN and Phosphoinositol. Interestingly, substituting a hydroxyl-group for NH_3+ is already sufficient to produce virtually the same effect. As a negative test, we excluded the mutual electrostatic interactions between the atom groups highlighted in Scheme S2 (c), indicating the sensitivity of the torsions in the vicinity of these groups (blue graph in Figure S7).

S3. Vertical placement of GPI anchor moieties.

In Figure S8 we display the vertical distribution $p(z)$ of selected atom groups in Ino-, GlcN-Ino- and GPI-core-PGL in DMPC (a, c, e, g) and POPC (b,d,f,h) bilayer systems, respectively, along the bilayer normal (z-direction). For (a,b), the structure of the phosphocholine (PC) head is shown to the right (black sticks), with the nitrogen N31 (orange) and phosphorous P31 (brown) highlighted as spheres. The distribution of the average position of the whole ensemble (64 atoms per leaflet) is given in (a,b) with the same color coding. Compared to the latter the distribution of PC is broadened, since all its constituting atoms contribute. In (b), the PC distribution from (a) has been reproduced with dashed lines to illustrate the difference in bilayer thickness. The respective PC distributions are reproduced as solid black lines in (e,f) and (g, h) to facilitate comparison. (c) and (d) show $p(z)$ for only P31 as part of the different GPI fragments. (e) and (f) show the distributions of all atoms belonging to GlcN-Ino highlighted to the right.

The distribution of Man3 and Man2 (g, h) at the non-reducing end of the glycan is only available for GPI-core-PGL. In (h), the $p(z)$ of (g) is reproduced to illustrate the strong intercalation of this mannose moiety with the head group region for the POPC lower leaflet ($z < 0$). All histograms have been normalized to a total weight of unity, except for those of N31 and P31 in (a,b), the $p(z)$ of which have been scaled down by a factor of 20 for visibility.

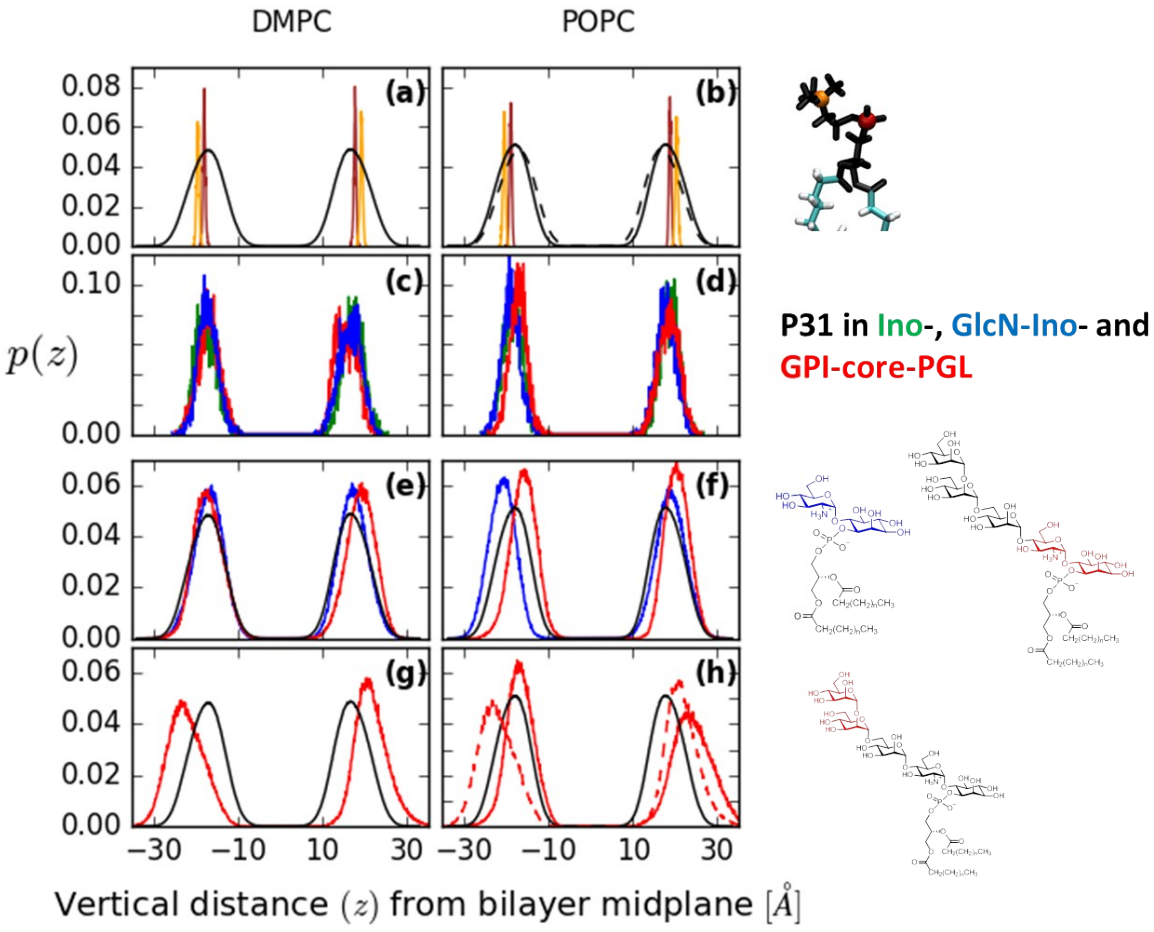
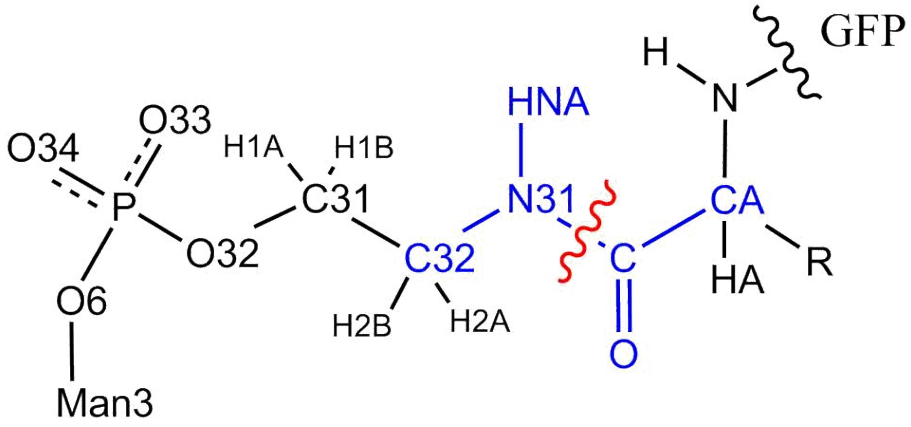


Figure S8. Vertical distribution $p(z)$ of selected atom groups in Ino-, GlcN-Ino- and GPI-core-PGL.

S4. Adding the phospho-ethanolamine linker and green fluorescent protein (GFP)

Scheme S3 shows how the phosphor-ethanolamine (PE) linker connects to the C-terminal amino acid residue (R denoting the side chain), with atom names annotated. The red wavy line indicates once more the transition from GLYCAM06 to the new domain of atom types (in this case parm10/ffSB12 for the protein).

Scheme S3: Phospho-ethanolamine linker with atom names annotated. On the phosphate side, the PE linker connects to the GPI-core via the O6 oxygen of Man3, see Figure 1 main text. R designates the threonine side chain (for example).



The linker comprises a non-trivial succession of torsions that can nevertheless be parametrized in a plausible fashion with existing parameters from GLYCAM06j and (the force field definition file) PARM10(.dat). We define the set of atoms highlighted in blue in Scheme S3 as the transition region, and treat it as a substituted, planar amide group as it occurs, e.g., in the description of GlcNAc or GalNAc in GLYCAM. The torsion potential around N31-C has two minima, with a preference for HNA-N31-C-O in *trans* (as shown). The parameters for atom sequences such as HNA-N31-C-O are identical in GLYCAM06 and Amber ffSB12 (PARM10) as well as the improper torsions for the sp2

atom groups C32-N31-HNA-C and N31-C-O-CA. Here we treat for simplicity the atom type of CA as Cg (generic sp3 carbon) and not CX (α-carbon in an amino acid), because we do not need to represent specificities that are important only for Φ/Ψ peptide torsions. In fact, in a series of test cases with smaller molecules containing the amide group, all-GLYCAM06j or all-PARM10 parameter sets yield similar results, a clear preference of *trans* over *cis* by more than 5kcal/mole. For this reason, the scaling factors assigned are largely immaterial, and so we opted for the GLYCAM scaling across the link (two GLYCAM-type atoms, two ffSB12/PARM10 type atoms). Note that the H-(Ng/N)-C-O scaling is 1.2/2.0, borrowed from the PARM10 (and earlier versions of PARMXX.dat). In general, we assign torsion parameters according to GLYCAM06j if a 4-sequence contains 3 or 4 GLYCAM atom types, and PARM10 values if there are 3 or 4 ffSB12 types. In the latter cases, of course, CA is poperly treated as CX. Table S6 shows how the atom names translate in atom types.

Table S6: Atom types involved in the link phospho-ethanolamine (PE) from Man3 to GFP at the latter's C-terminus (Threonine).

PE linker (GLYCAM06j)			Threonine residue (Amber ffSB12)	
Atomname	**Atomtype**		**Atomname**	**Atomtype**
O33, O34	O2 (carboxyl or phosphate group oxygen)		C	C (carbonyl carbon)
O6, O32	Os (Ether or ester oxygen)		O	O (carbonyl oxygen)
H1A, H1B, H2A, H2B	H1 (hydrogen connected to aliphatic carbon with one electron-withdrawing group)		HA	H1 (hydrogen connected to aliphatic carbon with one electron-withdrawing group)
C32	Cg sp3 aliphatic carbon		CA	CX (amino acid C-α carbon)
C31	Cp (carbon connected to oxygen connected to phosphorous atom)			
N31	Ng (sp2 amide nitrogen)			
HNA	H (hydrogen connected to nitrogen)			

Partial charges were assigned as follows: on the ffSB12 side, charges were taken from the predefined building blocks for amino acids; charges for the linker were calculated in a 1-stage RESP fit, a_{wt}=0.01, but with only one relaxed, all-trans configuration with methyl caps replacing CA and Man3, respectively (compare Scheme S3). The net charge on the whole linker was constrained to -1e. The aliphatic H1A to H2B on the GLYCAM side were each constrained to zero, and also the net charge on the group (O, C, CH3) on the ffSB12 side, such that the corresponding threonine (THR-) building block could be attached without further modification. It proved advantageous to impose no special constraints on the CH3 caps. Table S7 lists the charges obtained as described above.

Table S7: Charges on the capped PE linker. For atom names see Scheme S3.

Atomname		Charge [in units of e]
C	CA→CH3	-0.3658
H1		0.1041
H2		0.1029
H3		0.1045
O		-0.6312
C		0.6854
O34		-0.7635
O33		-0.7635
HNA		0.2848
N31		-0.6089
H2B		0.0000
H2A		0.0000
C32		0.2547
H1B		0.0000
H1A		0.0000
C31		0.2850
O32		-0.5815
P31		1.0868
C	Man3→ CH3	0.0035
H1		0.0386
H2		0.0630
H3		0.0633

The CH3 group replacing Man3 developed a net charge of +0.1684; to connect to Man3 it was removed together with O31 producing a charge deficiency of +0.249e. After removing the HO6 hydrogen, the Man3 moiety develops a deficiency of -0.194 according to the GLYCAM prescription. After attaching Man3 to the PE linker as indicated in Scheme S3, the residual deficiency is +0.055e, which was evenly spread across all atoms of the uncapped linker, resulting a rather small charge increment of 0.0069 each. A detailed investigation of the PE linker and its conformational characteristics will be presented elsewhere.[14]

In Figure S9 we finally summarize the distributions of the orientation and length of the end-to-end vector of the GPI-core glycan across the 6 trajectories respawned from the original 1μs simulation of GPI-GFP inserted into the POPC bilayer, compare Figure 10 of the main text for corresponding results of GPI-core-PGL. Figure S10 shows the corresponding hydrogen bonds between GFP and PE linker and all glycan moieties.

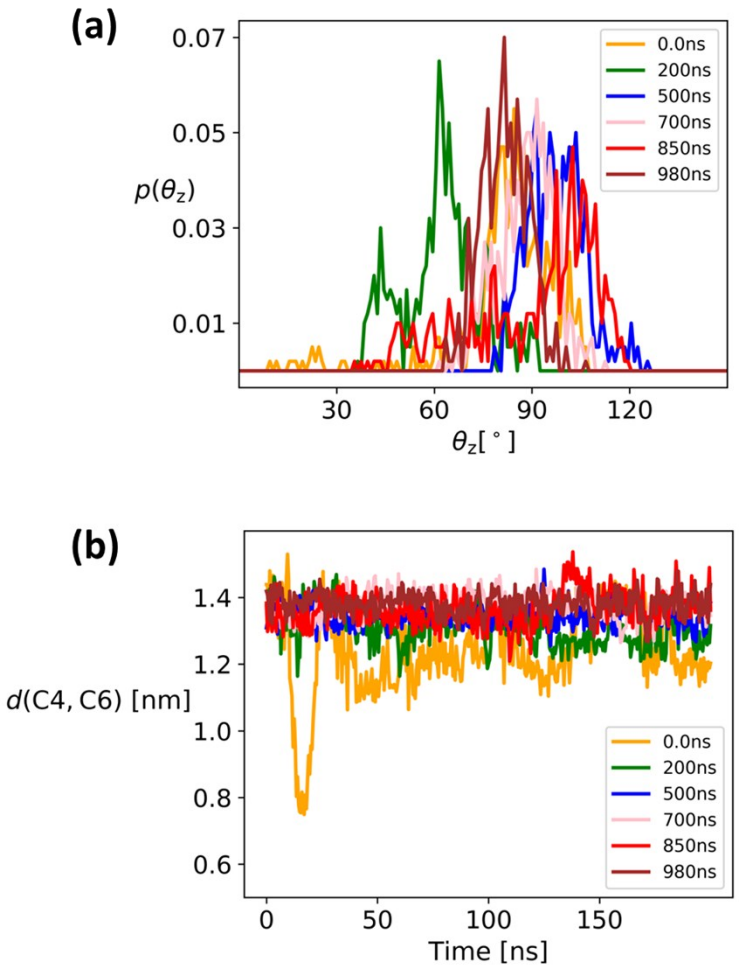


Figure S9: (a) distribution of the GPI-core orientation as defined by angle formed by the bilayer normal and the vector connecting C6 of inositol with C4 of Man3, compare Figure 10 in the main text. (b) distance distribution between these atoms. Bot plots comprise the six trajectories restarted from the initial trajectory of GPI-GFP inserted into the POPC lipid bilayer, same color coding.

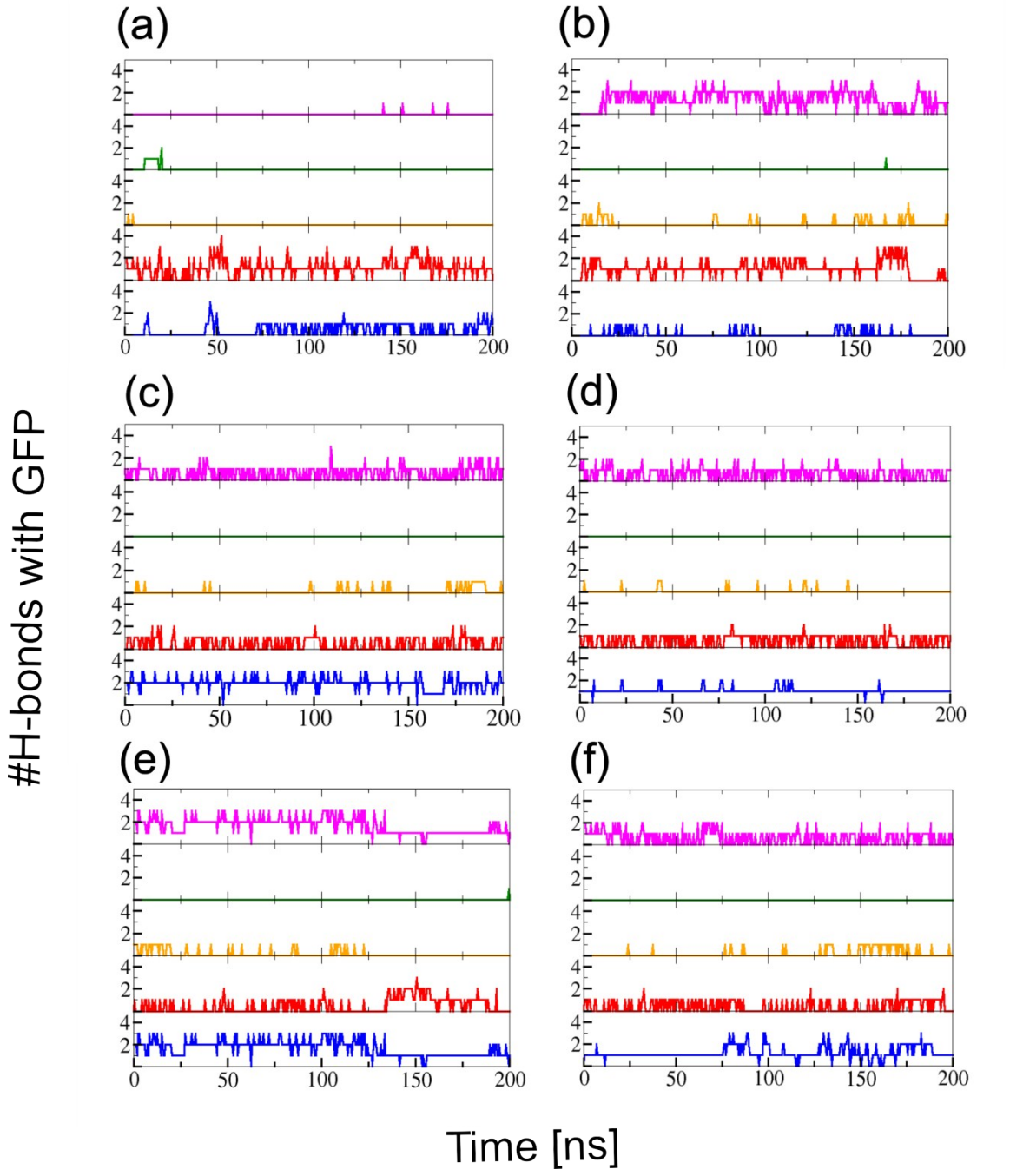


Figure S10: Number of hbonds formed between GFP and the glycan moieties of GPI core in the 6 trajectories displayed in Figure 11 of the main text and Figure S9. The data from (a) to (f) correspond to the different time points of simulation restart with freshly chosen initial velocities for configurations taken from the 1μs trajectory Figure 11(b) for (a) 0ns, (b) 200ns, (c) 500 ns, (d) 700ns, (e) 850ns and (f) 980ns, respectively. The color code is magenta: PE linker; green: Man3; yellow: Man2; red: Man1; blue: GlcN. Inositol has been omitted because it does not show any interaction whatsoever, because it is buried within the head group region. On the other hand, there is no reason why Man1 should not show any interaction with GFP at some point.

[1] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople. Gaussian 03, Revision E.01. **2004**. Gaussian, Inc., Wallingford CT Gaussian.

[2] Betz, R. M.; Walker, R. C. Paramfit: Automated optimization of force field parameters for molecular dynamics simulations. *J. Comp. Chem.* **2015**, *36*, 79-87.

[3] Dickson, C.; Rosso, L.; Betz, Robin M.; Walker, Ross C.; and Gould, Ian R. GAFFlipid: a General Amber Force Field for the accurate molecular dynamics simulation of phospholipid. *Soft Matter*. **2012**, *8*, 9617-9627.

[4] Breneman, Curt M.; and Wiberg, Kenneth B. Determining Atom –Centered Monopoles from Molecular Electrostatic Potentials. The Need for High Sampling Density in Formamide Conformational Analysis. *J. Comput. Chem.* **3**:361—373, **1989.**

[5] Bayly, C. I.; Cieplak, P.; Cornell, W. D.; and Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem*. **1993**, 97, 10269—10280.

[6] Cornell, W. D.; Cieplak, P.; Bayly, C. I.; and Kollman, P. A. Application of RESP Charges To Calculate Conformational Energies, Hydrogen Bond Energies, and Free Energies of Solvation. *J. Am. Chem. Soc.* **1993**, *115*, 9620—9631.

[7] Cieplak, P.; Cornell, W. D.; Bayly, C. I.; and Kollman, P. A. Application of the Multimolecule and Multiconformational RESP Methodology to Biopolymers: Charge Derivation for DNA, RNA, and Proteins. *J. Comp. Chem.* **1995, 16**(*11*), 1357—1377.

[8] Reynolds, C. A.; Essex, J. W.; and Richards, W. G. *J. Am. Chem. Soc*. **114**, pp. 9075, **1992.**

[9] Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem.* **2008**, *29*, 622-655.

[10] Becker, J.-P.; Wang, F.; Cieplak, P.; and Dupradeau, F.-Y. http://q4md-forcefieldtools.org/.

[11] Araz Jakalian, David B. Jack, Christopher I. Bayly. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parametrization and Validation. **2002**. *J. Comp.*

Chem. **23**(16), pp. 1623.

[12] Skjevik, Åge A.; Madej, Benjamin D.; Walker, Ross. C.; and Teigen, Knut. LIPID11: A modular Framework for Lipid Simulations Using Amber. *J. Phys. Chem. B.* **2012**, *116*, 11124-11136.

[13] Tessier, M.B.; DeMarco, M.L.; Yongye, A.B.; and Woods, R.J. Extension of the GLYCAM06 force field to lipids, lipid bilayers and glycolipids. Molecular Simulation. 2008, *34*(4), 349-364.

[14] P. Banerjee, R. Lipowsky, and Mark Santer, unpublished.