Glycosylphosphatidylinositols (GPIs) and GPI-anchored proteins tethered to lipid bilayers:

modelling a complex interplay of carbohydrates, proteins and lipids

von

Pallavi Banerjee

Kumulative Dissertation
zur Erlangung des akademischen Grades
"doctor rerum naturalium"
(Dr. rer. nat.)
in der Wissenschaftsdisziplin *Theoretische Biochemie*

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
Institut für Biochemie und Biologie
der Universität Potsdam
und
Max-Planck-Institut für Kolloid- und Grenzflächenforschung

Tag der disputation: 04. November 2020

1. Hauptbetreuerin: Dr. Stefanie Barbirz
2. Betreuer: Dr. Mark Santer
3. Gutachter*innen: Prof. Dr. Reinhard Lipowsky, Prof. Dr. Robert Woods

Published online on the
Publication Server of the University of Potsdam:
https://doi.org/10.25932/publishup-48956
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-489561

List of publications and authorship statement

This cumulative thesis is based on the following three publications, which are embedded in the thesis as Chapters 2-4.

- **Chapter 2**
  A molecular dynamics model for glycosylphosphatidyl-inositol anchors: "flop down" or "lollipop"? P. Banerjee, M. Wehle, R. Lipowsky and M. Santer, *PCCP* **2018**, *20*, 29314
  Authorship statement: My contributions focused on the molecular dynamics part of the paper. Quantum mechanical calculations were carried out by Dr. Marko Wehle.

- **Chapter 3**
  The importance of side chain branches of glycosylphosphatidylinositol anchors : A molecular dynamics perspective. P. Banerjee, R. Lipowsky and M. Santer [to be submitted]
  Authorship statement: This manuscript is the result of my own work with support from Dr. Mark Santer and Prof. Reinhard Lipowsky.

- **Chapter 4**
  Coarse-Grained Molecular Model for the Glycosylphosphatidylinositol Anchor with and without Protein. P. Banerjee, R. Lipowsky and M. Santer. *JCTC* **2020**, *16*, 6, 3889
  Authorship statement: This publication is the result of my own work with support from Dr. Mark Santer and Prof. Reinhard Lipowsky.

Author:
Pallavi Banerjee

Supervisor:
Dr. Mark Santer

Date:
16.06.2020

"A work of science is never finished, it is only abandoned."

Paul Valéry [co-opted]

Abstract

Glycosylphosphatidylinositols (GPIs) are highly complex glycolipids that serve as membrane anchors to a large variety of eukaryotic proteins. These are covalently attached to a group of peripheral proteins called GPI-anchored proteins (GPI-APs) through a post-translational modification in the endoplasmic reticulum. The GPI anchor is a unique structure composed of a glycan, with phospholipid tail at one end and a phosphoethanolamine linker at the other where the protein attaches. The glycan part of the GPI comprises a conserved pseudopentasaccharide core that could branch out to carry additional glycosyl or phosphoethanolamine units. GPI-APs are involved in a diverse range of cellular processes, few of which are signal transduction, protein trafficking, pathogenesis by protozoan parasites like the malaria-causing parasite *Plasmodium falciparum*. GPIs can also exist freely on the membrane surface without an attached protein such as those found in parasites like *Toxoplasma gondii*, the causative agent of Toxoplasmosis. These molecules are both structurally and functionally diverse, however, their structure-function relationship is still poorly understood. This is mainly because no clear picture exists regarding how the protein and the glycan arrange with respect to the lipid layer. Direct experimental evidence is rather scarce, due to which inconclusive pictures have emerged, especially regarding the orientation of GPIs and GPI-APs on membrane surfaces and the role of GPIs in membrane organization. It appears that computational modelling through molecular dynamics simulations would be a useful method to make progress. In this thesis, we attempt to explore characteristics of GPI anchors and GPI-APs embedded in lipid bilayers by constructing molecular models at two different resolutions – all-atom and coarse-grained.

First, we show how to construct a modular molecular model of GPIs and GPI-anchored proteins that can be readily extended to a broad variety of systems, addressing the micro-heterogeneity of GPIs. We do so by creating a hybrid link to which GPIs of diverse branching and lipid tails of varying saturation with their optimized force fields, GLYCAM06 and Lipid14 respectively, can be attached. Using microsecond simulations, we demonstrate that GPI prefers to "flop-down" on the membrane, thereby, strongly interacting with the lipid heads, over standing upright like a "lollipop". Secondly, we extend the model of the GPI core to carry out a systematic study of the structural aspects of GPIs carrying different side chains (parasitic and human GPI variants) inserted in lipid bilayers. Our results demonstrate the importance of the side branch residues as these are the most accessible, and thereby, recognizable epitopes. This finding qualitatively agrees with experimental observations that highlight the role of the side branches in immunogenicity of GPIs and the specificity thereof. The overall flop-down orientation of the GPIs with respect to the bilayer surface presents the side chain residues to face the solvent. Upon attaching the green fluorescent protein (GFP) to the GPI, it is seen to lie in close proximity to the bilayer, interacting both with the lipid heads and glycan part of the GPI. However the orientation of GFP is sensitive to the type of GPI it is attached to. Finally, we construct a coarse-grained model of the GPI and GPI-anchored GFP using a modified version of the MARTINI force-field, using which the timescale is enhanced by at least an order of magnitude compared to the atomistic system.

This study provides a theoretical perspective on the conformational behavior of the GPI core and some of its branched variations in presence of lipid bilayers, as well as draws comparisons with experimental

observations. Our modular atomistic model of GPI can be further employed to study GPIs of variable branching, and thereby, aid in designing future experiments especially in the area of vaccines and drug therapies. Our coarse-grained model can be used to study dynamic aspects of GPIs and GPI-APs w.r.t plasma membrane organization. Furthermore, the backmapping technique of converting coarse-grained trajectory back to the atomistic model would enable in-depth structural analysis with ample conformational sampling.

Zusammenfassung

Glykosylphosphatidyl-Inositole (GPIs) sind komplex Glykolipide, die insbesondere auf der Oberfläche eukaryotischer Zellen als Verankerung einer Reihe unterschiedlicher Proteine dienen. GPIs werden den Proteinen als post-translationale Modifikationen im endoplasmotischen Reticulum hinzugefügt. Die Verankerung in der Membran wird durch einen Phospholipidrest hergestellt, das Protein ist dann über ein sich daran anschließendes Pseudo-Pentasaccharid und einen Phospoethanolaminrest kovalent an den GPI Anker gebunden. Das Pseudo-Pentasaccharid ist dabei proteinunabhängig eine invariante Struktur, kann aber an bestimmten Stellen durch weitere Carbohydratseitenketten und/oder Phosphoethanolaminreste wesentlich erweitert werden.

GPI-verankerte Proteine (engl. GPI-anchored proteins, GPI-APs) sind an einer Reihe zellulärer Prozesse beteiligt; einige davon betreffen intra- und interzelluläre Signalübermittlung oder Proteintransport auf der Zelloberfläche; die Pathogenese vieler Parasiten, wie etwa *Plasmodium falciparum* (Malaria) wird entscheidend durch GPI-APs bestimmt; es können aber auch die bei vielen parasitischen Einzellern freien, ohne Protein auftretenden GPIs pathogene Wirkung entfalten wie etwa bei der Toxoplasmose (*Toxoplasma gondii*).

Der allgemeine Zusammenhang von Struktur eines GPI-AP und seiner Funktion ist allerdings bis heute zum größten Teil unbekannt. Dies liegt zum einen daran, dass sich kein klares Bild zeichnen lässt, wie ein GPI-AP relativ zur Zellmembran exponiert wird. Die relevanten Zeit- und Längenskalen sind experimentell unzugänglich, und entsprechende in vivo oder in vitro Untersuchungen liefern lediglich indirekte Hinweise. Der Fall GPI-verankerter Proteine ist daher ein Beispiel, in dem computergestützte Modellierung einen wesentlichen Beitrag zur Aufklärung leisten kann.

In der vorliegenden Arbeit wird zunächst ein atomistisches, molekulardynamisches Modell für GPIs und GPI-APs konstruiert und vorgestellt, mit dem sich GPI-APs auf der Längenskala einiger 10 Nanometer und einer Zeitskala von etwa 10 Mikrosekunden effizient untersuchen lassen. Modularität des Modells ist hierbei ein entscheidender Aspekt: mit den entwickelten Modellen lassen sich eine breite Palette von GPI Variationen darstellen. GPIs weisen, wie auch andere Proteinglykolysierungen eine sogenannte Mikroheterogenität auf; die Modifikation durch den Zucker kann sich zwischen den Kopien ein und desselben Proteins unterscheiden.

Die technische Umsetzung erfolgt im Rahmen der sogenannten AMBER- Familie atomistischer Kraftfelder, die nach einem bestimmten Schema für biomolekulare Simulationen entwickelt wurden. Dabei werden existierende Modelle für Zucker (GLYCAM06) und Lipide (Lipid14) durch die Optimierung und Herleitung fehlender Parameter so angepasst, dass sich ein vollständiges GPI-AP in einer Lipid-Doppelschicht darstellen lässt. Dabei zeigt sich, dass das Protein vermittelt über den flexiblen Anker über einen beachtlichen Bewegungsspielraum verfügt. Im Falle des hier betrachteten Green Fluorescent Protein (GFP) kann man daher das Bild einer festen Orientierung des Proteins in Bezug auf die Lipidoberfläche verwerfen; wie in der Mehrzahl der Simulationen beobachtet, kann das GFP sogar vollständig auf der Lipidschicht zu liegen kommen.

Weiterhin konnte nachgewiesen werden, dass eine Reihe möglicher Seitenketten des GPI Ankers, die zu Parasiten wie *Toxoplasma gondii* gehören und bei entsprechenden Immunreaktionen relevant sind,

tatsächlich so exponiert werden, dass ihre Rolle als Rezeptoren unterstrichen wird. Das Pseudopentasaccharid selbst ist dabei teilweise in die Kopfgruppenregion der Lipidschicht eingebettet.

Des Weiteren wurde hier das atomistische Modell auf eine vergröberte Darstellung im Rahmen des MARTINI Kraftfelds projiziert, um die zugänglichen Zeit- und Längenskalen noch einmal um einen Faktor 10 zu erweitern. Somit werden auch Studien GPI-APs möglich, bei denen sich ihre Dynamik in heterogenen Lipidschichten untersuchen lässt, etwa um Fragen zu beantworten, wie diese Proteine mit verschiedenen Membrandomänen assoziieren.

Insgesamt werden mit dieser Arbeit eine Reihe von Ansätzen aufgezeigt, wie sich GPI verankerte Proteine möglicherweise effektiver in speziell angepassten Experimenten und in größerem Detail untersuchen lassen, als dies bisher möglich war.

Acknowledgements

My deepest gratitude to all the people who have directly or indirectly, knowingly or unawares, helped me in this endeavour.

Foremost, I would like to thank my supervisor Dr. Mark Santer for his guidance and useful feedback throughout the project. Thank you for always being patient with me regarding the project's developments, no matter the pace, and allowing me the time and space to think independently. I am grateful to Prof. Lipowsky for lending his invaluable insights to the project. Besides, thank you for enabling an enriching and comfortable work environment at all times including the pandemic phase. With this I sincerely applaud his effective handling of the corona crisis. I thank Dr. Daniel Varon Silva for imparting his experimental expertise to the project, continued interest in the work and the fruitful discussions we have had. Special thanks goes to Dr. Alexander de Vries for the helpful discussions regarding MARTINI. In context of the project, I would also like to extend my thanks to Marko Wehle for his contribution to the atomistic model.

I thank Dr. Stefanie Barbirz for assuming the role of my supervisor at the University of Potsdam and to the IMPRS organization and the Max Planck Society for the funding and support. Thank you to Susann Weber and Angelina Schneider for all the administrative assistance, and to the IT staff especially Marko Ehlert and Rene Genz for warranting GPU resources, managing all the data, and always taking the trouble to promptly address my trouble tickets.

I can't thank all my friends enough who have helped me survive some difficult times of my PhD (including the writing phase during lockdown) with constant chatter, laughter and banter. A little humour a day does indeed keep depression at bay. I would like to especially thank my friend Aparna who went out of her way to help me during and after my surgery away from home. Thank you so much for the wonderful company and care during my days of disability and dependency.

Behlool, my homie and personal life-coach, thank you for making me stronger every day. Your love and support kept me going. I owe a big chunk of this achievement to you.

Finally, I must express my profound gratitude to my parents for their unflinching faith in me, the constant encouragement and for tolerating me at my worst. This accomplishment would not have been possible without their backing.

Contents

List of Figures

Abbreviations

GPI	Glycosylphosphatidylinositol
GPI-AP	Glycosylphosphatidylinositol-Anchored Protein
GFP	Green Fluorescent Protein
MD	Molecular Dynamics
FF	Force Field
AA	All-Atom
CG	Coarse-Grained
EtNP	ethanolamine phosphate / phosphoethanolamine
tEtNP	terminal phosphoethanolamine
mEtNP	middle phosphoethanolamine
PI-PLC	phosphatidylinositol-specific phospholipase C
VSG	Variant Surface Glycoprotein
mfVSG	membrane-bound form of Variant Surface Glycoprotein
sVSG	solution form of Variant Surface Glycoprotein
Man	mannose
GlcN	glucosamine
Ino	inositol
PGL	phosphoglycerol
GalNac	N-acetyl galactosamine
GlcNac	N-acetyl glucosamine
Glc	glucose
Gal	galactose
TNFα	Tumor Necrosis Factor-α
TLR	Toll-like receptor
DRM	Detergent-resistant membrane
FRET	Förster Resonance Energy Transfer

CGMD Coarse-grained Molecular Dynamics

Chapter 1

Introduction

Every living cell is enclosed by a membrane called **cell membrane** or **plasma membrane** that separates the cell's constituents from its surrounding environment, as well as, functions as a gate to allow transport of selective substances like ions, nutrients and waste products across. The plasma membrane is a complex system made up of a diverse group of biomolecules such as phospholipids, sphingolipids, glycolipids, cholesterol and different kinds of proteins, all of which, in one way or another, contribute towards maintaining the structural integrity of the membrane and facilitating various important cellular functions. Associated with all membranes are membrane proteins that can be broadly divided into two types based on the type of anchorage to host membranes: (a) Intrinsic/transmembrane proteins that are completely or partially embedded in the membrane, and (b) Peripheral proteins that are bound to the membrane by either a lipid tail or a special kind of glycolipid anchor called the Glycosylphosphatidylinositol (GPI) anchor (see Fig.1.1). The GPI anchor is assembled through a sequence of enzymatic reactions in the endoplasmic reticulum, and then the resulting pre-assembled structure is covalently added to the C-terminus end of the protein as a post-translational modification. This combined structure of GPI+protein is subject to further modification as it passes through the ER and Golgi complex, before being finally transported to the cell surface.

1.1 Timeline of the discovery of GPI anchor

Until the mid 1970s, it was widely believed that proteins associate with lipid membranes through non-covalent interactions. This dogma was challenged by the groups of Hiro Ikezawa and Martin Low through independent studies showing that the incubation of cells with bacterial phosphatidylinositol-specific phospholipase C (PI-PLC) led to the release of certain membrane proteins like alkaline phosphatase and 5'-nucleotidase[6, 7]. In the same study, Low showed that the release of alkaline phosphatase could not be reversed, thus pointing towards a covalent linkage between the protein and a phosphatidylinositol containing lipid[7]. However, owing to the lack of direct structural and chemical data, these

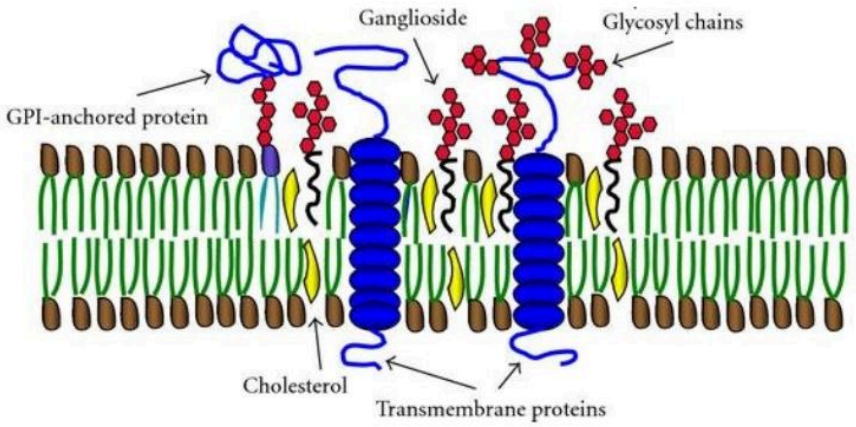


FIGURE 1.1: Schematic representation of a plasma membrane, containing phospholipids, gangliosides, cholesterol, transmembrane and GPI-anchored proteins. Figure adapted from ref.[1]

findings could not gain enough ground. In 1981, Alan Williams and coworkers provided the first chemical evidence of such a covalent bond from a study of rat Thy1 glycoprotein in which the C-terminus end of the protein was found to be unusually hydrophobic containing ethanolamine and fatty acids[8]. Subsequently, Anthony Holder and George Cross together showed that the C-terminus end of the soluble form of *Trypanosoma brucei* variant surface glycoprotein (VSG) is linked to ethanolamine which is further linked to carbohydrate residues, namely, mannose, glucosamine and galactose[9][10]. Almost around the same time, Mervyn Turner's group detected and isolated the membrane-bound form of VSG (mfVSG), and showed that trypanosomes contain an enzyme that rapidly releases mfVSG into solution in a different form (sVSG)[11]. They demonstrated that mfVSG is amphiphilic in nature and differs from sVSG both biochemically and immunochemically. In 1985, Michael Ferguson and coworkers confirmed all the aforementioned findings by performing a rigorous structural characterization of the glycolipid bound to mfVSG[12, 13], and the term glycosylphosphatidylinositol (GPI) anchor was thus coined[14]. Subsequently, in the same year, several groups demonstrated the covalent protein-lipid coupling in other proteins: *Torpedo* ACHe[15], Thy-1[16], and erythrocyte ACHe[17], with GPI anchor. Finally, in 1988 the first detailed and complete structure of GPI anchor of *T. brucei* VSG was established by Ferguson and colleagues through a combined strategy of proton NMR, mass spectrometry, and chemical and enzymatic modifications[18]. Since then, hundreds of GPI-anchored proteins have been discovered in eukaryotes, including humans[19].

1.2 Structure of the GPI anchor

All the GPI-anchors discovered till date share a common core structure, as well as the mode of covalent attachment to the protein. This pseudopentasaccharide core, Man-α(1->2)-Man-α(1->6)-Man-α(1->4)-GlcN-α(1->6)-*myo*Ino is attached to the protein via a phosphoethanolamine (EtNP) bridge, and at the other end is linked to a lipid tail at the C1 position of the inositol (see Fig.1.2). The presence of

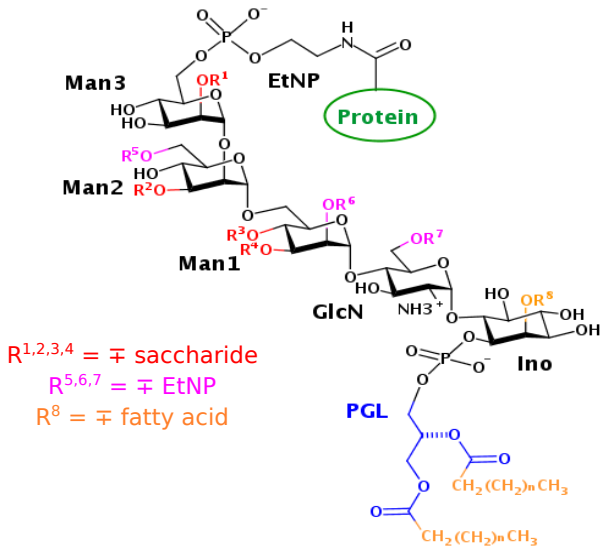


FIGURE 1.2: Chemical structure of the GPI core including possible structural modifications. The conserved portion of the GPI is in black. The colored portions represent possible variations. Here, PGL stands for phosphoglycerol and EtNP for phosphoethanolamine.

glucosamine (GlcN) is a rare and unique structural feature of GPIs because in nature, glucosamines are almost always found as either acetylated or sulfated[20]. GPIs can exist even without an attached protein (free GPIs) and abundantly so, such as those on the cell surfaces of protozoan parasites like *Toxoplasma gondii*, *Leishmania* etc. The core is conserved across all the GPIs (except one[21]), however the structure can differ through variations in the sugar side branches, the presence or absence of additional EtNP, and also the lipid tail. Branching-out typically occurs upwards from the first mannose (Man1) of the GPI core (Fig.1.2). Common modifications include branching at the Man1 residue to contain β-linked N-acetyl galactosamine (*Toxoplasma* SAG1, Rat Thy1, Human eCD59), α-linked galactosyl residues (*T.brucei*), or additional α-linked mannoses at Man3. Extra EtNP chains are found attached to Man1 or/and Man2 in only the GPIs of higher eurkaryotes. The phospholipid tail is also subject to variation, existing as diacylglycerol, alkyl-acyl glycerol, stearoyl-*lyso*glycerol, or ceramide. The tails can be of varying chain length (from 14 to 28 carbons) and saturation. Certain GPI anchors (Human eAChe, Human eCD59, *Plasmodium* MSP1) contain an additional fatty acid tail (usually palmitate) on the C2 of inositol that renders the GPI resistant to cleavage by PI-PLC. The structural variability of GPIs depends on the type of organism, tissue, and cell in which they are synthesized.

1.3 Functional diversity of GPIs and GPI-APs

More than 250 GPI-anchored proteins have been identified hitherto, however, there seems to be no characteristic similarity among them w.r.t. their functions or even sizes that range from 12 to 175 kDa [22]. GPI-APs have been associated directly or indirectly with a wide spectrum of biological functions.

The GPI anchor could affect the structure and conformation of the attached protein. The GPI-anchored protein procyclin of the insect form of *T.brucei* loses its antigenicity after removal of the lipid tail[23]. The vice versa is also true, as demonstrated in a binding study of an antibody raised against GPI-anchored Thy1 that failed to bind with the soluble form of the protein[24]. Change in antigenicity indicates change in conformation, possibly brought about by direct interactions of GPI with the protein[25], or indirect interactions between the protein and the membrane facilitated by the GPI[26]. GPI-APs are also involved in transducing signals from a cell's surrounding to its interior. Robinson and coworkers showed that on replacing the anchor component of a GPI-anchored antigen Qa-2 with a transmembrane domain in T cells, cell activation was inhibited [27]. This proves that GPI anchor plays an essential role in the signaling event of T cells. By cross-linking with antibodies and/or other transmembrane proteins, or by releasing signaling molecules like inositolphosphoglycans, GPIs can affect Ca^{2+} fluxes, secretion of cytokines, or protein phosphorylation[28, 29]. GPI-anchored neural cell adhesion molecule actively participates in the adhesion of neuronal cells, and regulates synapse formation[30]. Another putative function of GPIs is to transport the attached proteins to the apical sites of polarized membranes. The subject of apical sorting has proven to be particularly controversial. Initial attempts to understand the mechanism of apical sorting demonstrated that GPI-APs localize and agglomerate in transient glycolipid-rich detergent-resistant membrane domains, so called lipid rafts, while passing through the trans golgi network, and eventually end up in the apical membrane[31, 32]. The conclusion was that the association of GPI-APs with lipid rafts is a prerequisite for their apical delivery. This hypothesis was refuted by Zurzulo and colleagues who showed that in epithelial Fischer rat thyroid cells most GPI-APs transitioned to the basolateral domain, and some apical proteins did not even associate with lipid rafts[33]. Subsequent studies by other laboratories proposed that oligomerization or lectin-assisted cross-linking of GPI-APs, and not just raft association, is a requirement for apical sorting of the proteins[34, 35].

1.4 GPI anchors in diseases

The diversity of GPIs does not end at the normal, constructive physiological functions described in the above section. A growing body of evidence strongly suggests the first-hand involvement of GPI anchors in pathogenesis by several protozoan parasites. A subset of such GPI variants is shown in Fig.1.3. Parasites such as *Plasmodium falciparum, Toxoplasma gondii*, Trypanosomes like *T.brucei, T.congolense, T.cruzi* are enriched with GPI-APs and free GPIs on their cell surfaces[36]. Parasitic GPIs have been recognized as immunomodulators, as they can be detected by Natural killer T cells. As soon as they are introduced in host cells, an immune response is triggered, releasing macrophages to produce cytokines (inflammatory agents).

T. gondii is one of the most successful parasites that affects one-third of the world's population. *T. gondii* infection called Toxoplasmosis generally shows no lethal symptoms in immuno-competent individuals, whereas immunodeficient individuals are at risk of congenital disorders like severe encephalitis.

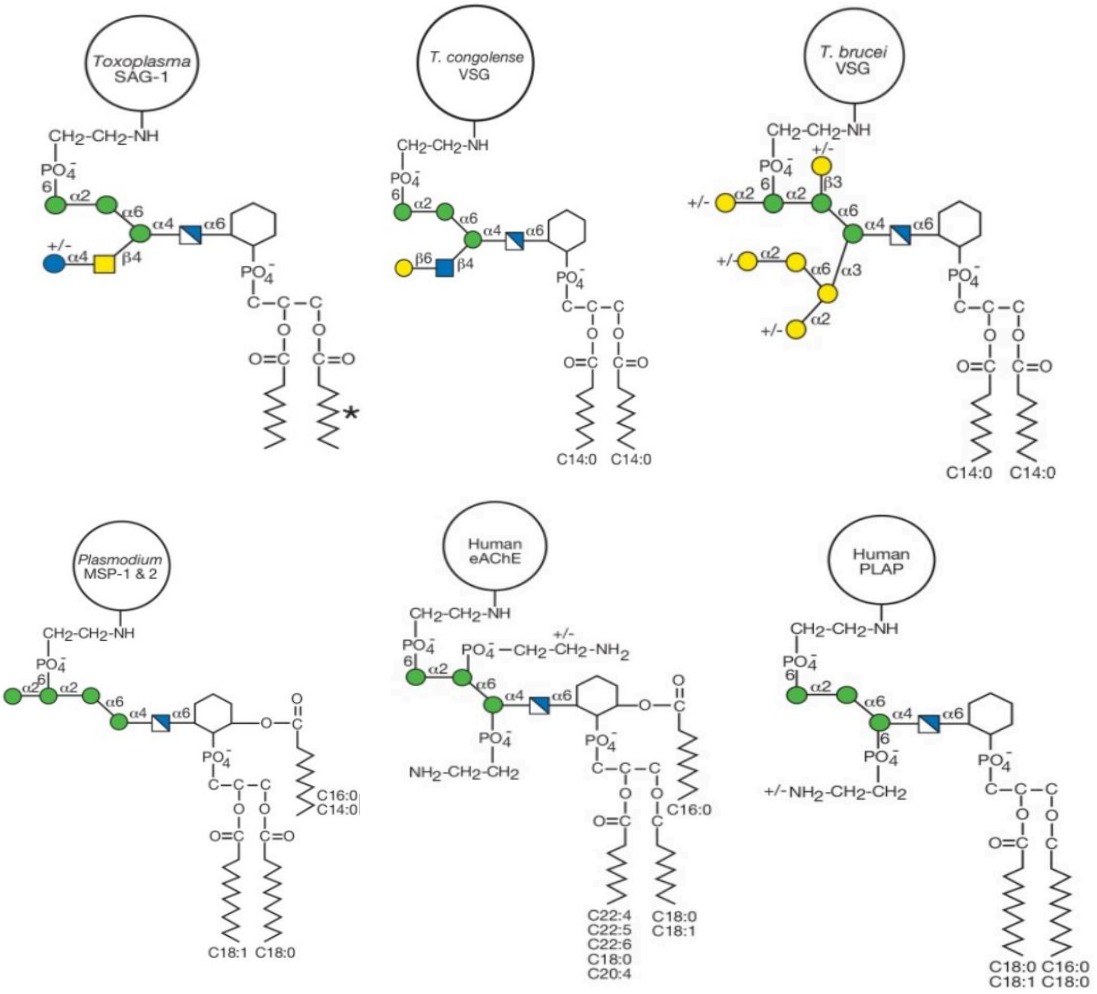


FIGURE 1.3: Chemical structures of some of the parasitic and human GPIs with the attached protein type. Chain lengths and levels of chain saturation are indicated. Glycosyl heterogenieity is expressed through '±', and '*' denotes unknown lipid type. The monosaccharide symbols follow the universal code: Symbol Nomenclature for Glycans (SNFG). Figure adapted from ref.[2]

Upon getting infected, previously unexposed pregnant women are at especially high risk of miscarriage. Debierre-Grockiego and team conducted thorough investigations on purified and synthesized *T. gondii* GPIs to show that both induce Tumor Necrosis Factor-α (TNFα) production in macrophages by activating the transcription factor NF-κB. Such an immune response involves the participation of toll like receptors TLR2 and TLR4[37, 38]. *P. falciparium* is the causative agent of malaria, a fatal disease that is of tremendous health concern especially in developing countries. Multiple *in vivo* and clinical studies have demonstrated that *P. falciparium* GPIs induce the production of cytokines and that raised levels of cytokine TNFα correlates with the severity of malaria[39]. Highly purified GPIs from Trypanosomes, that are responsible for causing Trypanosomiasis and Chagas' disease, also release potent inflammatory agents such as TNFα factor and nitric oxide[40].

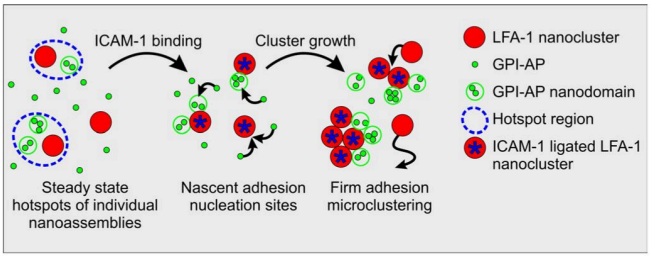
Inspired by the unambiguous set of data supporting the immunogenic nature of GPIs, scientists have recently been expending their efforts on making GPI-based vaccines. The GPI anchor is a more valuable candidate than the anchored protein for vaccine therapy because of its conserved structure. For example, the VSG protein of *T.brucei* is subject to frequent mutations in genetic code which enables the parasite to conveniently evade the host's immunity. The existing malaria vaccines are all peptide/protein based, and have failed to confer substantive protection[41]. The group of Seeberger showed the promising ability of synthetic malaria GPI to prevent the progress of fatal symtoms like pulmonary edema and acidosis in *Plasmodium berghei* infection of rodents[42]. Although, the vaccination did not protect against parasitaemia, yet the study establishes the toxic nature of parasitic GPI. Using a novel strategy of synthetic glycan arrays, they could identify the difference in anti-GPI response of two kinds of malaria GPIs that differ by one mannosyl residue[43]. The use of synthetic *T.gondii* GPI glycans as diagnostic tools has been proposed to detect and even differentiate between the latent and acute stages of Toxoplasmosis[44]. GPI vaccines when administered to infected mice could also alleviate immune complications, and prolonged survival[45]. Apart from parasitic infections, diseases involvong self-contained GPIs can afflict humans through GPI-anchor deficiency causing *Paroxysmal Nocturnal Hemoglobinuria* and misfolding of GPI-anchored prion protein causing prion diseases[36].

1.5 The role of GPIs in membrane organization

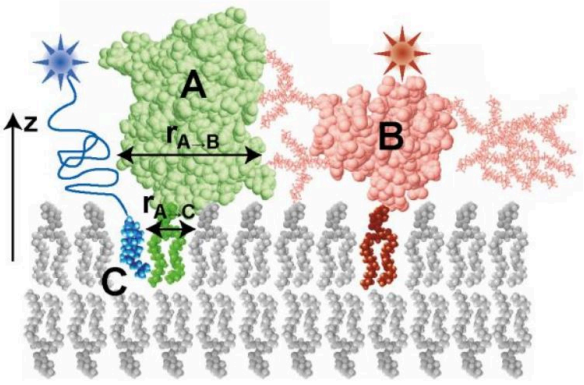
In 1972, Singer and Nicholson presented the fluid-mosaic model for the plasma membrane based on equilibrium thermodynamic principles. The model describes the cell membrane as a fluid matrix wherein membrane proteins are embedded. The fluid-like character of the membrane suggested a uniform distribution of lipids, lacking any heterogeneity[46]. However, over time findings on local lateral heterogeneities of lipids and polarized membranes of epithelial cells indicated that the plasma membrane could actually be heterogeneous[47, 48]. From polarized MDCK cells, Brown and Rose could precipitate highly-dense membrane extracts that were insoluble in non-ionic detergents, like TritonX-100, to show that these detergent-resistant membrane (DRM) patches were significantly enriched in GPI-anchored proteins along with sphingolipids and cholesterol[32]. Furthermore, it came to light that DRMs are sensitive to the levels of cholesterol, as cholesterol depletion by the same detergent resulted in complete solubilization of the domains[49]. The DRMs, later, came to be known as "lipid rafts". Thereby, the "lipid raft" hypothesis emerged, proposing that rafts, consisting primarily of saturated lipids, sphingolipids, cholesterol and GPI-APs, could be the functionally-active domains of the plasma membrane, plausibly serving as sorting stations for certain proteins[50]. Consequently, GPIs/GPI-APs were chosen to be raft markers in subsequent imaging experiments, for the purpose of understanding the organization and dynamics of raft domains.

However, the notion that GPI-APs associate with lipid rafts was put to test by Mayor and coworkers who reported that GPI-anchored folate receptor (GPI-FR) was seen to be uniformly distributed on the

cell surface[51]. They went on to show that detergent extraction redistributes GPI-FRs into caveolae sites through cross-linking with secondary antibodies[51, 52]. Subsequent investigations further bolstered the fact that detergent extraction introduces artifacts[53, 54]. This entailed devising different methods to study membrane organization. In 2004, Mayor and group studied the organization of GPI-APs in live cells, using a combination of homo- and hetero-FRET microscopy along with theoretical modelling. They reported that the cell surface exhibits a mixture of high-density clusters (<5nm) and monomers, the former being 20-40% of total GPI-APs[55]. These clusters were made up of at most 4 GPI-APs, some of them carrying different species of anchored proteins. Cholesterol depletion was seen to significantly affect clustering, adding further impetus to cholesterol's role in forming raft-like domains. However, only a few years later the same group, with the help of high-spatial and temporal resolution FRET and fluorescence anisotropy, presented a very novel picture. They observed that the formation of nanoclusters of GPI-APs in live cells is actually a non-equilibrium process governed by the dynamics of the cortical actin meshwork underlying the cell membrane[56]. Cholesterol depletion led to disruption of the actin network, and consequently of the nanoclusters. Therefore, the prevailing idea that phase partitioning of lipid membranes results in raft formation was severely challenged through the suggestion of active mechanisms being at play.



(a) Stages leading to integrin-mediated cluster formation. In resting cells, LFA-1 and GPI-APs preorganize in separate but proximal nanocompartments forming hotspot sites on the cell surface. (Center) Subsequent activation of LFA-1 by ligand binding drives the formation of larger supramolecular platforms that serve as nucleation sites for nascent cell adhesion. (Right) LFA-1 and GPI-AP microclustering brought about by recruitment of pools of mobile LFA-1 to the nucleation sites results in firm cell adhesion. Taken from Ref.[57]



(b) GFP-GPI (green) behaves as inert obstacles to the diffusion of approaching CD59 (red), and Cholesterol (blue). The reach of GFP-GPI is determined by its molecular size and shape. Taken from Ref.[58]

FIGURE 1.4: Opposing pictures of clustering of GPI-APs

Significant breakthroughs in super-resolution microscopy techniques have now enabled direct observation of cellular processes like molecular clustering. Aided by two different single-particle imaging techniques, two different groups arrived at absolutely contradicting pictures regarding the clustering behaviour of GPI-APs. The group of Garcia-Parajo employed single-molecule near-field scanning optical microscopy to directly study GPI-AP clustering in monocytes[57]. Their observation states that GPI-APs do form cholesterol-sensitive nanoclusters of 3-5 molecules each. Through associations with integrin

protein LFA-1, GPI-APs make larger clusters which serve as hot-spots for further recruitment of raft molecules. On the other hand, Schutz and team dismissed the existence of raft-type forces through a study of live cells using protein micro-patterning and single-molecule tracking[58]. They immobilized clusters of raft markers (GFP-GPI), and observed that GFP-GPI neither recruits other GPI-APs like CD59, nor does it increase membrane order in its vicinity. They attributed the reduced mobility of CD59 to steric hinderance posed by the cytoplasmic domain of GFP. Fig. 1.4 illustrates the two opposing pictures. The diverse results stemming from the use of different experimental tools imply that the quest to understand membrane organization and the mechanisms behind is still on. Using reductive approaches through molecular-level computational studies could be one useful way to go.

1.6 Experimental limitations and Computational Biology

Experimental studies of purified samples of GPIs are limited by the heterogeneous and amphiphilic character of these molecules. Extracted samples usually contain heterogeneous mixtures of different GPI anchors and GPI-anchored proteins that are difficult to isolate, often leading to inaccurate experimental interpretations. Recently, there have been considerable efforts towards chemically synthesizing pure GPIs, however, this process is tedious and not devoid of challenges[19]. Structural characterization of GPIs through X-ray crystallography or NMR is difficult because of flexibility of the constituent glycosidic linkages and lack of an appropriate solvent medium[59]. So far, no crystal structure of a GPI-anchored protein containing the GPI has been determined. Due to the inherent limitations of experimental tools and techniques, reports addressing the conformation and properties of GPIs are ambiguous. The orientation of GPIs and GPI-anchored proteins on membranes has been a matter of debate. Using FRET microscopy, Lehto and Sharom concluded that GPI-anchored alkaline phosphatase lies very close to the bilayer, nearly resting on the surface[26]. On the other hand, using a series of GPI analogues attached to green fluorescent protein (GFP), Bertozzi et al. conducted diffusivity studies to report high lateral mobility of these structures in lipid bilayers, thereby indicating that the anchor should be stiff such that GPI-APs lie away from the bilayer surface with no mutual interaction between the two[60]. Controversies surrounding the association of GPIs and GPI-APs with the differently ordered domains of membranes and their role in protein trafficking still persist (see Sec.1.5). Moreover, considering the varied structural and functional diversity of GPIs (Sec.1.2 and 1.3), addressal of a structure-function relationship is much warranted. Molecular-level understanding through computational studies of GPIs could significantly contribute towards clearing some of the prevailing controversies. Computer simulations act as a bridge between microscopic details and macroscopic properties, and between theory and experiments. Complementing experimental findings with simulation studies allows better understanding of real-life processes and helps in careful planning of experimental setup by making predictions of bulk properties. Among the existing computational methods, Molecular Dynamics (MD) is a powerful tool to provide invaluable data on the structure, dynamics and thermodynamics of biological macromolecules and complexes at different time (10^{-15} to 10^{-3}s) and (10^{-10} to 10^{-6}m) length scales.

1.6.1 Molecular Dynamics: atomistic and coarse-grained

Molecular Dynamics is a computational technique that is based on classical or Newtonian mechanics. In molecular mechanics, the smallest unit is an atom together with its nucleus and electrons. An atom carries its associated mass and partial charge. The atomic interactions are treated with analytic functions that describe the bonded and non-bonded potential energies, all of which together make up the **force-field** of the system. The force-field (FF) contains the necessary building blocks for calculating the energies and forces dictating the behaviour of the whole system. The generalized form of a typical force-field is the following:

$$V(r^N) = \sum_{Bonds} \frac{1}{2}K_l(l-l_0)^2 + \sum_{Angles} \frac{1}{2}K_\theta(\theta-\theta_0)^2 \sum_{Imp.Dihedral} \frac{1}{2}K_\omega(\omega-\omega_0)^2 + \sum_{Pr.Dihedral} K_\phi[1+cos(n\phi-\phi_0)] + \sum_i \sum_{j>i} \left(\frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}}\right) + \sum_i \sum_{j>i} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}^{12}\right) - \left(\frac{\sigma_{ij}}{r_{ij}}^{6}\right)\right] \tag{1.1}$$

The total potential energy of the system at a certain time point $V(r^N)$ is the sum total of bonded and non-bonded potential energy terms. The first four terms together make up the bonded potential energy. The first summation term represents energy from bond vibration, with the constants K_l as force constant and l_0 as the equilibrium bond length, the second is a 3-body angular potential energy with K_θ and θ_0 as the force constant and the equilibrium angle respectively. Both these terms are harmonic spring potentials. The next two terms are 4-body potentials, the former being a harmonic improper dihedral potential that maintains planarity in a molecular arrangement, where ω_0 is the equilibrium angle between a plane and a point, and K_ω is the corresponding force constant. The latter of the two is a cosine function representing proper dihedral that accounts for rotation around bonds, where K_ϕ is the force constant, ϕ_0 the equilibrium dihedral value, and n is the multipicity of the potential energy function. The last two terms in Eqn.1.1 make up the pairwise, non-bonded forces. The first of the two is Coulombic potential that accounts for interactions between charged species. The second is Lennard-Jones potential that represents van der Waals interactions where σ is the collision diameter and ϵ the minimum well depth of the potential between atoms i and j. All the constants that are fed into the force-field are obtained from quantum mechanical calculations.

Time evolution of the system is governed by Newton's laws of motion. The force acting on a particle i at a particular time point is the derivative of the potential energy obtained from Eqn.1.1:

$$F_i = -\frac{\delta}{\delta r_i}V = m_i\ddot{r}_i \tag{1.2}$$

The time integration of the system is implemented through various kinds of numerical algorithms like leap-frog, verlet, velocity verlet and langevin integrators. The most commonly-used atomistic force-fields for MD simulations of biomolecules including proteins, lipids, nucleic acids and carbohydrates are: AMBER, CHARMM, GROMOS[61].

Biological processes operate across a wide spectrum of temporal and spatial domains, right from the basic peptide bond formation all the way to multicellular signal transduction. Computational models serve to connect these different scales to enable their detailed understanding. Using a hierarchy of simulation approaches to treat systems across different scales is called **multiscale modelling**. Fig.1.5 shows the different simulation methods invoked at increasing time and length scales. The most fundamental and accurate description of matter is provided by quantum mechanics, in which the smallest unit is an electron. Quantum mechanics involves solving the Schrödinger equation to model processes like electron transfer in chemical reactions, obtaining molecular energies and orbitals, reaction pathways etc. All-atom molecular dynamics simulations can be applied to larger systems where the smallest particle is an atom. Here, classical mechanics is applied to propagate the system. Next up on the hierarchy ladder comes coarse-grained molecular dynamics (CGMD) where atoms are grouped together into pseudoatoms or beads, usually of diameter $\approx$ 0.5nm. Processes in the spatial regime of nano-to-micrometers functioning within microsecond timescales can be probed by CGMD through the generation of effective interaction potentials in a rather smooth potential energy landscape (see Fig.1.6). Processes like self assembly of membranes and micelles can be achieved in this domain. Further up is the mesoscale regime with bigger, coarser particles typically of diameter 1 nm. Dissipative Particle Dynamics (DPD) is the simulation algorithm built for this regime, where the chemical details of the system are lost, but physical attributes like hydrodynamics and thermal fluctuations are replicated well. Processes like budding, fusion and fission of membranes, dynamic behavior of polymers can be modelled with this technique. Beyond the mesoscale one enters the continuum regime where averaged-out effective fields replace explicit particles. Computational Fluid Dynamics (CFD) is one such technique that is used to model fluid flows using the Navier-Stokes' equations as the basis.

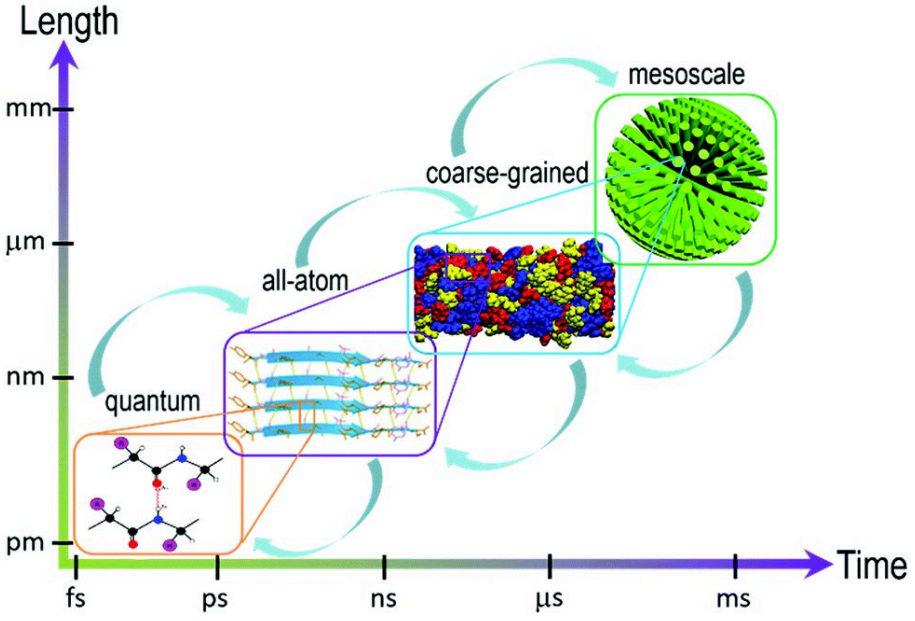


FIGURE 1.5: Illustration of multiple temporal and spatial scales covered by the relevant computational techniques[3]

CGMD is evolving as a powerful tool to model large scale biological complexes at biologically relevant temporal scales, up to few ms, through faster evolution of the system than traditional all-atom MD[62]. The speed-up is a consequence of the reduction in system-size through mapping of atoms to beads which leads to fewer degrees of freedom. Bridging the gap between atomistic simulations and macroscopic properties, CGMD offers a clever trade-off between simulation time and chemical specificity. There are three different ways of coarse graining a system: (i) energy-based, (ii) force-based and (iii) structure-based methods. Energy-based CG methods are based on reproducing energies of the underlying all-atom (AA) system or those from experiments for deriving interaction potentials for the CG landscape. For example, the MARTINI CG force field has been designed by parameterizing the building blocks (CG beads) with respect to the experimental oil/water partitioning free energies of small molecules[63]. MARTINI uses a four-to-one mapping scheme; i.e. four heavy atoms in the AA resolution are mapped to one pseudoatom or bead in the CG resolution. Membrane processes have been extensively studied and semi-quantitatively replicated by the MARTINI FF, e.g. phase separation and domain formation[64–66], cholesterol flip-flop in lipid bilayers[67], lipid desorption free energy[63], association of membrane proteins to lipids such as observed for rhodopsin[68], and many more that have been recently reviewed[69]. The MARTINI FF is being increasingly applied for investigating membrane systems, from simple, pure bilayer models to complex membranes with various asymmetries, progressing towards the goal of designing realistic models of the complex plasma membrane[5](see Fig.1.7).

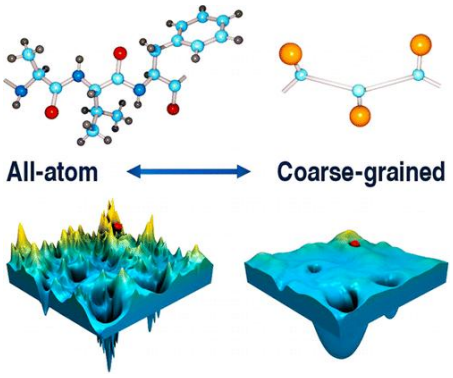


FIGURE 1.6: Visual representation of the difference in resolution between all-atom and coarse-grained systems with a depiction of their potential energy landscapes. Adapted from Ref.[4]

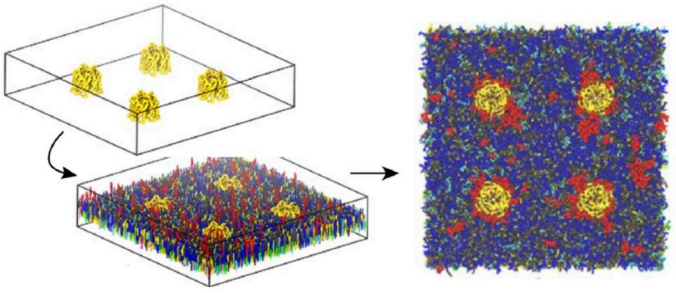


FIGURE 1.7: Example of a complex plasma membrane model consisting of 10 different membrane proteins embedded in a mixture of 63 different lipid types. The model was constructed using the MARTINI force-field. Adapted from Ref.[5]

Force-based CG approaches are based on deriving CG force-fields from the forces of the corresponding AA trajectory. Izvekov and Voth devised a multiscale coarse-graining method by employing systematic variational principle to obtain effective CG potentials from the forces of the underlying AA system[70]. This method has been used in modelling peptides[71], bilayers[72], carbohydrates[73], ionic liquids[74]. Force-matching is able to reproduce the atomistic structure excellently, however, its transferability is questionable. Moreover, its reliance on equilibrated all-atom systems limits its applications to only the trustworthy AA models. The third category, i.e, structure-based CG techniques are dedicated to reproducing the exact structure of the AA system. Methods such as Inverse Monte Carlo and Iterative

Boltzmann Inversion derive CG potentials targeting the radial distribution functions of the atomistic system[75–77]. These structure-based strategies are mostly custom-made for polymers where the chain connectivities and conformations are of particular importance, however, these methods are state-point dependent, and lack transferability.

1.7 Motivation

Computational modelling of GPIs and GPI-APs has been previously attempted in only a handful of studies. In the earliest study, Homans and coworkers arrived at a putative solution structure of the GPI anchor of *Trypanosoma brucei* VSG by combining two-dimensional ^{1}H NMR with molecular orbital calculations and restrained MD simulations[78]. Rademacher and colleagues presented a model of GPI-anchored Thy1 protein using NMR and theoretically obtained glycosidic linkages[25]. Both these studies show close association between the protein and GPI anchor. Zuegg and Gready simulated the human prion protein with its N-glycans and GPI anchor in a membrane model, where the oligosaccharides were parameterized using the GLYCAM94 force-field[79]. Through this model, they illustrated the flexibility of the GPI anchor. The flexibility of the GPI glycan core both in solution and micelle-bound forms was further demonstrated by a combined approach of NMR and MD simulations taken by Chevalier and coworkers[80, 81]. In another study of the prion protein, MD simulations revealed that the GPI anchor does not alter the conformation, however, influences the solvent-accessibility of certain regions of the protein[82]. They parameterized their model with the Levitt force-field[83] that was initially designed to model proteins and nucleic acids. The GPI anchor was parameterized with the GROMOS force-field[84] for the first time by Chiodi and Verli to study the conformation of GPI-anchored NETNES glycopeptide in membrane[85]. However, all the aforementioned studies are confined to simulation times of only a few nanoseconds, which considering the structural flexibility of GPIs, is not enough for sufficient sampling. Using MD simulations in the microsecond regime, Wu et al were able to capture the conformational differences between the soluble and the GPI-anchored form of prion protein that were modeled with the CHARMM force-field[86]. They noted that the mutual interactions among the anchored protein, N-glycans, GPI and the membrane stabilized the protein's secondary structure. Due to this dearth of theoretical studies of GPIs and GPI-anchored proteins in literature, the relation between the unique structure and diverse functions of the GPI is still missing.

In a part of this work (Chapters 2 and 3), we attempted to explore the structural features of the GPI anchor and GPI-anchored protein placed in lipid bilayers by constructing an atomistic model for MD simulations. The lipid bilayer was modeled using parameters from the Lipid14 force-field[87], which comes under the AMBER family of force-fields. As a sample GPI-anchored protein, we chose the Green Fluorescent protein (GFP) for this study. Although GFP is not a naturally occurring GPI-anchored protein, due to its easy availability and fluorescent nature, it has been extensively used to experimentally study properties of GPI-APs[55, 58, 88]. Therefore, to allow correspondence with existing and future

experiments, we modelled the GFP to be anchored to GPI. GFP was parameterized with the AMBER-ff14SB force-field[89], the latest improved version of the AMBER force-field for proteins. The glycan part of the GPI anchor was modeled with the GLYCAM06 force-field[90], which is also part of the AMBER family. GLYCAM06 has been designed to model carbohydrates, and is generalizable beyond common monosaccharides, containing parameters for glycoproteins[91] and glycolipids[92] too. We chose GLYCAM over other force-fields because, apart from showing excellent agreement with experimental observables, (a) GLYCAM is compatible with both Lipid14 and AMBER-ff14SB as they all belong to the AMBER super-set, and (b) the GLYCAM protocol allows a modular framework to build polysaccharides from various monosaccharides with different glycosidic linkages and branching. The latter is of particular concern while modelling GPI anchors as these molecules exist in variously branched forms (see Sec.1.2). In a previous work by our group, a GLYCAM-based model of the GPI tetrasaccharide backbone was constructed to explore its conformational characteristics in solution[93]. Using biased sampling techniques on constituent disaccharide moieties, it was revealed that the Man2-$\alpha(1 \rightarrow 6)$-Man1 linkage acts like a hinge in the backbone, agreeing with a previous preliminary report[80], and that the glycosidic torsions can be considered as independent degrees of freedom. Furthermore, it was concluded that at least 1μs of simulation time is required for adequate sampling. In this thesis, we extended the aforementioned GPI glycan model by attaching to it a dimyristoyl lipid tail to study the conformation of its membrane-embedded form. In addition, by attaching the GFP to the anchor via a phosphoethanolamine linker, we attempted to address the conformation and orientation of both the protein and the GPI when placed in membranes. We also try to understand the link between the structure and function of GPI by investigating the structure and presentation of differently branched parasitic GPIs and human GPI in lipid bilayer patches. All the MD simulations were conducted on the microsecond timescale.

One of the drawbacks of atomistic models of membrane systems is the limited size of membrane patches permissible by atomistic simulations. Dynamic properties of membranes like lateral diffusion and membrane undulations are sensitive to periodic boundary conditions, resulting in unrealistic behaviour[5]. Camley et al. pointed out that the relative error in diffusion coefficient is much higher in atomistic systems considering the small box sizes[94]. With recent developments in force fields and computational power, appreciable patches of atomistic lipid bilayers or membranes can be modelled. To date, the largest biomembrane modelled at the atomistic resolution was of size 90 nm that was simulated up to 150 ns[95]. However, the timescale is still a major bottleneck, when it comes to achieving sufficient conformational sampling of flexible molecules like GPIs and also to study dynamic membrane-related phenomena. In the latter part of the thesis (Chapter 4), we present a coarse-grained model of the GPI anchor and GPI-anchored GFP in lipid bilayers using a modified version of the MARTINI force-field. This makes the first coarse-grained GPI model in the scientific literature. The enhanced dynamics enabled by the coarse-graining would potentially aid in addressing intriguing questions regarding the role of GPI-APs in membrane processes (see Sec.1.5). Moreover, conformational investigation of GPI-APs will be significantly augmented by back-mapping[96] of coarse-grained trajectories to our atomistic model.

Chapter 2

A Molecular Dynamics Model for Glycosylphosphatidylinositol Anchors: "flop down" or "lollipop"?

2.1 Overview

In this paper, we elucidate the construction of an all-atom model of the glycosylphosphatidylinositol (GPI) anchor that is suitable for Molecular Dynamics studies. We design the model by merging two force-fields – GLYCAM06 for the GPI glycan and Lipid14 for the lipid tail, at the phopshoinositol-glycerol (Ino-PGL) unit. To ensure a smooth transition between force-fields, we assign the parameters for Ino-PGL through a meticulous mixing of atom types from the two force-fields. Bonded interactions for the bridging link involving mixed atom-types required renewed adaption. Thus, we create a molecular model of the GPI anchor that allows modularity at both ends of the Ino-PGL. On the one hand, the glycan part of the GPI can be extended to multiple variants of GPI with different side branch residues and on the other, lipid tails of variable saturation can be plugged to the model.

We study the conformational behavior of the GPI anchor by inserting it in pure lipid bilayers of DMPC and POPC separately. Firstly, by conducting a fragment-wise study of the GPI in bilayers, we demonstrate the role of the positively charged NH_3^+ group of glucosamine in bringing about a significant conformational change in the molecule, particularly at the juncture of the glycan head and the lipid tail. From 1 μs long simulations, it is revealed that the GPI displays a variety of conformations. However, inspite of substantial internal flexibility in the glycan region, GPIs predominantly flop down on the bilayer surface, as opposed to assuming an upright, lollipop-like conformation. The extent of embedding in the lipid headgroup region depends on the type of bilayer, as it was seen to be more in POPC than in DMPC.

Furthermore, we attach the Green Fluorescent Protein (GFP) to the GPI anchor as a case study for a GPI-anchored protein. The attachment is achieved by connecting the two with a phosphoethanolamine linker that was modelled separately according to the GLYCAM protocol in order to maintain force-field compatibility and modularity. GFP is parametrized using the AMBER protein force-field – ff14SB. We show that through interactions with the GPI, the protein is pulled close to the bilayer resulting in enhanced protein-bilayer interactions. The orientation of the anchored protein is, however, subject to variation, as the phosphoethanolamine linker is highly flexible. Regardless, GPI lies close to and is flopped on the bilayer surface, just like the protein-free GPI.

The paper thus presents a modular molecular model of the GPI anchor that can be extended to study many other GPI variants (see Sec.1.4) in a membrane set-up. Our finding refutes the possibility of an erect orientation of the GPI anchor on a bilayer (that was previously implicated in a few studies), and is in alignment with experiments and a few other computational studies conducted on parasitic GPIs that show a close association between GPI and the attached protein.

In this work, the construction of the phosphoglycerol-inositol bridge connecting the glycan head and lipid tail was achieved by Marko Wehle. My contributions were to the molecular dynamics simulations and analysis section of the paper.

PCCP

PAPER

View Article Online
View Journal

Cite this: DOI: 10.1039/c8cp04059a

A molecular dynamics model for glycosylphosphatidyl-inositol anchors: "flop down" or "lollipop"?†

Pallavi Banerjee,‡ Marko Wehle,‡§ Reinhard Lipowsky and Mark Santer *

We present a computational model of glycosylphosphatidyl-inositol (GPI) anchors for molecular dynamics studies. The model is based on state-of-the-art biomolecular force fields from the AMBER family, employing GLYCAM06 for carbohydrates and Lipid14 to represent fatty acid tails. We construct an adapted glycero-phosphatidyl-inositol unit to establish a seamless transition between the two domains of atom types. This link can readily be extended into a broad variety of GPI variants by applying either domain's building block scheme. As test cases, selected GPI fragments inserted into DMPC and POPC bilayer patches are considered. Our results suggest that the glycan part of the GPI anchor interacts strongly with the lipid head groups, partially embedding the carbohydrate moieties. This behaviour is supported by the conformational preferences of the GPI anchor, which in particular are conveyed by the strong interactions between the proximal amine and phosphate groups. In a similar way we can conclude that the extension of the anchor away from the lipid bilayer surface that could prevent the contact of the membrane with an attached protein ("lollipop picture") is quite unfavorable. Indeed, when attaching green fluorescent protein to the GPI anchor, it is found to reside close to bilayer surface all the time, and the rather flexible phosphoethanolamine linker governs the extent to which the protein directly interacts not only with the head groups, but also with its own GPI core.

Received 26th June 2018,
Accepted 6th November 2018

DOI: 10.1039/c8cp04059a

rsc.li/pccp

Introduction

Glycosylphosphatidylinositols (GPIs) are complex glycolipids present in eukaryotic cells, typically covalently bound to the C-terminus of proteins *via* a phospho-ethanolamine unit.[1]

They primarily serve to anchor proteins to the outer leaflet of the cell membrane, where a Manα(1→2)-Man-α(1→4)-GlcN-α(1→6)-*myo*-inositol pseudopentasaccharide core bridges towards a lipid tail, which in turn is inserted into the plasma membrane. Fig. 1 shows a schematic of a minimal complete structure of a GPI-anchored protein. The sequence of four monosaccharides consisting of three mannoses and one glucosamine plus the trailing *myo*-inositol (blue) is referred to as the GPI core, the sequence without inositol is termed the GPI backbone. The core is a conserved part in almost all naturally occurring types of GPI-anchors, and it can be modified by

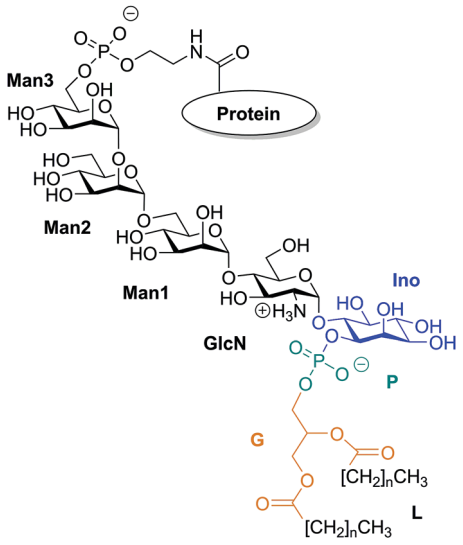


Fig. 1 Chemical structure of a GPI anchored protein. The GPI anchor consists of three mannoses Man1–Man3, a glucosamine- (GlcN) and an *myo*-inositol (Ino) residue linked to a phospho-*glycero*-lipid (PGL). The highlighted inositol (Ino), phosphate (P) and glycerol (G) moieties indicate the bridge between force field domains, see text.

Department of Theory and Biosystems, Max Planck Institute for Colloids and Interfaces, 14424 Potsdam, Germany. E-mail: mark.santer@mpikg.mpg.de; Fax: +49 (0)331 567 9602; Tel: +49 (0)331 567 9610

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c8cp04059a

‡ These authors contributed equally to this work.

§ Present address: ITK Engineering GmbH, Karl-Marx-Allee 90A, D-10243 Berlin, Germany. E-mail: marko.wehle@itk-engineering.de

various types of residues or side chains, the molecular weight of which may exceed that of the core itself.

Indeed, the vast number of possible side chain modifications including branched oligosaccharides that occur in nature suggests a rather broad functionality beyond the role of a mere anchoring device. GPIs are thought to be involved in localization of their proteins in membrane microdomains, commonly referred to as lipid rafts,[2,3] and bring proteins into close proximity with other raft-associated species to enable their interactions, which underline and determine diverse processes such as signal transduction,[4,5] cell adhesion protein trafficking and sorting,[6,7] and antigen presentation.[8] For all of these possible functions, a key aspect is how the GPI-anchored proteins distribute within the plasma membrane. In nerve cells, GPI-anchored Thy-1 and prion proteins have been suggested to form clusters in membrane domains with distinct lipid composition.[9] In general, the usually saturated alkyl chains of the anchor should favor association with membrane domains rich in cholesterol or other saturated phospho- and glycolipids, glycosphingolipids and sphingomyelin,[7] although there are indications that GPI-anchored proteins may also be distributed by non-equilibrium driving forces.[10] Although intuitively appealing, pictures such as raft association or trafficking are still intensely debated,[11,12] and in fact this conception has seriously been challenged by recent experiments of Schütz and coworkers.[13]

A major reason for the persisting controversies is that there is no clear molecular scale picture of how the complex of attached protein and the glycolipid anchor is embedded within and interacts with the local membrane environment. Experimental evidence for naturally occurring GPI-anchored proteins is scarce or ambiguous. From cryo-TEM studies with purified GPI-alkaline phosphatase (GPI-AP) inserted into liposomes[14] it was concluded that the glycan part should fill the space between protein and lipid, reminiscent of the protective coat of *Trypanosoma brucei*.[15] At ambient conditions in solution, a rather close proximity of AP to the lipid head group region has been inferred,[16] suggesting a mutual interaction of glycan, lipid and protein.[17–19] The seemingly opposing pictures are not necessarily a contradiction: in liposomes, the content of reconstituted protein or membrane composition and morphology are difficult to control. In addition, the glycan content of a GPI anchor is not exactly defined, for its composition may repetitively be remodeled during a protein's life cycle.[20]

Ideally a comprehensive characterization of GPI membrane embedding should proceed in a quasi-synthetic, bottom-up fashion starting from sufficiently simple and defined model systems. For instance, by creating a series of GPI-analogues attached to green fluorescent protein (GFP) and studying GPI–GFP diffusivities in supported lipid bilayers[21] and cells *in vivo*,[22] Bertozzi and coworkers conclude that high lateral mobility should be attributed to a rather stiff anchor preventing intermittent interactions with the membrane and increased protein–membrane contacts to more flexible anchor structures. In the former case GPI–GFP would roughly resemble a "lollipop"; in the latter, it can be thought to "flop down" onto the membrane, a picture introduced by Sharom and Lehto,[16] see also Fig. 2. Exploring molecular

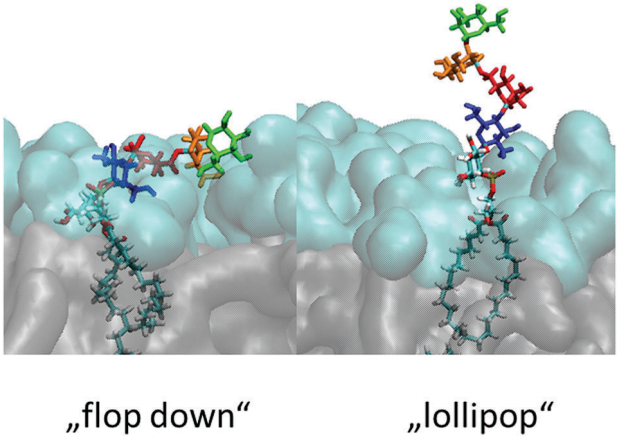


Fig. 2 Illustration of the "lollipop" *vs.* the "flop down" picture as snapshots taken from the same MD simulation in a DMPC bilayer (upper leaflet). Color code (cmp. Fig. 1) green: Man3, orange: Man2, red: Man1, blue: GlcN, followed by phosphoinositol and a di-myristoyl-acylglycerol lipid tail.

scale details of GPI placement, however, even with elaborate NMR approaches inevitably requires computational methodology.[23–25] With the present work we want to study the embedding of GPI anchors in lipid bilayers with atomistic, molecular dynamics (MD) simulations. To facilitate modular buildup and accurate representation of either molecular species, we shall construct a hybrid computational model by combining two state-of-the-art force fields from the AMBER family, GLYCAM06[26] (glycans) and Lipid14[27] (lipids). However, unlike simply "stitching together" a carbohydrate and a lipid *via* some *ad hoc* procedure, we shall specifically parametrise a molecular hub (highlighted in Fig. 1) that may readily be extended to various different GPI topologies simply by applying either force field's building block scheme. In this way, a coherent source of atomistic models is provided allowing to walk the "synthetic route" computationally. In order to keep the investigation at a manageable level in the spirit of a bottom-up approach, we shall restrict most of the exposition to anchor fragments up to and including Man3, compare Fig. 1. In the last section, we shall give a preliminary account of a fully fledged model of GPI-anchored green fluorescent protein (GFP) including the phosphoethanolamine (PE-) linker; the results largely support the major conclusions: the basic appearance of the embedded GPI-anchor can be understood in terms of rather simple molecular-mechanical arguments suggesting that a flop-down picture should dominate; in DMPC and POPC bilayers configurations with all glycan moieties stretching away from the solvent–headgroup interface are rare. Fig. 2 illustrates the "flop down" *vs.* the "lollipop" picture with specifically selected simulation snapshots of the GPI anchor model developed in this work. The flop down, however, can be realized in a number of ways, depending on the extent to which the glycan part intercalates with the head groups. Before analysing this behaviour in detail, we shall first discuss the computational model and how GPI conformations may be characterised before bilayer insertion. We will also comment on how the hybrid model may be validated and/or calibrated

further especially along with our results on GPI–GFP, which gives some clues to experimental validation.

Computational approach and model validation

To our knowledge there are only very few computational studies involving GPI anchors or GPI anchored proteins. Homans *et al.* gave an early structural account of the GPI anchor covering the surface of *Trypanosoma brucei*,[15] Zuegg and Gready[28] investigated the flexibility of an isolated prion protein attached to a simplified model of a lipid monolayer. Lomás and coworkers have studied the conformational characteristics of the GPI core,[23] and that of isolated GPI anchors inserted in small micelles,[24] with short simulations in support of NMR data.

In fact, providing a faithful model of a GPI anchored protein immersed in a cell membrane is a challenging endeavor, implying on the one hand that the characteristics of each type of biomolecule (protein, carbohydrate and lipid) must be represented with sufficient accuracy (with respect to stability, conformational preferences and phase behaviour, for instance).[29] On the other, the mutual nonbonded interactions of the three species in close proximity must be calibrated, and the experimental evidence for such situations is scarce.[30–33] A mere computational validation is rather difficult and subtle; in the scope of the present work we can only briefly indicate the state of the art of this subject (see, *e.g.*, the discussion preceding the Conclusion section).

The former issue mainly requires specialization. In the Amber family of force fields, Lipid14[27,34] has been developed to describe lipid bilayers in a stable and robust way; the focus of GLYCAM06 is devoted to modeling complex carbohydrates of almost arbitrary composition[26,35] (compare the corresponding developments in the CHARMM force field[36,37]). The transferrability of GLYCAM06 (particularly important to the case of GPI anchors) required the adaption of bonded interactions and how they are encoded by a suitable (re-)definition of atom types. For creating a hybrid link that may be extended with both, GLYCAM06 and Lipid14 building blocks, we need to establish a smooth transition between the two domains of atom types that define torsion, angle bending and stretching potential terms. This is accomplished by selecting a suitable bridging fragment (highlighted in Fig. 1) and turning it into a molecule with appropriate methyl cappings, see Fig. 3(a). The transition between atom types occurs across the bond from the C1 carbon (type Cp, GLYCAM) to the phosphate oxygen (oT, Lipid14). χ_0 denotes the corresponding torsion angle. Link (a) will be provided in two flavors: in addition to 6OMe-Ino-PGL(0) we define Ino-PGL(0) with bare inositol if the latter is the terminating head group moiety. This dual character is indicated in (b) with 6OMe-Ino-OMe and Ino-OMe, which play the key role for obtaining parameters of bonded interactions (*e.g.*, χ_0 and χ_1) involving a mixed set of atom types (Cg, Cp pertaining to GLYCAM06, oT, pA and oP to Lipid14), and for defining partial charges as to comply with either domains building block principle. The path from (b) to (a) involves a

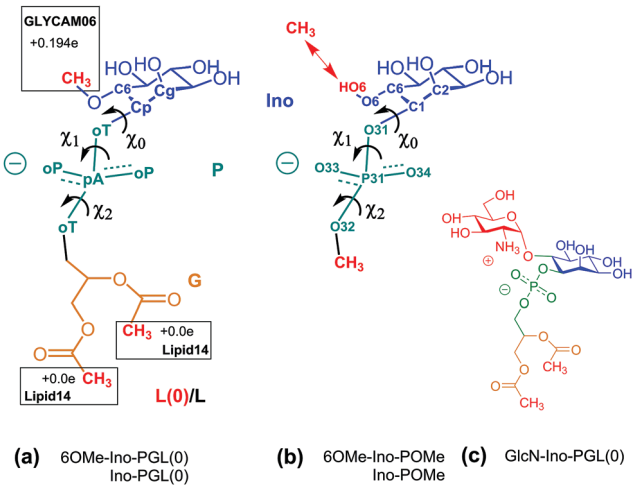


Fig. 3 (a) 6OMe-Ino-PGL(0) and Ino-PGL(0) representing the desired hybrid link highlighted in Fig. 1, here with appropriate methyl cappings to be substituted with GLYCAM06 carbohydrate building blocks and various types of (Lipid14-) alkyl chains. L(0) indicates the bare link with methyl caps (the nomenclature with "L" will be used for a general alkyl tail extension). Net charges on caps are set as to comply with the respective building block principle. (b) Reduced fragment from (a), 6OMe-Ino-POMe and Ino-POMe. The schematic also illustrates the atom numbering used throughout this text. (c) GlcN-Ino-PGL(0), GlcN has been substituted for the methyl cap at the 6 position.

number of somewhat lengthy and tedious steps and is outlined in detail in the ESI,† Section S1. To check for consistency, we compare the conformational preferences of χ_0–χ_2 obtained for Ino-POMe to either an all-GLYCAM06 or -Lipid14 parametrization presented in Section S2 (ESI†). These results prove to be rather valuable in order to categorize the plausible conformational preferences of the hybrid link, and then the full GPI anchor. We shall briefly summarize the essential findings:

(i) χ_0–χ_2 in the hybrid model of Ino-POMe are not affected by the type of partial charges assigned (ensemble averaged GLYCAM06 charges *vs.* AM1-BCC), see Fig. S3 (ESI†) for comparison. We use the AM1-BCC scheme as a convenient method to produce Lipid14-like partial charges, which rest on the standard two-stage RESP protocol in Amber.[38] This observation is useful as it simplifies the compilation of reference compounds to compare effects of GLYCAM06 *vs.* Lipid14 parameters. In addition it facilitates further validation of fully fledged models for GPI anchors (see concluding discussion of the following section).

(ii) The behaviour of χ_0 of the hybrid-, (obtained from MD in TIP3P solvent at ambient conditions), the all-GLYCAM06- and Lipid14 parametrisation is rather similar, major weights are given to intervals $\sim[+60^\circ,+90^\circ]$ (within *+synclinical* or *+sc* orientation), and $\sim[+150^\circ,+180^\circ]$ (within the *antiperiplanar* (*ap*) or *trans*-like domain of torsion angles). The distribution in χ_0 exhibits two neighboring peaks, compare Fig. S4 and S5 in the ESI† (or S6, for direct comparison).

(iii) The distribution of χ_1 in the hybrid model of Ino-POMe is very similar to that in the all-GLYCAM06 version exhibiting two peaks within *−sc*, *+sc* (close to $\sim\pm90^\circ$), with slight preference

for +*sc* (Fig. S4, ESI†). The all-Lipid14 version, by contrast, is different in that broad access is also given to torsions within *ap*. This can uniquely be attributed to the atom type assigned to the C1 carbon of the inositol ring: cA in Lipid14 (generic sp^3 bonded carbon) and Cp in GLYCAM06 (sp^3 carbon bonded to phosphate oxygen). The "cA–Cp exchange" does not impact χ_0 see Fig. S6 and the corresponding discussion in the ESI.†[39]

(iv) The torsions χ_0–χ_2 of the full link 6OMe-Ino-PGL(0) behave quite similarly to those of Ino-POMe; however, the substitution of glucosamine at the O6 oxygen (Fig. 3(c)) has a pronounced impact. The two peaks for χ_0 merge and shift into *ap*; for χ_1, major preference is now given +*sc*; for χ_2, an asymmetry is established with a preference for +*sc*/−*ac*. This is solely due to the mutual intramolecular interaction of the proximal amine and phosphate groups, elucidated by a series of auxiliary simulations reported in Fig. S7 and the accompanying text in the ESI,† Section 2; see also the discussion related to Fig. S4–S6 (ESI†).

Observation (iv) is very useful as it allows us to single out a few particular GPI conformations that can be expected to occur with high probability. Fig. 4 illustrates the corresponding geometries, which have been obtained in a "molecular mechanics"-like fashion as follows: from an equilibrated DMPC bilayer a representative phospholipid tail was selected with both alkyl chains pointing along the *z*-direction, and the O32–P31 bond tilted by 45° with respect to the bilayer normal. This tail is common to structures (a)–(d). Structure (a) was then obtained by adding inositol and GlcN with (χ_0,χ_1,χ_2) set to (175°,90°,90°) according to observation (iv). The (Ψ,Φ) glycosidic angles from GlcN to inositol exhibit only one narrow minimum around (70°,−150°). The distance distribution between amine and phosphate group (distance between nitrogen and phosphorous atom) has a maximum at 3.5 Å and a secondary one at 5.0 Å. The remaining three mannoses were added with glycosidic angles set according to their global free energy minima.[41] Selecting the lowest energy values for all torsions furnishes the anchor with a hook-like appearance shown in (a) or (b), where the latter is obtained simply by setting the Ω-glycosidic angle in the Manα(1→6)Man linkage (captured by the orange ellipse in Fig. 4(a and b)) to the *gt* configuration, which should be equally or a little less populated than *gg*. The transition *gg*→*gt* twists the hydroxy-methyl group of Man3 approximately from −*z* to +*z*-direction.

(c) is derived from (a) by setting (χ_0,χ_1,χ_2) to (175°,−90°,−90°) that is, choosing for both, χ_1 and χ_2 the values that are less likely (compare Fig. S7, ESI†). In (d) we have stretched out the GPI core as much as possible (Ω in Manα(1→6)Man linkage set to the unfavorable *trans gauche* and Manα(1→2)Man tweaked to a high energy conformation) and setting (χ_0,χ_1,χ_2) to (90°,90°,90°) that is, unfavorable for χ_0. In this conformation the amine and phosphate group are rather close (nitrogen–phosphorous distance about 3.1 Å) such that steric conflicts can be expected. Twisting the two charged groups away from each other as shown in (e) requires adjustment of the GlcN-Ino glycosidic angles, elevating the internal molecular energy as well. From these considerations one may predict that a lollipop-like conformation (d and e) should be rare or rather overwhelmed by variations of the hook-like appearance (a and b). In the following section we will compare the corresponding behavior with molecular dynamics and lipid bilayer systems.

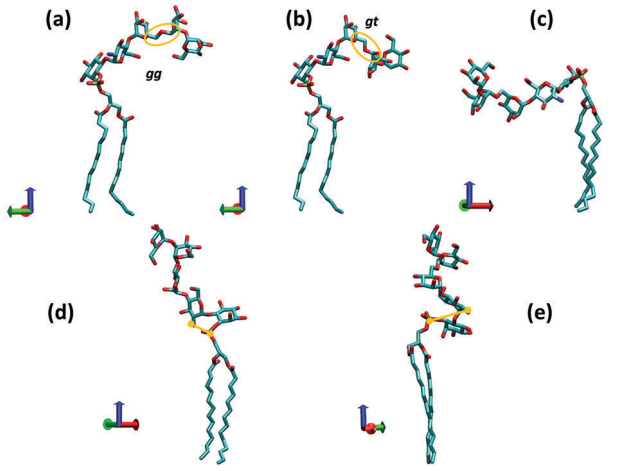


Fig. 4 Various conformations of a complete GPI anchor derived from the torsional preferences of the hybrid fragment 6OMe-Ino-PGL(0) and those of the backbone. The red, green and blue coordinate axes indicate positive *x*, *y* and *z* directions, respectively (visualization with VMD[40]) (*z* indicating the direction of the bilayer normal). The bond P31–O31 is oriented roughly at 45° w.r.t. the *z*-axis. In (a) and (c), the Ω glycosidic angle of the 1→6 linkage (orange ellipse) is set to the *gauche–gauche* (*gg*), in (b) and (e) to *gauche–trans* (*gt*) and in (d) to the *trans–gauche* (*tg*) orientation. In (d) and (e), the distance between phosphorous atom P31 and the nitrogen of GlcN is indicated.

GPI-anchors embedded in lipid bilayer patches

Simulation setup

For MD simulations in small DMPC- and POPC bilayer patches we consider the GPI-anchor fragments displayed in Fig. 5. (a) has been modeled from the link Ino-PGL(0), (b) and (c) from 6OMe-Ino-PGL(0). As simulation engines Amber14[42] as well as GROMACS v. 4.6.4 were employed[43,44] the latter in the majority of cases for running long (microsecond) bilayer simulations. All simulations pertaining to GROMACS were run using the CPU as well as the GPU version. To translate Amber/GLYCAM- to GROMACS molecule topologies and input files we utilized a script originally devised by Sorin and Pande[45] and adapted by us[41] to cope with the specificities of GLYCAM06. All simulations were carried out under standard conditions at 303 K with the TIP3P water model with pressure maintained at 1 bar using semi-isotropic Berendsen rescaling[46] in an orthorhombic simulation box. Removal of center of mass motion was applied to the bilayer system as a whole. The species (a–c) were inserted through a number of separate steps. First a quadratic bilayer patch of 2 × 64 (8 by 8) phospholipid molecules was prepared with surface area of 100 $Å^2$ per lipid. At this stage one lipid per leaflet was deleted and replaced with a corresponding GPI molecule, keeping the symmetry of the simulation setup.

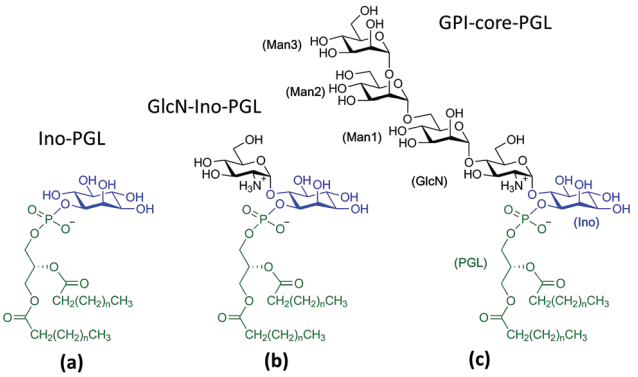


Fig. 5 GPI fragments studied within DMPC and POPC lipid bilayers, created by replacing the methyl cappings of the respective links with saturated myristoyl alkyl chains (corresponding to n = 11 in the schematic) taken from the Lipid14 database. Parameters for GlcN have been taken from Singh *et al.*[61] All other glycan building blocks have either been assembled from the GLYCAM06 force field modification file or taken from the GLYCAM web database (www.glycam.org).

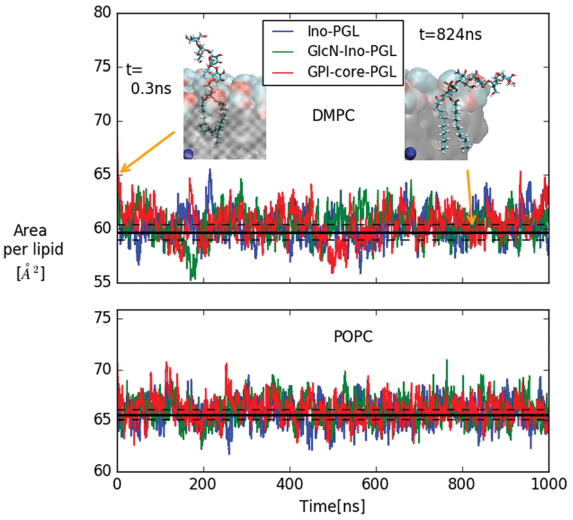


Fig. 6 Area per lipid of DMPC and POPC bilayer systems with embedded GPI-anchor fragments displayed in Fig. 5; color coding for POPC is the same as in the upper panel. The black lines denote average and error quoted by Dickson *et al.*[27] for the case of pure DMPC (59.7 ± 0.7 Å^2) and POPC (65.6 ± 0.5 Å^2) bilayers, respectively. For DMPC, the snapshots of GPI-core-PGL of the upper leaflet are displayed for the time points indicated by arrows. Blue spheres mark the bilayer center. To improve visibility, in the snapshot at 824 ns half of the bilayer has been removed.

The latter was pulled out a few Å in z-direction as to avoid any initial steric clashes of the glycan part with the PC head groups. In addition, the glycan portion was twisted away from the bilayer surface as much as possible by manually adjusting the torsion angles χ_0–χ_2 and all subsequent glycosidic linkages, obtaining a structure close to that shown in Fig. 4(e). In this way some excess internal energy is provided at the start of the simulation procedure. The whole system is then solvated with TIP3P waters taken from pre-equilibrated solvent boxes. Waters placed within the region of alkyl chains were subsequently removed. This fact and the different sizes of (a) to (c) lead to slight variations in the final system size with average extensions of 6.5 nm, 6.5 nm and 13 nm in x, y and z-direction, respectively, ~66 500 atoms in total, and ~17 500 water molecules. In the case of Ino-PGL (a) one Na^+ counterion was added to the system. We used a plain cutoff of 1 nm for vdW interactions and the same value was used as the real space cutoff in the particle mesh ewald (PME) electrostatics, which was used througout.

Initially, all lipids were subject to soft harmonic restraints with respect to their center-of-mass motion to allow the bilayer to soak with solvent during an initial period of 100 ps at 100 K during which bad contacts were resolved, using a Langevin-type stochastic integrator[47] with a collision time of 1 ps. Subsequently, the temperature was ramped up to 300 K during another 100 ps, again removing water molecules that penetrated into the domain of alkyl chains. Then the system was allowed to relax with pressure control on (pressure relaxation time of 2 ps) at 1 bar. The area per lipid quickly (well within 10 ns) relaxed to its equilibrium value, and remained there for the total simulation time in each case. Reference simulations with temperature controlled by a Nosé–Hoover chain thermostat[48] resulted in the same behavior. Snapshots were written out at an interval of 0.25 ns resulting in a total of 4000 frames per μs, of which we omit the first 50 ns (equilibration period) leaving 3801 for data analysis.

In Fig. 6 we show the time-dependent area per lipid for systems (a–c) in DMPC and POPC bilayers, respectively. The panel for DMPC indicates the evolution of GPI-core-PGL from the initial erected- (snapshot taken at 0.3 ns) to a flop-down conformation where the GPI backbone appears to point roughly parallel to the bilayer plane.

General appearance of GPI conformations within bilayers

In the previous section it was argued that for GPI-core-PGL hook-like conformations arise through the special preferences of torsion angles χ_0–χ_2, induced by intramolecular interactions between amine and phosphate group. We can essentially make the same observation as described in (iv) with GPI fragments (a–c) in Fig. 5, but now embedded within the bilayers, see Fig. 7. As soon as GlcN is included, the bimodal distribution for χ_0 merges and shifts into $+ap$, almost complete emphasis for χ_1 on $+sc$, and an asymmetry of χ_2 with emphasis on $+ac$/$+ap$ configurations, the orientation of the O31–P31 bond shows its maximum around 45° (see insets). Indeed, in the simulation runs we observe (with the exception of the "submarine", see below) a pronounced occurrence of hook-like conformations for GPI-core-PGL in each leaflet, that is, the GPI core mostly tends to "flop down" onto the bilayer surface. Fig. 8 displays how similar conformations to the ones shown in Fig. 4 are assumed within the lipid environment.

(a) reproduces the snapshot shown in Fig. 2, the conformation belongs to a class that we shall coin "swimmer", because all carbohydrate moieties are well exposed to the solvent, situated at the interface between PC head groups and aqueous subphase, see also (b) for the corresponding top view. The swimmer can be realized by a number of variations of the

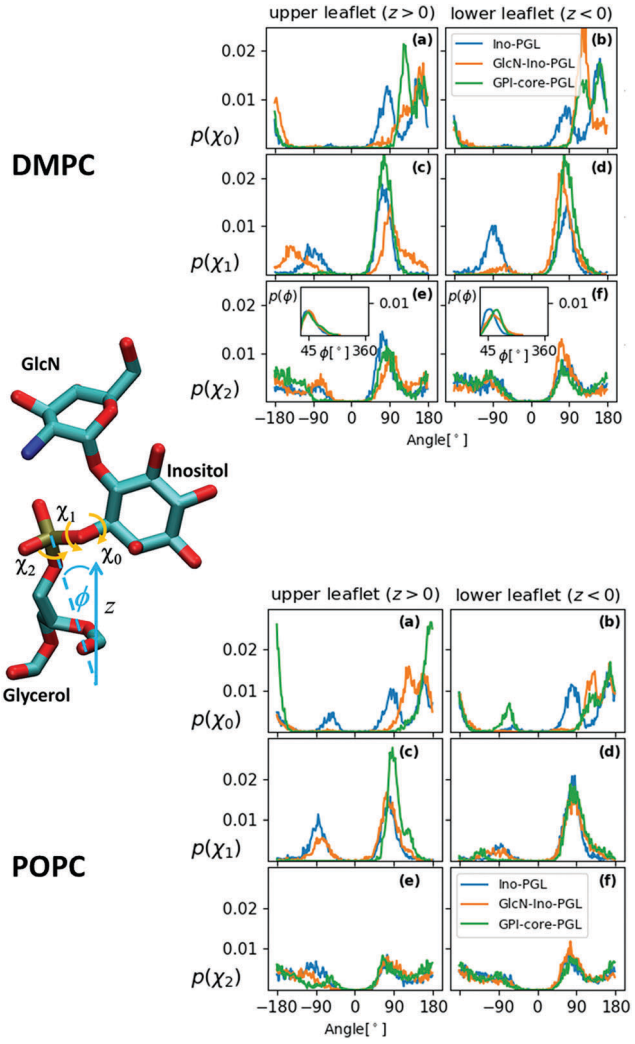


Fig. 7 Distribution of the torsion angles χ_0, χ_1 and χ_2 for all three GPI anchor derivatives embedded in the DMPC- (top panel) and POPC- (bottom panel) bilayer system. The bin width of all histograms in is 0.1°, each acquired from data of 3801 simulation snapshots. The insets in (e) and (f) in the DMPC case show the distribution of the inclination angle ϕ of the bond defining χ_2 with the z-axis (bilayer normal), see schematic to the left.

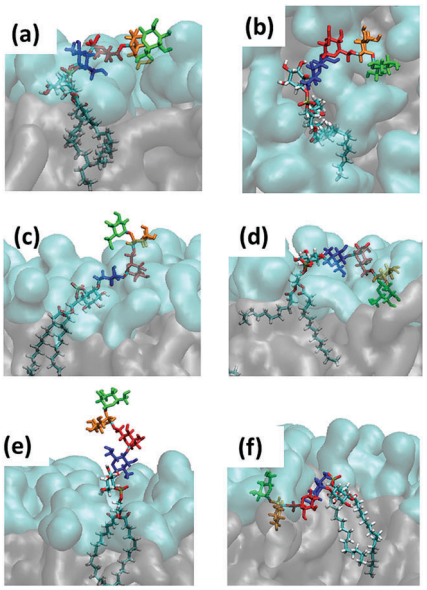


Fig. 8 Snapshots of the characteristic GPI configurations characterized in the text, the time points at which they were acquired are indicated in Fig. 9. Surface representation: lipid layer excluding GPI-core-PGL, with PC head groups in cyan and glycerol and fatty acid chains in grey. (a) "swimmer", viewed in vertical cross section through the bilayer, (b) same, but top view; (c) "submarine"; (d) "diver"; (e) phoenix; (f) "mole". Color coding of carbohydrate moieties: Man3/green, Man2/orange, Man1/red, GlcN/blue. The "flop down" and "lollipop" in Fig. 2 are reproduced by (a) and (e), respectively.

conformations (a)–(c) in Fig. 4, in combination with the flexible glycerol moiety. However, there are also embeddings with stronger contact to the surrounding lipids, see for instance (d) where Man3 and Man2 insert into the head group region ("diver"). (f) shows an extreme situation (with a conformation close to Fig. 4(c)) where the GPI core is completely buried within the head groups, contacting the hydrophobic core of the bilayer with minimum exposure to water. This "mole" actually persists over the full simulation time of the lower leaflet of the POPC bilayer, and appears to be stabilized or kinetically trapped. A similar situation is observed with extended conformations. Fig. 8(c) shows the "submarine": the stretched out GPI core persists for more than 500 ns, but only Man3 is well exposed to the solvent subphase. The submarine eventually turns into the "diver" at the end of the simulation run. The "phoenix" (e), which is close to the conformation (d) in Fig. 4, lasts as expected only for a few nanoseconds.

These (and in particular the last) examples show that it is not enough to distinguish between different conformations, but also between different levels of hydration. To fully capture the dynamics of the embedding the following observation is useful: the vertical positioning of all GPI species seems to be quite independent of conformation or carbohydrate content, see Fig. S8 and the corresponding discussion in the ESI,† Section S3. That is, in z-direction, the GPI phosphate resides at the level defined by all other phosphates in the lipid layers. We may thus picture the GPI core to move and reorient with respect to this reference level and to produce some characteristic time-dependence of the hydration pattern, complementary to conformational dynamics. This is analysed in the following.

Aspects of conformational dynamics

In addition to tracking the hydration level of all carbohydrate moieties we also consider the hydrogen bridges formed with the PC heads in each frame, see Fig. 9(a) and (b), respectively, where the time points of the snapshots in Fig. 8 have been annotated. The swimmer is characterized by an overall even exposure of all monosaccharide entities to the solvent, or very few h-bonds towards the PC heads. This situation is seen to prevail for the upper leaflets of the DMPC and POPC bilayer. The "phoenix" Fig. 8(e) observed in the upper DMPC leaflet (reproduced from Fig. 2) lasts for a few ns only; its hydration

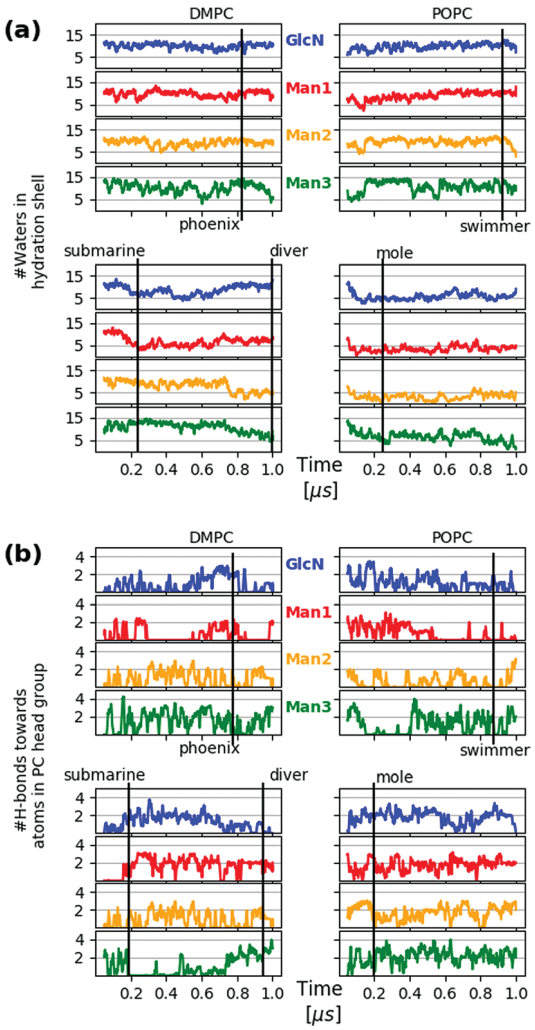


Fig. 9 Hydration level of GPI core moieties Man3, Man2, Man1 and GlcN, color coding as in Fig. 8. The upper two panels of each block refer to the upper leaflet ($z > 0$) of the respective bilayer system. The hydration level for a specific frame at time t is defined as the number of water molecules within a shell of 3 Å thickness around the selected residue. The distance criterion is applied to water oxygen only. (b) Number of hydrogen bonds of each monosaccharide towards the PC head group atoms per frame. Both quantities in (a) and (b) have been subjected to a running average with a 20-frame window in order to remove excessive noise.

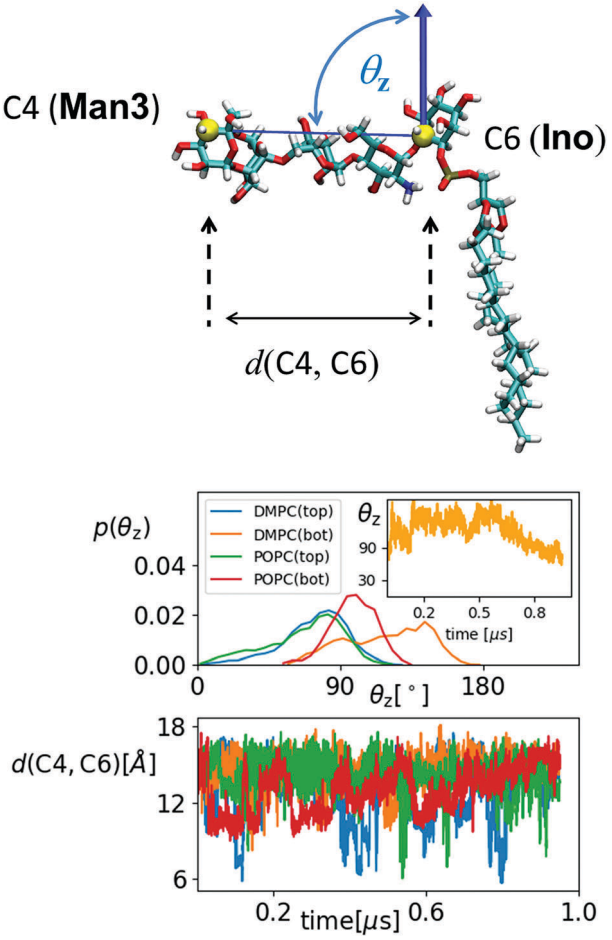


Fig. 10 Top: Schematic for defining the end-to-end distance d(C4,C6) and the tilt angle θ_z of the glycan GPI core with respect to the bilayer normal. The distribution of θ_z is shown below, separately for upper (top) and lower (bot) leaflet. The inset shows the trajectory for DMPC (bot). The lower panel shows the time dependence of d(C4,C6) for each case using the same color code.

level is similar to that of the swimmer, yet its placement is unfavorable not only because of the molecular conformation, but also because the alkyl chains have to enter the hydrophilic region to some extent.

The transition from submarine to diver is clearly visible in the lower left panels of (a) and (b), respectively: hydration is traded for h-bonds towards PC heads. The mole exhibits an even stronger interaction with the surrounding lipids and in both cases we expect that the conformational freedom of the GPI backbone itself is affected. This can be verified with the evolution of the end-to-end distance d(C4,C6) and the tilt angle θ_z of the core with respect to a vector defined by the C4 ring carbon of Man3, and C6 of inositol, see Fig. 10. For the submarine the tilt towards the z-axis persists for more than 0.6 μs (see inset) corresponding to the course of the hydration level in Fig. 9(a), until it gradually relaxes towards a lateral orientation. The mole shows little variation in tilt angle. Both configurations are additionally restrained with respect to end-to-end distance, which spans an interval of only 10–15 Å (see lower panel in Fig. 10) whereas the "swimmers" (green and blue graphs) can explore values of d(C4,C6) as low as 6 Å.

How to validate the force field for embedded GPIs?

In the previous assessment it became clear that a flop-down appearance of the GPI core prevails. In terms of steric/geometric accessibility by a protein the swimmer as a specific example naturally appears as a more plausible placement than the mole. But to determine the relative stability between a buried and a more flexible placement certainly requires advanced sampling strategies[49] beyond the scope of the present account. To judge whether the "mole" could eventually be ruled out requires yet additional work. As has already been pointed out in the computational approach, the mutual interaction of

aminoacids/protein, carbohydrates and lipids is extraordinarily difficult to characterise experimentally and thus poorly explored computationally. Consequently, one cannot expect every aspect to be calibrated with highest accuracy. The general problem of adequately tuning nonbonded interactions in crowded environments has been identified as one of the major challenges of current force field development.[50–52]

Variation of force field parameters within admissible limits is a one viable strategy to test the sensitivity (and thereby the plausibility) of a certain phenomenology.[53,54] For instance, in discussing how the conformational preferences of the hybrid link Ino-POMe emerge, extensive use has been made of the fact that use of either GLYCAM06-type or AM1-BCC charges led to indistinguishable results, see *e.g.* Fig. S3 in the ESI.† That is, without seriously affecting aspects of molecular mechanics, the hybrid link to bridge GLYCAM06 and Lipid14 domains of atom types may also be viewed as a vehicle to test nonbonded interactions systematically for instance, by altering the charge scheme on the glycan moieties, exploring different vdW combination rules between carbohydrates and lipids,[55] *etc.*

Another possibility is to immediately go on with the synthetic route. In what follows, we shall explore what we can learn from suitably extending GPI-core-PGL and provide a first account of a fully fledged model of GPI-anchored green fluorescent protein (GFP). GFP is not a natural candidate as it lacks the typical motifs of GPI anchored proteins around their C-termini,[56] but it is rather useful for several reasons. It is known to interact very little with other lipid components, and is frequently used as a marker for *in vivo* studies of cell membranes such as in the investigation of GPI analogues in supported lipid bilayers[21,22] or different re-assembled GPI–GFP into live cells.[57] Thus, GPI–GFP is a convenient starting point to study how the protein would act back on the GPI anchor and how the latter would influence orientation and proximity of the protein with respect to a membrane.

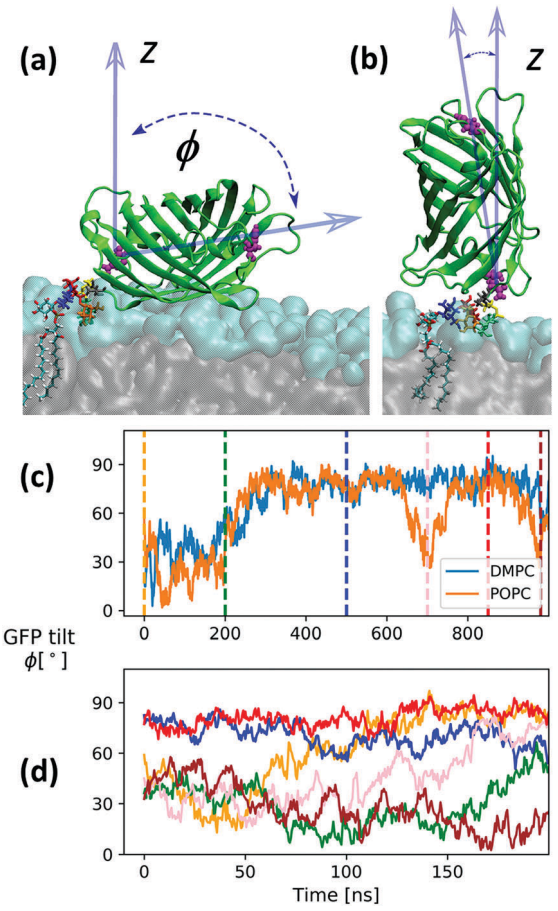


Fig. 11 Snapshots from a simulation with a full GPI–GFP inserted into a POPC bilayer as described in the text where the GFP "sleeps" on the head groups with its long axis pointing laterally (a) or stands upright (b), respectively. Panel (c) shows the corresponding evolution of long axis orientation for the GPI–GFP in POPC and DMPC, respectively. Panel (d) shows the dynamics of the long axis for trajectories restarted from the time points indicated in (c) (same color code). The times are 0 ns (initial, tweaked GPI configuration), 200 ns, 500 ns, 700 ns (snapshot (b)), 850 ns and 980 ns.

GPI-anchored green fluorescent protein

To construct a GPI anchored GFP we proceeded through the following steps. Coordinates for GFP were taken from the X-ray structure with PDB code 1EMA, and rendered a complete molecular topology using the tLeaP facility from the Amber suite. The phosphoethanolamine (PE) linker was parametrized with the GLYCAM06 force field defining a transition region between linker and GFP's C-terminus (Threonine) similar to the way the hybrid building block between GLYCAM06 and Lipid14 was derived. This is fully outlined in the ESI,† Section 4. The creation of the simulation setup followed along the same lines as described for the free GPI-anchor, except that now the bilayer patches are 16 by 16 phospholipids in size with only one phospholipid per bilayer (upper leaflet) replaced by a GPI-anchored GFP. The total number of atoms was 309 634 and 311 789 for the setup with DMPC and POPC lipids, respectively, and additional 7 Na^+ ions were introduced to compensate GFP's charged residues. All other simulation conditions and parameters are as in the case of free GPIs. Initially, the GPI–GFP points away from the bilayer surface with its long axis parallel to the bilayer normal (z-direction). The long axis is relatively well defined by a vector running through the center of masses of residues 76 (glutamine) and 135 (histidine), see Fig. 11. Since a twisted configuration of GPI-core-PGL was chosen as initial configuration (compare Fig. 6), GFP and phosphoethanolamine linker initially had no contact to the bilayer surfaces. In the course of 1 μs long simulations for each bilayer system, the GPI core within 200–300 ns flops down onto the bilayer assuming a hook-like conformation reminiscent of the "diver", with the GFP barrel resting on the head group region with the long axis pointing along the bilayer surface, compare Fig. 11(a); the "flop down", remarkably, also allows for a completely upright orientation conveyed by the PE linker (b). Panel (c) shows the evolution of GFP orientation for DMPC and POPC. Both simulations indicate that the parallel orientation (the "sleeping GFP")

should dominate, but the re-orientational dynamics is quite slow, occurring roughly over 200 ns. We therefore restarted the simulation of GPI–GFP embedded in POPC from 6 different time points along the trajectory indicated by the dashed lines in (c), with new initial velocities (simulation time 200 ns each). The corresponding trajectories are shown in (d). Four of these show a tilt angle of smaller than 60° initially. The trajectory in orange starts from the same initial configuration as the one in panel (c), but now relaxing somewhat faster to parallel orientation. After 200 ns, only in one run the protein still stands upright, in all other cases the sleeping GFP appears to evolve. In this way, our current model makes a rather clear prediction about the orientation of the GFP, and this could be verified experimentally by employing anisotropic fluorescence spectroscopy[58] because the GFP determines the orientation of the chromophore inside.

We note that we verified that interaction of GFP with lipid–water interface is weak. When brought into initial contact with the head group region (sleeping or upright orientation) the GFP detaches from the bilayer surface within 100–200 ns driven by diffusion and bilayer undulations.

By analysing all trajectories of Fig. 11(d), we can make further observations:

(i) Fig. S9(a) in the ESI† shows that taken together, the orientation distribution of the end-to-end vector from C6 (inositol) to C4 (Man3) is in accord with those of the swimmer-like configurations seen to dominate in the POPC and DMPC top leaflet of the free GPI anchor GPI-core-PGL, compare Fig. 10; the attachment to the protein appears to stabilizes the end-to-end distance d(C4,C6) at $\sim$1.4 nm, see Fig. S9(b) (ESI†). However, unlike the swimmer of GPI-core-PGL where the Ω glycosidic angle of the αMan1→6Manα linkage between Man2 and Man1 exhibits mostly the *gg* rotamer, in GPI–GFP it is exclusively *gt*, compare also Fig. 4. This actually causes Man1 to stick out of the headgroups (compare red-colored Man1 in Fig. 11(a)) more than Man2 or Man3.

(ii) The GPI core does not notably influence GFP's orientation; it is by contrast the PE linker and its flexibility that allows the GFP to explore a range of different orientations. It may even ''twist back'' to interact with in principle all glycan moieties. In Fig. S10 (ESI†) we characterise these interactions by displaying the dynamic formation of hydrogen bonds. The fact that an attached protein may quite extensively interact with its own GPI glycan could be a very important hint for possible NMR experiments where, for instance, a GPI-anchored protein is inserted into a well defined and suitable environment such as a small micelle[24] or a phospholipid nanodisc.[59]

Conclusions

We can state that the hybrid GPI anchor models developed in this work turn out to be quite useful. They exhibit a broad variety of conformational modes by which the GPI glycan can interact with the headgroup–solvent interface facilitated by the dominating hook-like appearance of GPI-core-PGL, which causes the GPI core glycan to flop down onto the lipid bilayer but nevertheless retains a high degree of internal flexibility. This may be contrasted to other glycolipid species such as the ganglioside GM3, which is inherently stiffer and rich in β-type glycosidic linkages, making the glycan headgroup point away from the bilayer surface.[25] In contrast we can conclude that a ''lollipop'' picture of a GPI anchor (represented, *e.g.*, by the ''phoenix'') is rather implausible. The preference of GPI-core-PGL for hook-like conformations can be expected to be a robust and generic phenomenon, as it is in particular conveyed by the proximity and mutual intramolecular interaction of amine- and phosphate groups, a mechanism retained after attaching GFP. The proximity to the bilayer surface and the flexibility of orientation conveyed by the phospho-ethanolamine linker permit the GFP to interact quite extensively with its own GPI-glycan. Naturally the question arises about the consequences of adding various side chains to Man1 such as in the case of, *e.g.*, *Toxoplasma gondii*.[60] Whether a specific side chain would impose a certain preference of orientation or stabilize a favorable protein fold are then indeed aspects that could be studied numerically and experimentally giving valuable insights into the general purpose of GPI anchoring.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We thank the International Max Planck Research School (IMPRS) for Multiscale Biosystems for support. Open Access funding provided by the Max Planck Society.

References

1 M. G. Paulick and C. R. Bertozzi, *Biochemistry*, 2008, **47**, 6991–7000.
2 K. Simons and D. Toomre, *Nat. Rev. Mol. Cell Biol.*, 2000, **1**, 31–39.
3 E. Klotzsch and G. J. Schütz, *Philos. Trans. R. Soc., B*, 2013, **368**, 20120033.
4 M. G. Low and A. R. Saltiel, *Science*, 1988, **239**, 268–275.
5 S. D. Tachado, S. D. Gerold, R. Schwarz, S. Novakovich, M. McConvill and L. Schofield, *Proc. Natl. Acad. Sci. U. S. A.*, 1997, **94**, 4022–4027.
6 S. Mayor and H. Riezmann, *Mol. Cell. Biol.*, 2004, **5**, 110.
7 D. A. Brown and J. K. Rose, *Cell*, 1992, **6**, 533–544.
8 Y. Maeda and T. Kinoshita, *Prog. Lipid Res.*, 2011, **50**, 411–424.
9 B. Brügger, C. Graham, I. Leibrecht, E. Mombelli, A. Jen, F. Wieland and R. Morris, *J. Biol. Chem.*, 2004, **279**, 7530–7536.
10 D. Goswami, K. Gowrishankar, S. Bilgrami, S. Ghosh, R. Raghupathy, R. Chadda, R. Vishwakarma, M. Rao and S. Mayor, *Cell*, 2008, **135**, 1085–1097.
11 S. Schuck and K. Simons, *J. Cell Biol.*, 2006, **172**, 963–965.

12 C. Eggeling, C. Ringemann, R. Medda, G. Schwarzmann, K. Sandhoff, S. Polyakova, V. N. Belov, B. Hein, C. von Middendorf, A. Schönle and S. W. Hell, *Nature*, 2009, **457**, 1159–1163.
13 E. Sevcsik, M. Brameshuber, M. Fölser, J. Weghuber, A. Honigmann and G. J. Schütz, *Nat. Commun.*, 2015, **6**, 6969.
14 F. Ronzon, S. Morandat, B. Roux and M. Bortolato, *J. Membr. Biol.*, 2004, **197**, 167–177.
15 S. W. Homans, C. J. Edge, M. A. J. Ferguson, R. A. Dwek and T. W. Rademacher, *Biochemistry*, 1989, **28**, 2881–2887.
16 M. T. Lehto and F. J. Sharom, *Biochemistry*, 2002, **41**, 8368–8376.
17 E. Barboni, B. P. Rivero, A. J. T. George, S. R. Martin, D. V. Renouf, D. F. Hounsell, P. C. Barber and R. J. Morris, *J. Cell Sci.*, 1995, **108**, 487–497.
18 P. Bütikofer, T. Malherbe, M. Boschung and I. Roditi, *FASEB J.*, 2001, **15**, 545–548.
19 F. J. Sharom, I. Lorimer and M. P. Lamb, *Can. J. Biochem. Cell Biol.*, 1985, **63**, 1049–1057.
20 M. Fujita and T. Kinoshita, *Biochim. Biophys. Acta*, 2012, **1821**, 1050–1058.
21 M. G. Paulick, M. B. Forstner, J. T. Groves and C. R. Bertozzi, *J. Am. Chem. Soc.*, 2007, **129**, 11543–11550.
22 M. G. Paulick, M. B. Forstner, J. T. Groves and C. R. Bertozzi, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 20332–20337.
23 F. Chevalier, J. Lopez-Prados, S. Perez, M. Martín-Lomas and P. M. Nieto, *Eur. J. Org. Chem.*, 2006, 3489–3498.
24 F. Chevalier, J. Lopez-Prados, P. Groves, S. Perez, M. Martín-Lomas and P. M. Nieto, *Glycobiology*, 2006, **16**, 969–980.
25 M. L. DeMarco and R. J. Woods, *Glycobiology*, 2008, **18**, 426.
26 K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. González-Outeiriño, C. R. Daniels, B. L. Foley and R. J. Woods, *J. Comput. Chem.*, 2008, **29**, 622–655.
27 C. J. Dickson, B. D. Madej, A. A. Skjevik, R. M. Betz, K. Teigen, I. R. Gould and R. C. Walker, *J. Chem. Theory Comput.*, 2014, **10**, 865–879.
28 J. Zuegg and J. E. Gready, *Glycobiology*, 2000, **10**, 959–974.
29 This is the reason why we did not stay with GLYCAM06 to describe lipids as well (presented in Tessier 2008): under ambient conditions, the GLYCAM lipid was awkward to work with. Bilayers frequently entered gel-like phases, even after careful preparation.
30 C. Stefaniu, I. Vilotijevic, M. Santer, D. Varon-Silva, G. Brezesinski and P. H. Seeberger, *Angew. Chem., Int. Ed.*, 2012, **51**, 12874–12878.
31 C. Stefaniu, I. Vilotijevic, M. Santer, G. Brezesinski, P. H. Santer and D. Varón-Silva, *Langmuir*, 2014, **30**, 5185–5192.
32 E. Schneck, T. Schubert, O. V. Konovalov, B. E. Q. T. Gutsmann, K. Brandenburg, R. G. Oliviera, D. A. Pink and M. Tanaka, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 9147–9151.
33 K. N. Kirschner, R. D. Lins, A. Maass and T. A. Soares, *J. Chem. Theory Comput.*, 2012, **8**, 4719–4731.
34 A. A. Skjevik, B. D. Madej, R. C. Walker and K. Teigen, *J. Phys. Chem. B*, 2012, **116**, 11124–11136.
35 R. J. Woods, R. A. Dwek, C. J. Edge and B. Fraser-Reid, *J. Phys. Chem.*, 1995, **99**, 3832–3846.
36 O. Guvench, E. Hatcher, R. M. Venable, R. W. Pastor and A. D. MacKerell Jr., *J. Chem. Theory Comput.*, 2009, **5**, 2353–2370.
37 R. W. Pastor and J. A. D. MacKerell, *J. Phys. Chem. Lett.*, 2011, **2**, 1526–1532.
38 A. Jakalian, B. L. Bush, D. B. Jack and C. I. Bayly, *J. Comput. Chem.*, 2000, **21**, 132–146.
39 The corresponding carbon (bonded to the phosphate O32) of the glycerol moiety is left as cA because the division between force fields is defined as one between atom types; the specialization to a type such as Cp would be part of a general revision of the lipid force field.
40 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics*, 1996, **14**, 33–38.
41 M. Wehle, I. Vilotijevic, R. Lipowsky, P. H. Seeberger, D. Varón-Silva and M. Santer, *J. Am. Chem. Soc.*, 2012, **134**, 18964–18972.
42 D. A. Case, V. Babin, J. T. Berryman, R. M. Betz, Q. Cai, D. S. Cerutti, I. T. E. Cheatham,. T. A. Darden, R. E. Duke, H. Gohlke, A. W. Goetz, S. Gusarov, N. Homeyer, P. J. J. Kaus, I. Kolossváry, A. Kovalenko, T. S. Lee, S. LeGrand, T. Luchko, R. Luo,. B. Madej, K. M. Merz, F. Paesani, D. R. Roe, A. Roitberg, C. Sagui,. R. Salomon-Ferrer, G. Seabra, C. L. Simmerling, W. Smith, J. Swails, R. C. W. J. Wang, R. M. Wolf, X. Wu and P. A. Kollman, *AMBER14*, University of California, San Francisco, 2014.
43 D. V. der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. C. Berendsen, *J. Comput. Chem.*, 2005, **26**, 1701–1718.
44 B. Hess, C. Kutzner, D. V. der Spoel and E. Lindahl, *J. Chem. Theory Comput.*, 2008, **4**, 435–447.
45 E. J. Sorin and V. S. Pande, *Biophys. J.*, 2005, **88**, 2472–2493.
46 H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684–3690.
47 W. F. van Gunsteren and H. J. C. Berendsen, *Mol. Simul.*, 1988, **1**, 173–185.
48 G. J. Martyna, M. L. Klein and M. J. Tuckerman, *J. Chem. Phys.*, 1992, **97**, 2635–2643.
49 L. Monticelli, E. J. Sorin, D. P. Tieleman and V. S. P. G. Colombo, *J. Comput. Chem.*, 2008, **29**, 1740–1752.
50 D. Petrov and B. Zagrovic, *PLoS Comput. Biol.*, 2014, **10**, e1003638.
51 S. Rauscher, V. Gapsys, M. J. Gajda, M. Zweckstetter, B. L. de Groot and H. Grubmüller, *J. Chem. Theory Comput.*, 2015, **11**, 5513–5524.
52 D. J. Cole, J. Z. V. J. Tirado-Rives, M. C. Payne and W. L. Jorgensen, *J. Chem. Theory Comput.*, 2016, **12**, 2312–2323.
53 S. Abel, F.-Y. Dupradeau, E. P. Raman, A. D. MacKerell Jr. and M. Marchi, *J. Phys. Chem. B*, 2011, **115**, 487–499.
54 J. A. Hadden, A. D. French and R. J. Woods, *Cellulose*, 2014, **21**, 879–884.
55 T. A. Halgren, *J. Am. Chem. Soc.*, 1992, **114**, 7827–7843.
56 M. A. J. Ferguson, T. Kinoshita and G. W. Hart, in *Essentials of Glycobiology*, ed. A. Varki, R. D. Cummings, J. D. Esko,

H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart and M. E. Etzler, Cold Spring Harbour, New York, 2nd edn, 2009, ch. 11, pp. 143–161.
57 D. F. Legler, M.-A. Doucey, P. Schneider, L. Chapatte, F. C. Bender and C. Bron, *FASEB J.*, 2005, 19.
58 J. V. Rocheleau, M. Edidin and D. W. Piston, *Biophys. J.*, 2003, **84**, 4078–4086.
59 I. G. Denisov, Y. V. Grinkova, A. A. Lazarides and S. G. Sligar, *J. Am. Chem. Soc.*, 2004, **126**, 3477–3487.
60 S. Götze, N. Azzouz, Y. H. Groß, A. Reinhardt, C. A. P. H. Seeberger and D. V. Silva, *Angew. Chem., Int. Ed.*, 2014, **53**, 13701–13705.
61 A. Singh, M. B. Tessier, K. Pederson, A. P. Wang, A. P. Venot, G.-J. Boons, J. H. Prestegaard and R. J. Woods, *Can. J. Chem.*, 2016, **94**, 927–935.

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.

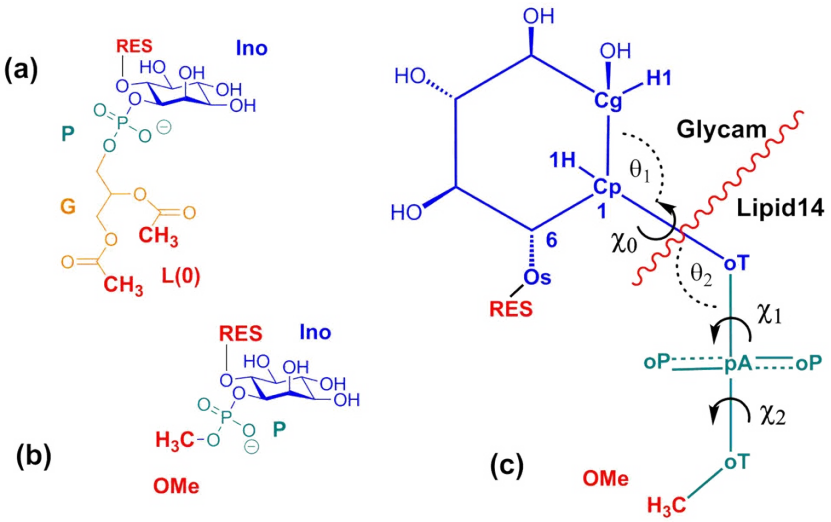


Figure S1. (a) Hybrid link reproduced from Fig. 3(a) of the main text, with RES=CH3 (6OMe-Ino-PGL(0) or RES=H (Ino-PGL(0)), and (b) reproduced from Figure3(b) with RES=CH3 (6OMe-Ino-POMe) or RES=H (Ino-POMe). (c) schematic version of (b) with atom names and types.

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 $\grave{V}_i$ 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 $\grave{V}_i$ 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

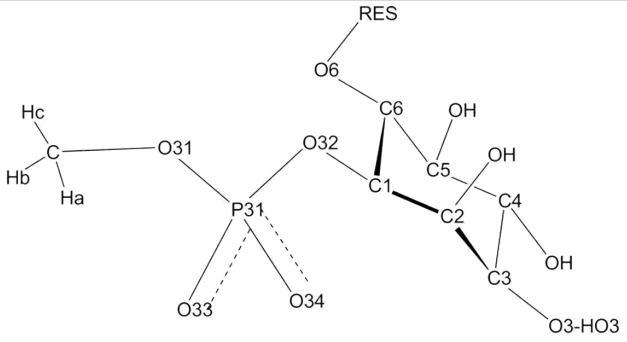
ambient conditions including solvent an average over a conformational ensemble is performed. Especially for molecules that can assume a number of conformations (such as the relative orientations of exocyclic torsions in carbohydrate rings), not only a single reference geometry (e.g. a minimum energy 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.

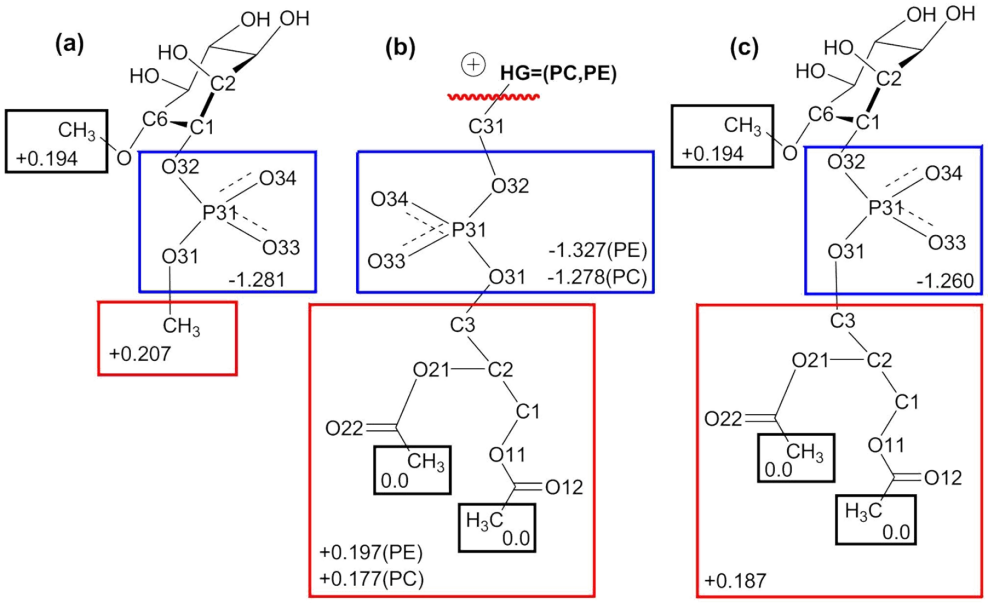


Figure S2. Chemical compounds used for allocation of partial charges on the hybrid link 6OMe-Ino-PGL(0), depicted in (c). (a) 6-OMe-Ino-POMe (b) Phosphocholine (PC-) and Phospho-Ethanolamine (PE)-glycerol. Black solid frames indicate capping groups: the CH_3 moiety at the O6 oxygen represents the link 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.

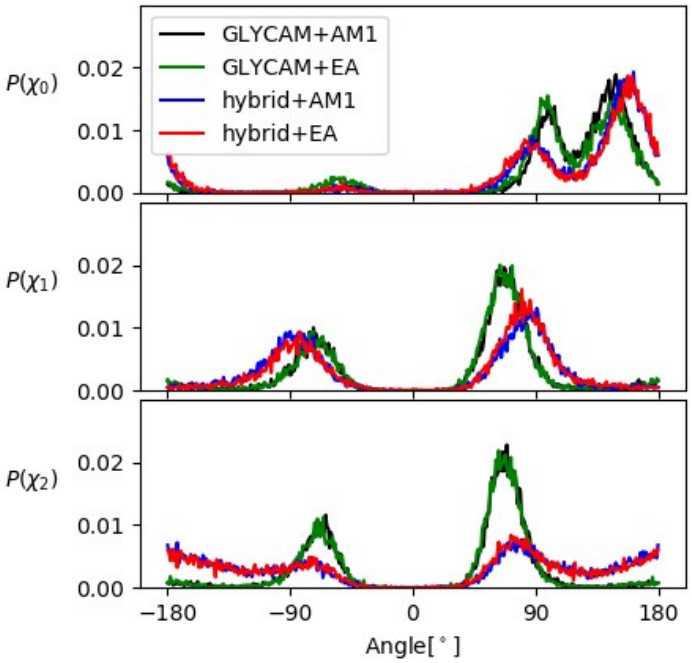


Figure S3: Density distributions $P(\chi_0)$ to $P(\chi_2)$ for Ino-POMe (structure (e) in Figure S1). Black and green graphs correspond to bonded interactions taken entirely from the GLYCAM06 force field, blue and red employ the rederived parameters of this work (Table S2). Simulation time is 100ns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

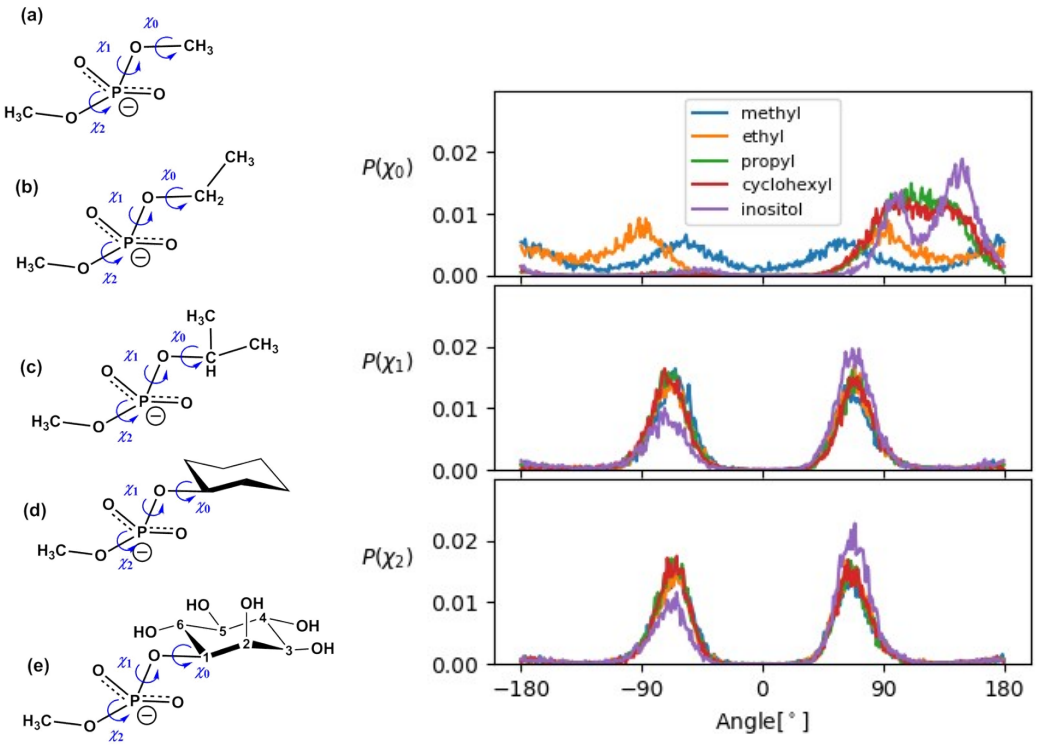


Figure S4: Left: phosphatidyl moieties with methyl capping and (a) methyl, (b) ethyl, (c) propyl, (d) cyclohexyl and (e) inositol head. The graphs to the right show the corresponding distribution of the torsion angles χ_0-χ_2 obtained from 100ns simulations at 303K in TIP3P water using the GLYCAM06 parameter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

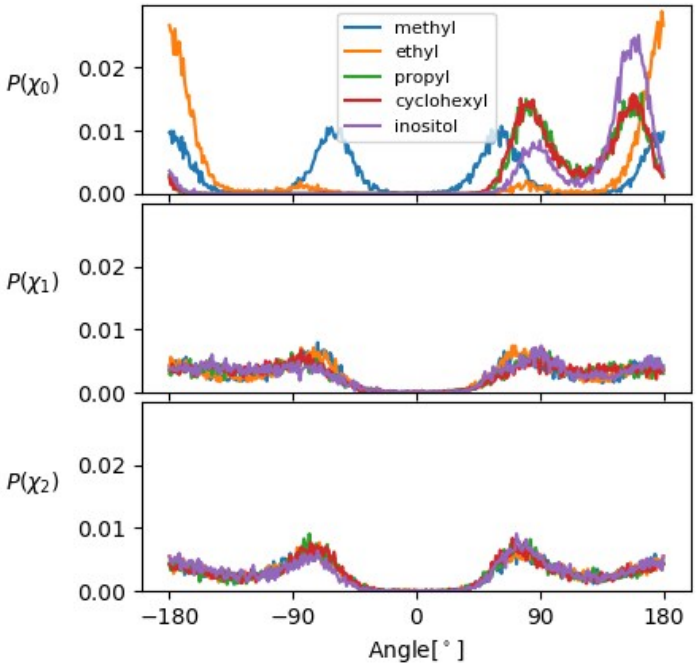


Figure S5: Torsion angle distribution of species (a) to (e) (see Figure S4) employing Lipid14 force field parameters.

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.

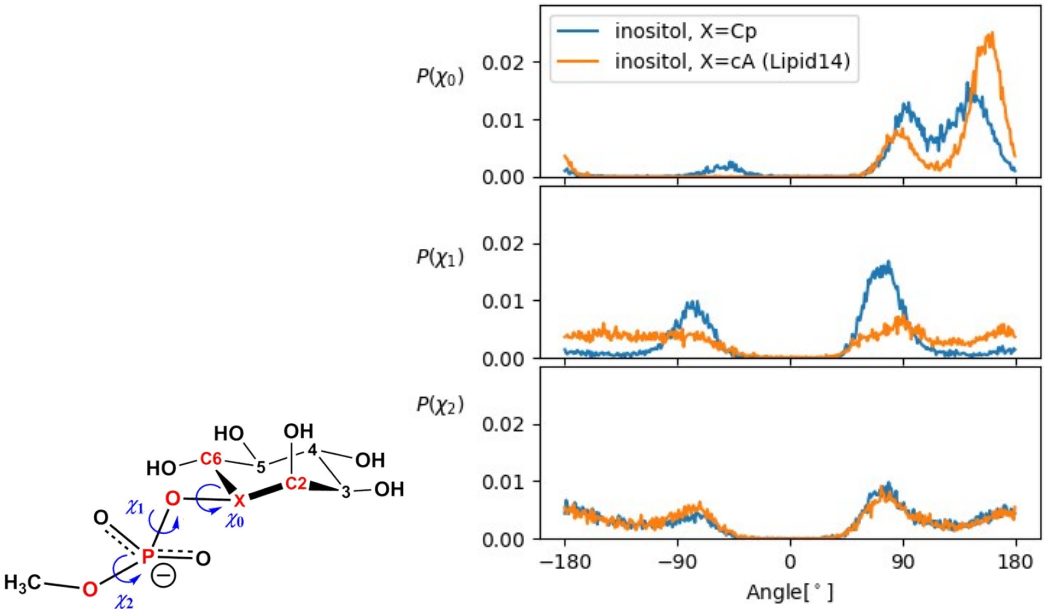
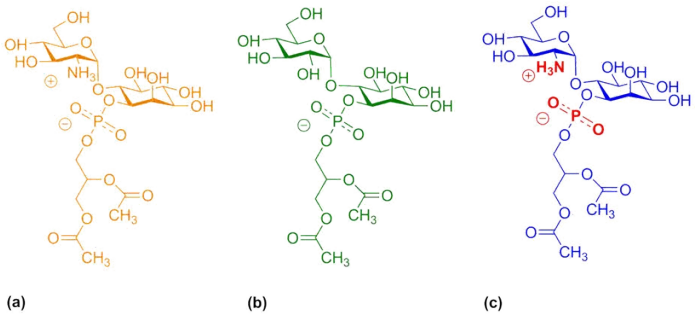


Figure S6: Torsion angle distribution in Ino-POMe, using the force field parameters for Lipid14 and two different atom types for the C1 carbon marked with X (see chemical schematic to the left). Orange solid lines: all-Lipid14 parameter set, X=cA. Blue solid lines: in all sequences containing X, (such 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.

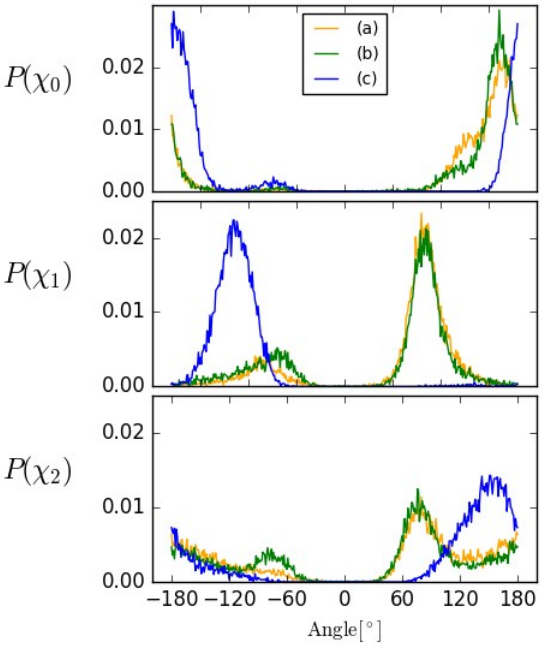


Figure S7: Density distributions $P(\chi_0)$ to $P(\chi_2)$ for the three molecules depicted in scheme S2 above. Simulation time: 100ns each, at 303K in TIP3P water.

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.

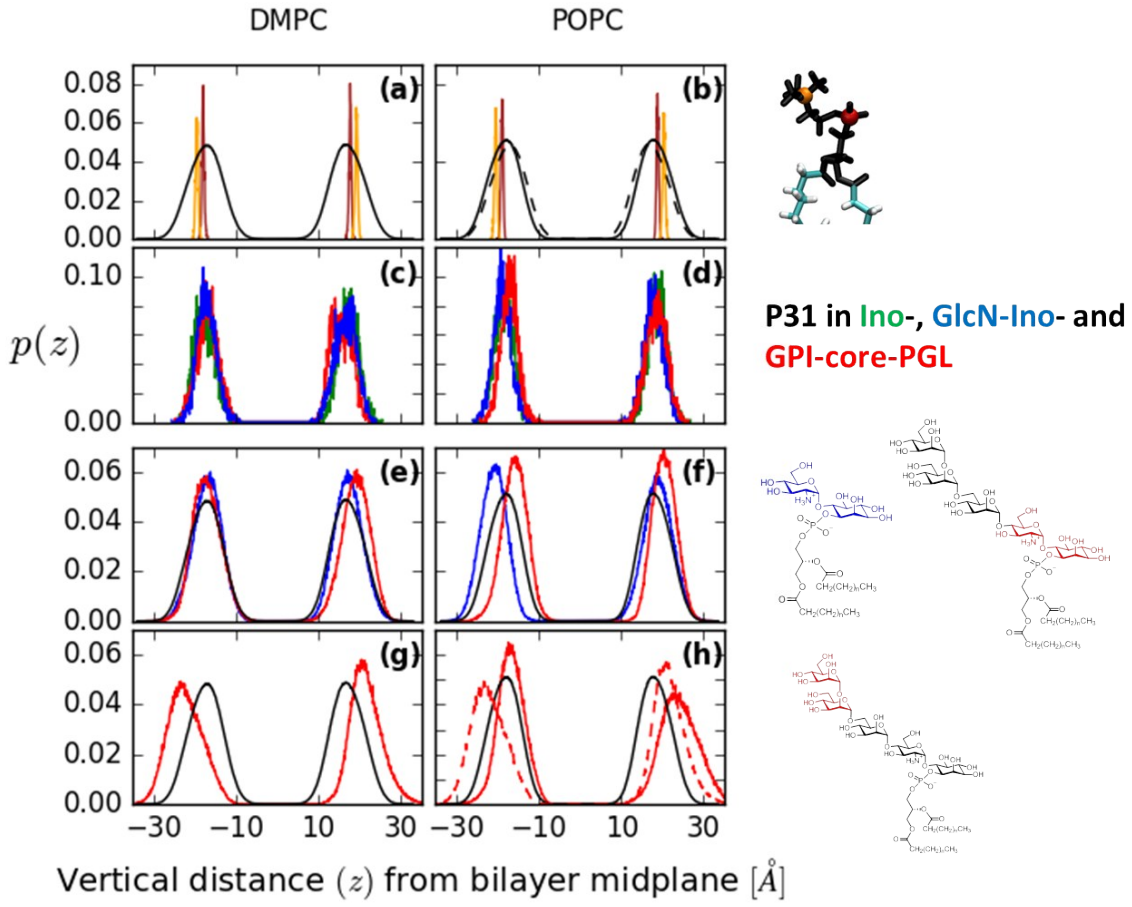
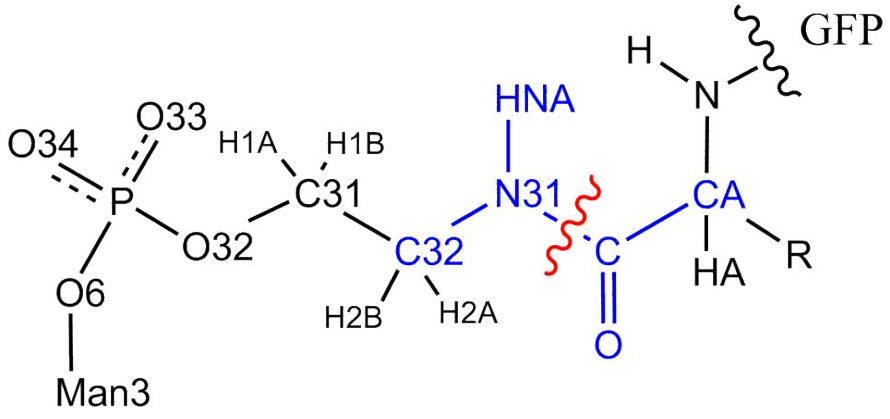


Figure S8. Vertical distribution $p(z)$ of selected atom groups in Ino-, GlcN-Ino- and GPI-core-PGL.

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.

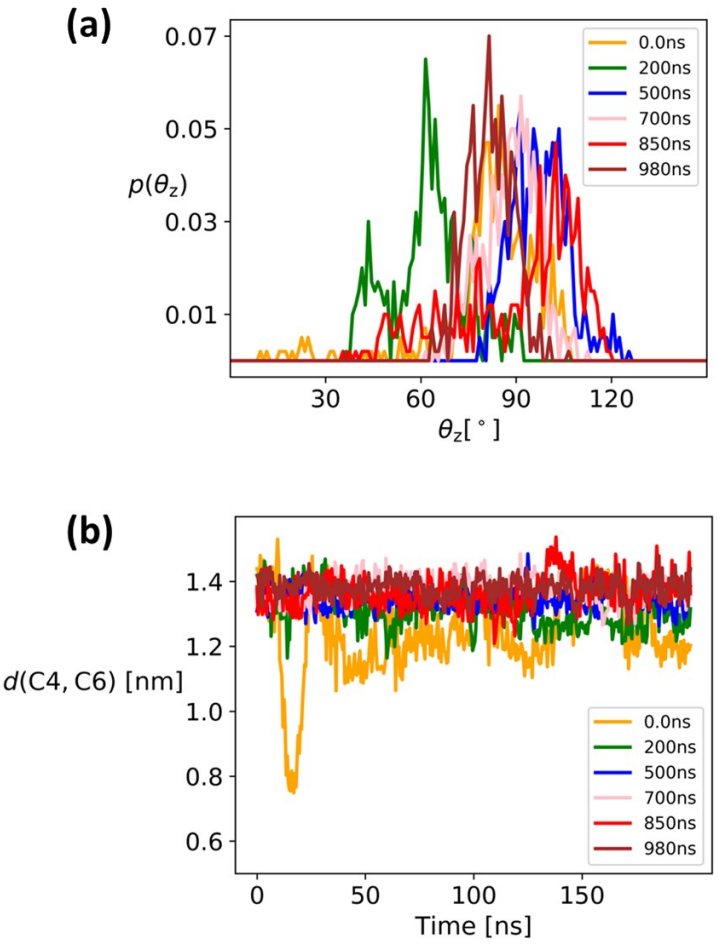


Figure S9: (a) distribution of the GPI-core orientation as defined by angle formed by the bilayer normal and the vector connecting C6 of inositol with C4 of Man3, compare Figure 10 in the main text. (b) distance distribution between these atoms. Bot plots comprise the six trajectories restarted from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

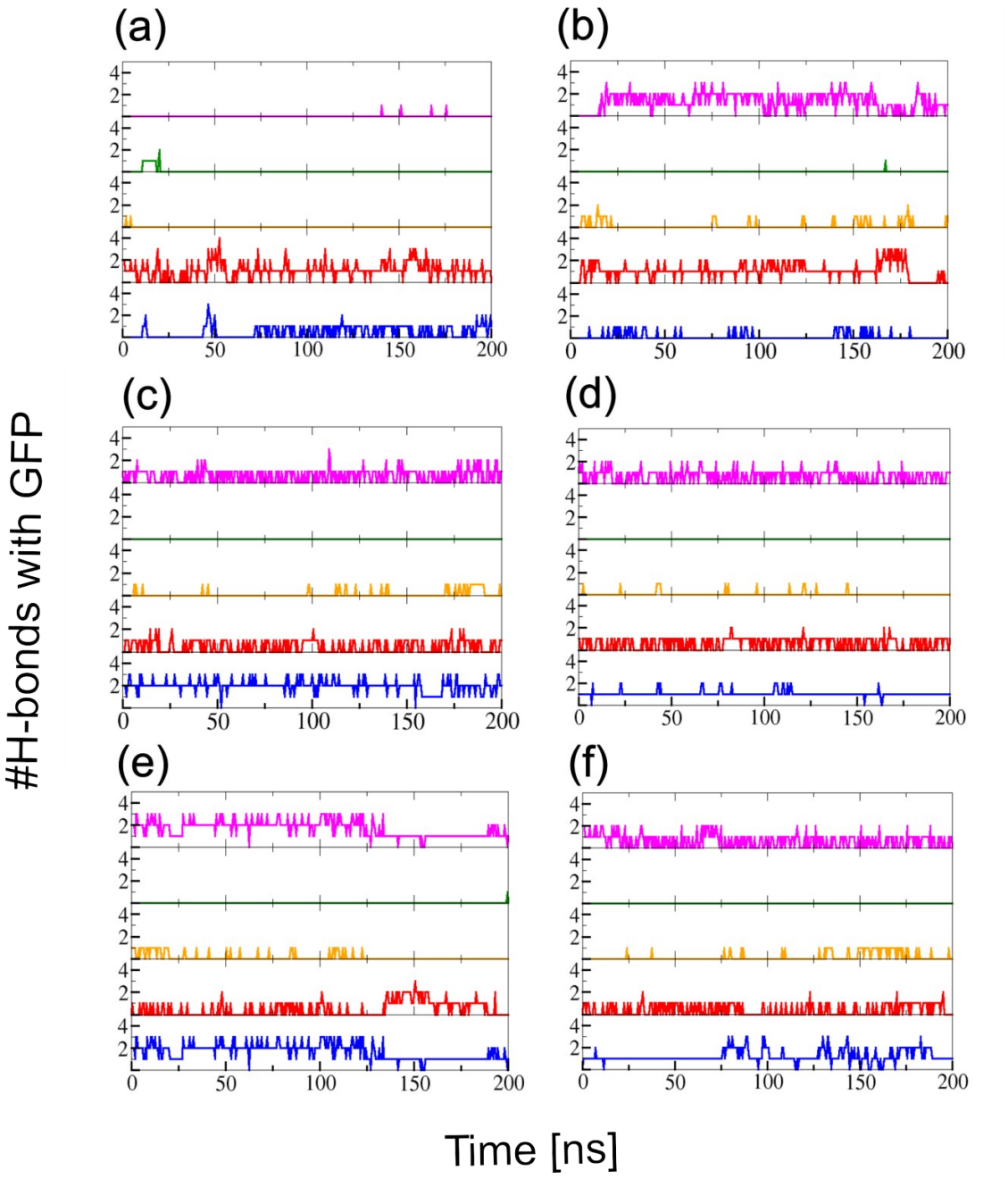


Figure S10: Number of hbonds formed between GFP and the glycan moieties of GPI core in the 6 trajectories displayed in Figure 11 of the main text and Figure S9. The data from (a) to (f) correspond to the different time points of simulation restart with freshly chosen initial velocities for configurations 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.

Chapter 3

The Importance of side branches of Glycosylphosphatidylinositol Anchors: A Molecular Dynamics perspective

3.1 Overview

Here, we build up on the model of the GPI core from the previous paper to systematically study three different parasitic GPI forms – *Toxoplasma gondii*, *Toxoplasma gondii* Low Molecular Weight, *Trypanosoma congolense* – and GPI found in humans. These GPIs share the conserved GPI glycan core but differ from each other in the chemical composition and the number of side chain sugar residues and phosphoethanolamine units. Simulations on the microsecond timescale were performed with the GPIs inserted in pure DMPC and POPC bilayers. We show that the flop-down conformation of the GPI glycan on the bilayer surface is maintained across all the GPIs, however, the human GPI shows a slightly higher tendency to stand upright.

By a comparative analysis of the glycosidic linkages, we show that the most notable differences in the internal conformation of the GPIs occur in the flexible Man2-$\alpha(1 \rightarrow 6)$-Man1 linkage, the molecular hinge of the GPI backbone. However, another highly flexible Gal-$\beta(1 \rightarrow 6)$-GlcNac linkage present in the side chain of *T.congolense* GPI imparts additional remarkable flexibility to its structure. The rotamer populations characterizing the Man2-Man1 linkage are significantly affected with increasing molecular size, in other words increased branching, of the GPIs. Through cluster analysis we show that all the glycosidic linkages of the most dominant clusters do not necessarily occupy their respective lowest energy regions. This is especially true for the higher branched and more flexible GPI structures. We then go on to demonstrate the importance of the side branches through our observations that these are the least buried and most solvent-accessible residues in the molecule, excepting Gal in *T.congolense* GPI that

is quite embedded in the lipid heads. We discuss how the general flop-down alignment of the GPIs relates to the presentation of the side branches to the solvent. Moreover, we compare the structural characteristics of human GPI with parasitic GPIs. Our results agree with several experimental findings that have reported the crucial role played by the side chain residues of GPIs in eliciting an immune response.

Finally, we attach the Green Fluorescent Protein (GFP) to all the parasitic GPIs to study the dependence of the orientation of the protein on different levels of GPI branching. We observe that the presentation of the protein on the bilayer surface is sensitive to not just the size of the GPI, but also the difference in chemistry and linkage type of the side chain residues. Moreover, the type of lipid bilayer and the region of the protein in contact with the GPI glycan also make a difference. Regardless of the orientation, it is clear that the protein is brought close to the bilayer through interactions with the GPI. We discuss the space-filling role of the GPI glycan, and the side branches thereof, between the bilayer and protein with particular relevance to the membrane-protective function of the Variant Surface Glycoprotein found in Trypanosomes.

This paper thus provides a molecular basis for the important role played by the side chain residues across different GPI variants, thereby, attempting to fill the gap in knowledge of the structure-function relationship of the GPI anchor.

The importance of side branches of glycosylphosphatidylinositol anchors: A molecular dynamics perspective

Pallavi Banerjee, Reinhard Lipowsky and Mark Santer

Abstract

A large family of peripheral proteins are anchored to the cell surface of eukaryotes with the help of a unique category of glycolipids called the glycosylphosphatidylinositol (GPI) anchor. These glycolipids can also exist without a protein especially on the cell surfaces of protozoan parasites where they are densely populated. GPIs together with the attached proteins participate in a wide array of vital cellular functions such as signal transduction, cell adhesion, protein trafficking and pathogenesis of diseases like Malaria, Toxoplasmosis, African sleeping sickness, prion diseases, to name a few. All GPI anchors share a common conserved glycan core, however, they structurally differ from each other in the number/type/point-of-attachment of side branches of additional sugar or phosphoethanolamine residues. Experiments have shown that both GPI isolates and synthetic GPIs act as toxins and activate immune responses to the effect of prolonging survival. In this work, we extend the model of the GPI core to carry out a systematic study of the structural aspects of GPIs carrying different side chains – *Toxoplasma* GPI, *Toxoplasma* Low Molecular Weight GPI, *Trypanosoma congolense* GPI and human/mammalian GPI inserted in pure lipid bilayers. Our results demonstrate the importance of the side branch residues as these are the most accessible, and thereby, recognizable epitopes. This finding aligns well with experimental observations that highlight the role of the side branches in recognition of GPIs by macrophages and in lending specificity to immune responses. The overall flop-down orientation of the GPIs with respect to the membrane surface presents the side chain residues to face the solvent. We investigate protein-GPI interactions by attaching the Green Fluorescent Protein (GFP) to the parasitic GPIs. By lying in a reclining manner on the bilayer surface, the GPI pulls down the attached GFP close to the bilayer. However due to flexibility of the phosphoethanolamine linker and of the loop residues of the protein that make intermittent contacts with GPI, the orientation and thus, the accessibility of GFP is subject to fluctuations and is sensitive to both the type of GPI and lipid bilayer. Our GPI model can be extended to study many other parasitic GPIs and GPI-anchored proteins, to potentially contribute to the immunoprophylaxis of related diseases.

1 Introduction

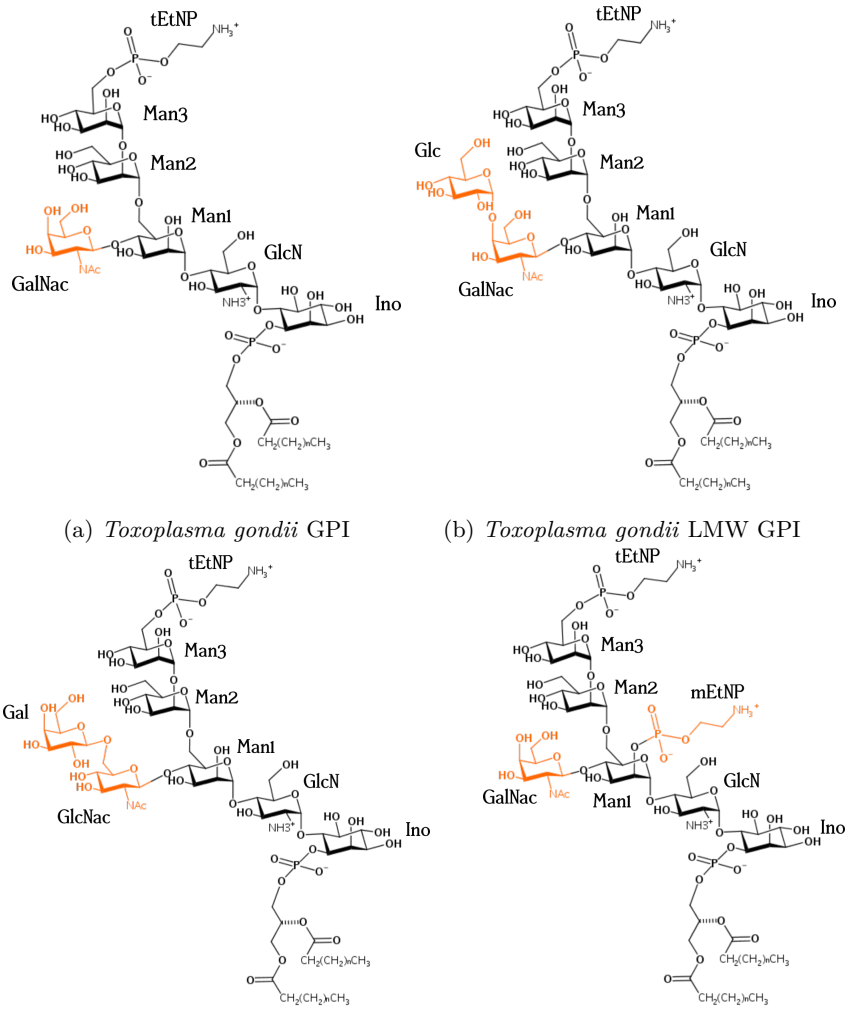
Glycosylphosphatidylinositols (GPI) are glycolipids of a complex chemical nature that are found as peripheral decorations in most eukaryotic cells, from protozoa through to humans, and have been ascribed to a large spectrum of cellular functions. They can exist either with or without an anchored protein, but invariably tethered to the outer leaflet of the cell membrane. GPIs attach to proteins at the C-terminus end by way of a post-translational modification that occurs in the endoplasmic reticulum. A phosphoethanolamine moeity bridges GPI-anchored proteins (GPI-APs) to GPIs. The structure of a GPI consists of a unique, conserved pseudopentasaccharide core, termed as the GPI core: Man$-\alpha(1\rightarrow 2)-$Man$-\alpha(1\rightarrow 6)-$Man$-\alpha(1\rightarrow 4)-$GlcN$-\alpha(1\rightarrow 6)-$myo-inositol, which extends to a lipid tail which in turn injects into the cell membrane. This core could be variably branched to include more sugars or additional phosphoethanolamine units as extending side chains. Although the glycan core remains the same in all GPIs, its many branched variations, different lipid tails and anchored proteins together contribute to a diverse set of cellular phenomena. Apart from being a stable anchoring device for the protein, GPIs along with GPI-APs have been implicated in key cellular functions such as signal transduction, cellular adhesion and communication, protein sorting and trafficking, and antigen presentation. Such multipotentiality of the GPI anchor makes it a subject of great interest.

Since its discovery more than 30 years from now [1, 2, 3], GPI anchor has consistently been a hot subject among biochemists. Several valuable questions are intensely debated, like, the conformation, the role in raft formation and protein sorting, and antigenic activity of parasitic GPIs. There has been a dichotomy in conclusions over the orientation of GPIs with respect to cell membranes. Some suggest that the GPI stands upright avoiding interaction with the membrane[4, 5], whereas others indicate that it lies in close proximity to the membrane consequently

pulling down the attached protein too[6, 7, 8]. The subject of preferential partitioning of GPI-anchored proteins into liquid-ordered domains/rafts of cell membranes is also rife with controversy [9, 10, 11]. One of the main reasons behind the prevailing controversies is that experimental studies on the GPI are rather limited. This is because the amphiphilic and heteregeneous nature of the GPI makes it difficult to isolate its pure forms, that prevents an accurate functional analysis. Synthesis of GPIs, especially of GPIs bearing unsaturated lipid tails and GPI-anchored protein, is challenging as the chemical routes are convoluted and enzymes not well-characterized. However, promising efforts are underway in that direction[12, 13]. GPIs are known to actively participate in the development of a slew of diseases caused by protozoan parasites, like Toxoplasmosis, Malaria, Chagas, Prion-protein abnormalities among many others[14]. The involved parasites are enriched with GPI-anchored proteins, and a lot more of free GPIs on their cell surfaces serving as weapons for invasion. One way of tackling GPI-related pathologies is by generating synthetic GPI anchors to serve as vaccines. By way of experimental studies conducted both *in vitro* and *in vivo*, it has been shown that GPIs, even without the protein, trigger immune responses. Through SPR experiments on human-galectin3 and *Toxoplasma gondii* GPIs, Debierre-Grockiego and coworkers concluded that galectin-3 dependent recruitment of effector cells is an important step in the process of recognition of parasitic GPI glycans by macrophages[15]. Synthetic GPI form of *T.gondii* has been demonstrated to act as a diagnostic biomarker, along with the ability to differentiate between stages of the disease[16]. In case of trypanosomes, it was shown that the administration of the galactose-modified GPI glycan core to infected mice induces TNF production by macrophages, thereby abrogating pathology and prolonging survival[17]. Synthetically prepared malarial GPI injected to infected mice conferred substantial protection against fatalities, hence, qualifying as a suitable candidate for vaccine therapy[18].

The functional variety and the importance of GPI anchors in regulating vital cellular phenomena are well known by now. However, due to the challenges associated with their experimental investigation, the structure-function relationship hasn't been elucidated yet. Molecular dynamics (MD) serves as a useful complementary technique in unravelling the structural and dynamical aspects of a system at the atomic resolution. There have been considerable efforts made towards modelling the GPI anchor in membrane-like environments like that of the GPI of *Trypanosoma brucei* Variant Surface Glycoprotein[2], GPI of NETNES glycopeptide of *Trypanosoma cruzi*[19], and of the human prion protein[20, 21]. Nonetheless, these studies have been conducted within only the nanosecond timescale. Considering the conformational flexibility of GPIs, longer simulation times are required for adequate sampling, at least to the microsecond timescale. Using microsecond-long MD simulations, Wu et al. discovered that the secondary structure of the prion protein is stabilized through mutual interactions of the protein, its GPI anchor and N-glycans, and the membrane[22]. Li et al conducted biased MD simulations to report that GPIs spontaneously insert into liquid-ordered phase of the membrane[23]. However, a thorough understanding of the differing structural features of various GPI types and their impact on the attached protein is still lacking. In our previous work, we constructed the solution form of the GPI glycan (without the tail) using the GLYCAM06 force-field, wherein we reported that simulation time of at least 1μs is required to achieve sufficient sampling[24]. In a succeeding paper, we attached the aforementioned GPI glycan to a lipid tail, and further to the Green Fluorescent Protein (GFP), to study the behaviour of GPI anchors and GPI-anchored GFP in bilayers. GPI was seen to flop down on the bilayer surface making extensive interactions with the lipid headgroup region[25].

In the current work, we extended our model of the GPI core to study three parasitic GPI types – (a) *Toxoplasma gondii*, (b) *Toxoplasma gondii* LMW and (c) *Trypanosoma congolense* – and (d) the GPI found in humans, without and with an attached GFP and inserted into bilayers of DMPC and POPC (see Figure 1). Note that both the GPI variants (a) and (b) are found in the same organism *T.gondii* where (a) always exists with an attached protein and (b) without. The model was constructed by combining three AMBER force-fields – GLYCAM06[26] for the GPI glycan, Lipid14[27] for the lipid tail, and AMBER-ff14SB[28] for GFP. From simulations at the microsecond timescale, we show that the GPI types lie close to and largely along the plane of the bilayer, making extensive interactions with the lipid heads, although exhibiting considerable flexibility internally especially at the $(1 \rightarrow 6)$ linkages. The observation that the flop-down conformation enables the side branches to be the most accessible epitopes provides a plausible explanation behind the specificity in immune responses trigerred by parasitic GPIs. We also show a significant difference in the surface presentation of parasitic and human GPIs. Results from the GPI-anchored GFP simulations demonstrate that the orientation and accessibility of the attached protein is sensitive to the type of GPI. The host bilayer also influences the exposedness of the GPI residues.



(a) *Toxoplasma gondii* GPI (b) *Toxoplasma gondii* LMW GPI

(c) *Trypanosoma congolense* GPI (d) Human GPI

Figure 1: Chemical structures of parasitic GPIs – (a) *T.gondii*, (b) *T.gondii* LMW, (c) *T.congolense* – and (d) human GPI considered in this study. The conserved portion is shown in black and the additional structural branching, including sugar and phosphoethanolamine (EtNP) residues, at the middle mannose (Man1) are in orange. The terminal EtNP at Man3 is denoted as tEtNP and the middle EtNP at Man1 is called mEtNP here.

2 Results

2.1 The "flop-down¨ conformation is consistent

In our previous work, we reported that the GPI core when embedded in bilayer patches prefers flopping down on the membrane heads, instead of standing erect [25]. From 3 parallel 1 μs long MD simulations of all the GPI variants (see Figure 1) inserted in (8x8) DMPC and POPC bilayer patches, similar flopping-down behaviour was observed. Figure 2 shows the tilt angles of each of these GPIs w.r.t. the bilayer normal. The schematic (Figure 2a) illustrates the definition of the tilt angle θ_z, which is the angle between the length vector connecting the end points of the GPI core – atom C6 of Ino and atom C4 of Man3 – and the bilayer normal (z axis). For the parasitic GPIs with longer branches, we also calculated an alternative tilt angle ζ_z, where the end points for the length vector are atom C6 of Ino and atom C4 of Glc/Gal depending on the GPI type (Figure 2d). Figures 2b and 2c show the distributions of tilt angle θ_z for all three parasitic and human GPIs inserted in pure DMPC and POPC bilayers respectively. The values for *T.gondii* (black) and *T.congolense* (green) GPIs peak around $70° - 80°$, whereas the distributions for *T.gondii* LMW (red) and Human (blue) GPIs are rather quite broad. The alternative tilt angle ζ_z distributions of *T.gondii* LMW and *T.congolense* GPIs are shown in Figures 2e, 2f. Both the distributions neatly

peak around $70° - 80°$, with *T.congolense* tending towards higher values. The tendency of the side branch of GPI to lie on the membrane heads is higher than the GPI core in both. Therefore, all the parasitic GPIs largely prefer to flop down on the membrane either through the GPI core or the side branch. However, human GPI makes an exception with its broad distribution suggesting high degree of flexibility.

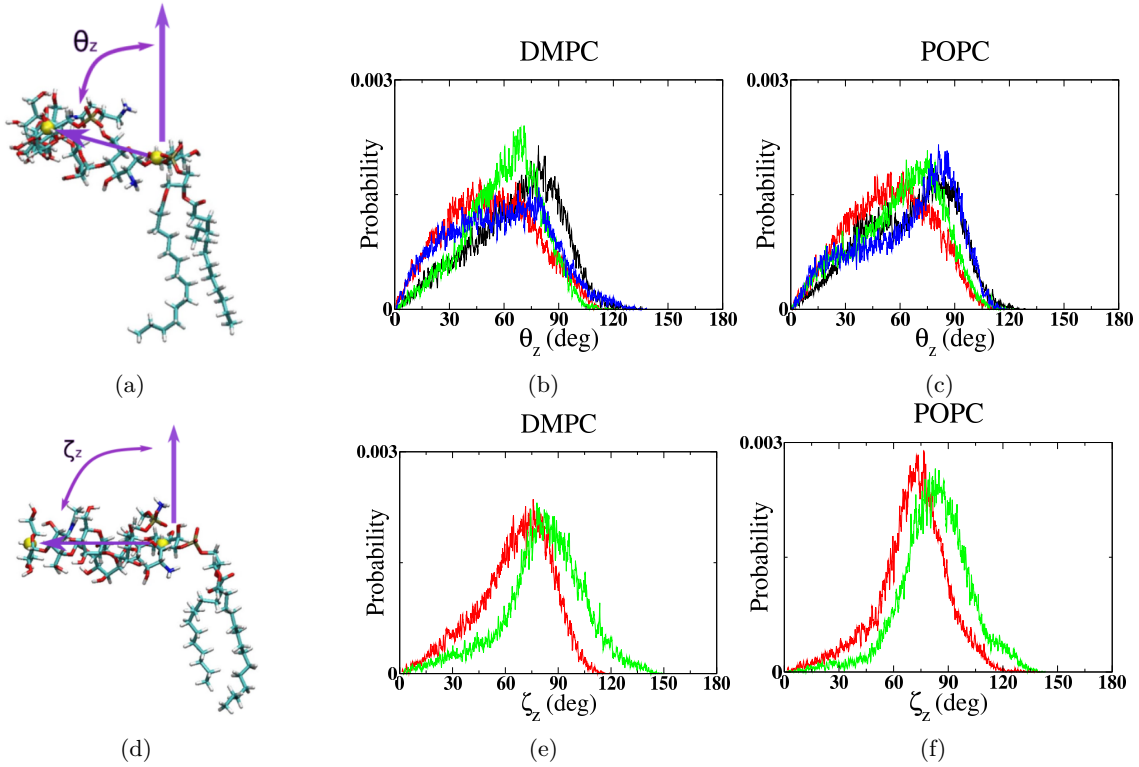


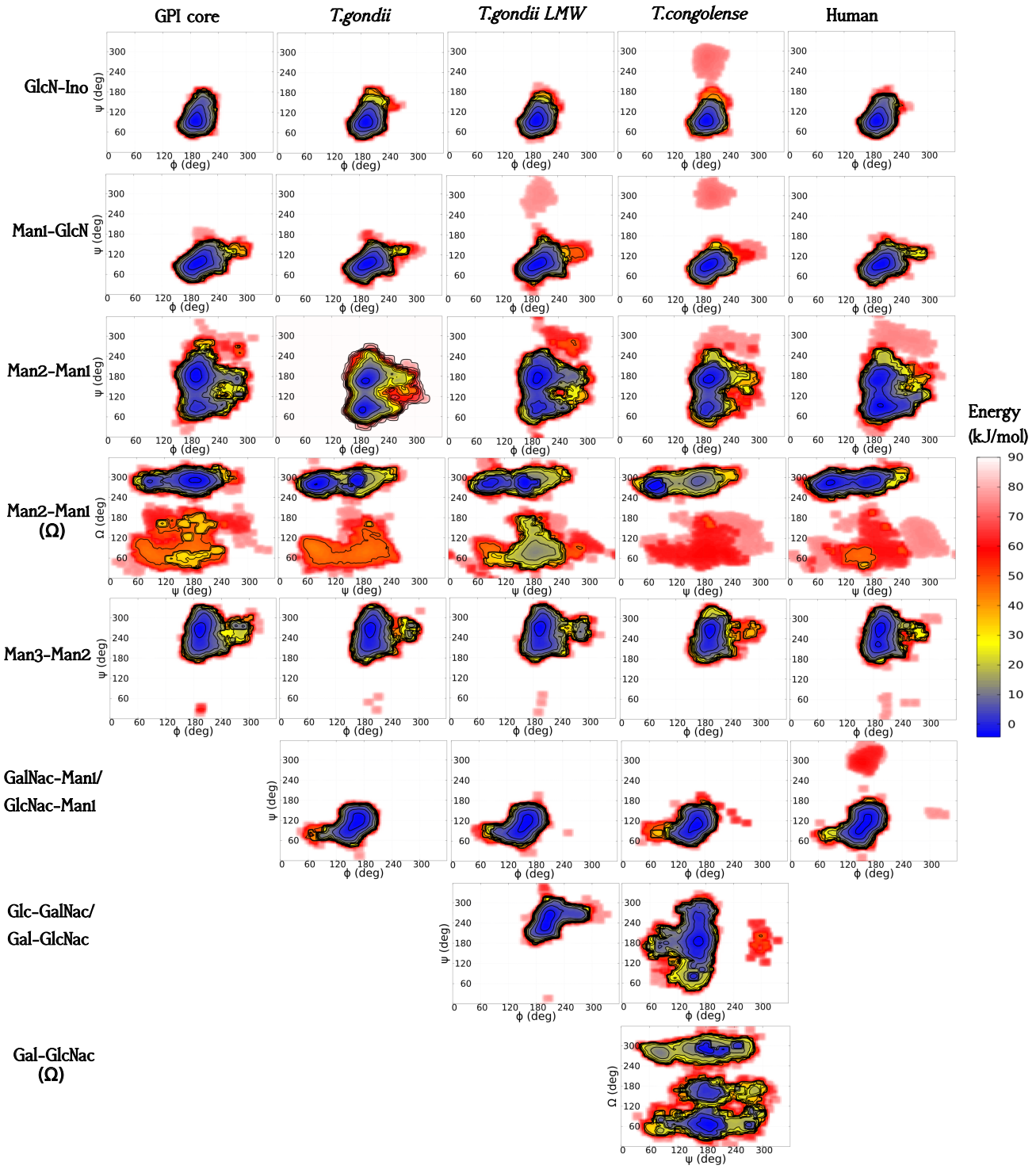
Figure 2: Comparison of tilt angles between all four GPI variants inserted in (b,e) DMPC and (c,f) POPC bilayers. The schematic of the tilt angle θ_z of the GPI core is shown in (a) and that of the alternative tilt angle ζ_z comprising the side branch is shown in (d). The yellow beads depict the end points of the length vector. The distributions of angle θ_z for *T.gondii* (black), *T.gondii* LMW (red), *T.congolense* (green), and human (blue) GPIs are placed against each other in (b) and (c). The distributions of ζ_z for *T.gondii* LMW (red) and *T.congolense* (green) GPIs are shown in (e) and (f).

2.2 Glycan topology

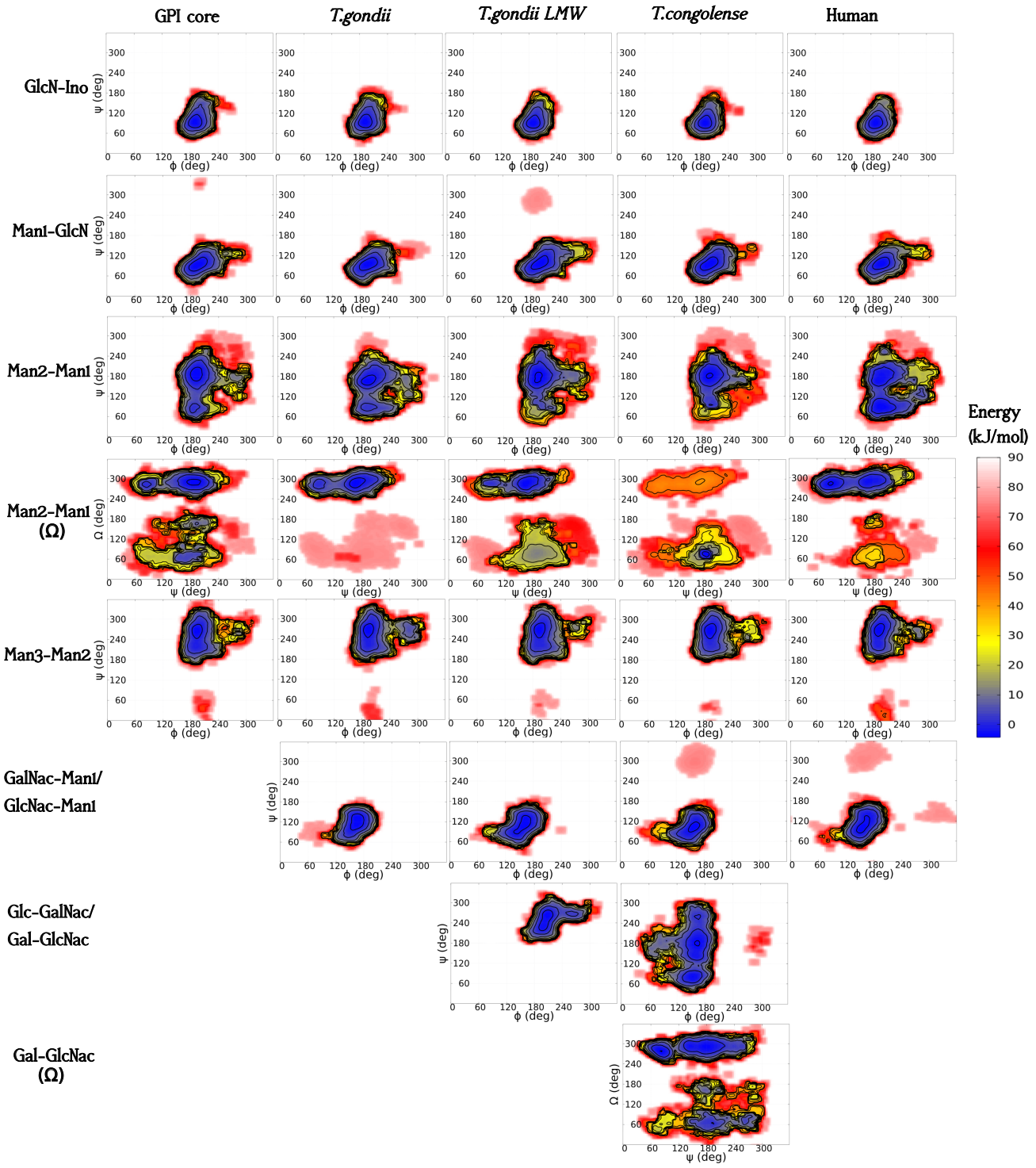
The overall three-dimensional conformation of a glycan is largely governed by the glycosidic linkages. These linkages can adopt different values depending on the type of monosaccharide, the anomer type, and also the type of linkage ($1 \rightarrow 2, 1 \rightarrow 3, 1 \rightarrow 4, 1 \rightarrow 6$). A glycosidic linkage connecting two sugar residues is represented with usually two torsion angles – $\phi(C_2, C_1, O'_x, C'_x)$ and $\psi(C_1, O'_x, C'_x, C'_y)$ – and an additional angle $\Omega(O'_6, C'_6, C'_5, O'_5)$ (relevant to $(1 \rightarrow 6)$ linkages). Figure 3 shows Ramachandran-like (ϕ, ψ) and (ψ, Ω) free energy plots for all the glycosidic linkages of protein-free GPI variants studied in (a) DMPC and (b) POPC bilayers. The free energy associated to a population distribution of a dihedral pair (ϕ, ψ) is obtained as follows:

$$F(\phi, \psi) = -k_B T log P(\phi, \psi) \tag{1}$$

where k_B is the Boltzmann constant, T is temperature, and $P(\phi, \psi)$ is the probability of occurrence of a dihedral pair.



(a) GPI in DMPC bilayer



(b) GPI in POPC bilayer

Figure 3: Two dimensional free energy landscapes as functions of dihedral pairs (ϕ, ψ) or (ψ, Ω) for all the glycosidic linkages in the protein-free GPI variants embedded in pure (a) DMPC and (b) POPC bilayers. Each of these plots is obtained from 6μs worth MD simulation data. Note that GalNac-Man1 belongs to *T.gondii*, *T.gondii LMW* and Human GPIs, whereas GlcNac-Man1 belongs to *T.congolense*. Similarly Glc-GalNac belongs to *T.gondii LMW* side branch, and Gal-GlcNac occurs in *T.congolense*

In comparison with the GPI conserved core, the distributions of all the corresponding torsions across all the

GPI variants are quite similar, except for the $\alpha(1 \rightarrow 6)$ Man2-Man1 linkage. As $(1 \rightarrow 6)$ linkages are known to be flexible due to the presence of an additional carbon, these differences are not surprising. In our previous study of the GPI tetrasaccharide in solution, we demonstrated the flexibility of the $(1 \rightarrow 6)$ linkage connecting the residues Man1 and Man2, serving as a hinge between the two relatively rigid disaccharides on either side (Man3-Man2 and Man1-GlcN)[24]. Out of the possible staggered rotamer states, *gg* was seen to be the most populated, with only minor traces of *gt* and *tg*, as a consequence of the gauche effect[29]. The rotamer populations for Man2-Man1 of all GPIs are listed in Table 1. For the GPI core, *T.gondii* and Human GPIs, the population ratios are pretty much consistent with *gg* being the highest, followed by *gt* and negligible traces of *tg*. As we move towards GPIs with an extra side branch residue, *T.gondii* LMW and *T.congolense*, the population of the *gt* rotamer relatively goes up, even to the same level as that of *gg* for *T.gondii* LMW in POPC and more than *gg* for *T.congolense* in POPC. These results suggest that the lipid membrane could also make a difference to the conformation of the GPI anchor, particularly at the flexible linkages. The $\beta(1 \rightarrow 6)$ Gal-GlcNac linkage presents a profile different from the Man2-Man1 linkage in both the ψ torsion and rotamer populations. The Gal-GlcNac rotamer population distribution *(gg/gt/tg)* in DMPC is (56/36/9)% and in POPC is (78/20/2)%, again showing highest preference for *gg*. Both ψ and Ω are more flexible in Gal-GlcNac than in the Man2-Man1 linkage, possibly because Gal is a terminal residue and so is subject to greater degrees of freedom. Rotamer populations are dominated by the *gauche* conformers over *anti* in both Man2-Man1 and Gal-GlcNac linkages, as has been observed to occur in manno- and glucopyranosides[30].

System	DMPC			POPC		
	gg%	gt%	tg%	gg%	gt%	tg%
GPI core	66	25	9	75	22	3
T. gondii	76	23	1	91	7	2
T. gondii LMW	59	33	8	41	41	18
T.congolense	83	13	4	37	54	9
Human	78	19	3	92	8	0

Table 1: Rotamer populations of the $\alpha(1 \rightarrow 6)$ linked Man2-Man1 linkage across all the protein-free GPI variants inserted separately in DMPC and POPC bilayers.

2.3 Effect of the terminal phosphoethanolamine (EtNP) linker

The terminal phosphoethanolamine that is attached to Man3 (see Figure 1) is found on all GPI variants, whether with or without protein. It serves to attach proteins to the GPI through an amide bond formation between the C terminal end of the protein and the amine group of EtNP. To understand how the presence of the phosphoethanolamine linker affects the conformation of the GPI, we simulated protein-free GPIs – GPI core and the three parasitic GPIs with the terminal EtNP group removed. All the GPIs were inserted in DMPC and POPC bilayers for a one-to-one comparison with the EtNP-fixed GPIs. These systems were also subjected to 3 parallel 1 μs long runs, just as the EtNP-fixed GPI systems. We observed that the conformation, by and large, remains the same as the EtNP-fixed GPIs, except for the terminal Man3-Man2 linkage. We reported in our preceding papers on the GPI anchor [24, 25] that the ψ torsion of the Man3-Man2 linkage fluctuates between two values that are separated by only a small energy barrier. This behaviour is a manifestation of the external anomeric effect that lowers the rotational energy barrier to a secondary resonance form of the Man-Man disaccharide[31]. Figures 4a, 4c show that the bimodal distribution of ψ(Man3-Man2) was seen to be consistent across all the GPI variants without the EtNP linker, with ψ values peaking at -90° and -140°. Note that the bimodality is somewhat diminished, that is, the distributions are shifted more towards $\psi \approx -90^\circ$, in the POPC systems. Upon attachment of EtNP, the distributions almost completely shift towards $\psi \approx -90^\circ$, with the curves only slightly shouldering at the other minor value (see Figures 4b, 4d). Upon observing the intramolecular interactions of GPI in detail, it was revealed that the positively charged NH_3^+ group of the EtNP linker forms H-bonds with Man2 which results in stabilization of the ψ(Man3-Man2) at $\approx -90^\circ$.

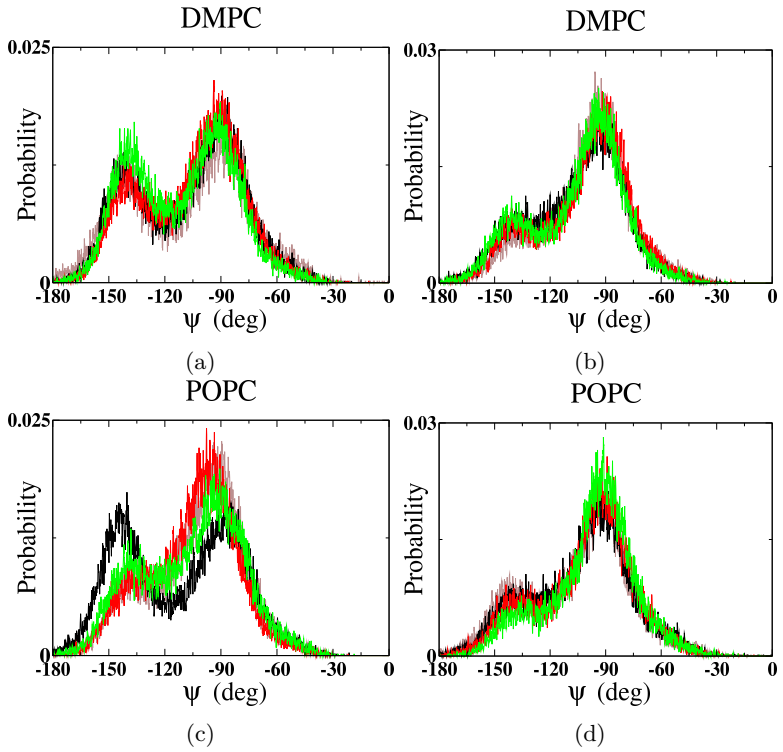


Figure 4: Dihedral distributions of ψ (C1-O2'-C2'-C1') of Man3-Man2 from (a) GPI without EtNP in DMPC, (b) GPI with EtNP in DMPC, (c) GPI without EtNP in POPC, and (d) GPI with EtNP in POPC bilayer. Results from 4 GPI variants are placed against each other – GPI core in light brown, *T.gondii* GPI in black, *T.gondii* LMW in red, and *T.congolense* GPI in green.

2.4 Importance of the side branches

The internal structure is a useful, yet, inadequate measure of the overall conformation for membrane-bound glycolipids. The presentation, including the orientation with respect to the membrane and the insertion depth, of these glycolipids also influences their recognition mechanism. Through a thorough structural investigation of GM3 in a lipid bilayer using NMR and MD simulations, DeMarco and Woods observed that the receptor protein recognizes the most easily accessible head-group residues of the glycolipid[32]. Similarly, with the help of MD simulations, another study on GM1 provided an explanation for certain carbohydrate-binding proteins having evolved to recognize the most frequently exposed glycolipid epitopes[33]. The extent of accessibility of the glycan residues can be estimated by their degree of interaction with the lipid heads of the bilayer/membrane. We calculated the probability distribution of the location of each residue with respect to the bilayer center along the bilayer normal for all the GPIs under study, as shown in Figure 5. The first three residues of the GPI core – Ino, GlcN, and Man1 – lie in a sequential order in all the systems, i.e., they lie at increasing distance from the point of attachment to the lipid bilayer. The succeeding residues – Man2, Man3, EtNP and the side branches – appear to occupy nearly the same distance from the bilayer center. However, the residue directly attached to Man1 – GalNac/GlcNac (shown in magenta) – is the outermost residue in most of these GPIs, for e.g., *T.gondii*, *T.gondii* LMW and *T.congolense*. The terminal side branch residue of *T.congolense*, Gal (shown in yellow), is located quite close to the lipid heads, almost at the level of GlcN. Human GPI is distinct from the parasitic GPIs in that there is an additional EtNP residue attached to the middle mannose Man1, which could alter the conformation of the molecule. In Figure 5, we can see that the terminal EtNP residue (tEtNP) is pushed inwards whereas the middle EtNP (mEtNP) lies distinctly facing the solvent.

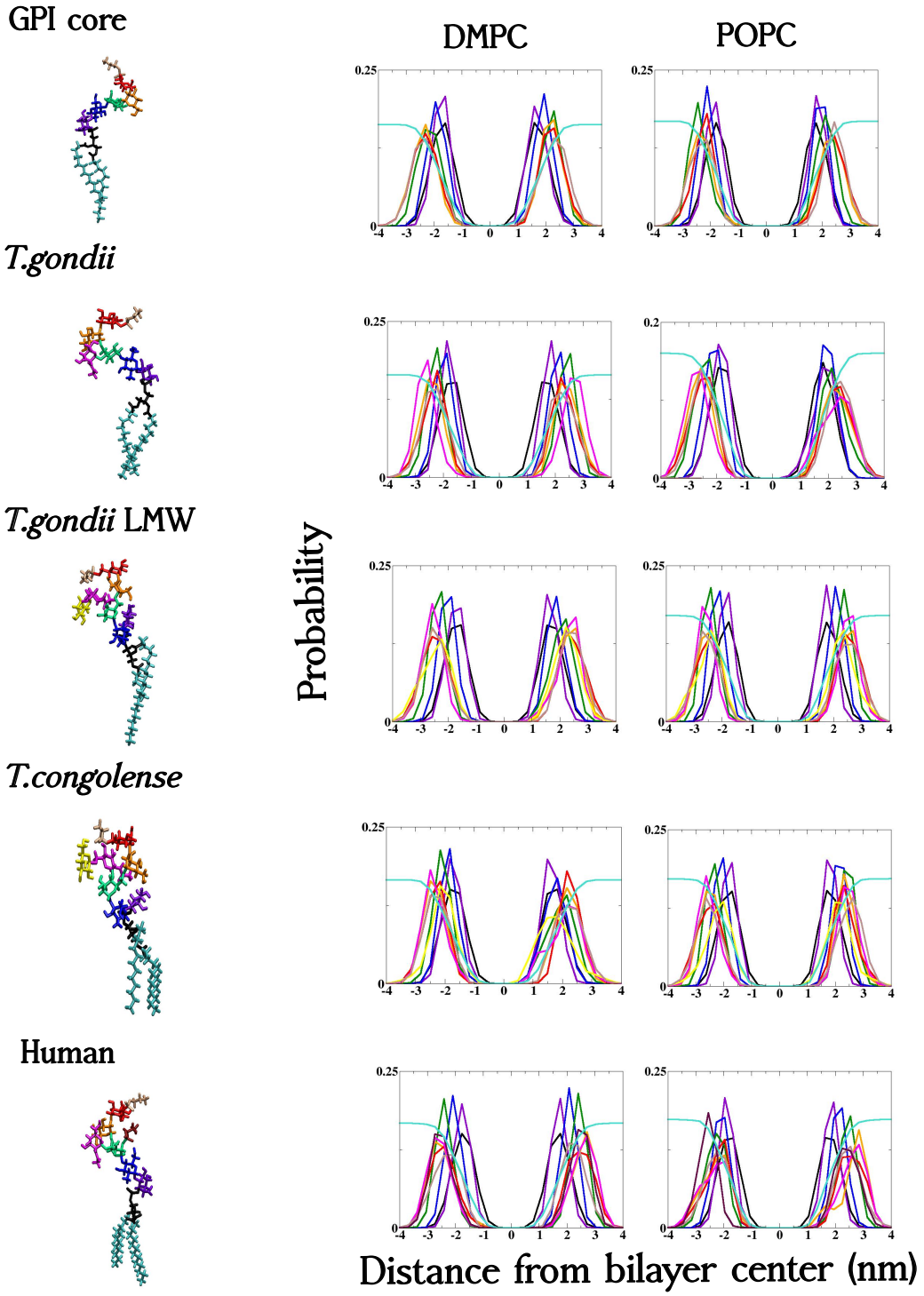
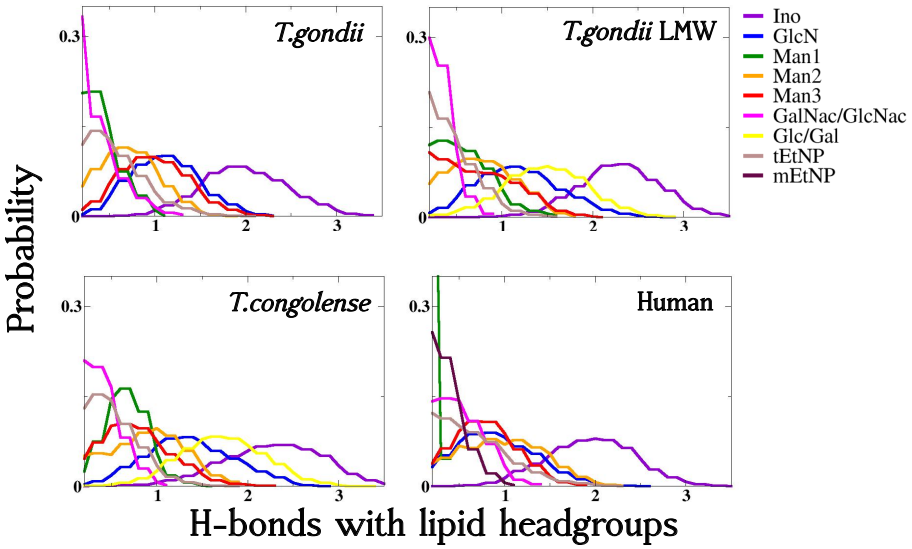


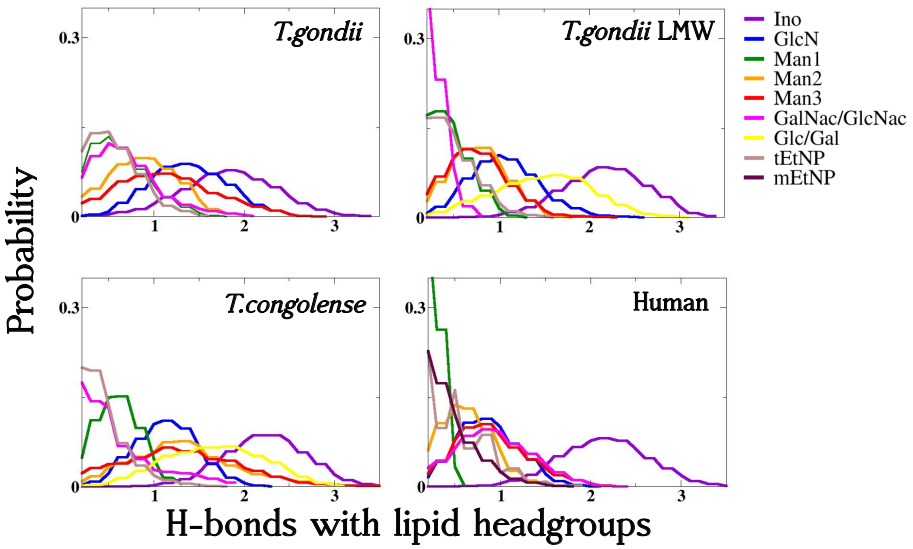
Figure 5: Distance distribution profiles for each residue of the GPI variants from the bilayer center along the normal (z axis). Plots are shown for GPI core, parasitic GPIs – *T.gondii*, *T.gondii* LMW and *T.congolense*, and human GPI embedded in DMPC and POPC bilayer. The plot colors are coordinated with the residues shown in adjacent figures where Ino is shown in violet, GlcN in blue, Man1 in green, Man2 in orange, Man3 in red, GalNac/GlcNac in magenta, Glc/Gal in yellow, tEtNP in light brown and mEtNP in maroon.

The presence of several hydroxyl groups in carbohydrates makes them highly prone to hydrogen bonding (H-bonding). Therefore, for additional insight into the interactions between GPI and the host lipid bilayer, we looked at the H-bond formations of each residue of the GPIs with the phosphocholine lipid heads. Figure 6 shows that

in all parasitic GPIs, GalNac/GlcNac, Man1 and tEtNP hardly H-bond with the lipids. The terminal side branch residue Glc/Gal interacts strongly with the bilayer, even more than GlcN does. In human GPI, similar behavior was observed except that Man1 and mEtNP interact significantly less with the bilayer than does GalNac. From the H-bond profiles, it is apparent that the type of membrane also makes a difference to the accessibility of the epitopes. All GPIs interact with the lipid headgroup region more in POPC than in DMPC, with the interactions being significantly more for Man2 and Man3 in parasitic GPIs, and for GalNac in human GPI. This discrepancy arises from the different areas-per-lipid of DMPC and POPC bilayers, resulting in less packing of lipids in POPC than in DMPC. Such discrepancy was previously observed with the GPI core simulations in our previous work as well[25]. Figure 7 shows the hydration number of each residue for every GPI variant, where hydration number means the number of water oxygens lying within a radius of 0.3 nm from every atom of the residue being considered. In alignment with the results from Figure 6, GalNac/GlcNac and EtNP are the most hydrated amongst all residues, however, Man1 is the least hydrated despite the observation that it scarcely interacts with the membrane. It is interesting to note that the terminal side branch residue Glc in *T.gondii* LMW is highly hydrated irrespective of its strong H-bonding with the membrane as shown in Figure 6. However, the equivalent side chain residue Gal in *T.congolense* is relatively less solvated. In human GPI, both the EtNP residues are almost equally hydrated. Upon comparing the hydration plots with the H-bonding profiles of each residue with water (see Figure S2 of SI), similar interaction pattern was observed. However, in the *T.gondii* LMW GPI, the terminal side chain residue Glc showed more H-bonding with water than its preceding residue GalNac, whereas Gal in *T.congolense* showed no difference whatsoever. The difference between the two EtNP linkers was revealed more clearly in Figure S2 in that mEtNP H-bonds more with water than does tEtNP. The middle mannose Man1 neither interacts much with the membrane nor with water. This is suggestive of its significant role in controlling and determining the conformation and presentation of GPI on the membrane surface via intramolecular interactions. Summing up the findings from Figure 7 and Figure S2, we conclude that through combined interactions via H-bonds and vDW contacts with water, the side chain residues and EtNP linkers are the most solvent-interactive, and therefore the most accessible epitopes for recognition by antibodies. The side chain Gal residue in *T.congolense* makes an exception though, as it turns out to be quite embedded in the lipid heads. A plausible reason behind this behaviour could be the longer and flexible $\beta(1 \rightarrow 6)$ linkage which allows it the conformational freedom to interact with the lipid heads. These findings suggest that the side chain residues, whether carbohydrates or EtNP linkers, that branch out from Man1 may have an important part to play in GPI recognition.

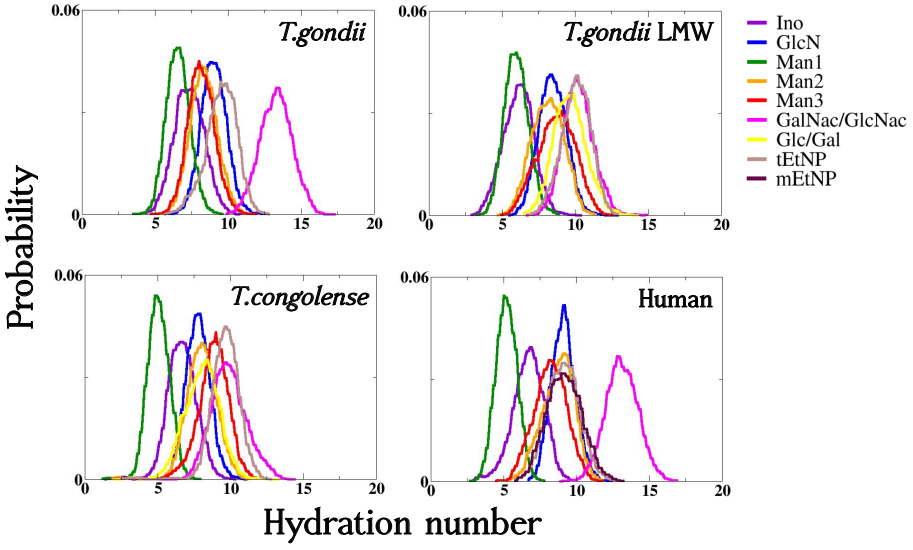


(a) GPI in DMPC bilayer

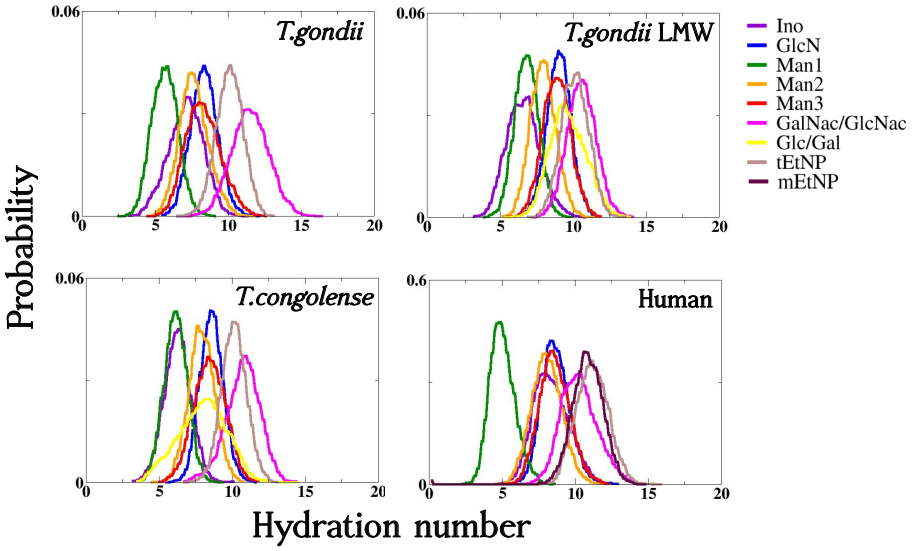


(b) GPI in POPC bilayer

Figure 6: Probability distributions of number of H-bonds formed between each residue of the GPI variant – *T.gondii*, *T.gondii* LMW, *T.congolense*, Human GPI – and the lipid headgroup region of (a) DMPC and (b) POPC bilayer



(a) GPI in DMPC bilayer



(b) GPI in POPC bilayer

Figure 7: Probability distributions of hydration number of each residue of the GPI variants – *T.gondii*, *T.gondii* LMW, *T.congolense*, Human GPI – inserted in (a) DMPC and (b) POPC bilayer

2.5 Cluster analysis

To filter out GPI conformations that predominantly occur in the simulations, we performed cluster analysis on each of the concatenated 3μs long MD trajectories, including the GPIs from both the top and bottom leaflet of the bilayer. The distribution of cluster populations for each GPI system is shown Figure 8. The variety in conformations displayed is further proof for the structural flexibility of GPIs. Considering the flexibility, only the clusters with more than 40 members have been considered for analysis. Out of all the GPIs, *T.congolense* GPI shows extraordinary structural variability, as is evident from the broad distribution in its cluster sizes. This is because of the presence of two highly flexible and dynamic (1 $\rightarrow$ 6) linkages, one α (main chain) and the other β (side chain) in its structure (see Figure 1c). Representative structures from each cluster of the GPI variants are displayed in Figure 9. For better visualization only the three most populated clusters, which together make up the majority, are shown for *T.gondii*, *T.gondii* LMW and Human GPI, excepting *T.congolense* GPI for which four most populated clusters are shown considering its wide spread in conformations. The cluster conformations have been super-positioned against each other by aligning them with respect to the residues GlcN-Ino-PGL. The most dominant conformation of each GPI is shown to the right of the cluster populations, excepting Figure 9c and 9d(ii) where two most dominant conformations are shown given their wide cluster size distributions. The side chain residue GalNac is seen to distinctly project out of the membrane heads facing the solvent in *T.gondii* GPI inserted in DMPC (Figure 10a(i)), while it is relatively less exposed in POPC (Figure 10a(ii)). Similar behaviour is observed for *T.gondii* LMW GPI where residues GalNac-Glc are well-exposed to the solvent when the GPI is in DMPC (Figure 10b(i)) compared to in POPC (Figure 10b(ii)). Here, the conformations in DMPC and POPC are almost opposing in appearance. The main-chain terminal branches Man2-Man3 face the lipids in DMPC whereas they face the solvent in POPC, and vice-versa for the side branches GalNac-Glc. Large conformational variability is exhibited by *T.congolense* GPI, a little more so in DMPC where both the main chain residues Man2-Man3 and side chain residues GlcNac-Gal display many degrees of freedom (Figure 10c). The side branches point towards the solvent, albeit not too prominently, in DMPC (Figure 10c(i)), however, they are quite embedded in the lipids in POPC (Figure 10c(ii)), thereby, following a similar pattern as that of *T.gondii* and *T.gondii* LMW GPIs. Although the side branch residue GalNac in Human GPI is relatively more hydrated (Figure 7), it occupies nearly the same level (z axis) on the bilayer (Figure 5) and interacts quite similarly with the lipid heads (Figure 6) as do Man2 and Man3. These observations are in alignment with the cluster populations (Figure 10d) where Man2, Man3 and GalNac seem to occupy similar positions on the lipid surface. Two dominant conformations are shown for Human GPI in POPC (Figure 10d(ii)) as they are equally populated. As the chemical composition of Human GPI is the same as *T.gondii* GPI except for the middle EtNP, this additional residue is the natural object of interest here

when comparing the two. As has been noted in Figures 6 and 7 that mEtNP is more solvent-accessible than tEtNP, the same observation is consistent in the cluster populations where mEtNP faces the solvent a little more than does tEtNP, despite the fact that tEtNP lies at the terminal end of the GPI.

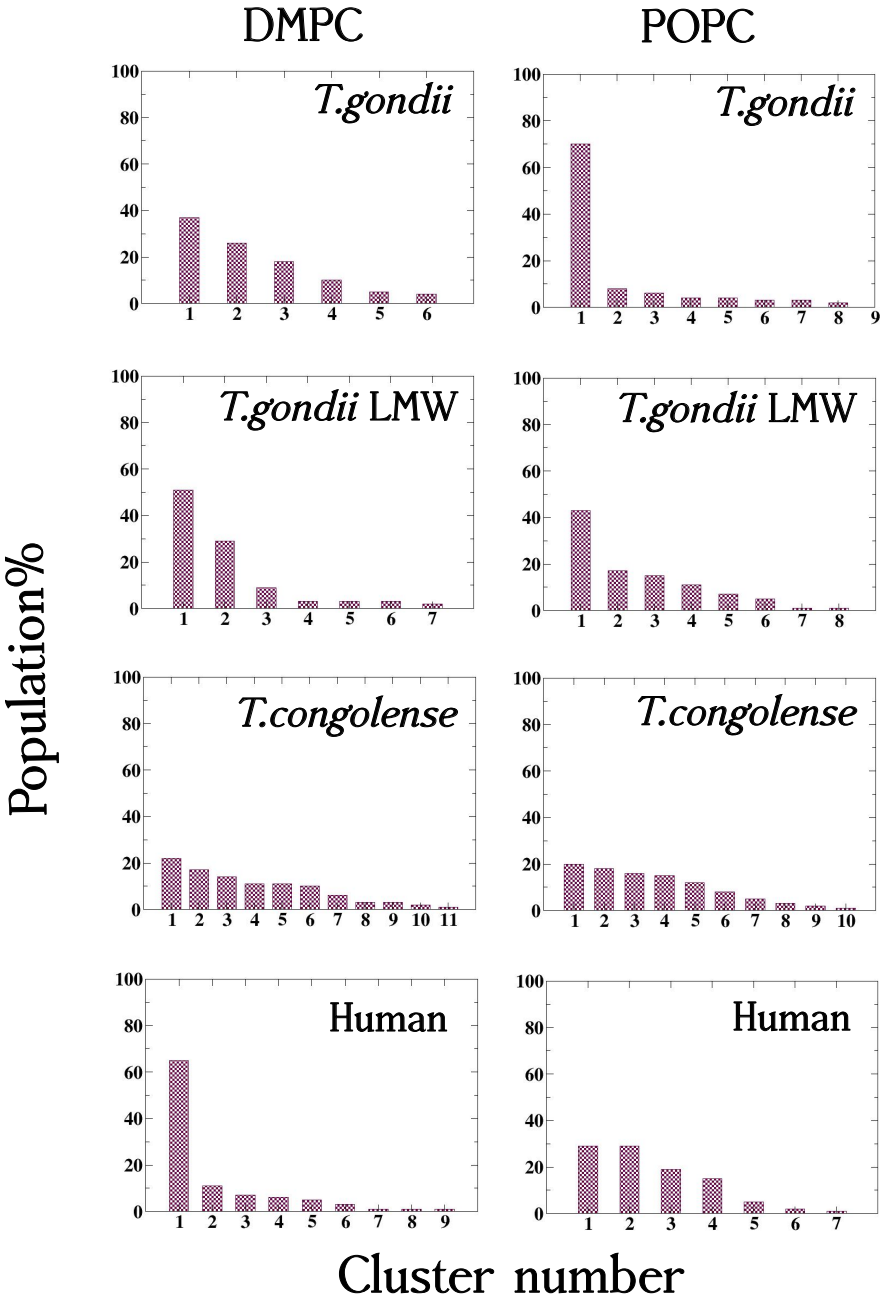
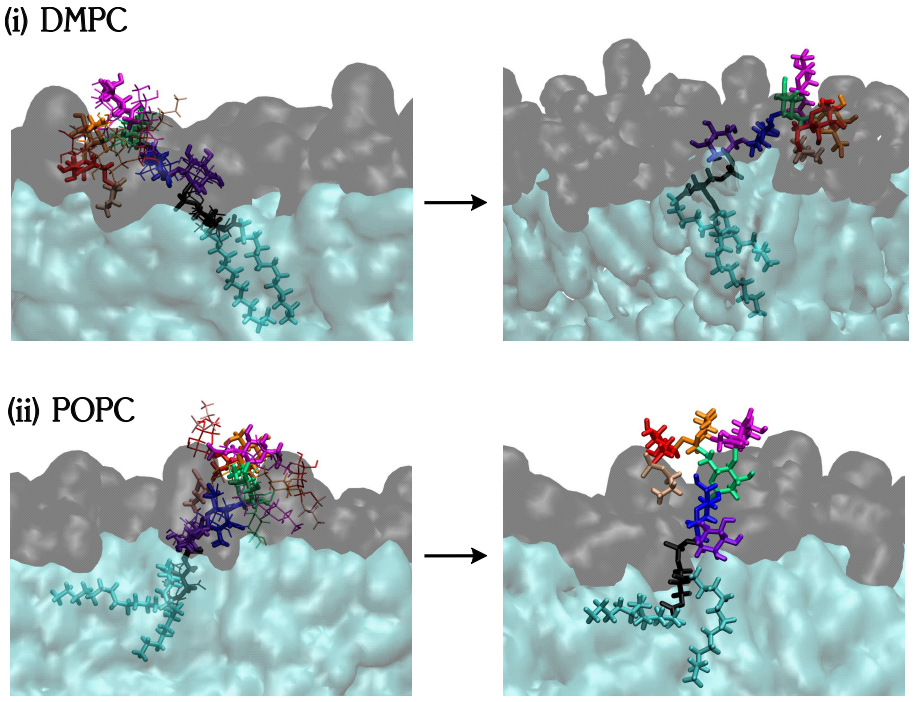


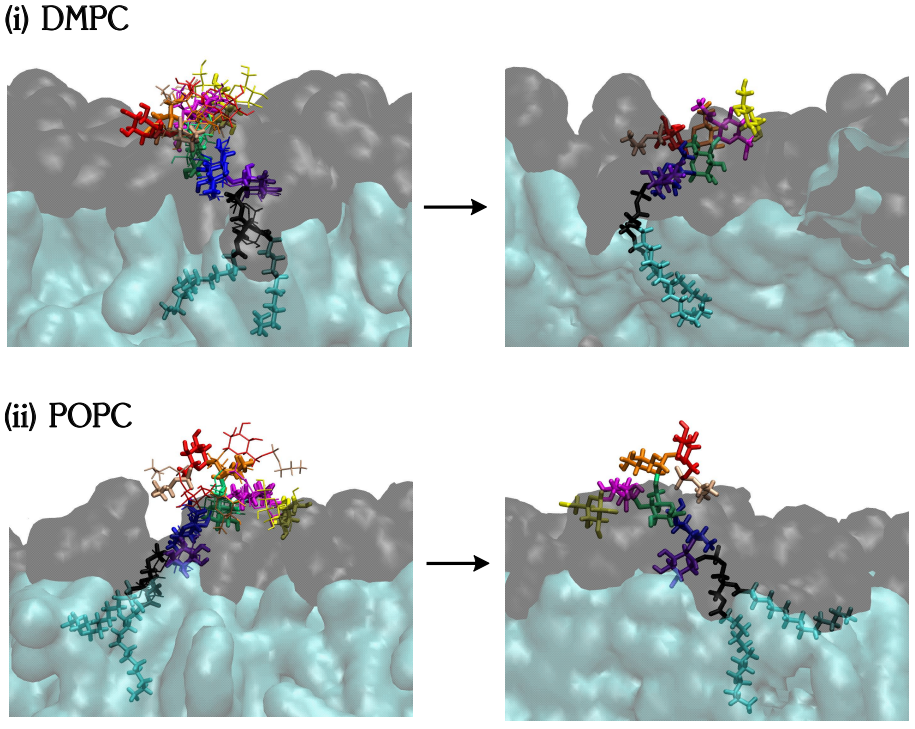
Figure 8: Cluster size populations obtained from 6μs worth of simulation data of *T.gondii*, *T.gondii* LMW, *T.congolense* and Human GPI in DMPC and POPC bilayer. Note that GPIs from both the top and bottom leaflet have been included in the calculation.

Upon measuring the glycosidic torsions of the clusters, it was revealed that the most dominant cluster conformations of the GPIs, except for *T.congolense*, bore torsion values from all the lowest free energy wells together, i.e., the overall GPI conformation correlated with the well depths of the glycosidic torsions (Figure 3). In the case of *T.congolense* GPI, although substantial conformational differences were expected at the flexible linkages Man2-Man1 and Gal-GlcNac, even the relatively rigid linkages showed considerable variation in both DMPC and POPC. The most dominant cluster (population = 22%), cluster 1, of *T.congolense* in DMPC carried torsional val-

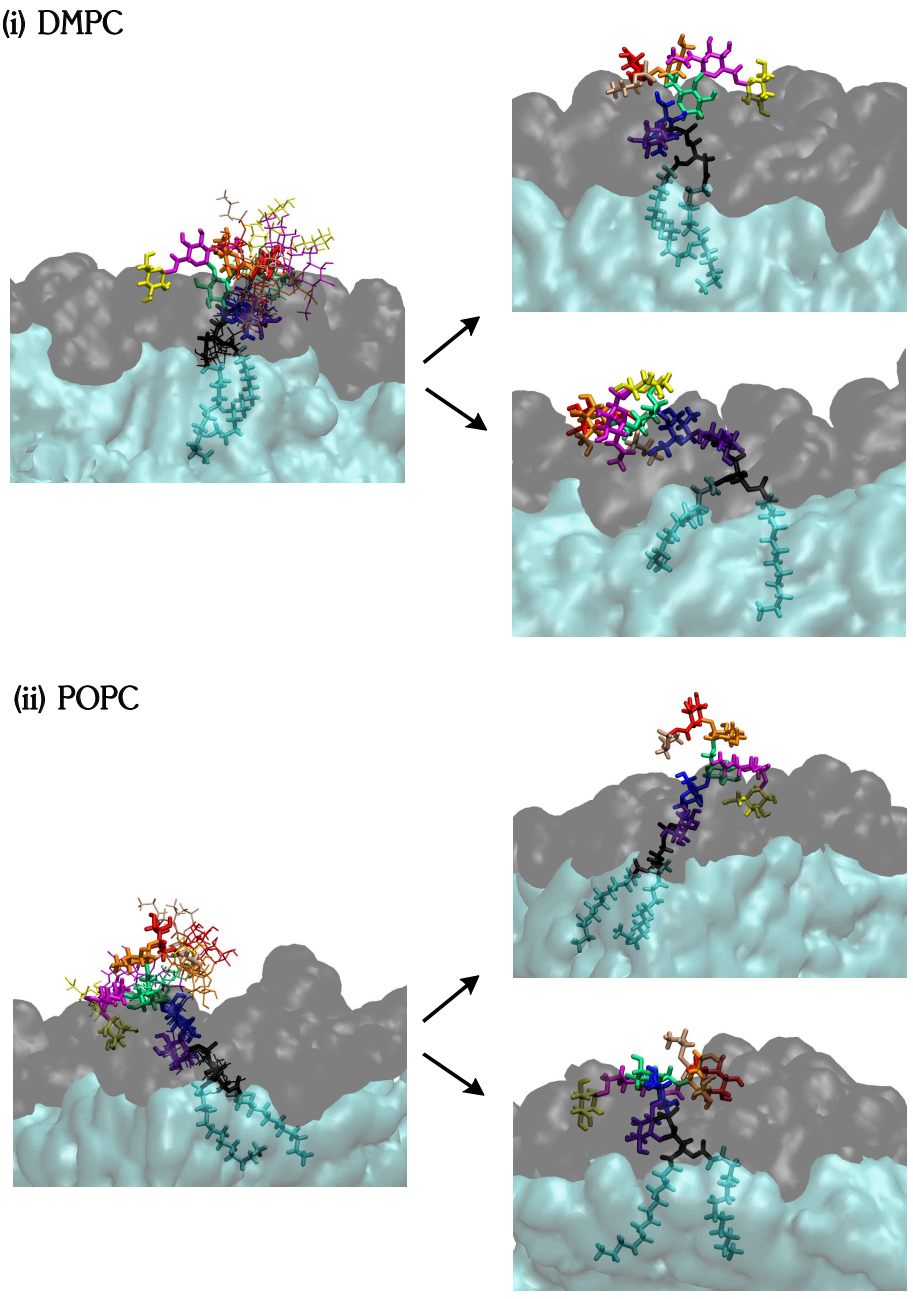
ues (ϕ, ψ) = (192,281) for GlcN-Ino and the second most populated cluster (population = 17%), cluster 2, carried (ϕ, ψ) = (209,300) for Man1-GlcN, both from seemingly minor energy wells (Fig.3a). Similarly, the fourth largest cluster (population = 15%), cluster 4, of *T.congolense* in POPC measured (ϕ, ψ) = (174,292) for GlcNac-Man1 belonging to a minor (ϕ, ψ) population (Fig.3b). The Man2$-\alpha(1 \rightarrow 6)-$Man1 linkage consistently occupied the *gg* rotamer in the largest clusters, i.e. cluster 1, of all GPI variants, even including *Tcongolense* in POPC for which the Ramachandran plot shows *gt* rotamer as the most populated rotamer (Fig.3b). However, clusters 2, 3 and 4 do indeed contain **gt** rotamers. Significant populations of *gt* rotamers also occur in cluster 3 of *T.gondii* in both DMPC and POPC, clusters 2 and 3 of *T.gondii* LMW in DMPC, cluster 4 of *T.congolense* in DMPC and cluster 3 of Human GPI. The *tg* rotamer also shows up in clusters 2 and 3 of *T.gondii* LMW in POPC.



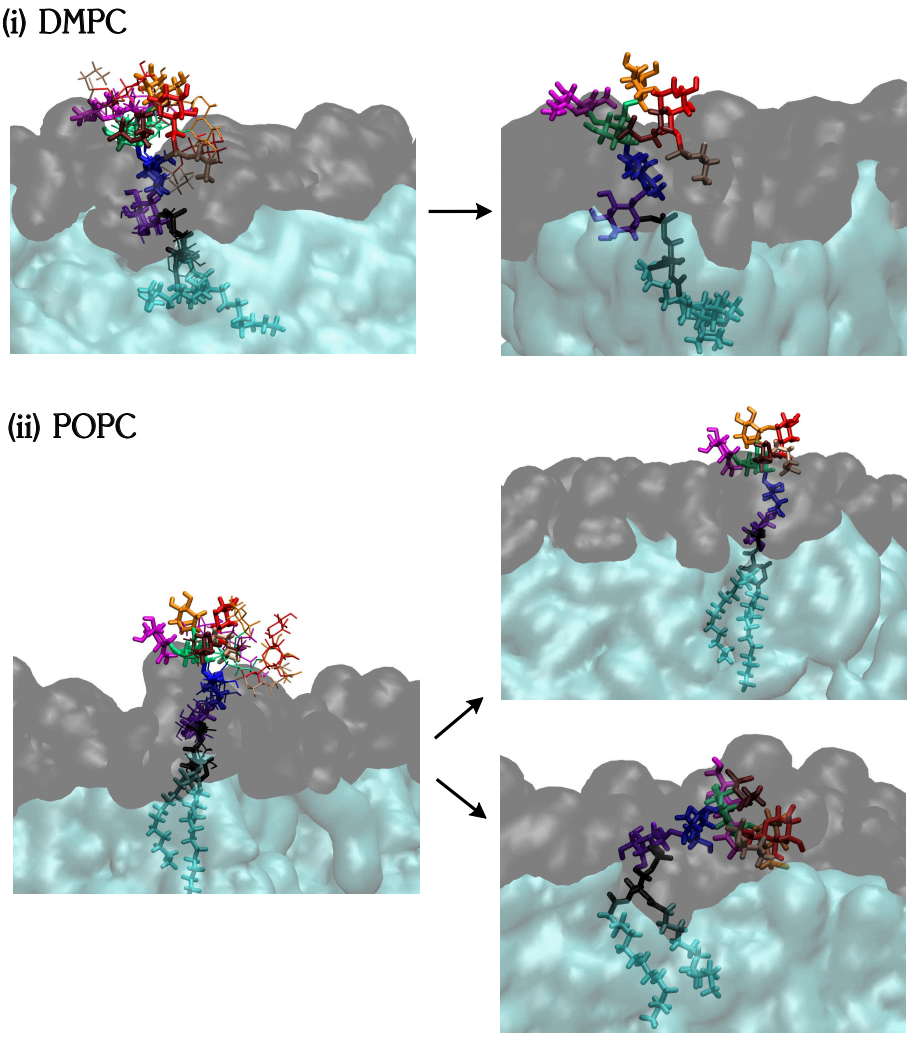
(a) *T. gondii* GPI



(b) *T. gondii* LMW GPI



(c) *T. congolense* GPI



(d) Human GPI

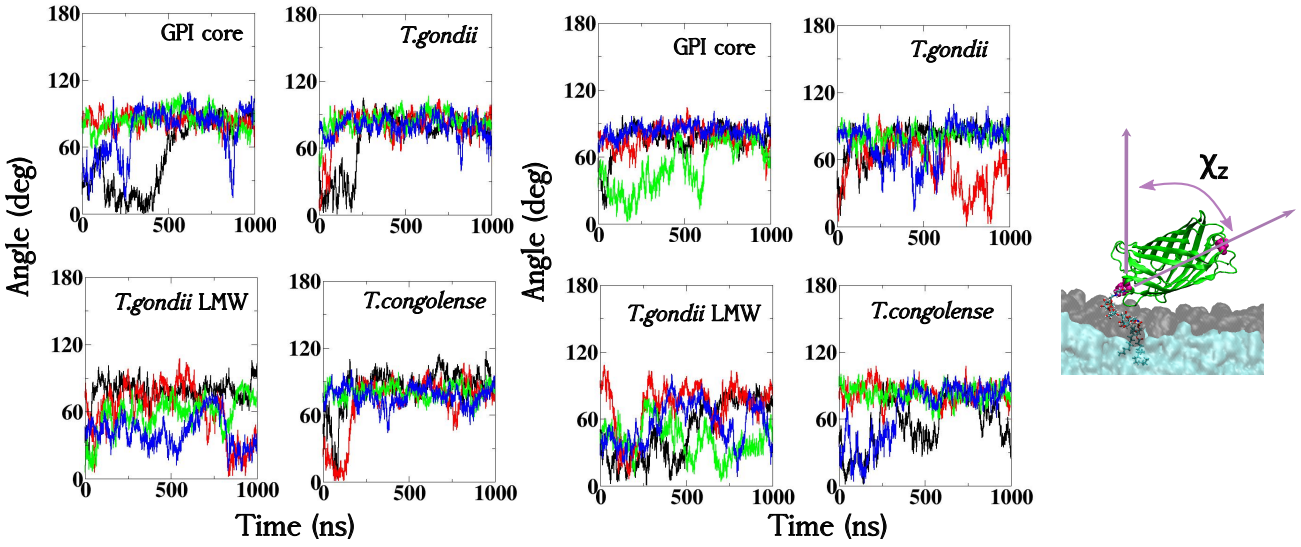
Figure 9: Conformations obtained from clustering of structures of (a) *T.gondii*, (b) *T.gondii* LMW, (c) *T.congolense*, and (d) Human GPIs in (i) DMPC and (ii) POPC bilayers. The left column shows the first three most populated clusters for (a),(b) and, (d), and first four most populated clusters for (c), aligned together w.r.t the residues GlcN-Ino-PGL. Here PGL stands for the phosphoglycerol head of GPI. The right column shows the most dominant cluster conformations for each, excepting (c) and (d) where the first two most dominant clusters are shown owing to the relatively broad cluster size distributions. Color coding for residues follows the same pattern as in Figures 5, 6, 7.

2.6 GPI-anchored GFP

2.6.1 Orientation of anchored GFP

GPI anchors are more commonly found attached to extracellular proteins than found without them, the latter kind being more prevalent in protozoan parasites. To understand the nature/extent of interactions of an attached protein with GPI, we chose the Green Fluorescent Protein (GFP). GFP is not a naturally-occurring GPI-anchored protein, however, due to its easy availability and fluorescent nature, it has been extensively used for experimental studies of GPIs[11, 34, 35]. The protein was attached to the GPI through a phosphoethanolamine (EtNP) linker, at the amine group of the molecule. The orientation of the protein with respect to the bilayer was determined by measuring the angle of tilt of the axis of the GFP barrel formed with the bilayer normal (see schematic in Figure 10). The time evolution of the tilt angle in each of the four 1μs long trajectories for each GPI variant in DMPC and POPC bilayers is displayed in Figure 10. The tilt angle largely occupies high values $\approx 80°$, except in *T.gondii* LMW GPI, where the orientation of GFP was seen to highly fluctuate. Interestingly, *T.congolense* GPI, despite

having the same size as *T.gondii* LMW, behaves very differently, in the sense that the tilt angle is quite stable at high values. This indicates that the orientation of the attached protein could be affected not just by the size of GPI, but also the different chemical composition of the side chain residues. Moreover, GPIs inserted in POPC bilayer exhibit greater fluctuation in tilt angle than those in DMPC bilayer, suggesting that the host lipid bilayer also influences the protein's presentation. Here note that the *T.gondii* LMW GPI does not naturally come with a protein, but only exists in the protein-free form on the cell surface of the parasite. Nevertheless, it was attached to the protein in this study to facilitate a systematic investigation.



(a) GFP-GPI in DMPC bilayer

(b) GFP-GPI in POPC bilayer

Figure 10: Time evolution of the tilt angle of GFP when attached to four different GPI types – GPI core, *T.gondii*, *T.gondii* LMW, and *T.congolense* GPIs – placed in (a) DMPC and (b) POPC bilayers. For each system, results from four independent 1μs long trajectories (black, red, green, blue) are placed against each other. The schematic shows the definition of the tilt angle χ_z as the angle between the vector connecting GFP residues GLN(76) and HIE(135) (shown in pink), and the bilayer normal.

2.6.2 Cluster analysis

Cluster analysis on protein-free GPIs revealed that all the GPI variants lie in a reclining manner on the lipid head-group region of bilayers, with different sugar residues exposed to the solvent depending on the GPI type (Sec 2.5). The same analysis was carried out on protein-attached GPIs to observe the conformation of GPI relative to the orientation of the protein (GFP). Figure 11 shows the cluster sizes, in other words, the population distribution of clusters of GFP-attached-GPIs simulated separately in DMPC and POPC bilayers. Only the clusters with more than 30 members have been included in the analysis. When these distributions are compared with those from the protein-free GPIs (Figure 8), it is clear that the conformation of GPI is more stable, i.e., it is less flexible, with an attached protein. This difference is most prominent in *T.congolense* GPI, where with an attached protein, not only did the number of clusters reduce, but also the size of the largest cluster significantly rose. All in all, upon attaching GFP, cluster size distributions narrowed down for all the GPI variants. This restricted flexibility in conformation of GFP-attached-GPIs is also apparent in the Ramachandran-like plots for the glycosidic torsions shown in Figure S3 of the SI, where the energy wells are less scattered than in protein-free GPIs. Dominant conformations from the most populated clusters of each GPI can be found in Figure S4 of the SI. Protein attachment was seen to inflict maximum conformational deviation w.r.t. free GPIs at the flexible $\alpha(1 \rightarrow 6)$ Man2-Man1 linkage. For example, the rotamer population was the highest for *gg* in protein-free GPI core both in DMPC and POPC bilayer, whereas with attached GFP, the population shifted towards *gt*. There was considerable deviation observed at ψ of Man2-Man1 too, going from two prominent energy wells in protein-free GPIs to essentially one in GFP-GPIs. Cluster 1 of GPI core in both DMPC and POPC shows Man1 distinctly peeking out of the membrane, a conformational feature of the *gt* rotamer, while the rest of the sugar residues swim in the headgroup region (see Figure S4a). Cluster 2 of GPI core in DMPC, being a *gg* rotamer, appears different. Cluster 1 of *T.gondii* GPI occupies a clear majority with

cluster size population being more than 80% in both DMPC and POPC (Figure 11). The conformation of cluster 1 in DMPC is similar to that of its protein-free counterpart (*gg* rotamer in both), however in POPC the conformation shifts from *gg* in protein-free GPI to *gt* in protein-attached GPI (see Figure S3). Similar pattern emerges in the other GPI, *T.gondii* LMW, of the same parasite. However, it is to be noted here that in the protein-free form of *T.gondii* LMW GPI, a significant amount of *gt* rotamer is already populated. Upon protein attachment, the population further shifts towards *gt* both in DMPC and POPC, with cluster 1 in POPC containing the *gt* rotamer. Cluster 1 in Figure S4c(i) contains *gg* and cluster 2 contains *gt*, whereas cluster 1 in Figure S4c(ii) carries *gt* and cluster 2 carries *gg* rotamer. The orientation of GFP is also strikingly different in cluster 1 and 2 in Figure S4c(i), however, upon further investigation it was revealed that there is no correlation between Man2-Man1 rotamer and orientation of the protein. The shift towards *gt* is, however, not consistent in *T.congolense* GPI where the dominant clusters in both DMPC and POPC carry *gg* rotamer for both $(1 \rightarrow 6)$ linkages of Man2-Man1 and Gal-GlcNac. The ψ values for the same linkages differ significantly from their protein-free counterparts. The terminal side branch residue Gal-GlcNac in *T.congolense* GPI is highly flexible, irrespective of the presence or absence of protein. From the cluster conformations it is clear that irrespective of the orientation of the attached protein, whether erect or reclining, the GPI lies close to the membrane with the glycan head spanning across the membrane surface, although the internal arrangement of GPI residues, particularly at the flexible linkages – Man2-Man1 and Gal-GlcNac – may vary.

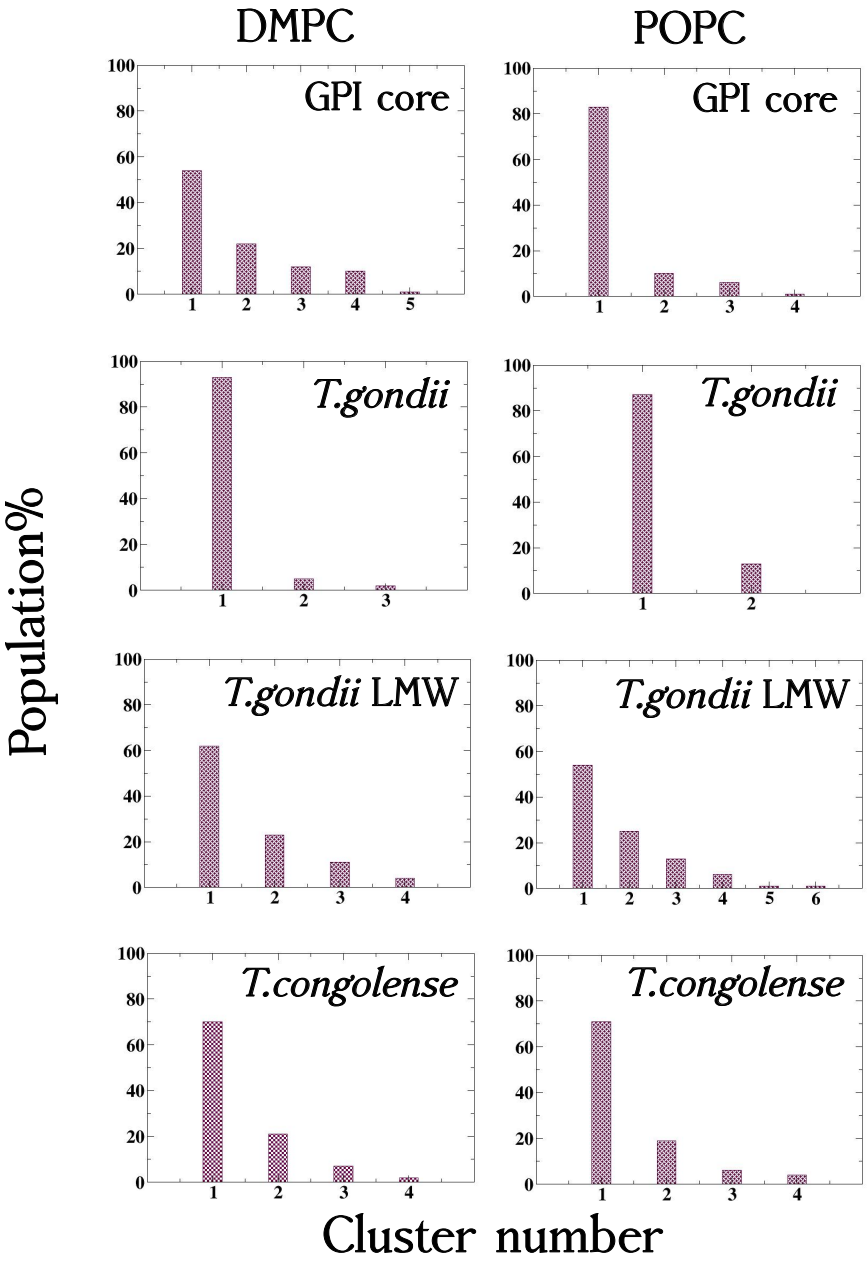
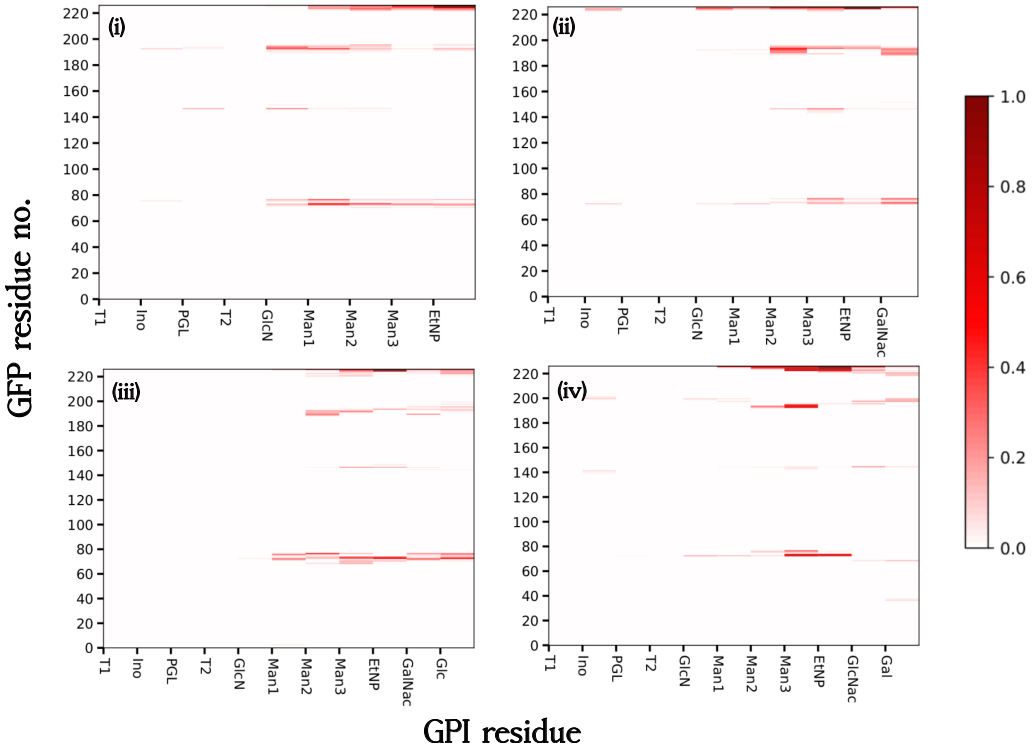


Figure 11: Cluster size populations obtained from 4μs worth of simulation data of GFP-attached GPIs – GPI core, *T.gondii*, *T.gondii* LMW, and *T.congolense* – in DMPC and POPC bilayer.

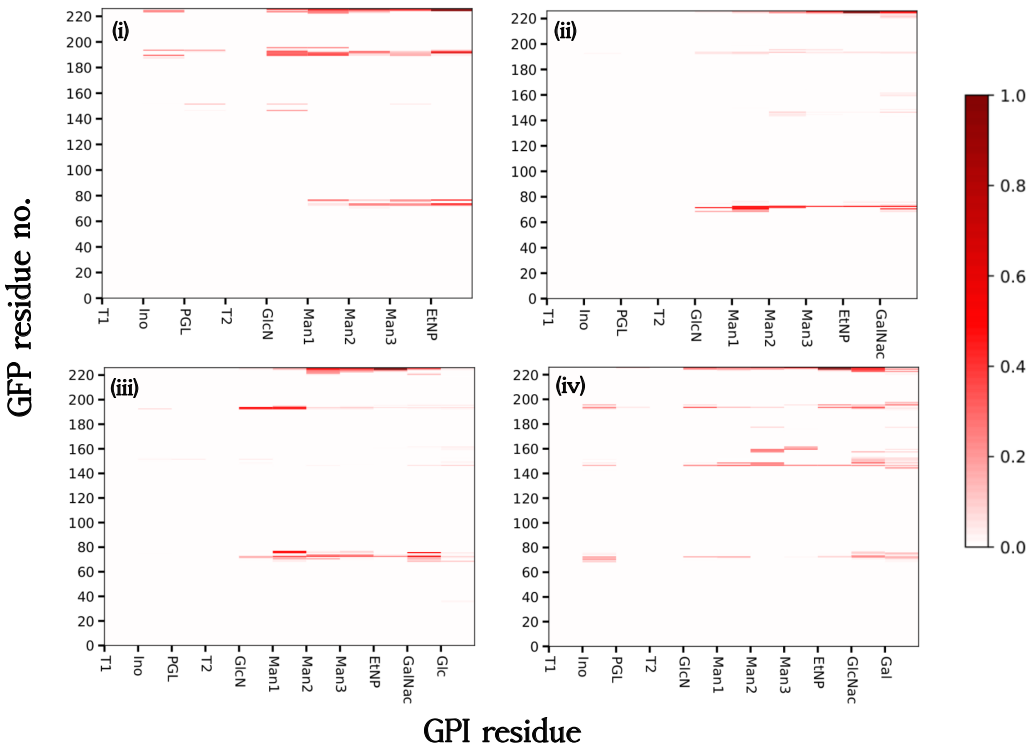
2.6.3 Interaction of GFP with GPI

Understanding the interactions between the GPI anchor and its attached protein is imperative to their conformational analysis. Just as GPIs form H-bonds with the membrane headgroup region, they also do so with the GFP (see Figure S5). This is expected as GPI is full of H-bond-forming hydroxyl groups. Because H-bonding is not the only possible way for intermolecular interactions, for a more comprehensive analysis we calculated the number of contacts formed between every possible pair of residues of GFP and GPI by counting the atoms of the protein lying within a cutoff of 0.5 nm from the reference GPI residue. A python library called *contact-map.py* developed by Swenson and Roet was used in this exercise[**swenson**]. Figure 12 shows contact maps for residue-wise contacts formed between GFP and the following GPI variants – (i) GPI core, (ii) *T.gondii*, (iii) *T.gondii* LMW, and (iv) *T.congolense* GPI – in (a) DMPC and (b) POPC bilayers. There is extensive interaction between GFP and all the GPI types which even goes so far as the innermost GPI sugar residue, inositol, and sometimes even to the phosphoglycerol (PGL) head too. There are substantial contacts made with GFP from Man1 upwards till

Man3 including the side branches, with GlcN, Ino, and PGL interacting only seldom (in decreasing order) with the protein. The contact maps show that certain regions of the protein consistently interact with all the GPI types. These regions, or residue patches, that are frequently in contact with the GPI are illustrated in Figure 13. A closer scrutiny into the individual trajectories brought out a relation between tilt angle of GFP and contact profiles with GPI. Lower tilt angle values corresponded with increased contacts of GPI with GFP at the residues 69 to 77 (blue) and decreased contacts with residues 190 to 196 (yellow) and the terminal residues 222 to 224 (magenta). The vice versa was also true. However, residues 225 (isoleucine) and 226 (threonine) constantly interact with all the GPIs, these being the penultimate and ultimate C-terminal protein residues respectively. This relation is particularly distinct in *T.gondii* LMW GPI. The frequency of contacts formed between GFP and each GPI variant is presented in Figure 14, with the frequency values mapped onto the respective residue of contact on the protein. Alongside these, contact frequency maps are displayed for the GPI residues making the most and second-most number of contacts with attached GFP. The most protein-interactive GPI residues include at least one of the side branch residues in all the parasitic GPIs. GPI core has Man1, the residue that branches out to side chains, making the maximal contacts with GFP in both DMPC and POPC. Interestingly, the residue-wise hydration number profiles show that the side branches are quite distinctly the most hydrated out of all residues (see Figure S6), inspite of substantial communication with the protein. These results, like those from the protein-free GPIs, are again suggestive of the importance of the side chain residues in GPI anchors. The total number of atom-atom contacts between GPIs and attached GFPs are listed in Table 2.



(a) GPI-anchored GFP in DMPC bilayer



(b) GPI-anchored GFP in POPC bilayer

Figure 12: Contact maps showing the relative fraction of frames containing contacts between every possible pair of residues of GFP (y axis) and GPI (x axis). On the x axis, T1 and T2 refer to myristoyl tails of GPI. The scale of contact fractions (colorbar) ranges from 0 to 1. Contact maps were obtained for GFP-attached-GPI variants – (i) GPI core, (ii) *T.gondii*, (iii) *T.gondii* LMW and, (iv) *T.congolense* GPIs inserted in (a) DMPC and (b) POPC bilayers. Each color box represents the GPI residue to its left and GFP residue under it.

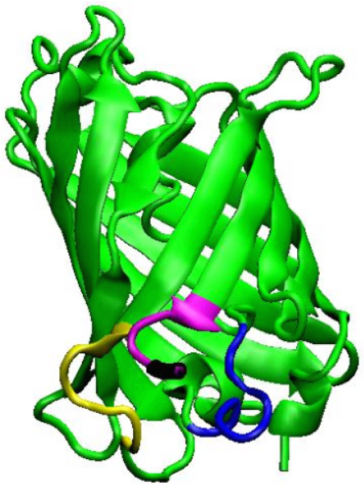
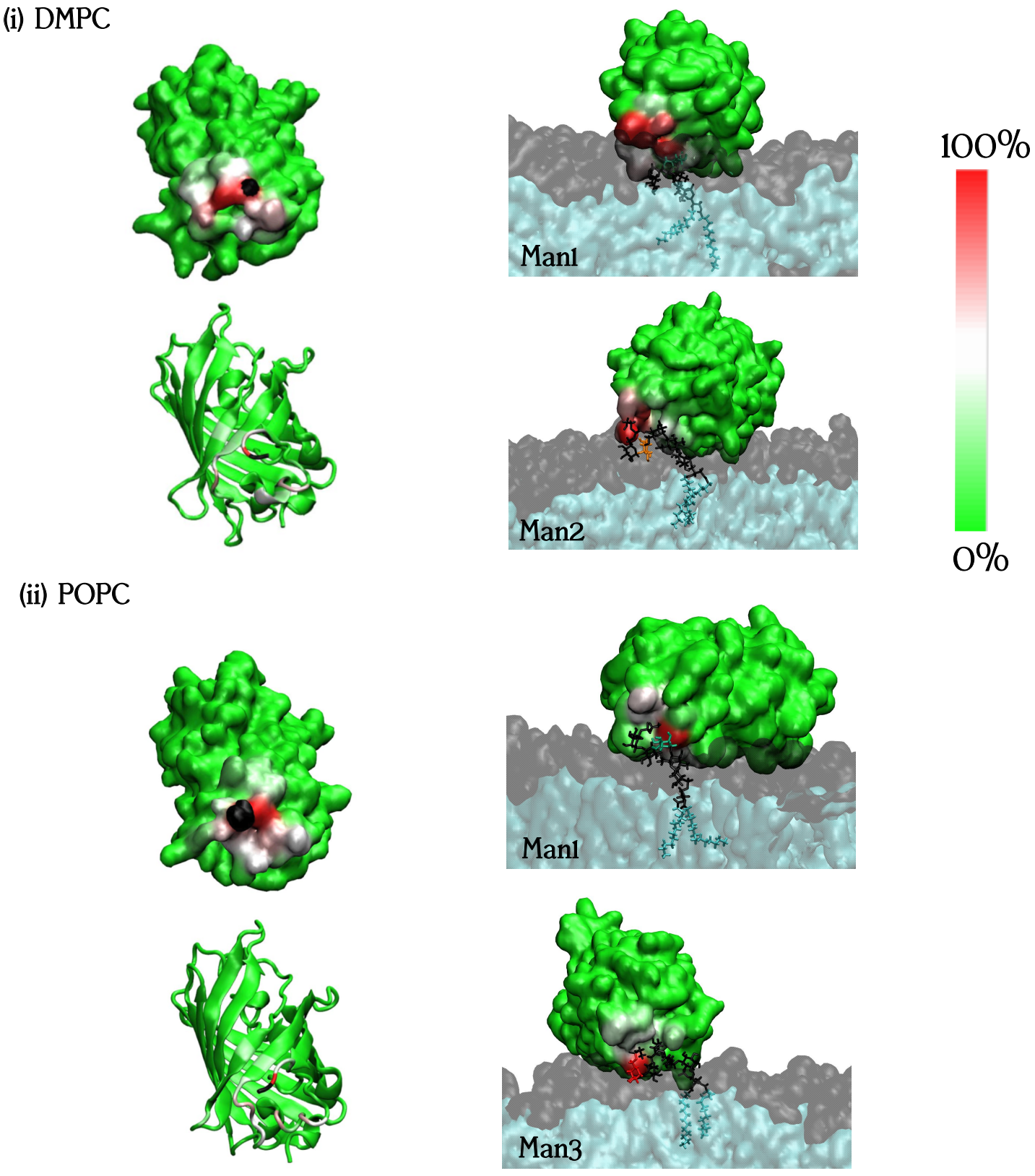
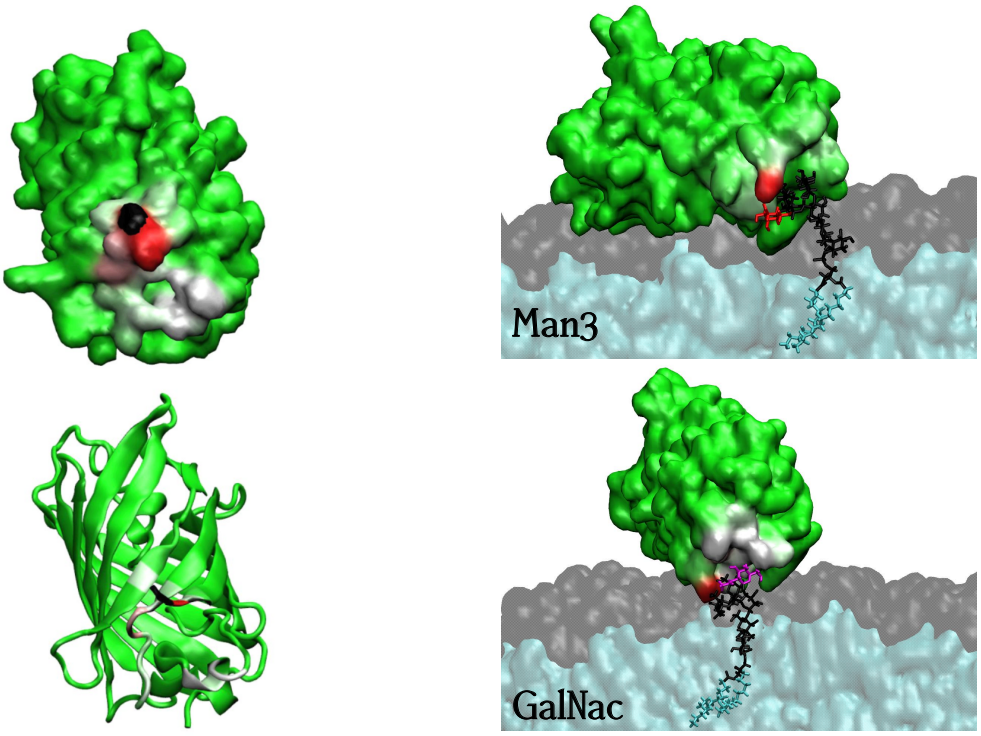


Figure 13: GFP residues that are in frequent contact with GPI. Different colors indicate different residue ranges, with residues 69-77 shown in blue, residues 190-196 in yellow, and residues 222-226 in magenta. The C terminal threonine residue is colored black to indicate the point of attachment to GPI.

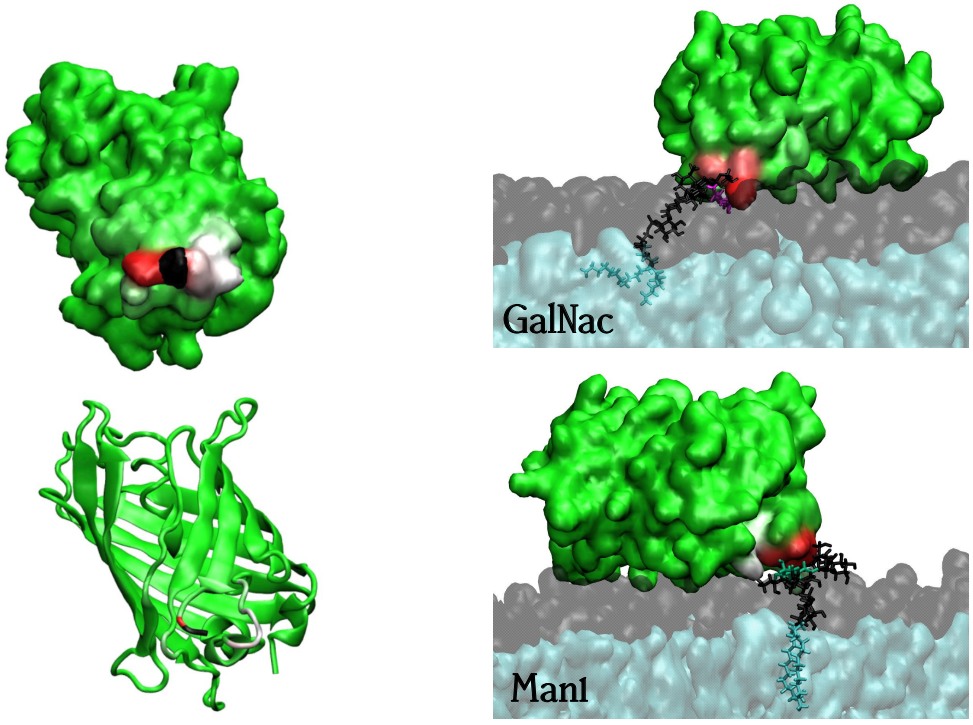


(a) GFP-attached GPI core

(i) DMPC

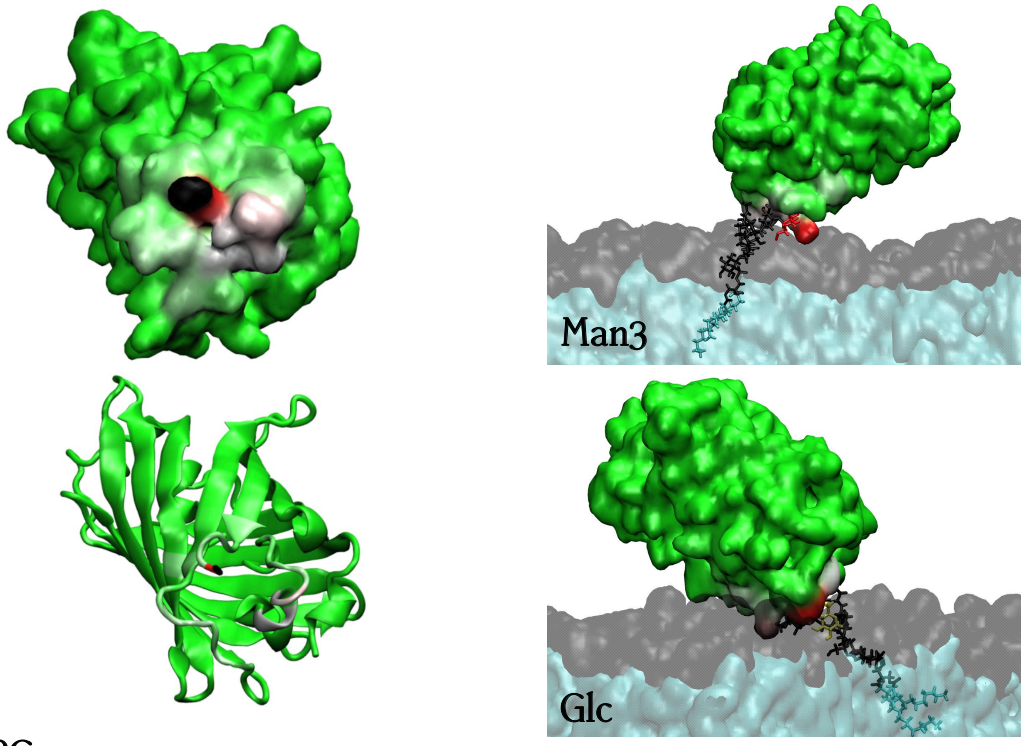


(ii) POPC

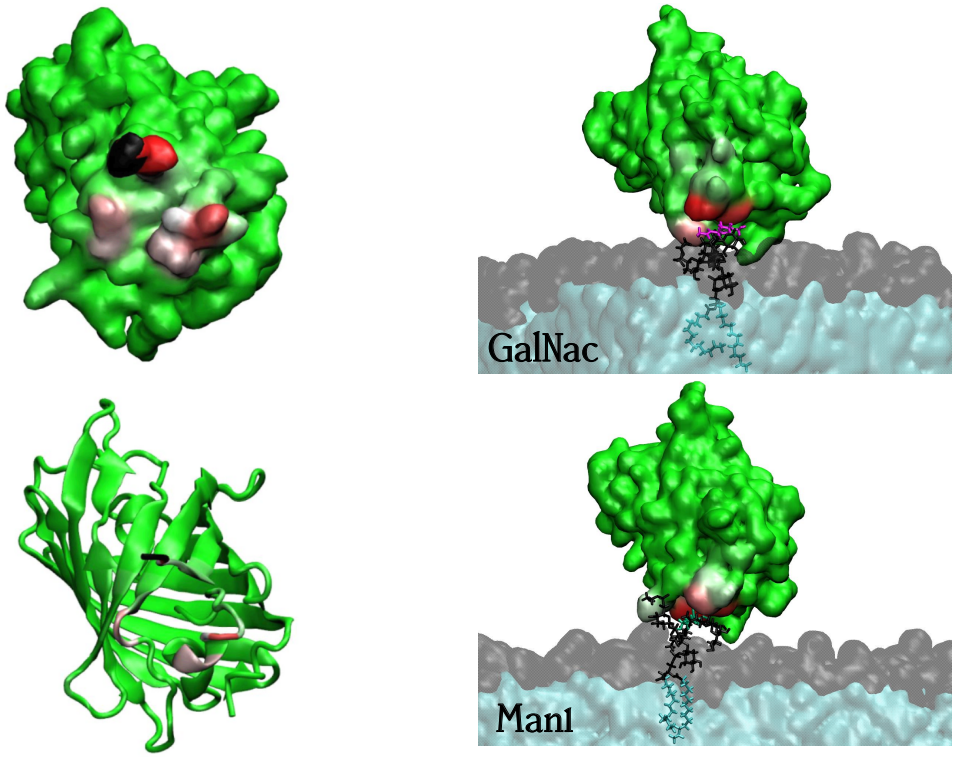


(b) GFP-attached *T.gondii* GPI

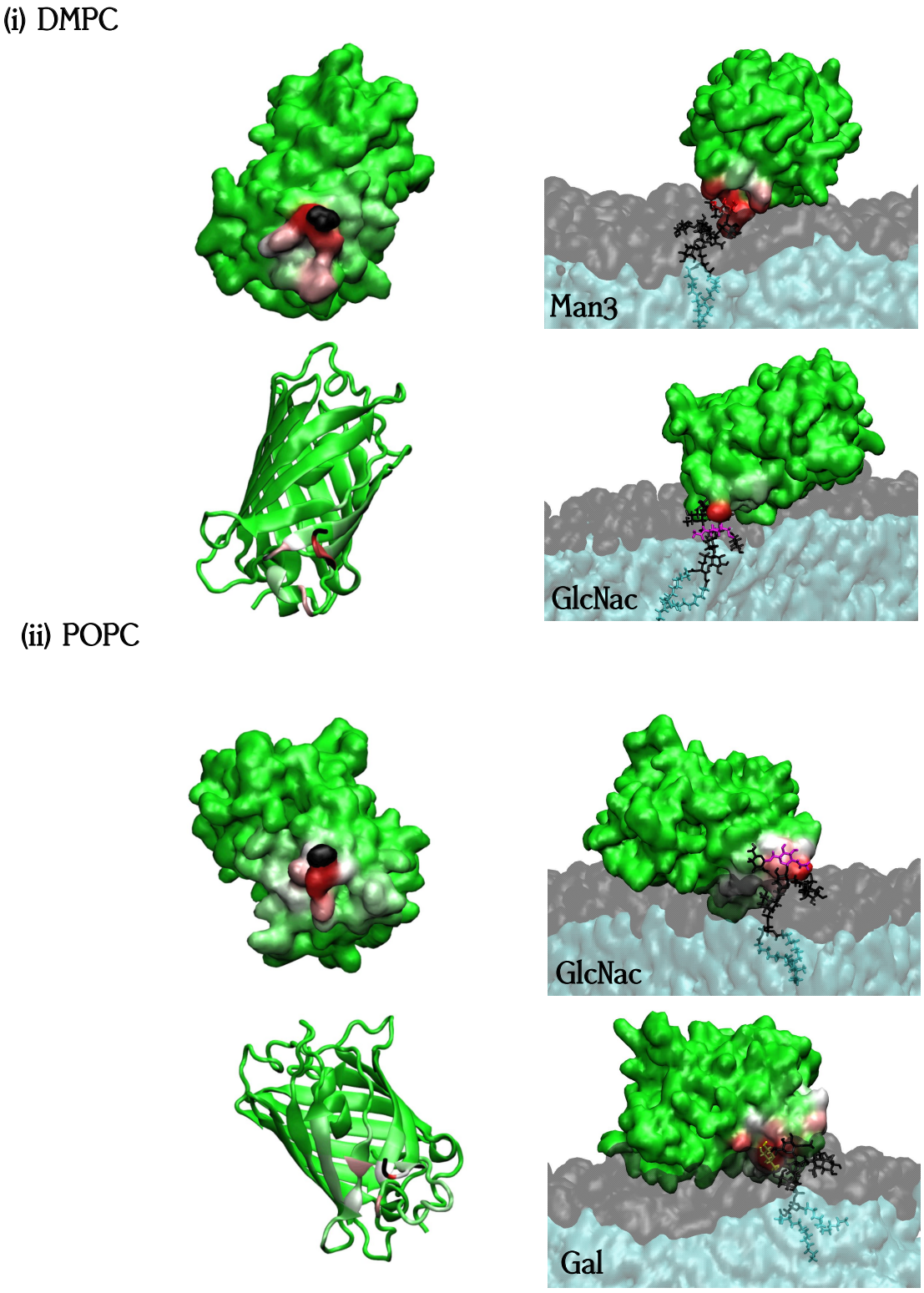
(i) DMPC



(ii) POPC



(c) GFP-attached *T.gondii* LMW GPI



(d) GFP-attached *T.congolense* GPI

Figure 14: Frequency of contacts made by atoms of the whole GPI glycan with GFP mapped onto the structure of GFP in two different representations (quicksurf and cartoon). On the right, contact frequency for the two most protein-interactive residues, as mentioned on the images, of the GPI variants is mapped onto representative conformations of the corresponding GPI-anchored GFP in bilayers. The C-terminal residue, threonine, is colored in black to depict the point of attachment of GPI. This analysis is presented for GFP-attached-GPI variants (a) GPI core, (b) *T.gondii*, (c) *T.gondii* LMW, and (d) *T.congolense* GPI anchored in (i) DMPC and (ii) POPC bilayers.

System	DMPC	POPC
GPI core	486.66±239.26	702.34±166.7
T.gondii	631.83±235.51	343.9±195.78
T.gondii LMW	514.22±204.88	970.48±247.2
T.congolense	673.6±233.23	709.35±152.27

Table 2: Average number of atom-atom contacts formed between GFP and GPI variants anchored in DMPC and POPC bilayers

3 Discussion

Chemical heterogeneity and conformational flexibility of glycolipids like the GPI anchor pose limitations on high-resolution structural investigation of these molecules in the environment of membranes experimentally. As GPI anchors are found abundantly across protozoan parasites, these molecules are of particular relevance to vaccine development and therapeutics, for which study of the molecular basis of recognition mechanism serves as an important supplement. Molecular dynamics has proven to be an indispensable tool in probing interactions at the atomic-level of resolution. In our previous work, we constructed a model of the GPI anchor (GPI core) by combining the force-fields GLYCAM06 and Lipid14 for molecular dynamics studies. In the present work, we have expanded this model by adding additional saccharide and phosphoethanolamine units to study three different parasitic GPI variants – *Toxoplasma gondii*, *Toxoplasma gondii* LMW, *Trypanosoma congolense* – and Human GPI. We have previously demonstrated the flexibility in the solution structure [24] and the membrane form[25] of the GPI glycan core, where large torsional fluctuations were observed at the Man2-$\alpha(1 \rightarrow 6)$-Man1 linkage, as has also been reported by Chevalier et al. through NMR and molecular mechanics[36]. Although the rest of the glycosidic linkages are comparatively rigid, they are still rather flexible when pitted against other biomolecules like lipids or proteins. The GPI variants in this study show a similar conformational pattern, with the differences in the overall three-dimensional conformation of the GPIs largely arising from differences in rotamer populations at the Man2-Man1 linkage, as well as at the Gal-$\beta(1 \rightarrow 6)$-GlcNac linkage in the case of *T.congolense* GPI. In the relatively smaller GPI types, the relative rotamer populations of the Man2-Man1 hinge followed the same order of preference as was seen previously in GPI glycan core in solution and membrane – *gg>gt>tg*. However, in the GPIs with longer branches (two side chain residues), the population of *gt* rotamer significantly increased at the expense of *gg*, even resulting in their equal proportions in *T.gondii* LMW in POPC (Table 1). The presence of the terminal phosphoethanolamine linker stabilized the torsion ψ of the terminal Man3-$\alpha(1 \rightarrow 2)$-Man2 linkage to essentially one value which was largely brought about by the H-bonding of EtNP with Man2. No other notable conformational deviations arose from the presence of EtNP. Out of all the protein-free GPIs considered in this study, *T.congolense* is the most flexible owing to the presence of two flexible $(1 \rightarrow 6)$ linkages. Moreover, all the glycosidic pairs in the most populated cluster of *T.congolense* GPI are not necessarily located at their corresponding lowest energy regions in the Ramachandran plots, even for the rather rigid linkages.

Parasitic GPIs differ from each other in the number/composition of the sugar side chains and phosphoethanolamine residues branching out from the conserved core. An interesting aspect of GPIs is that the side chain decorations are most often found directly bonded to the middle mannose Man1 (especially heavy branches), except when the additional residues are mannoses in which case they branch out from Man3 [37]. Our results are suggestive of a peculiar role played by Man1 in that it is the only residue that barely interacts with either the membrane or water, implying that it is extensively involved in intra-molecular interactions. Therefore, Man1 could be an important determinant of the conformation of GPI through its interactions with the other GPI residues. We also observed a consistent pattern across all the GPIs in the presentation of the side branches on the membrane surface. The side chain residues – GalNac in *T.gondii*, GalNac and Glc in *T.gondii* LMW, GlcNac in *T.congolense*, and GalNac and mEtNP in Human GPI – are the least membrane-bound, most hydrated and solvent-interactive residues, and therefore, could be the most accessible epitopes for recognition. However, note that the galactose (Gal) residue in *T.congolense* makes an exception by interacting profusely with the membrane heads, more so when inserted in POPC. The importance of the side branches has also been demonstrated in experimental studies. Using surface plasmon resonance, Debierre-Grockiego and coworkers showed that *T.gondii* GPIs strongly bind to galectin-3, a protein whose expression is a necessary precursor to TNF-α production by GPI-stimulated macrophages. Because galectin-3 binds specifically to β-galactosides, this finding implies that it associates with the residue GalNac in *T.gondii* GPI[15]. No difference in binding was observed in the presence of an additional Glc residue (*T.gondii* LMW GPI). However, other studies report a striking difference in the immunogenicity of the two GPI variants of

T.gondii. Azzouz et al. conducted *in vitro* macrophage activation with *T.gondii* isolates to reveal that the protein-free *T.gondii* LMW GPIs generated four to five times more TNF-α than the protein-bound *T.gondii* GPIs[38]. This finding was further corroborated by another independent study using synthetic, and so purer, protein-free *T.gondii* GPIs whereupon difference between the two GPI variants was observed in the intensity of binding to IgG and IgM antibodies from sera of infected patients. *T.gondii* LMW GPI bound more strongly to both antibody types than did *T.gondii* GPI, thereby, qualifying as a suitable biomarker to differentiate between latent and acute phases of infection that differ in the levels of IgM antibodies[16]. Interestingly, they also showed that a minimal epitope comprising the full side branch – Man1-GalNac-Glc – bound to the antibodies comparably as the full length variant. Our results are in agreement with these findings in that the side branches of *T.gondii* LMW GPI, GalNac-Glc, are the most solvent-accessible residues, and thereby, are potential antigenic epitopes. We were also interested in the differences in conformation and presentation of *T.gondii* and human/mammalian GPIs as they are very similar in structure, differing only by the presence of the additional EtNP residue (mEtNP) located at Man1 in human GPI. There was considerable difference seen in the orientation of human GPI w.r.t the membrane surface. On the one hand, the GPI core and parasitic GPIs lie along the plane of the membrane either through the main chain or the side chain, but on the other hand, human GPI is relatively erect on the surface, however still lying close to the membrane heads like the rest. In humans, an enzyme called GPI transamidase (GPIT) catalyzes the attachment of the GPI anchor to its protein. In a study by Vainauskas and Menon, it was shown that the middle mEtNP is required for the recognition of GPI by human GPIT, whereas the terminal tEtNP does not play a crucial role[39]. In fact, the minimal GPI epitope for recognition by GPIT comprised Ino-GlcN-Man1-mEtNP, although binding was slightly enhanced using the full GPI anchor. These observations align well with our findings as we show that mEtNP is more solvent-accessible than tEtNP, irrespective of the host lipid bilayer. Upon scrutinizing the glycosidic torsion populations of Human GPI, we noted a stark difference from that of *T.gondii* GPI at the Man2-Man1 linkage, as the dominant cluster conformation carried $\psi \approx 90°$, as opposed to $\psi \approx 170°$ in *T.gondii* GPI. Therefore, it is clear that the conformation of human GPI differs from that of *T.gondii* GPI, regardless of the large similarity in their chemical built. There is experimental proof for the absence of cross-reactivity between parasitic and human GPIs, *i.e.*, antibodies raised against parasitic GPIs do not bind with human GPI[16], thereby ruling out the possibility of auto-immune responses. Moreover, the observation that side chain residues act as antigenic epitopes through their solvent-exposure hints towards why immune responses to different GPI variants are specific in nature[13]. The different side chain compositions makes GPI variants different from each other, implying that specific immune responses could be directed against these differing residues.

As GPI anchors are often found attached to proteins, it was worthwhile to investigate the interactions between our model of GPI and an attached protein, for which we chose the GFP as the substitute for natural GPI-anchored proteins. GFP by itself does not favourably interact with zwitterionic membranes (such as DMPC and POPC) as is known experimentally[40] and also through our control simulations of GFP without the GPI on membrane (data not shown). However, upon attachment to a GPI anchor, protein-membrane interactions were enhanced in case of GPI core, *T.gondii* and *T.congolense*, but not much in case of *T.gondii* LMW GPI. The orientation of GFP largely fluctuates with an attached *T.gondii* LMW GPI, thus destabilizing protein-membrane interactions. Irrespective of the GPI type, all GPIs interact profusely with GFP, consequently affecting its orientation. The highly flexible phosphoethanolamine linker that bridges the GFP to GPI has much to contribute to the oscillating tilt values. Note that the tendency towards smaller tilt angles for the GFPs attached to the GPI variants increases when the lipid bilayer is of POPC. A plausible reason behind this trend could be competing interactions of GPI with the lipid heads and GFP, because GPIs have been shown to be more embedded into the membrane headgroup region when inserted in POPC compared to in DMPC. The GFP-vs-GPI contact maps also reveal that contacts are made mostly at the flexible and disordered loops at the base of the rigid barrel of GFP, and so these interactions would be weak, thus, contributing to the unstable orientation of the protein. Cluster conformations show that irrespective of the orientation of GFP, GPIs lie close to and along the plane of the membrane, consequently pulling the protein close to the membrane, even with *T.gondii* LMW GPI where although the orientation of the protein may not be stable, but the protein is anchored close to the membrane surface. It has been reported that GPI anchor brings about conformational change in Thy1[41], but not in PrP protein[21]. As GFP has quite a rigid structure we did not observe any significant conformational changes. Our model of GPI-anchored GFP in bilayer appears similar to the structure of GPI-anchored VSG of Trypanosomes in a membrane deduced by Homans et al. where the GPI anchor, by spanning across the membrane surface, acts as a space filler between the protein and the membrane[6]. Another study on the GPI-anchored VSG reported that the galactose side branch increases the volume of the C-terminal domain by associating closely with the protein[42]. These findings coupled with ours (of *T.congolense*) together suggest that the membrane-protective property of the VSG is assisted by close contacts of the GPI with

both the membrane and the protein. Our results also bring to light an interesting dual characteristic of the GPI side branches when attached to GFP, where on the one hand they are in close proximity to the membrane, and on the other are the most solvent-accessible residues. This suggests that despite protein-attachment, the side chain residues are still exposed, and could be potential chemical targets for drug therapies.

4 Conclusions

In summary, using molecular dynamics we have investigated and compared the conformational behaviour of parasitic and human/mammalian protein-free and protein-attached (with GFP) GPIs inserted into pure bilayers of DMPC and POPC by extending our previously reported model of the GPI anchor core. Our results from the simulations of protein-free GPIs indicate that the side chain residues that branch out from the middle mannose Man1, whether sugar residue or phosphoethanolamine, are the most solvent-accessible epitopes along the whole length of GPI, and thus, are potential targets for recognition by macrophages (parasitic GPIs) or GPIT enzyme (human GPI). The residue Man1 stands out to serve an important constructive purpose in dictating the overall conformation of the GPI through extensive intramolecular interactions, with barely any communication with either the membrane or the solvent. This study provides a rationale behind the importance of the side chain residues that has been demonstrated in experiments. The orientation and presentation of the GPI anchor w.r.t. lipid bilayers is qualitatively consistent across all GPI variants in that the GPIs lie along the plane of the bilayer spanning the surface. Such a flop-down orientation allows for the side chain residues to project out into the solvent making them readily approachable. An erect conformation of GPI would have, instead, exposed the terminal residues, Man2 and Man3, the most. Thus, here we have further attempted to rest the case of the controversial subject of the orientation of membrane-embedded GPIs. The lipid bilayer also makes a marked difference in the recognition process, as it was observed that the GPIs are more buried, and so less exposed, in the lipid headgroup region in POPC compared to DMPC. We conclude from the GPI-anchored GFP simulations that the orientation of the protein on the membrane depends on several factors – (a) the molecular size of GPI, (b) the type of glycosidic linkage between residues, (c) the region of the protein in contact with GPI (whether a rigid or flexible patch), (d) the flexibility of the EtNP linker, and (e) the type of lipid bilayer. Nevertheless, the protein is brought close to the membrane through interactions mediated by the GPI anchor. However, the extent of interactions depends on the GPI-type. As is evident from this study that the presentation of the protein depends even on small differences in chemistry between the GPI variants (e.g. between *T.gondii* LMW and *T.congolense* GPI), using GPI analogs to study the behavior of true GPIs should be avoided. Through this investigation, we have demonstrated the applicability of our GPI model in understanding the molecular basis for recognition of parasitic and human GPIs. Our model serves as a suitable candidate for studying many other parasitic GPIs so as to assist in the development of vaccine-strategies and therapeutics against GPI-related diseases.

5 Computational Methods

5.1 MD simulations

All the MD simulations were conducted using the simulation package GROMACS version-2018.3[43]. Lipid bilayers, including the lipid tail of the GPI anchors, were modeled with Lipid14[27] parameters. The glycan heads of all the GPI anchors in this study were designed with GLYCAM06[26], an AMBER-compatible force-field designed exclusively for carbohydrates. The glycan head and the lipid tail were linked together via a hybrid inositol-phosphoglycerol moiety that was constructed in our previous work[25]. The protein GFP was constructed with AMBER's latest force-field for proteins called AMBERff14SB[28]. These three force-fields belong to the AMBER family and are known to be compatible with each other. The structure file of GFP was taken from its crystal structure, RCSB id – 1EMA. Explicit waters modeled with TIP3P[44] were used to represent the aqueous phase. Simulations of bilayers were set up in rectangular boxes of dimensions (6.5x6.5x17) nm^3 for the free GPI systems and (12.5x12.5x22)nm^3 for the GPI-anchored GFP systems. The smaller protein-free GPI systems consisted of one GPI anchor per leaflet in pure bilayers of (8x8) lipids, whereas the bigger systems had one GPI-anchored GFP embedded in one leaflet of pure (16x16) lipid bilayers. The pure bilayers considered in this work were of DMPC and POPC. Charged systems were neutralized by adding Na^+ or Cl^- counter ions. System construction was achieved with the LEap facility of AMBER, following which the topology and structure files were converted to GROMACS' format using a modified version[24] of the script originally written by Sorin and Pande[45]. For every protein-free GPI system, 3 independent 1μs long simulations were performed resulting in 3μs worth of sampling. Similarly 4 independent 1μs long trajectories amounting to 4μs were carried out for the GFP-GPI systems.

To begin with, each system was subjected to energy-minimization through 10000 steps of the conjugate-gradient method with steepest descent invoked every 1000 steps. Next, NPT equilibration was performed for 100ps at temperature 100K while restraining the positions of the solute molecules so as to relax the water molecules around the solute. The temperature was then ramped up to 303K for another 100 ps while still restraining the solute. After equilibration, the restraint was released to carry out production run. Temperature was maintained at 303 K by a Langevin thermostat with a coupling constant of 1ps. Semiisotropic pressure coupling was applied with a time constant of 1 ps by the berendsen barostat[46] to maintain the pressure at 1 bar. The linear constraint solver algorithm[47] was employed to constrain all the bonds containing hydrogen. The leap-frog stochastic dynamics integrator[48] was used at a time-step of 2 fs to solve the equations of motion to propagate the system. The Particle-Mesh Ewald (PME) algorithm was used to describe electrostatic interactions[49]. The cut-off for both Coulombic and van der Waals potentials was 1 nm.

5.2 Charge derivation for phosphoethanolamine

To ensure force-field compatibility with the attached GPI glycan, we followed the approach of the original GLYCAM paper[26] to derive charges for the phosphoethanolamine (EtNP) residue that is attached to the terminal mannose Man3 of GPI. EtNP alongwith Man3, shown in Figure S1 was considered for the calculation of atomic partial charges for the EtNP residue. Geometry optimization was conducted at a high level of theory – B3LYP/6-31++g(2d,2p) – including diffuse orbitals to account for the charges. The main chain bonds of the molecule were maintained in an all-trans arrangement throughout the calculation. This was to avoid transfer of positively charged H^+ from NH_3^+ to negatively charged O^- of PO_4^-. Partial charges were then derived by applying the restrained electrostatic potential (RESP) method for charge fitting. The electrostatic potential was obtained from the optimized geometry at the HF/6-31*G level of theory. As the designed system is meant for condensed-phase simulations, a restraint weight of 0.01 was applied for the RESP charge fitting. Charges on alkyl hydrogens were fixed at 0 to maintain consistency with GLYCAM parameters. GLYCAM's charge derivation protocol offers modular blocks of sugar residues for building long glycans with variable branching. As a result, every terminal sugar residue carries a net charge of 0.194. The total charge on the EtNP linker was also 0.194, consistent with GLYCAM's formula. The atom types and bonded parameters (bonds, angles and dihedrals) assigned to the linker were taken from GLYCAM's database. In our previous work, partial charges were derived similarly for the EtNP linker that bonds with GPI at one end and to protein on the other. The details of the methodology can be found in ref[25].

5.3 Cluster size calculation

Cluster analysis on the conformation of GPIs was performed using the method described in the work of Daura et al [50]. The analysis was carried out on a concatenated trajectory of all independent simulations of each GPI variant under study. For each protein-free GPI system, 5000 structures were used for the calculation. 2500 structures from each protein-attached GPI system were subjected to clustering. In this method, the root mean square deviation (RMSD) of atom positions between every pair of structures is calculated. The number of neighbours of a cluster is counted based on RMSD value within a given cut-off. The structure with maximum neighbours is picked as the center of that cluster. This process of filtering is done on repeat until there are no structures left in the pool. We applied a cut-off of 0.3 nm, which is higher than the usual cut-off (0.1 nm) used for proteins. Carbohydrates are a lot more flexible than proteins, and a tight cut-off value would not give meaningful results. The GROMACS utility *gmx cluster* was employed for this calculation.

References

[1] MAl Ferguson, MG Low, and GA Cross. "Glycosyl-sn-1, 2-dimyristylphosphatidylinositol is covalently linked to Trypanosoma brucei variant surface glycoprotein." In: *Journal of Biological Chemistry* 260.27 (1985), pp. 14547–14555.

[2] Steve W Homans et al. "Complete structure of the glycosyl phosphatidylinositol membrane anchor of rat brain Thy-1 glycoprotein". In: *Nature* 333.6170 (1988), p. 269.

[3] Michael A Ferguson. "The structure, biosynthesis and functions of glycosylphosphatidylinositol anchors, and the contributions of trypanosome research". In: *J Cell Sci* 112.17 (1999), pp. 2799–2809.

[4] Margot G Paulick et al. “Synthetic analogues of glycosylphosphatidylinositol-anchored proteins and their behavior in supported lipid bilayers”. In: *Journal of the American Chemical Society* 129.37 (2007), pp. 11543–11550.

[5] Margot G Paulick et al. “A chemical approach to unraveling the biological function of the glycosylphosphatidylinositol anchor”. In: *Proceedings of the National Academy of Sciences* 104.51 (2007), pp. 20332–20337.

[6] SW Homans et al. “Solution structure of the glycosylphosphatidylinositol membrane anchor glycan of Trypanosoma brucei variant surface glycoprotein”. In: *Biochemistry* 28.7 (1989), pp. 2881–2887.

[7] Marty T Lehto and Frances J Sharom. “Proximity of the protein moiety of a GPI-anchored protein to the membrane surface: a FRET study”. In: *Biochemistry* 41.26 (2002), pp. 8368–8376.

[8] TW Rademacher, CJ Edge, and RA Dwek. “Dropping anchor with the lipophosphoglycans”. In: *Current Biology* 1.1 (1991), pp. 41–42.

[9] Sebastian Schuck and Kai Simons. “Controversy fuels trafficking of GPI-anchored proteins”. In: *J Cell Biol* 172.7 (2006), pp. 963–965.

[10] Christian Eggeling et al. “Direct observation of the nanoscale dynamics of membrane lipids in a living cell”. In: *Nature* 457.7233 (2009), p. 1159.

[11] Eva Sevcsik et al. “GPI-anchored proteins do not reside in ordered domains in the live cell plasma membrane”. In: *Nature communications* 6 (2015), p. 6969.

[12] Taroh Kinoshita and Morihisa Fujita. “Biosynthesis of GPI-anchored proteins: special emphasis on GPI lipid remodeling”. In: *Journal of lipid research* 57.1 (2016), pp. 6–24.

[13] Yu-Hsuan Tsai, Xinyu Liu, and Peter H Seeberger. “Chemical biology of glycosylphosphatidylinositol anchors”. In: *Angewandte Chemie International Edition* 51.46 (2012), pp. 11438–11456.

[14] David R Taylor and Nigel M Hooper. “GPI-anchored proteins in health and disease”. In: *Post-translational modifications in health and disease*. Springer, 2011, pp. 39–55.

[15] Françoise Debierre-Grockiego et al. “Binding of Toxoplasma gondii glycosylphosphatidylinositols to galectin-3 is required for their recognition by macrophages”. In: *Journal of Biological Chemistry* 285.43 (2010), pp. 32744–32750.

[16] Sebastian Götze et al. “Diagnosis of toxoplasmosis using a synthetic glycosylphosphatidylinositol glycan”. In: *Angewandte Chemie International Edition* 53.50 (2014), pp. 13701–13705.

[17] Stefan Magez et al. “VSG-GPI anchors of African trypanosomes: their role in macrophage activation and induction of infection-associated immunopathology”. In: *Microbes and infection* 4.9 (2002), pp. 999–1006.

[18] Louis Schofield et al. “Synthetic GPI as a candidate anti-toxic vaccine in a model of malaria”. In: *Nature* 418.6899 (2002), pp. 785–789.

[19] Carla G Chiodi and Hugo Verli. “Structural characterization of NETNES glycopeptide from Trypanosoma cruzi”. In: *Carbohydrate research* 373 (2013), pp. 28–34.

[20] In: ().

[21] Johannes Zuegg and Jill E Gready. “Molecular dynamics simulation of human prion protein including both N-linked oligosaccharides and the GPI anchor”. In: *Glycobiology* 10.10 (2000), pp. 959–974.

[22] Emilia L Wu et al. “Insight into early-stage unfolding of GPI-anchored human prion protein”. In: *Biophysical journal* 109.10 (2015), pp. 2090–2100.

[23] Jing Li et al. “Spontaneous insertion of GPI anchors into cholesterol-rich membrane domains”. In: *AIP Advances* 8.5 (2018), p. 055210.

[24] Marko Wehle et al. “Mechanical compressibility of the glycosylphosphatidylinositol (GPI) anchor backbone governed by independent glycosidic linkages”. In: *Journal of the American Chemical Society* 134.46 (2012), pp. 18964–18972.

[25] Pallavi Banerjee et al. “A molecular dynamics model for glycosylphosphatidyl-inositol anchors:“flop down” or “lollipop”?” In: *Physical Chemistry Chemical Physics* 20.46 (2018), pp. 29314–29324.

[26] Karl N Kirschner et al. “GLYCAM06: a generalizable biomolecular force field. Carbohydrates”. In: *Journal of computational chemistry* 29.4 (2008), pp. 622–655.

[27] Callum J Dickson et al. “Lipid14: the amber lipid force field”. In: *Journal of chemical theory and computation* 10.2 (2014), pp. 865–879.

[28] James A Maier et al. “ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB”. In: *Journal of chemical theory and computation* 11.8 (2015), pp. 3696–3713.

[29] Karl N Kirschner and Robert J Woods. “Solvent interactions determine carbohydrate conformation”. In: *Proceedings of the National Academy of Sciences* 98.19 (2001), pp. 10541–10545.

[30] Dhilon S Patel et al. “Conformational properties of α-or β-(1$\rightarrow$ 6)-linked oligosaccharides: Hamiltonian replica exchange MD simulations and NMR experiments”. In: *The Journal of Physical Chemistry B* 118.11 (2014), pp. 2851–2871.

[31] Norman L Allinger. *Molecular structure: understanding steric and electronic effects from molecular mechanics*. John Wiley & Sons, 2010.

[32] Mari L DeMarco and Robert J Woods. “Atomic-resolution conformational analysis of the GM3 ganglioside in a lipid bilayer and its implications for ganglioside–protein recognition at membrane surfaces”. In: *Glycobiology* 19.4 (2009), pp. 344–355.

[33] Mari L DeMarco et al. “Presentation of membrane-anchored glycosphingolipids determined from molecular dynamics simulations and NMR paramagnetic relaxation rate enhancement”. In: *Journal of the American Chemical Society* 132.4 (2010), pp. 1334–1338.

[34] Pranav Sharma et al. “Nanoscale organization of multiple GPI-anchored proteins in living cell membranes”. In: *Cell* 116.4 (2004), pp. 577–589.

[35] Ian A Prior et al. “Direct visualization of Ras proteins in spatially distinct cell surface microdomains”. In: *The Journal of cell biology* 160.2 (2003), pp. 165–170.

[36] Franck Chevalier et al. “Conformational study of GPI anchors: the common oligosaccharide GPI anchor backbone”. In: *European journal of organic chemistry* 2005.16 (2005), pp. 3489–3498.

[37] Michael AJ Ferguson, Gerald W Hart, and Taroh Kinoshita. “Glycosylphosphatidylinositol anchors”. In: *Essentials of Glycobiology. 3rd edition*. Cold Spring Harbor Laboratory Press, 2017.

[38] Nahid Azzouz et al. “Toxoplasma gondii grown in human cells uses GalNAc-containing glycosylphosphatidylinositol precursors to anchor surface antigens while the immunogenic Glc–GalNAc-containing precursors remain free at the parasite cell surface”. In: *The international journal of biochemistry & cell biology* 38.11 (2006), pp. 1914–1925.

[39] Saulius Vainauskas and Anant K Menon. “Ethanolamine phosphate linked to the first mannose residue of glycosylphosphatidylinositol (GPI) lipids is a major feature of the GPI structure that is recognized by human GPI transamidase”. In: *Journal of Biological Chemistry* 281.50 (2006), pp. 38358–38364.

[40] Hyunjin Kim, Hamid Samareh Afsari, and Wonhwa Cho. “High-throughput fluorescence assay for membrane-protein interaction”. In: *Journal of lipid research* 54.12 (2013), pp. 3531–3538.

[41] E Barboni et al. “The glycophosphatidylinositol anchor affects the conformation of Thy-1 protein”. In: *Journal of cell science* 108.2 (1995), pp. 487–497.

[42] Nicola G Jones et al. “Structure of a glycosylphosphatidylinositol-anchored domain from a trypanosome variant surface glycoprotein”. In: *Journal of Biological Chemistry* 283.6 (2008), pp. 3584–3593.

[43] Mark James Abraham et al. “GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers”. In: *SoftwareX* 1 (2015), pp. 19–25.

[44] William L Jorgensen et al. “Comparison of simple potential functions for simulating liquid water”. In: *The Journal of chemical physics* 79.2 (1983), pp. 926–935.

[45] Eric J Sorin and Vijay S Pande. “Exploring the helix-coil transition via all-atom equilibrium ensemble simulations”. In: *Biophysical journal* 88.4 (2005), pp. 2472–2493.

[46] Herman JC Berendsen et al. “Molecular dynamics with coupling to an external bath”. In: *The Journal of chemical physics* 81.8 (1984), pp. 3684–3690.

[47] Berk Hess et al. “LINCS: a linear constraint solver for molecular simulations”. In: *Journal of computational chemistry* 18.12 (1997), pp. 1463–1472.

[48] Wilfred F Van Gunsteren and Herman JC Berendsen. “A leap-frog algorithm for stochastic dynamics”. In: *Molecular Simulation* 1.3 (1988), pp. 173–185.

[49] Darden Tom, York Darrin, and Pedersen Lee. “Particle mesh Ewald: an N [center-dot] log (N) method for Ewald sums in large systems”. In: (1993).

[50] Xavier Daura et al. “Peptide folding: when simulation meets experiment”. In: *Angewandte Chemie International Edition* 38.1-2 (1999), pp. 236–240.

Supporting information

The importance of side branches of glycosylphosphatidylinositol anchors: A molecular dynamics perspective

Pallavi Banerjee, Reinhard Lipowsky and Mark Santer

S1 Charges on PE linker

Table.S1 presents the derived charges on the phosphoethanolamine linker as protein-free and protein-attached.

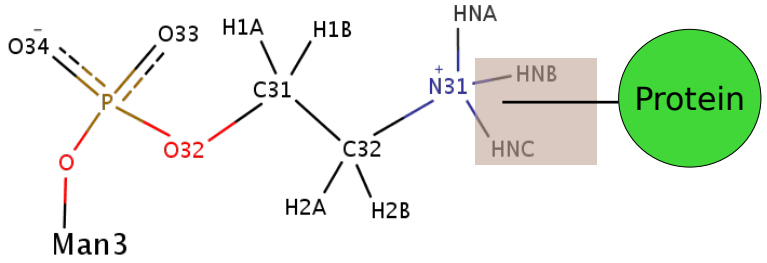
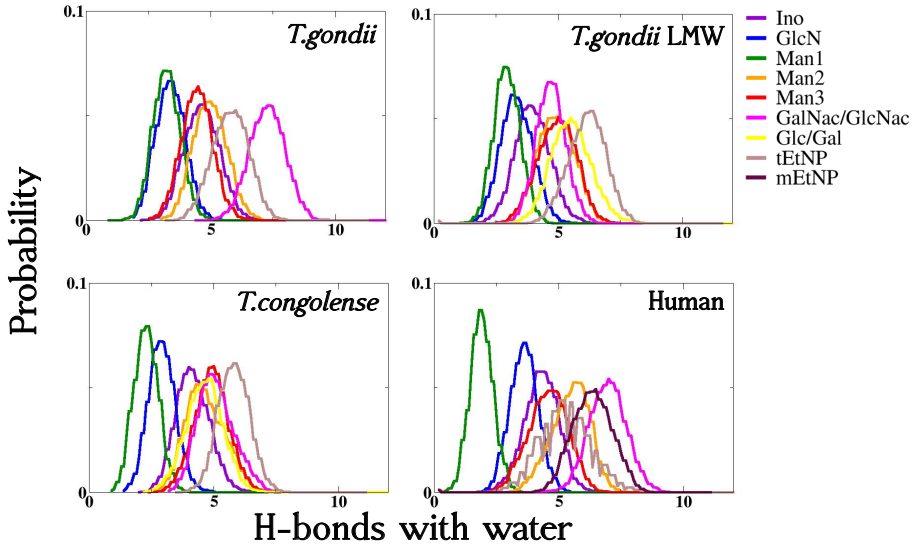


Figure S1: Chemical structure of the phosphoethanolamine linker that attaches to free GPIs. Atoms are depicted with their atom names. The fuzzy brown box indicates the presence or absence of ammonium hydrogens when the linker exists without (former) and with protein (latter).

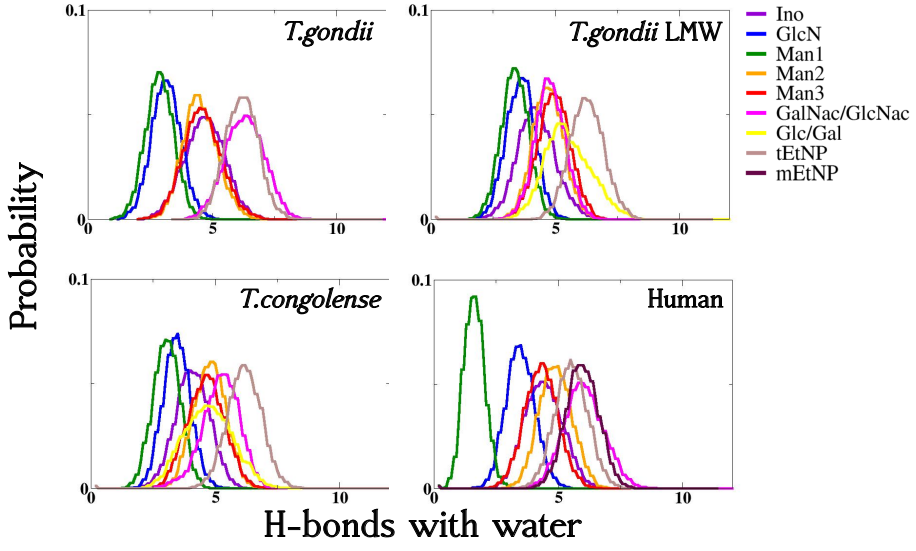
	Protein-free		Protein attached	
Atom name	Atom type	Charge	Atom type	Charge
P31	P	1.0136	P	1.0868
O32	Os	-0.5495	Os	-0.5815
O33	O2	-0.7108	O2	-0.7635
O34	O2	-0.7108	O2	-0.7635
C31	Cp	0.262	Cp	0.285
H1A	H1	0	H1	0
H1B	H1	0	H1	0
C32	Cg	0.3882	Cg	0.2547
H2A	Hp	0	H1	0
H2B	Hp	0	H1	0
N31	N3	-0.4957	Ng	-0.6089
HNA	H	0.3322	H	0.2848
HNB	H	0.3322		
HNC	H	0.3322		

Table S1: Atom types and charges on the atoms of the phosphoethanolamine unit, when protein-free and protein-attached. Atom names correspond with those in Figure S1

S2 Interaction of GPI with water



(a)

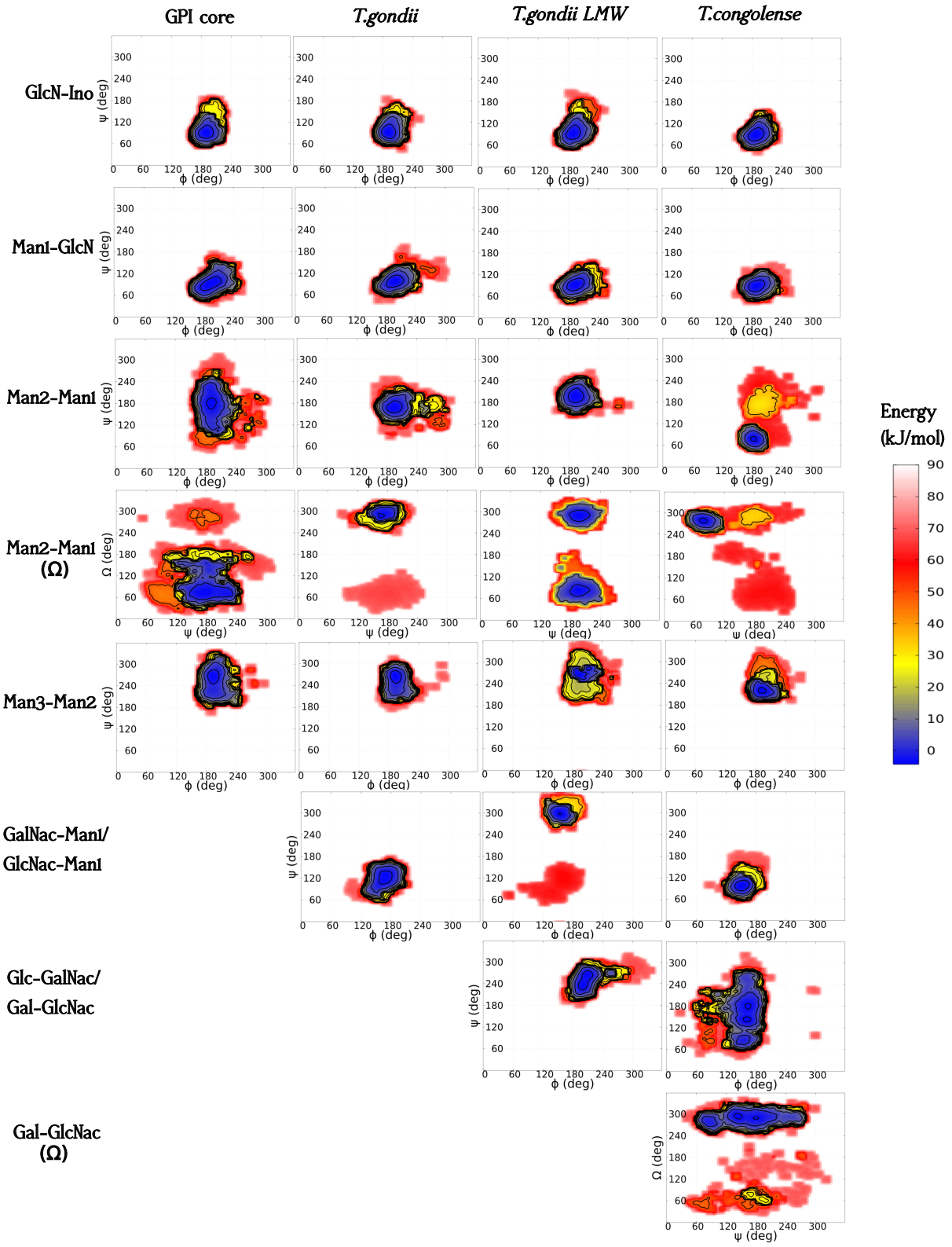


(b)

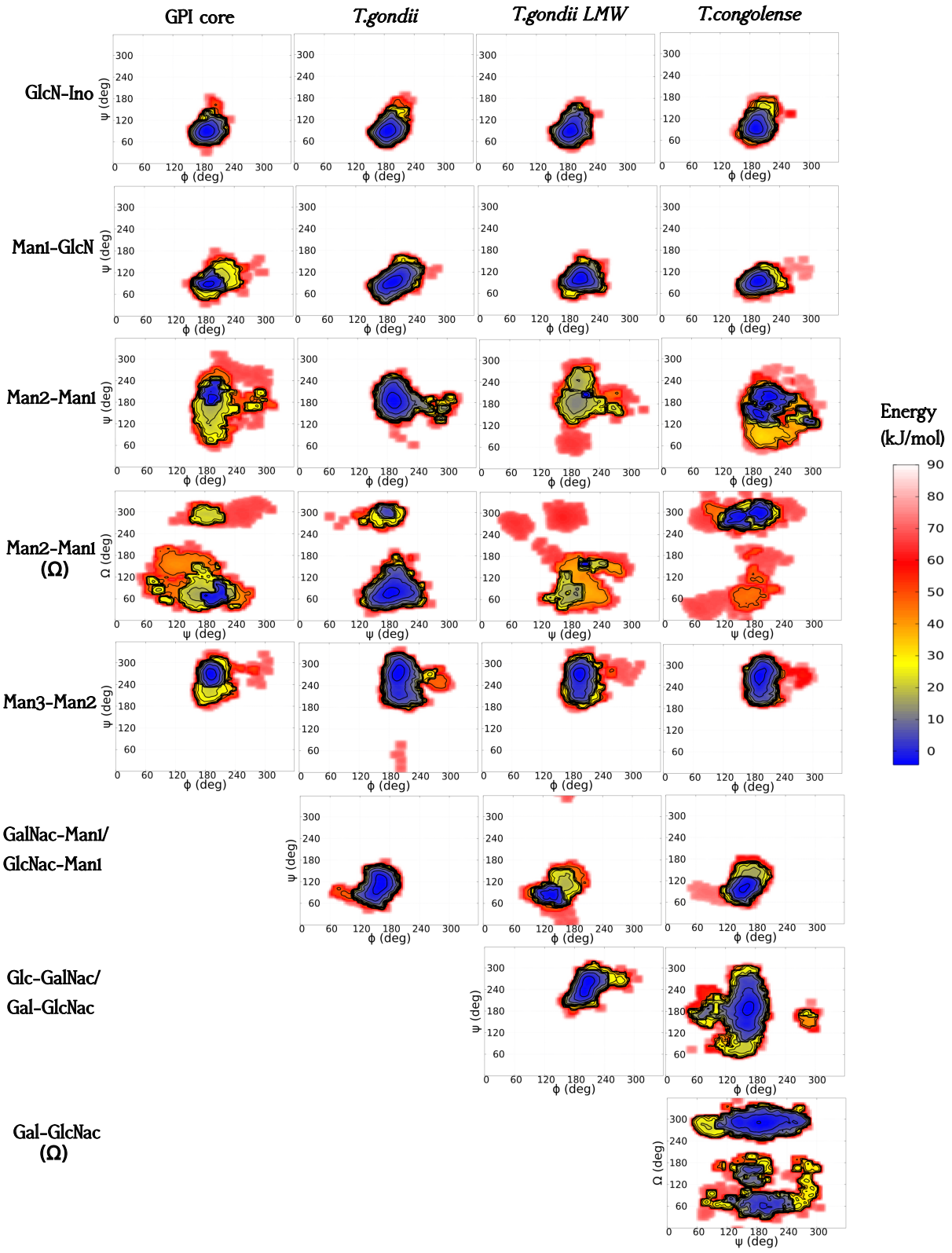
Figure S2: Probability distributions of the number of H-bonds formed between each residue of the GPI variants – *T.gondii*, *T.gondii* LMW, *T.congolense*, Human GPI – and water, with GPIs inserted in (a) DMPC and (b) POPC bilayer

S3 GPI-anchored GFP

S3.1 Dihedral analysis



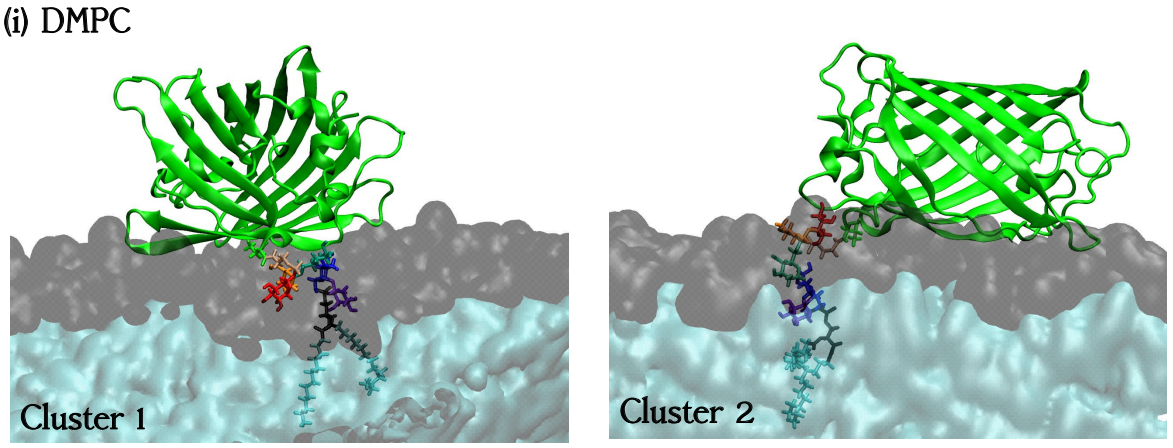
(a) GPI-anchored GFP in DMPC bilayer



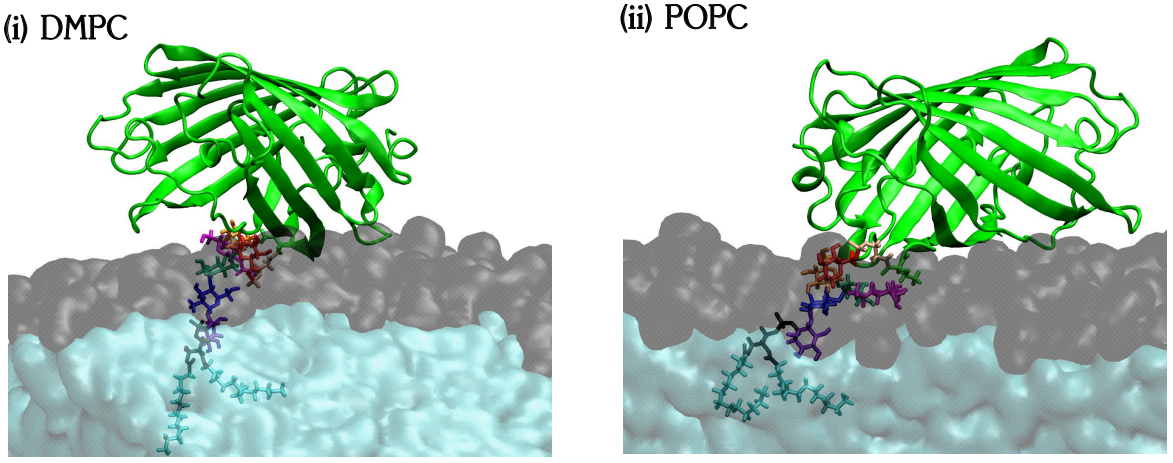
(b) GPI-anchored GFP in POPC bilayer

Figure S3: Two dimensional free energy landscapes as functions of dihedral pairs (ϕ, ψ) or (ψ, Ω) of all the glycosidic linkages in the GFP-attached-GPI variants embedded in pure (a) DMPC and (b) POPC bilayers. Each of these plots is obtained from 4μs worth of MD simulation data. Note that GalNac-Man1 belongs to *T.gondii* and *T.gondii* LMW, whereas GlcNac-Man1 belongs to *T.congolense*. Similarly Glc-GalNac belongs to *T.gondii LMW* side branch, and Gal-GlcNac occurs in *T.congolense*

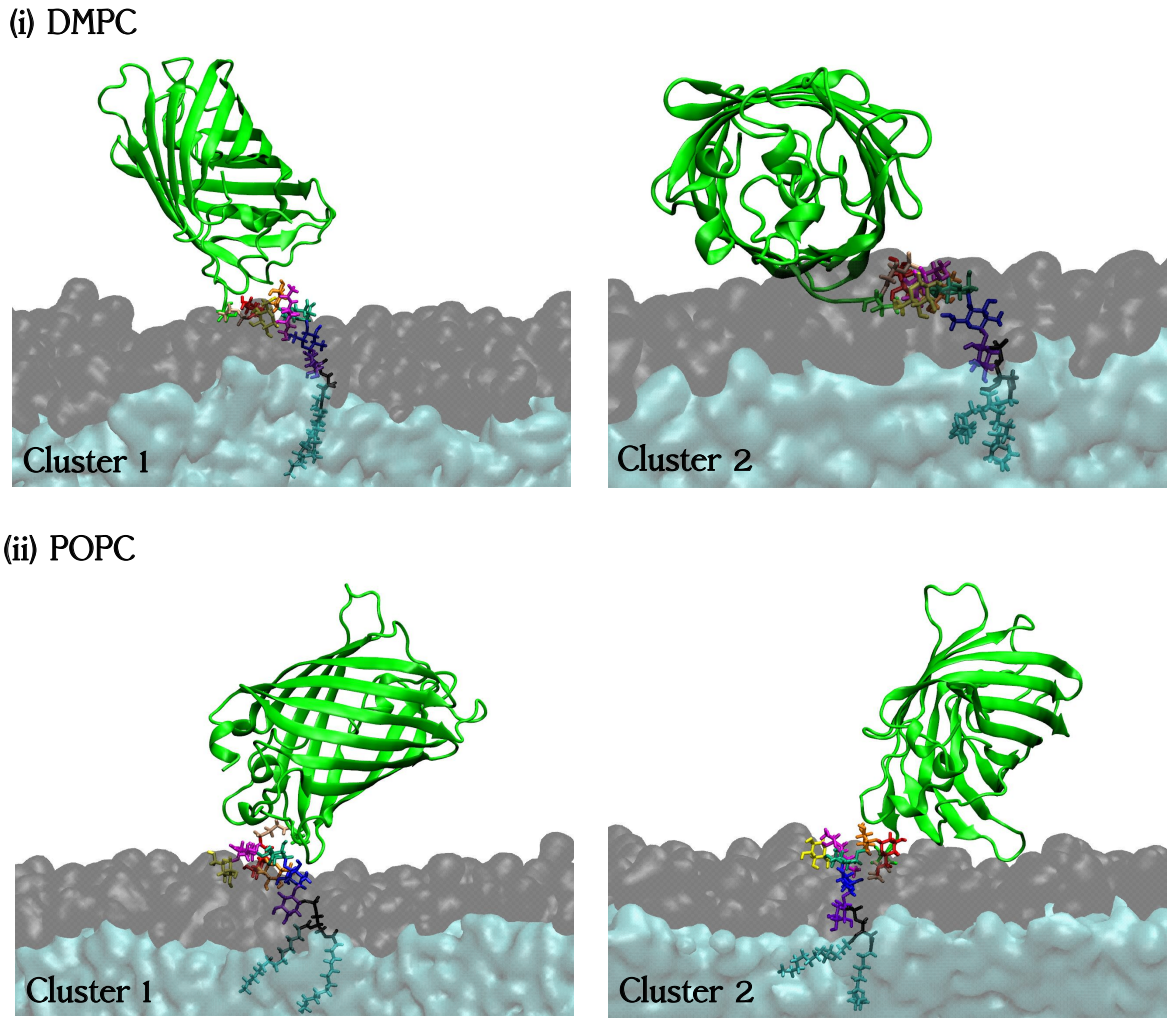
S3.2 Cluster analysis of GPI-anchored GFP



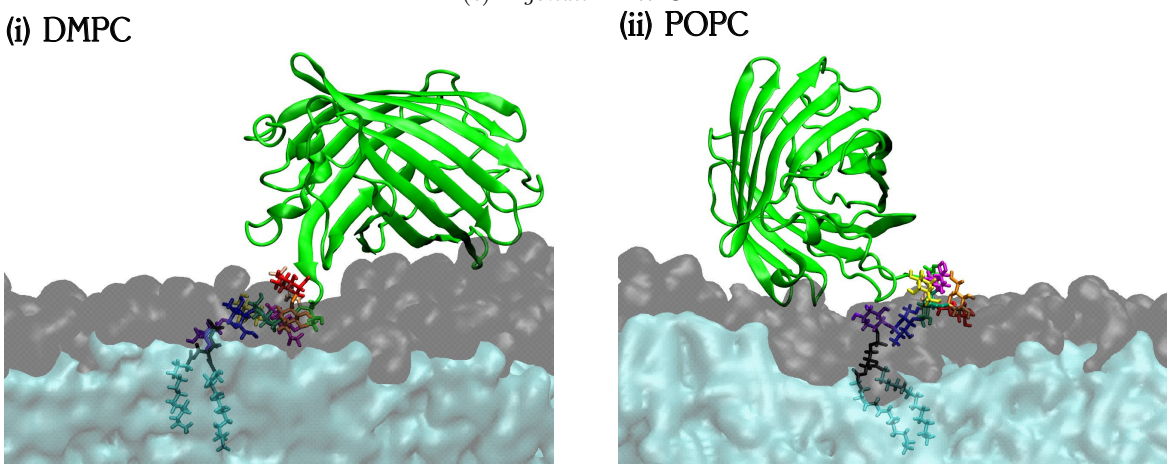
(a) GPI core



(b) *T.gondii* GPI



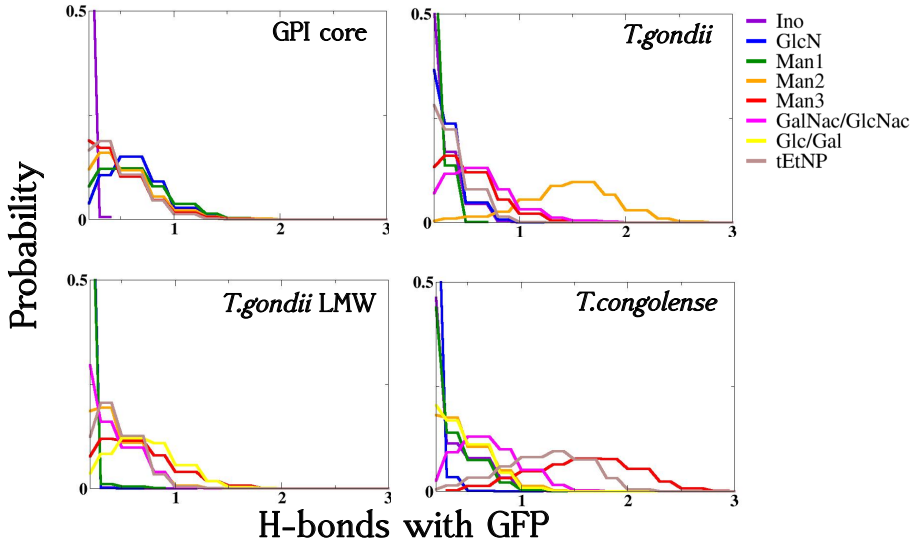
(c) *T.gondii* LMW GPI



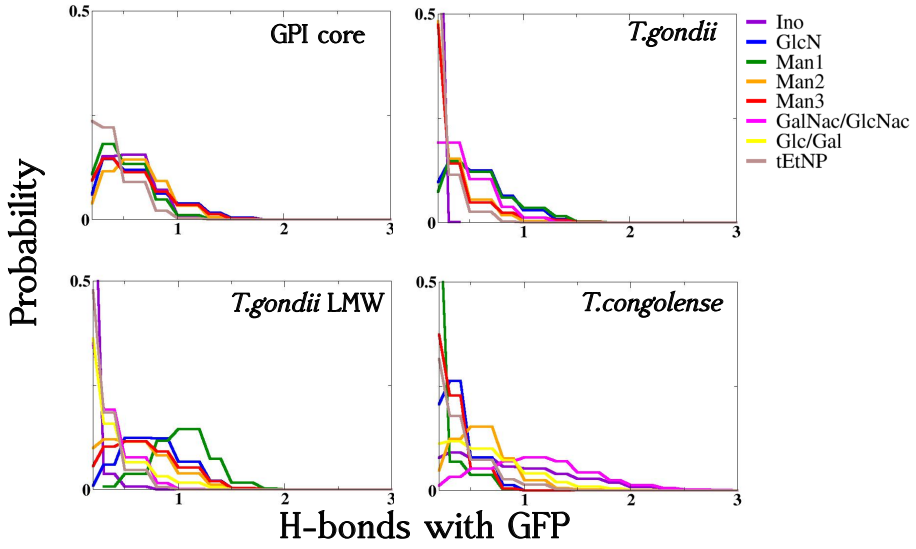
(d) *T.congolense* GPI

Figure S4: Conformations obtained from clustering of structures of GFP-attached GPIs – (a) GPI core, (b) *T.gondii*, (c) *T.gondii* LMW, and (d) *T.congolense* – in (i) DMPC and (ii) POPC bilayers. The most dominant cluster conformations (cluster 1) are shown for all GPI variants except (a)(i) GPI core in DMPC and (c) *T.gondii* LMW where the largest two cluster conformations(cluster 1 and cluster 2) are displayed considering their relatively broad cluster distributions. Color coding for residues follows the same pattern as in Figures 6, 7, 8. of the manuscript.

S3.3 Interaction of GFP with GPI

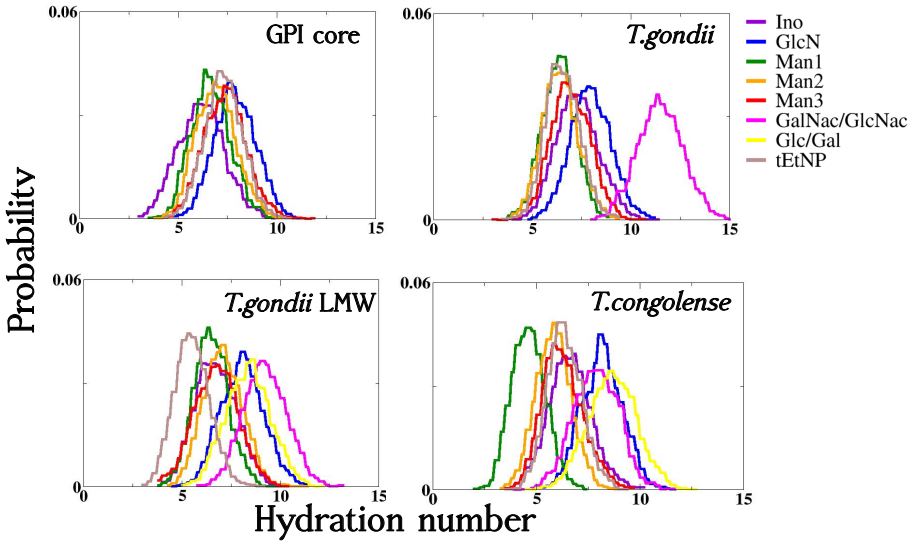


(a)

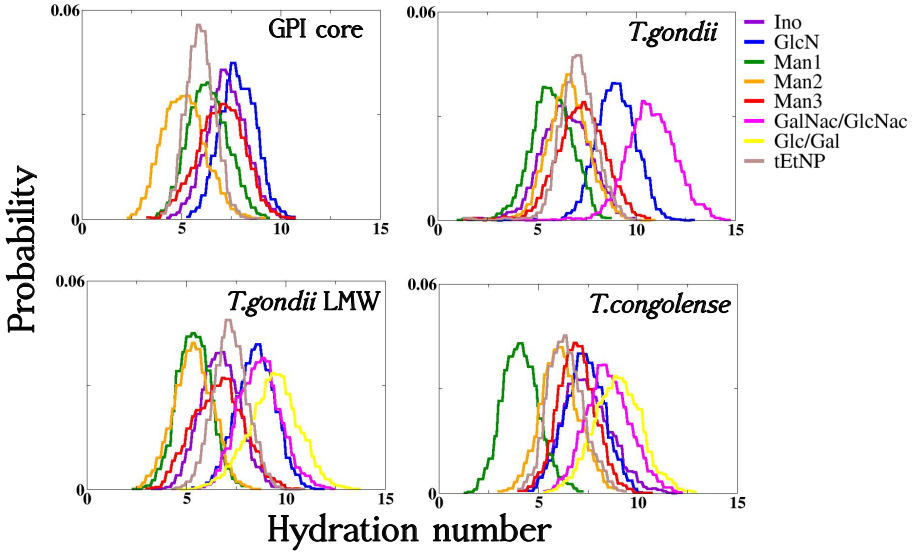


(b)

Figure S5: Probability distributions of the number of H-bonds formed between each residue of the GPI variants – GPI core, *T.gondii*, *T.gondii* LMW, and *T.congolense* – and attached GFP, inserted in (a) DMPC and (b) POPC bilayer



(a)



(b)

Figure S6: Probability distributions of hydration number of each residue of the GFP-attached-GPI variants – GPI core, *T.gondii*, *T.gondii* LMW, and *T.congolense* – inserted in (a) DMPC and (b) POPC bilayer

Chapter 4

Coarse-Grained Molecular Model for the Glycosylphosphatidylinositol Anchor with and without Protein

4.1 Overview

In this paper, we present a coarse-grained model of the GPI anchor core that has been constructed in an aqueous environment of polarizable water using a modified version of the MARTINI force-field. We build the model using a combined bottom-up and top-down approach, where each coarse-grained bead is mapped from four heavy atoms.

First, we model simple sugars – glucose, sucrose, and trehalose – in polarizable water. Bonded potentials (bonds, angles and dihedrals) for these sugars are obtained from atomistic simulations of 1 solvated molecule. Non-bonded parameters (or bead types) for each sugar molecule are derived using thermodynamic integration to calibrate their octanol-water partitioning free energies against experimental values. In order to have a realistic representation of sugar-sugar interactions, we calculate the second virial coefficient (B_{22}) of osmotic pressure for the sugar solutions. Thereby, we obtain a scaling factor $\gamma = 0.85$ to apply on the Lennard Jones parameter ϵ defining sugar-sugar interaction potentials. On applying this scaling factor, we reach experimental B_{22} values and also override sugar-sugar aggregation, a persisting pitfall of the original MARTINI force-field. Thus, we assign bead types to the GPI glycan core by building up from the simple sugar models and estimating the polarity of each bead according to Martini's interaction matrix.

Our results show excellent match of structure of the solvated GPI glycan with the atomistic model. We attach a dimyristoyl lipid tail to the GPI glycan and subsequently insert it to a pure DMPC bilayer.

The conformation and orientation of the GPI anchor with respect to the bilayer match well with that of the atomistic system. Moreover, we overcome the problem of aggregation among multiple GPIs.

We then extend the coarse-grained GPI to model a GPI-anchored protein by attaching to it the Green Fluorescent Protein (GFP). We model the GFP independently in solution using the ELNEDYN approach of MARTINI. The phosphoethanolamine linker is also modeled separately in water. Finally, we attach all three molecules to simulate the whole entity in DMPC bilayer. Because GPI and GFP displayed overestimated attraction when compared to the atomistic model, we apply the same scaling factor $\gamma = 0.85$ to the sugar-protein non-bonded potentials. We show that upon imposing the scale-down, GFP and GPI make similar mutual contacts as do their atomistic counterpart.

Our coarse-grained model of GPI and GPI-anchored protein allows a speed-up in the runtime by at least an order of magnitude compared to the atomistic model. Along with the GPI model, we present an improved version of the MARTINI force-field to simulate three types of biomolecules together – cabohydrates, protein and lipids – with polarizable water as the solvent medium. We finally discuss how our coarse-grained GPI model can be combined with Martini's lipidome to study dynamic cellular processes like raft-partitioning and protein-trafficking.

JCTC Journal of Chemical Theory and Computation

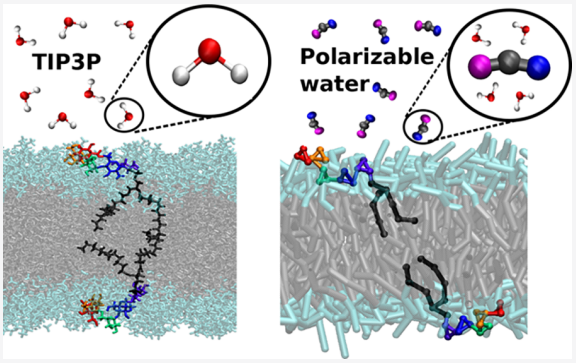
pubs.acs.org/JCTC Article

Coarse-Grained Molecular Model for the Glycosylphosphatidylinositol Anchor with and without Protein

Pallavi Banerjee, Reinhard Lipowsky, and Mark Santer*

Cite This: https://dx.doi.org/10.1021/acs.jctc.0c00056 Read Online

ABSTRACT: Glycosylphosphatidylinositol (GPI) anchors are a unique class of complex glycolipids that anchor a great variety of proteins to the extracellular leaflet of plasma membranes of eukaryotic cells. These anchors can exist either with or without an attached protein called GPI-anchored protein (GPI-AP) both *in vitro* and *in vivo*. Although GPIs are known to participate in a broad range of cellular functions, it is to a large extent unknown how these are related to GPI structure and composition. Their conformational flexibility and microheterogeneity make it difficult to study them experimentally. Simplified atomistic models are amenable to all-atom computer simulations in small lipid bilayer patches but not suitable for studying their partitioning and trafficking in complex and heterogeneous membranes. Here, we present a coarse-grained model of the GPI anchor constructed with a modified version of the MARTINI force field that is suited for modeling carbohydrates, proteins, and lipids in an aqueous environment using MARTINI's polarizable water. The nonbonded interactions for sugars were reparametrized by calculating their partitioning free energies between polar and apolar phases. In addition, sugar−sugar interactions were optimized by adjusting the second virial coefficients of osmotic pressures for solutions of glucose, sucrose, and trehalose to match with experimental data. With respect to the conformational dynamics of GPI-anchored green fluorescent protein, the accessible time scales are now at least an order of magnitude larger than for the all-atom system. This is particularly important for fine-tuning the mutual interactions of lipids, carbohydrates, and amino acids when comparing to experimental results. We discuss the prospective use of the coarse-grained GPI model for studying protein-sorting and trafficking in membrane models.



INTRODUCTION

The plasma membrane of eukaryotic cells contains a large variety of functionally active proteins, such as transmembrane proteins acting as ion channels or RAS proteins which have a simple fatty acid tail tethering them to the plasma membrane. The so-called glycosylphosphatidylinositols (GPIs) provide a particularly intriguing anchoring mechanism. They are covalently added to the C-terminus of proteins through post-translational modification in the endoplasmic reticulum. The structure of GPI consists of a highly conserved pseudopenta-saccharide glycan core **Man-α(1→2)-Man-α(1→6)-Man-α(1→4)-GlcN-α(1→6)-*myo*-inositol** that is further connected to a lipid tail which inserts into the plasma membrane. GPI-anchored proteins (GPI-APs) are involved in many cellular functions such as signal transduction,[1,2] adhesion,[3] and apical sorting.[4,5] GPIs are also found on the cell surfaces of protozoan parasites such as *Toxoplasma gondii*, *Trypanosoma brucei*, and *Plasmodium falciparum*,[6] either with or without an attached protein, as an end product of metabolic processes in the latter case. Figure 1 shows the chemical structure of a GPI with its attached protein. At the trailing mannose (Man3), a phosphoethanolamine bridge (EtNP) connects the protein to the GPI. In spite of the conserved core, GPIs are of heterogeneous structure through various types of sugar side branches, the composition of which can vary even with the very same protein (microheterogeneity).

Ever since the discovery of GPIs, the question of the relationship between their exceptional structure and functions has been a matter of debate until today. One of the many controversial subjects is the conformation of GPIs and the orientation and placement of GPI-APs relative to the membrane they are embedded in. Conclusions vary with the type of experiment conducted. One scenario is that GPI-APs lie in close proximity to the membrane, almost flopping down on it.[8,9] Lehto and Sharom conducted a FRET-based study on lipid bilayer vesicles to conclude that the fluorescent tag on a GPI-anchored placental alkaline phosphatase (PLAP) is at most 10−14 Å away from the lipid−water interfacial region,

Received: January 18, 2020
Published: May 11, 2020

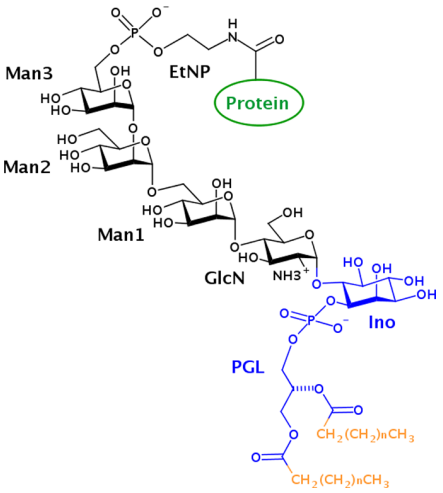


Figure 1. Chemical structure of a GPI anchor. The GPI core consists of Man3-Man2-Man1-GlcN-Ino. The core is connected to a phosphoglycerol (PGL) head which further connects to the lipid tail. A phosphoethanolamine linker (EtNP) attaches the protein to GPI. Ino+PGL are shown in blue to indicate the transition between the two force-field domains of GLYCAM06h (black) and Lipid14 (orange) that have been merged[7] to provide a molecular model of the full structure.

implying that the protein could be resting on the surface of the bilayer.[8] On the other hand, through diffusivity studies of synthetic GPI analogs in both supported bilayers and live cells, Paulick and co-workers concluded that the rapidly diffusing GPI analogs do not interact much with the membrane, thereby preventing interactions between the membrane and the attached protein.[10,11] On the other hand, a combined experimental and computational investigation of GPI-anchored Thy1 protein showed that GPI could substantially influence the conformation of the attached protein, suggesting considerable interactions between the two.[9] One may also wonder about how much this would impact another persisting controversy regarding the localization and partitioning of GPIs in functionally active, dense membrane microdomains, frequently referred to as lipid rafts. Some of the experiments conducted to address the partitioning behavior of GPIs in different lipid domains are quite contradictory.[12−14]

However, structural and dynamical details at the atomic level and high temporal resolution, for instance the femto-second time scale, are difficult to assess through experiments. Computer simulations of atomistic models provide powerful tools for filling these gaps, but only a few numerical studies have been conducted for GPIs so far, most of them advertising the idea that a GPI-AP may essentially be viewed as a rather rigid molecular arrangement rather than a vivid, dynamically changing object.[15−17] In our previous work, we devised an all-atom model of GPI using GLYCAM06h, Lipid14, and AMBER-ff14SB force fields to elucidate the conformational flexibility of the glycan core in solution[18] and to study a full GPI-AP embedded in lipid bilayer patches.[7] Through plain and biased MD simulations, the GPI core was revealed to behave effectively as a hinge, with two rather rigid disaccharide units connected via a flexible Man-$\alpha(1\rightarrow6)$-Man linkage. With a lipid tail attached and inserted into a bilayer, GPIs tend to assume a hooklike conformation with the glycan core partially immersed in the lipid headgroup region. In the simulations, all three species—lipids, proteins, and GPI (carbohydrates)—were seen to mutually interact. In general, one may envisage several avenues to further develop the hybrid model of GPI-AP via a reasonable refinement of force-field parameters. We want to recall, however, that the situation of three disparate, mutually interacting biomolecular species is not covered by the usual process of force-field development to begin with. The effect of a reparametrization will, however, experimentally only be visible in an extended context such as the dynamic behavior of GPI-APs in heterogeneous membrane patches, and the lack of sufficient statistical sampling will inevitably impose a strict limit on how an all-atom model can be tested. The mapping of our atomistic GPI model to a numerically efficient coarse-grained representation is thus highly desirable.

The MARTINI force field is a coarse-grained representation for biomolecular systems composed of lipids, proteins, glycolipids, and nucleotides, as well as, nanoparticles and a variety of polymers.[19] The MARTINI model is designed based on mapping 3 to 4 heavy atoms to one spherical superatom (bead). The interaction potentials between beads are inferred from the partitioning free energy of small coarse-grained molecules determined from their relative distributions in polar and apolar phases. MARTINI performs well in mimicking various types of lipids and replicating protein−lipid interactions as demonstrated for processes such as formation of pores and nanodisks, lipid-mediated protein clustering, and protein-mediated lipid flip-flop.[20]

In the present work, we devise a coarse-grained model of a full GPI and GPI-anchored green fluorescent protein (GFP) based on the MARTINI force field with polarizable water which has been proven to work consistently for modeling membranes in aqueous environment. After exhibiting our parametrization strategy and the definition of new parameters, we compare the behavior of the coarse-grained free-GPI and GPI-anchored-GFP with corresponding all-atom simulation results. We then discuss how the coarse-grained model may be used to study GPI-anchored proteins in membrane environments and how to deal with the situation that an optimally balanced parameter set for a GPI molecule is a priori unknown.

PARAMETRIZATION STRATEGY

Mapping Scheme, Bonded Interactions, and Bead Types. Due to microheterogeneity of naturally occurring GPIs and the inherent difficulties of synthesizing sufficient amounts of pure GPI species,[21] molecular-level studies of GPIs are difficult, and concise experimental data are lacking. To build a coarse-grained model of GPI, parametrization of simple sugars (mono- or disaccharides) was necessary. Parametrizing glucose (monosaccharide), sucrose, and trehalose (disaccharides) was sufficient to model the whole GPI glycan in a building-block manner as the mapping strategy was consistent across all these saccharides, entailing similar bead types. Moreover, properties such as partitioning free energies, which will be used in turn to derive nonbonded interactions, are well-known for these species. Note that at the coarse-grained level, there is no difference between the nonbonded parameters of different epimers of sugars such as glucose, mannose, and galactose. The differences are contained in the bonded parameters that are derived straight from the atomistic systems. The MARTINI coarse-grained force field is based on mapping 3 or 4 heavy atoms of the underlying atomistic system to one coarse-grained bead. To coarse-grain monosaccharides, we followed a similar mapping scheme as in the original work of the MARTINI team where the model for carbohydrates[22] was introduced. One

https://dx.doi.org/10.1021/acs.jctc.0c00056
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

saccharide unit is composed of three coarse-grained beads, connected together like a triangle. Unlike in ref 22 where polysaccharides were mapped linearly, we adopted a triangular mapping protocol (see Figure 2). The glycosidic linkages were

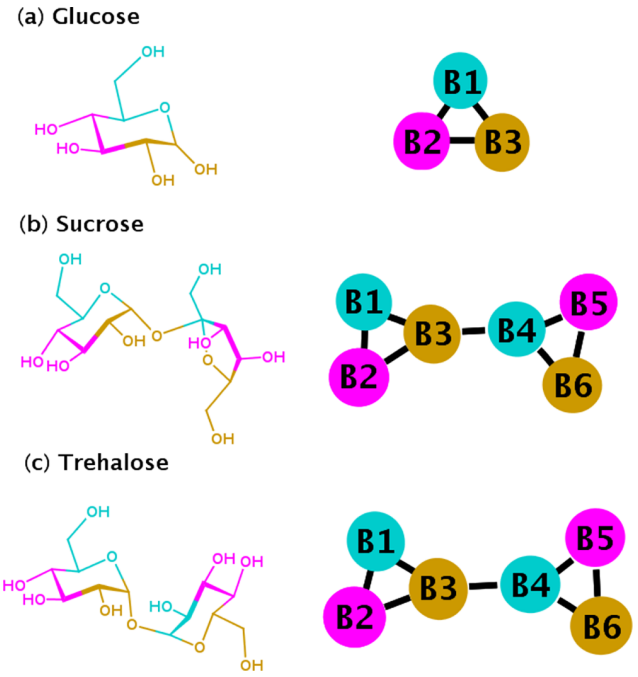


Figure 2. Mapping scheme for coarse-graining sugars: (a) glucose, (b) sucrose, and (c) trehalose. The colors of the coarse-grained beads encode the mapped groups of the atomistic molecules.

represented by just one bond in the coarse-grained landscape. There appears to be no general advantage of preferring the linear mapping over the triangular in the MARTINI scheme. The choice is usually made according to numerical stability of the simulation. In the present study, the triangular mapping scheme with a time step of 5 fs worked consistently for all the simulations. All systems in our study were parametrized against MARTINI's polarizable water as the aqueous medium. The polarizable water model implements the dielectric screening of bulk water through the orientational polarizability induced by its three-bead water model.[23] This water model is known to give more realistic and closer to atomistic results for processes involving membranes, such as pore formation,[24] phase transition,[25] and adsorption of charged peptides on membranes.[26]

Bonded Interactions. Bonded potentials for the simple sugars were obtained from 200 ns all-atom trajectories of one sugar molecule in water. GPIs were mapped from the atomistic structure using the same triangular mapping scheme as for the simple sugars (see Figure 3). Potentials for bonds, angles, and dihedrals were derived from a 1 μs long all-atom trajectory of one GPI glycan in water. In this way, bonded parameters as a function of just the intramolecular interactions and the effect of the solvent were captured. The potentials were obtained from the all-atom trajectories through simple Boltzmann inversion. Bonds between coarse-grained beads were imposed by harmonic potentials

$$V_b(r) = \frac{1}{2}K_b(r - r_0)^2 \tag{1}$$

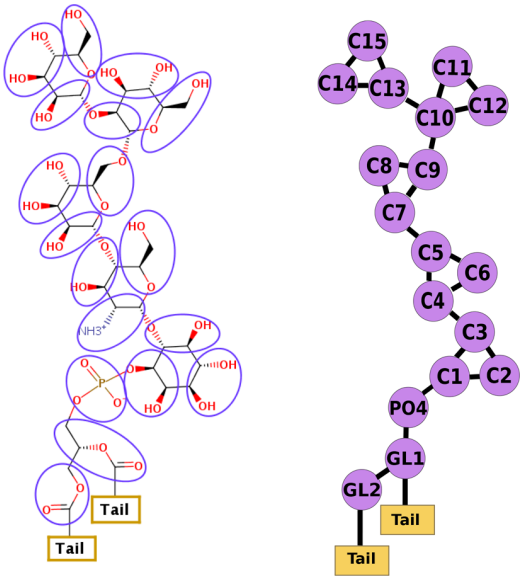


Figure 3. Mapping of GPI anchor from atomistic to coarse-grained representation.

where K_b is the spring force constant, and r_0 is the equilibrium bond length. Similarly, an angle connecting three consecutively placed beads is defined by a cosine-harmonic potential

$$V_a = \frac{1}{2}K_a[\cos(\theta) - \cos(\theta_0)]^2 \tag{2}$$

where K_a and θ_0 are the force constant and equilibrium angle, respectively. Lennard-Jones interactions between beads connected by bonds and angles were excluded from the nonbonded force calculation. This exclusion was necessary in order to incorporate all the crucial bonded potentials while avoiding numerical instabilities. The same strategy was employed by Gu et al.[27] to model the glycolipids GM1 and GM3. Torsions were incorporated through a proper dihedral potential with multiplicity(m) = 1, unless otherwise specified

$$V_d = K_d[1 + \cos(m\phi - \phi_0)] \tag{3}$$

where K_d is the force constant, and ϕ_0 is the equilibrium dihedral angle. Improper torsions were included wherever explicitly mentioned, the potential energy of which is described by a harmonic function, with K_i as the harmonic force constant and ξ_0 as the equilibrium dihedral angle

$$V_i = \frac{1}{2}K_i(\xi - \xi_0)^2 \tag{4}$$

Equilibrium values of the potentials for all bonds, angles and dihedrals were picked from target distributions at the atomistic level. The bonded parameters of the coarse-grained sugars and GPIs are listed in Table 2.

Partitioning Free Energy. Nonbonded or Lennard-Jones parameters of the coarse-grained molecules are contained in the assigned bead types. The bead types of the simple sugars—glucose, sucrose, trehalose—were assigned by considering the octanol—water partition coefficient (log P_{OW}) obtained from free energy calculations. Free energies of solvation of the sugars in (polarizable) water and water-saturated octanol were calculated separately to obtain P_{OW}. The amount of water in water-saturated octanol was 25 mol %. Only one sugar molecule was coupled/decoupled with the solvent. Solvation

https://dx.doi.org/10.1021/acs.jctc.0c00056
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

free energy (ΔG), i.e., the free energy difference (ΔF) of the solute in vacuum (F_X) and in the condensed phase (F_Y), was calculated using thermodynamic integration according to

$$\Delta G = \Delta F_{YX} = F_Y - F_X = \int_{\lambda_X}^{\lambda_Y} d\lambda \left\langle \frac{\delta U(\lambda)}{\delta \lambda} \right\rangle_\lambda \quad (5)$$

The coupling parameter λ defines the strength of the potential energy U between the solute and the solvent. λ lies in the range between 0 (no interaction) and 1 (full interaction between the two). A soft core approach was used to couple nonbonded interactions in order to remove singularities from the potential energy calculation.[28] Bonded interactions were linearly interpolated. $\delta U/\delta\lambda$ was calculated at 25 regularly spaced λ intervals between 0 and 1. The simulation time at each such window was 30 ns. The free energy curve was then integrated by the trapezoidal rule to obtain the final value of ΔG. Block averaging was done at every λ value to calculate the statistical error in free energy. Partition coefficients were obtained from the difference in the two solvation energies, given by

$$\Delta\Delta G_{OW} = \Delta G_O - \Delta G_W = -2.3RT \log P_{OW} \quad (6)$$

Here, the subscript O refers to water-saturated octanol, and the subscript W refers to water. The obtained free energy values are listed in Table 1, and the solvation free energy profiles from which these values were derived are shown in Figure 4. The calculated partition coefficients compare well with experiments.

Table 1. Octanol−Water Partitioning Coefficients (log P_{OW}) of Glucose, Sucrose, and Trehalose Compared to Experimental Values

	$\Delta\Delta G$ (K_BT)	log P_{OW} (calc)	log P_{OW} (exp)[29]
glucose	6.81	−2.95	−2.8
sucrose	7.28	−3.16	−3.3
trehalose	9.37	−4.06	−3.78

Bead Types. To arrive at the final bead types comprising the simple sugars, an iterative process of trial-and-error was carried out to arrive at their respective experimental octanol−water partitioning coefficients. The bead types examined here were taken from the database of MARTINI's polarizable force field and assigned through the parametrization procedure described in the section above. The distribution of bead types within the same sugar ring was determined based on the polarities of the beads relative to each other. For example, the two GP3 beads (B2, B3) in glucose have two free OH groups making them more polar than the GP2 bead (B1) that contains one free OH and one ether oxygen (see Figure 2 and Table 2). The bead types of GPI were assigned on the basis of the newly devised bead types of simple sugars, the chemical nature of the bead, and the interaction matrix of MARTINI. The glycan was constructed in a modular fashion from the models of mono- and disaccharides. Charged beads were used to represent the groups containing PO_4^- and NH_3^+. The bead types together with the bonded parameters making up the simple sugars and GPI are listed in Table 2. Alessandri et al. pointed out that short bonds in MARTINI could give rise to discrepancies in the hydrophilic/hydrophobic interactions of the molecule.[30] To take this possibility into account, we have used small (S) beads wherever short bonds (<0.3 nm) had to be included to facilitate finer mapping (see Table 2).

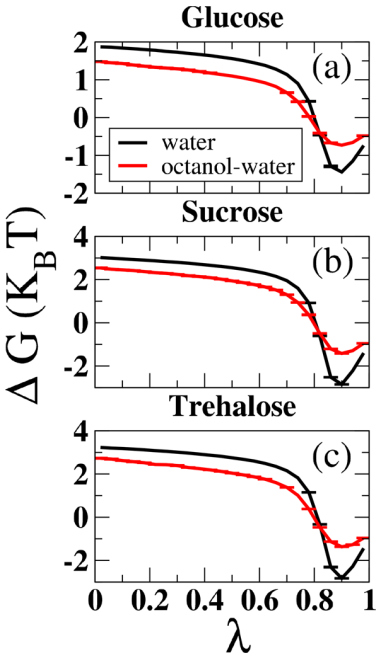


Figure 4. Free energy profiles ΔG as a function of the coupling parameter λ for (a) glucose, (b) sucrose, and (c) trehalose obtained from the thermodynamic integration of one sugar molecule in water (black) and in water-saturated octanol (red) separately.

Parametrizing EtNP Linker. To study the behavior of GPI-anchored GFP placed in lipid bilayers, a crucial step was to model the linker connecting protein and GPI. In all the GPI-APs discovered so far, this bridging linker is the same−phosphoethanolamine (EtNP). The EtNP linker was individually coarse-grained in an aqueous environment of polarizable water. Coarse-grained bonded parameters of the EtNP linker were derived from 200 ns all-atom simulations of the molecule shown in Figure 5a. The simulations were conducted in an aqueous medium of TIP3P water. For the GPI-anchored GFP, the EtNP linker is the bridge between the protein GFP and GPI. Therefore, in order to maintain the same connectivities, the linker was connected to amino acid residues: Threonine-Isoleucine-Glycine-Terminal Cap (THR-ILE-GLY-T), in the same order as in GFP, as shown in Figure 5. The terminal cap (T) is an acetyl group that was added to end the amino-acid chain. At the other end, the linker was connected to the last two mannose residues (Man3-Man2) of GPI. The EtNP linker was represented by two beads: a neutral L1 bead to substitute for ethanolamine and a negatively charged L2 bead to represent the phosphate group. The bead definitions and bonded parameters of the entire molecule in Figure 5b are listed in Table S1 of the SI. Coarse-grained simulations of the molecule in Figure 5b were also conducted for 200 ns to compare with the all-atom system. The derived bonded parameters involving beads L1 and L2 were plugged into the coarse-grained model of GPI-anchored GFP.

Coarse Graining GFP. GFP was modeled based on the ELNEDYN[31] framework of MARTINI. ELNEDYN, or the elastic network approach, is built on the philosophy of combining a structure-based coarse-grained model with a thermodynamics-based coarse-grained force field to model a protein. Secondary and tertiary structures of proteins are stabilized to a large extent by h-bonds, but this vital information is lost in the coarse representation. Therefore, to

https://dx.doi.org/10.1021/acs.jctc.0c00056
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 2. Bead Definitions and Bonded Parameters for the Carbohydrates Incorporated in Our Study: Glucose, Sucrose, Trehalose, and GPI[a]

molecule	bead name	bead type	bonds	r_0 (nm)	K_b (kJ/mol)	angle	θ_0 (deg)	K_a (kJ/mol)	dihedral	ϕ_0(deg)	K_d (kJ/mol)
glucose	B1	GP2	B1−B2	0.328	35000						
	B2	GP3	B1−B3	0.375	35000						
	B3	GP3	B2−B3	0.311	50000						
sucrose	B1	GP2	B1−B2	0.325	30000	B1−B3−B4	85	10	B1−B3−B4−B5	108	14
	B2	GP3	B2−B3	0.311	35000	B2−B3−B4	143	160	B1−B3−B4−B6	166	15
	B3	GP2	B1−B3	0.379	35000	B3−B4−B5	93	165	B2−B3−B4−B5	143	8
	B4	GSN0	B3−B4	0.335	5000	B3−B4−B6	80	280			
	B5	GP3	B4−B5	0.327	10000						
	B6	GP2	B5−B6	0.302	10000						
			B4−B6	0.406	10000						
trehalose	B1	GP2	B1−B2	0.329	20000	B1−B3−B4	77	150	B1−B3−B4−B5	2.9	50
	B2	GP3	B2−B3	0.311	35000	B2−B3−B4	107	300	B1−B3−B4−B6	−54	28
	B3	GP2	B1−B3	0.379	35000	B3−B4−B5	96	300	B2−B3−B4−B5	44	50
	B4	GSP1	B3−B4	0.376	30000	B3−B4−B6	69	250			
	B5	GP3	B4−B5	0.299	50000						
	B6	GP2	B5−B6	0.329	25000						
			B4−B6	0.399	30000						
GPI	C1	GP2	C1−C2	0.325	40000	C1−C2−C3	55	600	C1−PO4−GL1−GL2	39.3	2.5
	C2	GP3	C1−C3	0.307	35000	C1−C3−C2	60.5	600	C2−C1−PO4−GL1 (m = 2)	23	5
	C3	GP2	C2−C3	0.34	40000	C1−C3−C4	88	200	C3−C1−PO4−GL1	15.4	3
	PO4	GQa	C3−C4	0.37	20000	C3−C1−PO4	112	70	C3−C4−C5−C7	− 32.3	6,2
	C4	GSQd	C4−C5	0.30	40000	C2−C1−PO4	144	450	C1−C3−C4−C5	−5.7	20
	C5	GP2	C4−C6	0.40	35000	C2−C3−C4	142	400	C1−C3−C4−C6	−54.4	25
	C6	GP2	C5−C6	0.32	20000	C3−C4−C5	90	500	PO4−C1−C3−C4	−163.3	80
	C7	GSP1	C5−C7	0.35	20000	C3−C4−C6	63	550	C4−C5−C7−C8	12.6	10
	C8	GP3	C7−C8	0.28	35000	C4−C5−C6	80	400	C4−C5−C7−C9	−44	7.8
	C9	GNa	C7−C9	0.34	20000	C4−C6−C5	48	500	C5−C7−C9−C10	46	25
	C10	GSN0	C8−C9	0.32	20000	C4−C5−C7	172	500	C7−C9−C10−C11	114	4.7
	C11	GP3	C9−C10	0.40	15000	C5−C7−C8	114	350	C7−C9−C10−C12	57.7	6
	C12	GP2	C10−C11	0.28	40000	C5−C7−C9	90	250	C7−C9−C10−C13	−91.14	4
	C13	GSP1	C10−C12	0.35	30000	C6−C5−C7	109	300	C9−C10−C13−C14	−80	14
	C14	GP3	C11−C12	0.33	30000	C7−C8−C9	69	300	C9−C10−C13−C15	−132	15
	C15	GP2	C10−C13	0.36	20000	C7−C9−C8	50	400			
	L1	GNa	C13−C14	0.28	40000	C7−C9−C10	126	50			
	L2	GNa	C13−C15	0.35	30000	C8−C9−C10	118	80			
			C14−C15	0.33	30000	C9−C10−C11	100	120			
			C1−PO4	0.30	3000	C9−C10−C12	82	90			
			PO4−GL1	0.40	5000	C9−C10−C13	140	30			
			GL1−GL2	0.34	3000	C10−C11−C12	69	400			
						C10−C12−C11	49	500			
						C10−C13−C14	94	300			
						C10−C13−C14	67	300			
						C11−C10−C13	120	100			
						C12−C10−C13	128	120			
						C13−C14−C15	69	200			
						C13−C15−C14	48	150			
						C1−PO4−GL1	112	20			
						PO4−GL1−GL2	96	50			

[a]MARTINI bead types are prefixed with 'G' to indicate the redefined nonbonded parameters.

replicate the secondary, tertiary, and quaternary structures more realistically, an elastic network was imposed on the protein through the ELNEDYN approach. Mapping of amino acids and assignment of bead types is done according to the same protocol as in ref 31, where the center of the backbone bead is located on the C_α atom of the respective all-atom amino acid. When the distance between the nearest-neighbor beads was less than the imposed cutoff R_C, a harmonic spring potential of force constant K_S was turned on between the two. The equilibrium lengths of these artificial bonds were set to the distances obtained from an equilibrated structure of atomistic GFP in water, and the values of R_C and K_S were kept uniform across all such pairs of beads. Nonbonded potentials among the backbone beads connected through a spring force are excluded from the calculation of the system potential. Bonded parameters (bonds, angles) were derived straight from the

https://dx.doi.org/10.1021/acs.jctc.0c00056
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

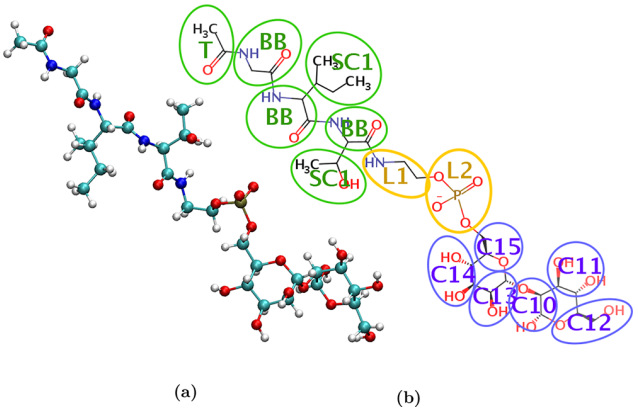


Figure 5. (a) All-atom representation of EtNP linker. (b) Mapping of the all-atom model in (a) to a coarse-grained parametrization consisting of beads, with the green beads representing the amino-acid residues in the following order: THR-ILE-GLY-T, starting from the linkage at L1. BB beads are backbone beads, and SC are side chain beads. The yellow beads make up the EtNP linker, and the blue beads represent GPI's last two mannose residues. Beads are shown with their bead names.

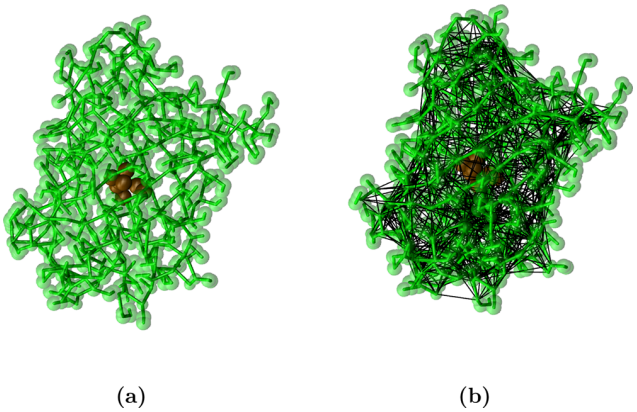


Figure 6. Coarse-grained representation of GFP (a) without and (b) with elastic bonds. The black mesh in (b) depicts the elastic network imposed on the backbone beads of GFP. The chromophore is shown as brown beads in the center of the barrel.

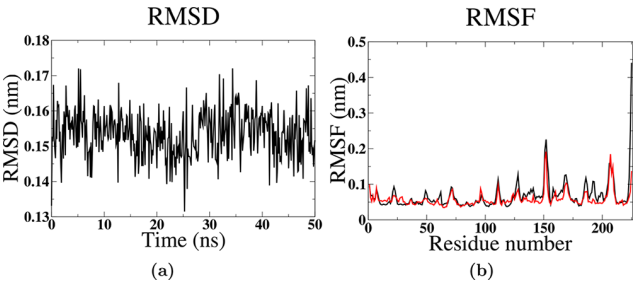


Figure 7. (a) Root-mean-square deviation (RMSD) of coarse-grained GFP compared to the crystal structure. (b) Comparison of root-mean-square fluctuation (RMSF) of the all-atom (black) and coarse-grained (red) GFPs in water.

corresponding atomistic simulations of GFP in water. Along with the protein, a coarse-grained representation of the chromophore situated inside the barrel of GFP was also modeled from the all-atom system. We observed that the presence of the chromophore affected the size of GFP and hence was important to model the protein more realistically. Details of the chromophore model are provided in the SI with the mapping scheme illustrated in Figure S3, and the corresponding bonded parameters are listed in Table S2.

As per the work of Periole and co-workers,[31] the optimal values of the elastic scaffold parameters could range from 0.8 to 1.0 nm for R_C and from 500 to 1000 kJ/mol for K_S. We observed that for our system of GFP in polarizable water, the combination of R_C = 1.0 nm and K_S = 500 kJ/mol replicates the atomistic system sufficiently well. Mapping of atoms to coarse beads was conducted on an equilibrated structure of GFP from the atomistic simulations. Note that the crystal structure of protein should not be directly mapped to coarse-grained representation, as the protein changes in size upon solvation and equilibration. As the elastic network ensures that the structure and size of the protein are maintained throughout the simulation, the atomistic system to be mapped should be chosen carefully. Figure 6 shows the coarse-grained representation of the protein with and without the elastic network.

To compare with the crystal structure, we calculate the root-mean-square deviation (RMSD) of GFP. RMSD is a metric used to quantify the degree of similarity between two corresponding, superimposed structures. It is calculated by the following relation

$$\mathrm{RMSD}(t) = \sqrt{\frac{1}{M}\sum_{i=1}^{N} m_i |r_i(t) - r_i^{ref}|^2} \tag{7}$$

where $M = \sum_i m_i$, the sum of masses of all atoms, $r_i(t)$ is the position of atom i at time t of the simulation trajectory, and r_i^{ref} is the position of atom i in the reference structure. For the calculation of RMSD, only the backbone beads are taken into account. As shown in Figure 7a, RMSD stays well within the resolution of determination of crystal structure, i.e., 0.19 nm,[32] throughout the trajectory, suggesting that the protein is structurally stable. The flexibility of each residue of a protein can be measured by root-mean-square fluctuation (RMSF). RMSF is useful for characterizing local changes along the protein chain. It is calculated for the C_α atoms in the all-atom case and backbone beads in the coarse-grained case. The RMSF for residue i is

$$\mathrm{RMSF}_i = \sqrt{\langle (r_i - \langle r_i \rangle)^2 \rangle} \tag{8}$$

where r_i is the position of atom i in the residue after superposition with the reference structure, and $\langle r_i \rangle$ is the average position of atom i. Figure 7b shows the comparison of root-mean-square fluctuation of each residue of the protein between the all-atom and coarse-grained systems. The local fluctuations/dynamics of the all-atom and coarse-grained GFPs turn out to be quite similar. We also compare the global structure of the protein in the two resolutions by calculating the radius of gyration of the backbone beads in Figure 8 and Table 3. Both RMSF and R_g plots show good overlap between the two resolutions, further validating the coarse-grained force field.

Solute−Solute Adapted Nonbonded Interactions. Nonbonded interactions between neutral beads in MARTINI are described by a Lennard-Jones 12-6 potential energy function

$$V_{Lennard\text{-}Jones}(r) = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r}\right)^{12} - \left(\frac{\sigma_{ij}}{r}\right)^{6}\right] \tag{9}$$

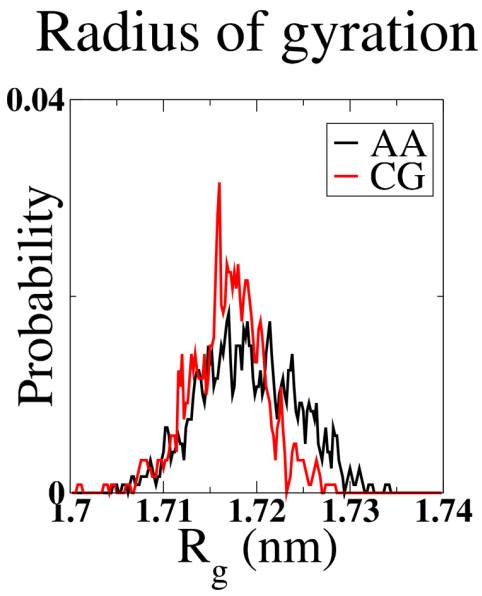


Figure 8. Radius of gyration R_g for GFP as obtained with the atomistic (AA) (black) and coarse-grained (CG) (red) models.

Table 3. Average Values of Radius of Gyration R_g for Atomistic and Coarse-Grained GFP

	atomistic	coarse-grained
R_g (nm)	1.725 ± 0.005	1.717 ± 0.004

where r is the distance between two particles i and j, σ_{ij} is the distance between them at which potential energy is zero, and ϵ_{ij} is the well depth which is a measure of the strength of their interaction. Interaction between charged beads is represented both by aforementioned Lennard-Jones potential and a Coulombic potential energy function to describe the electrostatics

$$V_{Coulomb}(r) = \frac{q_i q_j}{4\pi\epsilon_0\epsilon_{rel} r} \tag{10}$$

where q is the charge on the particle, ϵ_0 is the dielectric permittivity of vacuum, and ϵ_{rel} is the relative dielectric permittivity of the medium. Charged nonbonded interactions are determined by the charge on the beads, and uncharged nonbonded Lennard-Jones interactions are dictated by the bead types, the parameters of which have been fit to reproduce partition coefficients of small organic molecules in polar−apolar solvent phases.[33] In accordance with the MARTINI parametrization, we did not alter the sugar−lipid interaction parameters because these interactions are taken care of through the octanol/water partitioning coefficients. A couple of studies have reported that MARTINI sugar−lipid parameters obtained through this parametrization scheme are well-characterized. Lopez et al. demonstrated the cryo- and anhydro-protective effect of MARTINI sugars on lipid bilayers.[22] In another study, MARTINI nonreducing disaccharides were shown to disrupt phase segregation in mixed membranes, whereas monosaccharides and reducing disaccharides had no such effect, as was also observed in experiments.[34]

The strategy of using octanol−water partitioning free energies to define nonbonded interactions naturally addresses carbohydrate-lipid or amino acid-lipid interactions, but it is quite plausible that it cannot cover all conceivable situations met in biochemical modeling. Solute−solute interactions with sugars[35] and proteins[36,37] have previously been reported to turn out overestimated, leading to unnatural aggregation. The degree of aggregation, or stickiness, increases with the increase in length/size of the solute, as observed by Schmalhorst and co-workers.[35]

The MARTINI force field has already been extended to carbohydrates including simple sugars[22] and glycolipids;[38] however, their self-interactions are overestimated leading to unnatural aggregation both in solution and in membranes. Gu et al. proposed to use the small (S) beads of MARTINI which reduced the clustering propensity of glycolipids GM1 and GM3 when placed in membranes to better reproduce the clustering observed in the atomistic system.[27] Here, note that badly parametrized intermolecular vdW interactions are a general problem in force-field development, whether coarse-grained or atomistic.[39,40] Therefore, a coarse-grained model parametrized on the basis of atomistic cluster sizes cannot be trusted.

To fix this imbalance in interactions, a few strategies have been proposed based on the incorporation of solution observables in the parametrization process such as Kirkwood/Buff integrals,[41−43] osmotic pressure,[44,45] and osmotic coefficient.[46] Yet another way of optimizing potentials in MD simulations is by calculating the second virial coefficient of osmotic pressure B_{22}, a quantity that describes the deviation of a solution from ideality. It is related to the osmotic pressure π in the following way

$$\pi(T, c) = RT(c + B_{22}c^2 + B_{23}c^3 + ...) \tag{11}$$

where c is the solution concentration, T is the temperature, R is the gas constant, and B_{ij} are coefficients of the virial expansion of osmotic pressure. The nonbonded forces between aggregating solutes can be scaled down by scaling down the pairwise amplitudes ϵ_{ij}s of the Lennard-Jones potentials (eq 9) to match the experimental B_{22} values. This method has been applied on MARTINI for proteins by Elcock et al.[36] and for polysaccharides by Schmalhorst et al.,[35] in the environment of antifreeze water of MARTINI. We followed the same protocol to optimize the nonbonded interactions of simple sugars and GPIs in polarizable water as polarizability of the aqueous medium is essential to our study.

Based on the assumption that the total solute potential energy can be approximated as the sum of pairwise solute−solute interactions, McMillan and Mayer[46] derived a relation for B_{22} from the potential of mean force ($w(r)$) between two particles separated by distance r

$$B_{22} = -2\pi N_A \int_0^\infty \left[\exp\left(-\frac{w(r)}{RT}\right) - 1\right] r^2 \, dr \tag{12}$$

with N_A being Avogadro's constant. At thermodynamic equilibrium, $w(r)$ can be approximately related to the radial distribution function (RDF) $g(r)$ in the following way

$$g(r) = \exp\left(-\frac{w(r)}{RT}\right) \tag{13}$$

In order to calculate B_{22} from simulations, the integral in eq 12 needs to be finite

$$B_{22}(r') = -2\pi N_A \int_0^{r'} [g(r) - 1] r^2 \, dr \tag{14}$$

https://dx.doi.org/10.1021/acs.jctc.0c00056
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

The value of r' should be high enough where the solute−solute interactions vanish and $B_{22}(r') \rightarrow B_{22}(\infty)$. In our systems, we found that a value of $r' = 5$ nm worked consistently for all three sugar systems. For a two-component system, subscript 1 in B_{ij} stands for solvent, subscript 2 stands for solute. Thereby, B_{22} denotes solute−solute interactions. Positive values of B_{22} indicate net repulsion, and negative values indicate attraction between solute molecules. Its magnitude denotes the extent of aggregation. Experimentally, B_{22} can be obtained from static light scattering, and in an MD simulation it is derived from cumulative solute−solute RDF. Aqueous solutions of 100 mM sugar solutions were prepared and simulated for 1 μs for monosaccharide (glucose) and 2 μs for disaccharides (sucrose and trehalose). Cumulative RDFs were calculated for every 200 ns segment of the trajectories. Using eq 14, B_{22} was obtained by an integration over the solute−solute RDFs. Solute−solute interactions were varied by scaling down the ϵ_{ij} of all the sugar−sugar pairwise nonbonded potentials of MARTINI, using a simple relation

$$\epsilon_{ij,new} = 2 + \gamma(\epsilon_{ij,old} - 2) \tag{15}$$

with γ as the scaling factor. This ansatz was also used by Schmalhorst et al. The constant, 2 kJ/mol, is the lowest value of ϵ_{ij} in the MARTINI database. After systematically testing different scaling factors, we arrived at $\gamma = 0.85$ that worked consistently for all the sugars in achieving more realistic osmotic pressure coefficients and eliminating aggregation in sugars. As can be seen in Figure 9, unscaled/original MARTINI resulted in B_{22} values in the attractive regime, whereas the experimentally obtained values suggest somewhat repulsive interactions. The B_{22} profiles obtained after the scale-down resulted in positive values with the averages close to those from experiments (see Table 4).

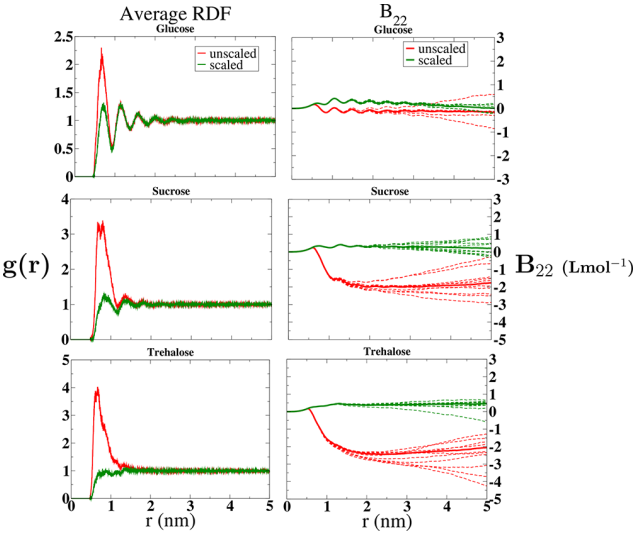


Figure 9. Sugar−sugar radial distribution functions (RDFs) $g(r)$ as a function of distance r averaged over all 200 ns segments and corresponding B_{22} vs r profiles of all 200 ns segments put together for solutions of glucose, sucrose, and trehalose. In the B_{22} plots, the dotted lines come from the 200 ns intervals, and the solid line is the averaged profile over all the intervals. Profiles from unscaled $\gamma = 1$ are shown in red, and profiles from scaled $\gamma = 0.85$ are shown in green. The averaged constant value at the far end (at 5 nm) is the reported B_{22} value.

Table 4. B_{22} Values Collected at the Tail End of B_{22} vs r Profiles Calculated from Averaged RDFs

	B_{22}(L mol^{-1})		
	$\gamma = 1.0$	$\gamma = 0.85$	exp
glucose	−0.171	0.012	0.117[47]
sucrose	−1.765	0.206	0.305[47]
trehalose	−2.059	0.451	0.51[48]

Simulation Details. All the Molecular Dynamics (MD) simulations in this work were performed with the simulation engine: GROMACS-2018.3.[49]

All-Atom. The all-atom models of the simple sugars considered in this study, glucose, sucrose, and trehalose, were built with the GLYCAM06h force field[50] with TIP3P water[51] in the background. Only one sugar solvated in water in cubic boxes was simulated for 200 ns each, so as to extract bonded information (bonds, angles, dihedrals) to build their coarse-grained representations.

The all-atom model of the GPI anchor was constructed by merging two force-fields: GLYCAM06h to represent the glycan head and Lipid14[52] for the lipid tail. Figure 1 shows the transition between the two force-field domains. The inositol-together-with-phosphoglycerol (Ino+PGL) part of the molecule, shown in blue, is the hybrid, bridging moiety connecting the glycan head and the lipid tail. The atom types for this bridging residue were chosen through a careful mixing of the atoms from GLYCAM06h and Lipid14. Partial charges, angles, and torsions of this bridge were derived using quantum mechanical calculations, as described in our previous work.[7] We consider only pure DMPC lipid bilayer in this study, which was modeled with Lipid14. The lipid tail of the GPI is also a dimyristoyl. GFP was parametrized using AMBER's protein force field: ff14SB.[53] The aqueous phase was represented by TIP3P waters. The construction of the systems was achieved using the LEaP facility of AMBER. AMBER and GLYCAM topologies were converted to GROMACS format using a script that was originally written by Sorin and Pande[54] and was further modified by us to accommodate the specifics of GLYCAM06h.[18] One μs long simulations were conducted for free GPIs in water and in 8*8 DMPC bilayers each, and 4 sets of 1 μs long simulations amounting to a total of 4 μs of simulation time were performed for GFP-GPIs embedded in larger 16*16 DMPC bilayers. The detailed methodology of the all-atom model development has been described in our previous paper.[7]

Coarse-Grained. The coarse-grained GPI glycan was attached to a dimyristoyl lipid tail, the parameters of which were directly taken from the MARTINI lipid parameter set.[55] Bonded parameters to define the link between the phosphoinositol of GPI and the lipid tail were also taken from MARTINI's database. One GPI was inserted into each leaflet of an 8*8 bilayer of pure, hydrated DMPC and simulated for 1 μs. The system was assembled using the *insane* script of the Wassenaar group.[56] A single GFP-GPI was inserted into a 16*16 pure, hydrated bilayer of DMPC to study its conformational behavior w.r.t. lipid bilayers. All the aforementioned coarse-grained systems were solvated in MARTINI's polarizable water. Nonbilayer systems were set up in cubic boxes with a minimum distance of 1.2 nm between the edges of the solute and the box. Bilayer systems were constructed in orthorhombic boxes. Counterions, represented as hydrated Na+ beads, were added to the GFP-GPI-bilayer

Table 5. Technical Details of Simulation Settings for All the All-Atom and Coarse-Grained Systems Included in This Study

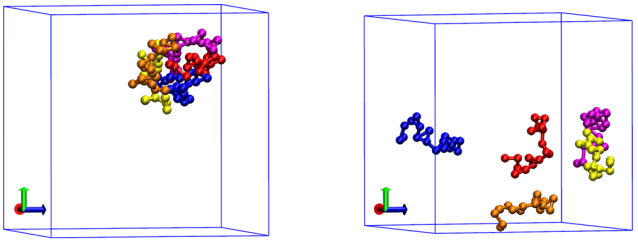
system	species	number	box size (nm)	time (ns)
all-atom (AA) mapping (aqueous systems)				
glucose	glucose	1	4 × 4 × 4	200
	water	876		
sucrose	sucrose	1	4 × 4 × 4	200
	water	2178		
trehalose	trehalose	1	4 × 4 × 4	200
	water	2170		
GPI	GPI glycan	1	5.3 × 5.3 × 5.3	1000
	water	4753		
GFP	GFP	1	8.2 × 8.2 × 8.2	200
	water	16608		
	Na+	7		
EtNP	EtNP molecule	1	5.2 × 5.2 × 5.2	200
	water	4592		
	Na+	1		
membrane systems				
GPI in DMPC	GPI	2	8.4 × 8.4 × 15.4	1000
	DMPC	126		
	water	17095		
GFP-GPI in DMPC	GFP-GPI	1	15 × 15 × 19	4 × 1000
	DMPC	511		
	water	81846		
	Na+	7		
coarse-grained (CG) mapping (aqueous systems)				
glucose	glucose	1	5 × 5 × 5	200
	water	338		
sucrose	sucrose	1	5 × 5 × 5	200
	water	545		
trehalose	trehalose	1	5 × 5 × 5	200
	water	543		
GPI	GPI	1	6 × 6 × 6	1000
	water	1188		
GFP	GFP	1	10 × 10 × 10	50
	water	4626		
	Na+	7		
EtNP	EtNP molecule	1	5 × 5 × 5	200
	water	617		
	Na+	1		
calculation of B_{22}				
glucose	glucose	420	28 × 28 × 28	1200
	water	59289		
	Na^+Cl^-	420		
	$Ca^{2+}Cl_2^-$	42		
sucrose	sucrose	420	28 × 28 × 28	2200
	water	58738		
	Na^+Cl^-	420		
	$Ca^{2+}Cl_2^-$	42		
trehalose	trehalose	420	28 × 28 × 28	2200
	water	58699		
	Na^+Cl^-	420		
	$Ca^{2+}Cl_2^-$	42		
membrane systems				
GPI in DMPC	GPI	2	6.5 × 6.5 × 16	1000
	DMPC	126		
	water	4657		
GFP-GPI in DMPC	GFP-GPI	1	12.5 × 12.5 × 19.5	4000
	DMPC	511		
	water	20581		
	Na+	7		

system to neutralize the net charge of −7 on the protein. Energy minimization was performed for 10000 steps using the steepest descent algorithm, followed by an NPT equilibration for 1 ns. Postequilibration, the production run was carried out in an NPT ensemble. Protein-free GPIs in bilayers were simulated for 1 μs. GFP-GPI-bilayer systems were simulated for 4 μs. The first 10 ns of the production run of each system were excluded from analysis. The time step used for GPI simulations was 5 fs, which is relatively small compared to the typical range of time steps (10−40 fs) used in MARTINI models. Since GPI is structurally quite flexible, we avoided imposing constraints on the molecular conformation. The inclusion of rather tight bonds, some of them with force constants around 40000 kJ/mol, and the crucial glycosidic dihedrals made it necessary to limit the time step to 5 fs so as to avoid numerical instabilities. Besides, the choice of time step is in agreement with the study of MARTINI glycolipids where the small time step was required to avoid numerical instabilities arising from the tight force constants and a large number of angle and dihedral potentials used to maintain the complicated conformation of the atomistic glycolipids.[38] The cutoff (both vdW and Coulomb) for all the systems was 1.1 nm, imposed by the Verlet scheme.[57] The PME method[58] was employed for electrostatics, and the plain cutoff method was employed for vdW interactions. The vdW potential was shifted in energy to smoothly reduce it to zero at the cutoff. The relative dielectric constant was fixed at 2.5, the default value for polarizable water in MARTINI. The leapfrog stochastic dynamics (sd) integrator[59] was used to integrate Newton's equations of motion. Temperature was controlled by the sd integrator with a time constant of 1 ps. For equilibration, the Berendsen barostat[60] was used to maintain the pressure at 1 bar, whereas for the production run the Parrinello−Rahman barostat[61] was employed. A time constant of 5 ps was used for the former, and a time constant of 12 ps was used for the latter. For all the cubic boxes, isotropic pressure coupling was applied, but for the bilayer systems semi-isotropic coupling was used, that is, isotropically only in x and y directions. Detailed information on the simulation settings can be found in Table 5.

RESULTS AND DISCUSSION

Scaled Solute−Solute Interactions: GPI and GFP-GPI.

The scaling factor, $\gamma = 0.85$, that was derived from simulations of sugar solutions was applied to nonbonded interactions between GPIs. To observe the aggregating tendencies of GPIs before and after scaling, 5 GPIs (GPI core + PGL) were solvated in water and simulated for 1 μs. Figure 10 shows the snapshots taken at the end of the simulations with (a) unscaled and (b) scaled MARTINI parameters. With the original MARTINI parameters, all the GPIs ball up to form a globule which remains stable throughout the simulation. Upon scaling down the sugar−sugar interactions using the same scaling law (eq 15), we observed that GPIs freely float in water and



(a) $\gamma = 1$ (b) $\gamma = 0.85$

Figure 10. Snapshots taken at the end of 1 μs long simulations of five GPI glycans in water modeled with (a) unscaled MARTINI at γ = 1 and (b) scaled MARTINI at γ = 0.85. Each GPI molecule has a different color.

intermittently associate with each other. At no point do they aggregate into a solid, compact globule.

The combined model of GPI+EtNP+GFP was inserted into a pure 16*16 lipid bilayer of DMPC to study the conformational behavior of GFP w.r.t. the bilayer. From a 4 μs long simulation, it became apparent that the interactions between GFP and GPI were significantly stronger compared to the atomistic system. This is not surprising since the parametrization of nonbonded interactions in atomistic and coarse-grained systems follows different routes. We recall that the issue of overestimation of solute−solute interactions has been reported for both all-atom and coarse-grained systems (MARTINI in particular). In order to be consistent, we must, of course, make the coarse-grained model reflect the one at the atomistic level of a single molecular species (GFP-GPI) and weaken the sugar−protein interactions. Due to the lack of explicit experimental data on mixtures of sugars and amino acids, we tentatively use the scaling factor obtained for sugar−sugar interactions. Since the issue of aggregation has been reported both in proteins and sugars, it is not surprising that the interactions between proteins (GFP) and sugars (GPI) would also be similarly affected. To beat the excessive attractive force down, we applied the same scaling factor, γ = 0.85, to the Lennard-Jones potential between GFP and GPI beads. The scale-down presented results comparable with the all-atom system. The extent of interaction between molecules in close proximity can be quantified by the number of contacts formed between the two. We counted the number of contacts made by GFP as a whole with every atom of GPI. Figure 11 shows how the unscaled and scaled coarse-grained versions compare with the all-atom system. Results of four different 1 μs long atomistic trajectories are placed against those of 4 μs long coarse-grained trajectories. A number of contacts made were counted within a shell of radius 0.6 nm. For a 1-to-1 comparison between the all-atom and coarse-grained resolutions, we mapped the atomistic GFP-GPI system to the coarse-grained form prior to calculating the frequency of contacts. The scaled coarse-grained force field (orange) covers the same range of contact frequencies as the all-atom system, whereas the unscaled coarse-grained force field lies far on the higher side, an unchartered regime (15−20) of the all-atom system. This shows that the interactions between GFP and GPI are overly strong in the regular MARTINI force field.

Comparison to All-Atom Simulations. *GPI.* Having validated the modified MARTINI force field for simple sugars, the study was extended to model our system of interest, the GPI anchor, as outlined in the *Parametrization Strategy*. All the

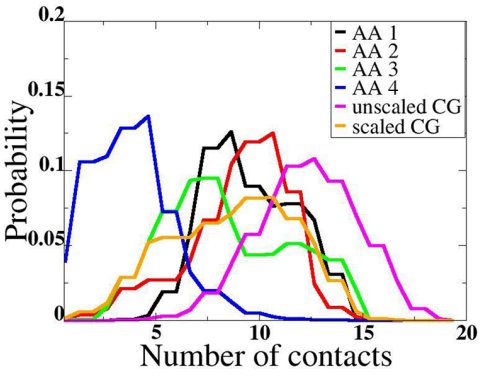


Figure 11. Comparison of distributions of number of contacts made within a radius of 0.6 nm between GFP and GPI glycan between four different all-atom (AA) trajectories (black, red, green, blue) and coarse-grained (CG) trajectories (magenta for the unscaled and orange for the scaled force field). The plots show running averages over five neighboring data points to enhance legibility.

bonded parameters were derived from an all-atom system of 1 GPI core (without the lipid tail) in water. The comparison of the bonded potentials is shown in the SI in Figures S4 and S5. All the comparisons between the all-atom and coarse-grained systems were conducted between the mapped atomistic (in other words, pseudo-CG) and actual coarse-grained trajectories. To compare the global structures of the GPIs between atomistic and coarse-grained descriptions, we calculated their radius of gyration and end-to-end distance. Radius of gyration, R_g, gives an estimate of the size and conformation of a chainlike molecule, for, e.g., if the chain is coiled up or extended. End-to-end distance, R_{ee}, describes how much the polymer is stretched in structure. The comparison along with the values are shown in Figure 12 and Table 6, respectively. As

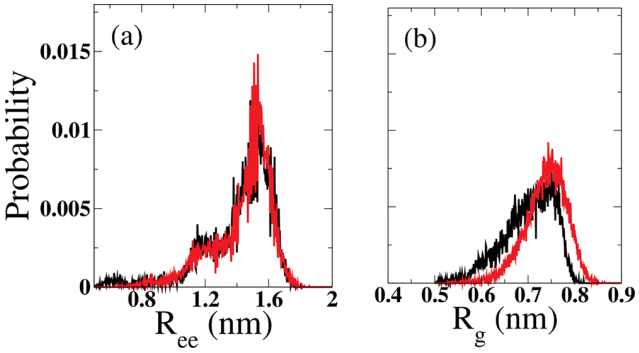


Figure 12. Comparison of structural properties (a) end-to-end distance R_{ee} and (b) radius of gyration R_g between the all-atom (black) and coarse-grained (red) representations of a single GPI core in water.

is evident from the overlapping plots and values, our coarse-grained GPI structurally represents its atomistic counterpart

Table 6. Average Values of End-to-End Distance R_{ee} and Radius of Gyration R_g between All-Atom (AA) and Coarse-Grained (CG) GPI Core in Water

	R_{ee} (nm)	R_g (nm)
AA	1.41 ± 0.22	0.70 ± 0.05
CG	1.44 ± 0.19	0.74 ± 0.04

https://dx.doi.org/10.1021/acs.jctc.0c00056
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

really well. The R_{ee} values match perfectly, whereas the coarse-grained R_g distribution (red) is slightly right-shifted, even though the modes are the same. This is because of the bigger sized coarse-grained particles that experience a basal LJ repulsion, which is absent in the pseudo-CG trajectory (black).

The GPI anchor was inserted into pure lipid bilayers of DMPC maintaining the same setup as in the corresponding atomistic system (see Table 5). Global structural properties, i.e., radius of gyration and end-to-end distance, were again compared between the all-atom and coarse-grained systems as shown in Figures 13a and 13b. Plots are shown for GPIs both

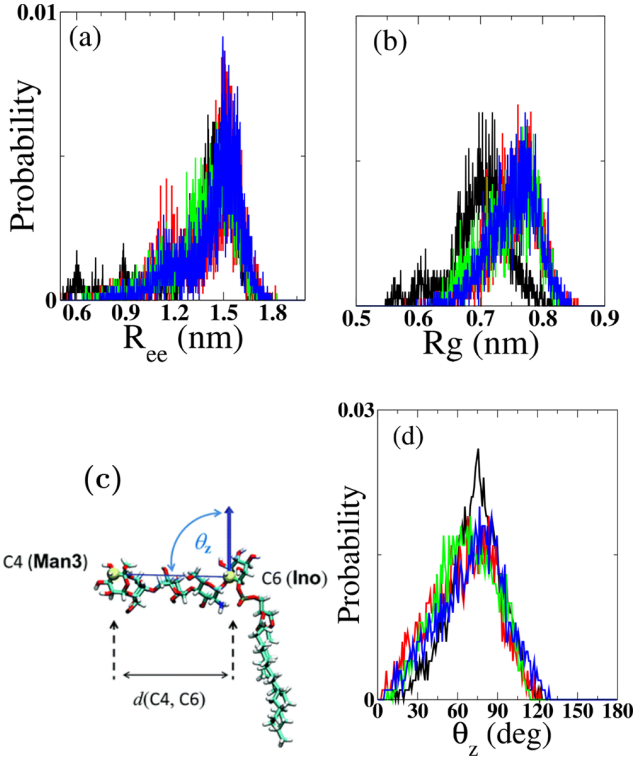


Figure 13. Comparison of structural properties (a) end-to-end distance, R_{ee}, and (b) radius of gyration, R_g, between all-atom and coarse-grained GPIs in a pure DMPC bilayer. Part (c) shows the description of tilt angle θ_z of the GPI core, and its corresponding distribution profiles are displayed in (d). Profiles of all-atom GPI in the top leaflet are shown in black, in the bottom leaflet is shown in red, the coarse-grained GPI in the top leaflet is shown in green, and in the bottom leaflet it is shown in blue.

in upper and lower leaflets. To study the conformation adopted by the GPI with respect to the lipid bilayer, we calculate the angle of tilt formed by the GPI core with the bilayer normal. Figure 13c shows the definition of the tilt angle. In the atomistic system, it is the angle formed by the vector connecting the end points: C4 atom of Man3 and C6 atom of Ino, with the bilayer normal (z axis in this case). In the coarse system, this vector connects the beads containing the aforementioned atoms in the atomistic system. The distribution of the tilt angle of the coarse-grained GPI largely overlaps with that of the atomistic GPI (see Figure 13d). The peak value is ≈80 degrees, which implies that in both all-atom and coarse-grained representations GPIs flop down on the membrane, with the whole GPI core almost swimming in the headgroup region of the lipid bilayer (see Figure 14).

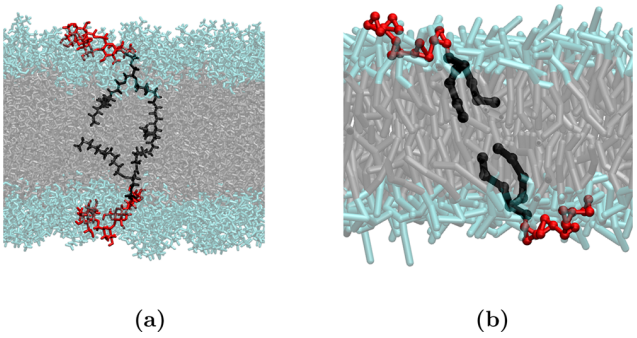


Figure 14. Snapshots at the end of 1 μs long simulations of GPIs in DMPC bilayers for (a) the all-atom and (b) the coarse-grained model.

We characterized the embedding of the GPI within the lipid headgroup region by calculating the hydration number for each of the five sugar residues of the GPI, which is the number of water molecules lying within a radius of 5.5 Å from the atoms of the sugar residues. This distance criterion was applied only to the oxygen atoms of the waters in the all-atom system and to the central, neutral beads of the three-bead-waters in the coarse-grained system. Hydration numbers of each saccharide ring were compared between the purely aqueous system (N_{soln}) where only 1 GPI is solvated in water and the bilayer system (N_{mem}) with 1 GPI inserted into each leaflet. Figure 15 shows the hydration ratios $\frac{N_{mem}}{N_{soln}}$ for the all-atom and coarse-grained GPIs. The relative hydration is lowest for the Ino (violet) and GlcN (blue) residues in both the all-atom and coarse-grained cases and highest for Man1 (green) in the all-atom and for Man2 (orange) in the coarse-grained systems. When comparing hydration ratios to the density profiles of each residue along the bilayer normal (see Figure 16), it is observed that either Man1 or Man2 can be the outermost residue or, in other words, the most solvent-exposed residue in the all-atom system. Ino and GlcN lie at about the same distance away from the bilayer center in the coarse-grained system, whereas a small difference can be seen in the all-atom system. In both the all-atom and coarse-grained systems, Man3 (red) lies closer to the bilayer head than either Man1 or Man2. The same is conveyed by the hydration ratio plots of Man3, indicating that GPIs flop down on the bilayer in both representations. Note that the embedding of GPI into the lipid head is more pronounced in the all-atom than the coarse-grained system. This difference arises from the differences in size of the all-atom and the coarse-grained particles they define. Atoms can percolate more easily into gaps between lipid heads than coarse-grained beads, thereby exposing them less to the solvent phase. Layering effects tend to occur for the coarse-grained system at a larger length scale compared to the all-atom system. Regardless, the overall qualitative picture of the conformation of GPI and its interaction with the membrane is retained in the coarse-grained representation.

GPI-Anchored GFP. We recall that from the all-atom simulations from our previous work[7] we could convincingly infer the following properties of the mutual interaction of the three different molecular species: (i) the GPI core undergoes similar conformational changes as if free in solution; (ii) the GPI core lies in close contact with the lipid head groups for both the free GPI and with the GPI-AP; (iii) the GPI core makes contacts with the attached protein; and (iv) the EtNP-

https://dx.doi.org/10.1021/acs.jctc.0c00056
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

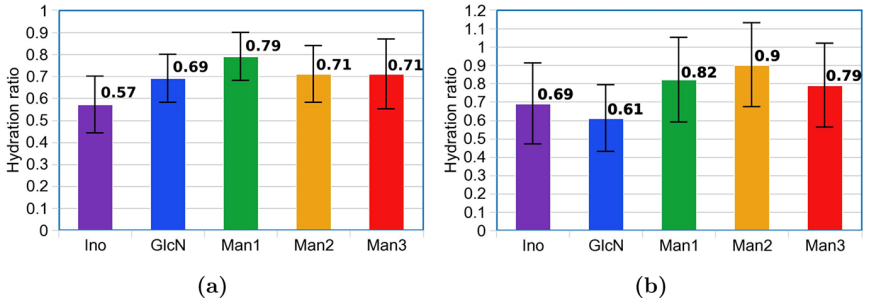


Figure 15. Hydration ratios for each carbohydrate residue of the GPI in the (a) all-atom and (b) coarse-grained system.

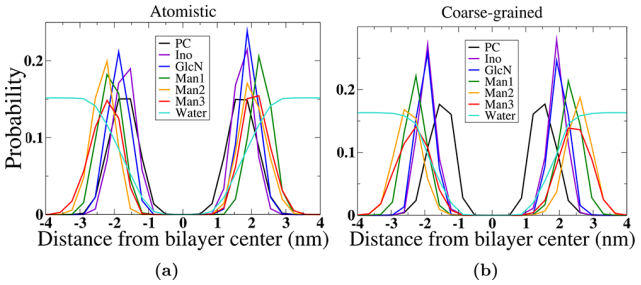


Figure 16. Comparison of density distributions of each residue of the GPI away from the bilayer center along the bilayer normal between the (a) all-atom and (b) coarse-grained systems.

linker conveys extraordinary flexibility to the orientation of the protein w.r.t the bilayer.

We now verify the aforementioned properties with our coarse-grained model. In Figure 17, we compare the structural

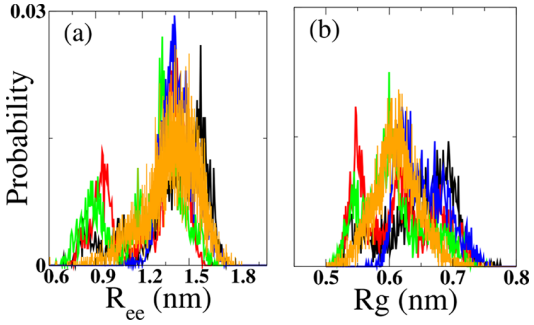


Figure 17. Comparison of end-to-end (R_{ee}) distance and radius of gyration (R_g) of GFP-attached-GPI between four different 1 μs long all-atom (black, red, green, blue) and a 4 μs long coarse-grained (orange) trajectories.

properties, end-to-end distance, R_{ee}, and radius of gyration, R_g, of GPI when attached to GFP between the two resolutions. For both properties, the values from the coarse-grained system average around the peak values of the all-atom plots. The angle of tilt of both the GFP and GPI from the bilayer normal is a way of quantifying the extent of their communication with the lipid bilayers and of overall conformation in general. The definition of the tilt angle, along with the plots of comparison of the values, is illustrated in the schematic in Figure 18. The results from all four all-atom trajectories show that GFP eventually ends up reclining on the membrane, with its tilt angle saturating around 70°. The coarse-grained profile shows similar behavior of GFP until 1 μs, beyond which the protein fluctuates greatly in its orientation. It is to be noted that the

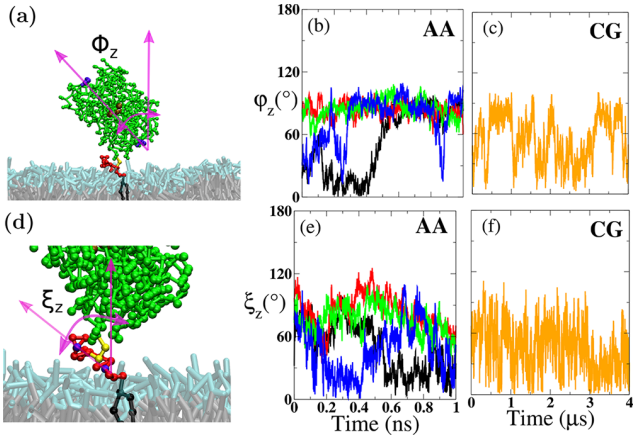


Figure 18. Comparison of tilt angle of (a) GFP and (d) GPI between four independent all-atom (black, red, green, blue) and coarse-grained (orange) systems. Parts (b) and (c) show tilt angles of GFP, and parts (e) and (f) show tilt angles of GPI. Tilt angle ϕ_z of GFP is defined as the angle between the bilayer normal (z axis) and the vector connecting the purple residues (glutamine and histidine). (d) Tilt angle ξ_z of GPI is defined in the same way as in Figure 13c.

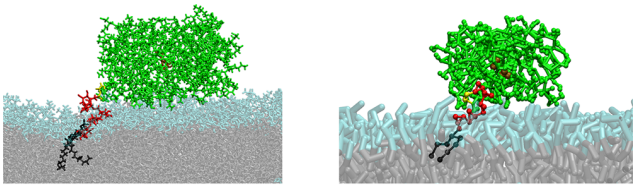
dynamics of a coarse-grained system is always faster than atomistic, about 4 times faster as has been reported for MARTINI. This is because of reduced degrees of freedom in the coarse-grained landscape that leads to loss of friction and hence faster dynamics. This implies that 1 μs of coarse-grained simulation is equivalent to 4 μs of all-atom simulation. Up until the same time frame as the atomistic simulations, coarse-grained GFP-GPI shows similar profiles of tilt angle. On running the simulation longer, it is revealed that, in fact, GFP does wobble a fair deal, instead of lying consistently flat on the membrane, a deceptive picture presented by the all-atom simulations as an offshoot of slow dynamics. The tilt angle of GPI also fluctuates between 20 and 100° in both the all-atom and coarse-grained systems. This shows that the GPI is equally flexible in structure in both the all-atom and coarse-grained systems.

Figure 19 shows snapshots of all-atom and coarse-grained simulations after 700 ns when GFP lies flat on the membrane.

CONCLUSIONS

We developed a coarse-grained model of simple sugars—glucose, sucrose, and trehalose—, GPI and GPI-anchored GFP with a combined bottom-up and top-down approach to parametrize the bonded and nonbonded interactions, respectively. The model development is based on a modified version of the MARTINI force field that is suitable for

https://dx.doi.org/10.1021/acs.jctc.0c00056
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX



(a) Atomistic GFP-GPI (b) Coarse-grained GFP-GPI

Figure 19. GFP-GPI inserted into DMPC lipid bilayers at atomistic and coarse-grained resolutions in (a) and (b), respectively.

modeling carbohydrates in the environment of polarizable water. The interaction potentials of lipid−lipid, sugar−lipid, and protein−lipid were retained from the MARTINI polarizable force field, but the potentials describing sugar−sugar and sugar−protein were altered by scaling down the amplitudes ϵ_{ij}s of the Lennard-Jones potentials to match the experimental and atomistic behavior. A scaling factor of $\gamma = 0.85$ was sufficient to reproduce the experimental osmotic virial coefficients (B_{22}) of simple sugars, which was extended to the bead types of GPI core. Using polarizable water was essential to the study because our objective was to study the conformational characteristics of GPI and GPI-AP inserted in lipid bilayers for which the interfacial interplay of interactions among lipid heads, carbohydrates, protein, and water needed to be well characterized. On comparing our model of GPI in polarizable water versus in standard MARTINI water, we observed that GPIs interact a great deal with the membrane in polarizable water, just as the atomistic case, whereas they barely interacted with the lipids in standard water causing the glycan, for the most part, to project out of the lipids like a brush. This shows that the water model has a strong effect on the GPI conformation.

GFP was individually coarse-grained in water with the ELNEDYN force field and was subsequently attached to GPI in a modular fashion with a EtNP linker, which also was separately coarse-grained from the atomistic system. GPI proves to be flexible both in the atomistic and coarse-grained landscapes, and the orientation of the attached protein (GFP) with respect to the lipid membrane fluctuates significantly. A plausible reason for this unsteady behavior could be the absence of specific adhesive interactions between GFP and the lipid bilayer. This phenomenon was observed in our control simulations where upon forcing GFP to lie in contact with the bilayer headgroups through a biased force for 300 ns and subsequently releasing the force, GFP moved away from the bilayer after about 500 ns. It has also been reported in experiments that GFP only negligibly binds to membranes.[62] The analysis of the similar number of contacts formed between GFP and GPI at the atomistic and coarse-grained resolutions suggests that they interact similarly in the two representations, providing further validation to our coarse-grained model. Our coarse-grained model of GPI along with its EtNP linker, both of which together form a conserved entity, can be combined with other GPI-anchored proteins like alkaline phosphatase, Thy1, MSP1 of *Plasmodium falciparum*, or even prion protein to address crucial questions concerning their general orientation, mechanisms of action, or pathogenesis.

The speed-up obtained from the coarse-graining was 16-fold in the GPI simulations and 10-fold in the GFP-GPI simulations. With this fast dynamics, we can further address challenging questions that entail larger systems and longer simulation runtime, like the role of GPIs in protein trafficking which can be studied by observing their partitioning tendencies toward liquid-ordered or liquid-disordered regions of heterogeneous membranes consisting of a variety of lipids including gangliosides and cholesterol. The coarse-grained model of GPI presented herewith can be used in conjunction with the ever-expanding library of MARTINI lipid types to add another component toward building a complex plasma membrane.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.0c00056.

Parameters of coarse-grained EtNP linker containing associated bead types and bonded potentials, Table S1; dihedral (Figure S1) and angle distributions (Figure S2) compared between atomistic and coarse-grained EtNP; mapping scheme of chromophore inside GFP, Figure S3; bead types and bonded parameters, Table S2; comparison of dihedral distributions for GPI in water between atomistic and coarse-grained systems (Figure S4) and of angle distributions (Figure S5); and procedure for backmapping coarse-grained system of GPI-anchored GFP, complemented by list of residue names of lipid DMPC (Table S3) and of GPI (Table S4) (PDF)

gpi.amber.map − mapping file for GPI; dmpc.amber.map − mapping file for DMPC; gfp-gpi-aa.top − atomistic topology of GPI-anchored GFP inserted in DMPC bilayer; and gfp-gpi-cg.gro − coarse-grained structure file of GPI-anchored GFP inserted in DMPC bilayer (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

Mark Santer − *Max Planck Institute of Colloids and Interfaces, Potsdam 14476, Germany;* orcid.org/0000-0002-5545-5155; Email: mark.santer@mpikg.mpg.de

Authors

Pallavi Banerjee − *Max Planck Institute of Colloids and Interfaces, Potsdam 14476, Germany; Institute of Biochemistry and Biology, University of Potsdam, Potsdam 14469, Germany;* orcid.org/0000-0002-1714-2780

Reinhard Lipowsky − *Max Planck Institute of Colloids and Interfaces, Potsdam 14476, Germany; Institute of Biochemistry and Biology, University of Potsdam, Potsdam 14469, Germany;* orcid.org/0000-0001-8417-8567

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.0c00056

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the International Max Planck Research School (IMPRS) for Multiscale Biosystems. Open Access funding was provided by the Max Planck Society. We thank Marko Wehle for his contribution to the atomistic GPI model. We are grateful to Alex H. de Vries, Mateusz Sikora and Philipp S. Schmalhorst for the helpful discussions.

REFERENCES

(1) Low, M. G.; Saltiel, A. R. Structural and functional roles of glycosyl-phosphatidylinositol in membranes. *Science* **1988**, *239*, 268−275.

(2) Tachado, S. D.; Gerold, P.; Schwarz, R.; Novakovic, S.; McConville, M.; Schofield, L. Signal transduction in macrophages by glycosylphosphatidylinositols of Plasmodium, Trypanosoma, and Leishmania: activation of protein tyrosine kinases and protein kinase C by inositolglycan and diacylglycerol moieties. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 4022−4027.

(3) Karagogeos, D. Neural GPI-anchored cell adhesion molecules. *Front. Biosci., Landmark Ed.* **2003**, *8*, s1304−20.

(4) Brown, D. A.; Rose, J. K. Sorting of GPI-anchored proteins to glycolipid-enriched membrane subdomains during transport to the apical cell surface. *Cell* **1992**, *68*, 533−544.

(5) Mayor, S.; Riezman, H. Sorting GPI-anchored proteins. *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 110.

(6) Taylor, D. R.; Hooper, N. M. *Post-translational modifications in health and disease*; Springer: New York, NY, USA, pp 39−55, DOI: 10.1007/978-1-4419-6382-6_2.

(7) Banerjee, P.; Wehle, M.; Lipowsky, R.; Santer, M. A molecular dynamics model for glycosylphosphatidyl-inositol anchors:flop down or lollipop? *Phys. Chem. Chem. Phys.* **2018**, *20*, 29314−29324.

(8) Lehto, M. T.; Sharom, F. J. Proximity of the protein moiety of a GPI-anchored protein to the membrane surface: a FRET study. *Biochemistry* **2002**, *41*, 8368−8376.

(9) Barboni, E.; Rivero, B. P.; George, A.; Martin, S.; Renoup, D.; Hounsell, E.; Barber, P.; Morris, R. The glycophosphatidylinositol anchor affects the conformation of Thy-1 protein. *J. Cell Sci.* **1995**, *108*, 487−497.

(10) Paulick, M. G.; Wise, A. R.; Forstner, M. B.; Groves, J. T.; Bertozzi, C. R. Synthetic analogues of glycosylphosphatidylinositol-anchored proteins and their behavior in supported lipid bilayers. *J. Am. Chem. Soc.* **2007**, *129*, 11543−11550.

(11) Paulick, M. G.; Forstner, M. B.; Groves, J. T.; Bertozzi, C. R. A chemical approach to unraveling the biological function of the glycosylphosphatidylinositol anchor. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 20332−20337.

(12) Schroeder, R.; London, E.; Brown, D. Interactions between saturated acyl chains confer detergent resistance on lipids and glycosylphosphatidylinositol (GPI)-anchored proteins: GPI-anchored proteins in liposomes and cells show similar behavior. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 12130−12134.

(13) Varma, R.; Mayor, S. GPI-anchored proteins are organized in submicron domains at the cell surface. *Nature* **1998**, *394*, 798.

(14) Sevcsik, E.; Brameshuber, M.; Fölser, M.; Weghuber, J.; Honigmann, A.; Schütz, G. J. GPI-anchored proteins do not reside in ordered domains in the live cell plasma membrane. *Nat. Commun.* **2015**, *6*, 6969.

(15) Homans, S.; Edge, C.; Ferguson, M.; Dwek, R.; Rademacher, T. Solution structure of the glycosylphosphatidylinositol membrane anchor glycan of Trypanosoma brucei variant surface glycoprotein. *Biochemistry* **1989**, *28*, 2881−2887.

(16) Chevalier, F.; Lopez-Prados, J.; Perez, S.; Martín-Lomas, M.; Nieto, P. M. Conformational study of GPI anchors: the common oligosaccharide GPI anchor backbone. *Eur. J. Org. Chem.* **2005**, *2005*, 3489−3498.

(17) Zuegg, J.; Gready, J. E. Molecular dynamics simulation of human prion protein including both N-linked oligosaccharides and the GPI anchor. *Glycobiology* **2000**, *10*, 959−974.

(18) Wehle, M.; Vilotijevic, I.; Lipowsky, R.; Seeberger, P. H.; Varon Silva, D.; Santer, M. Mechanical compressibility of the glycosylphosphatidylinositol (GPI) anchor backbone governed by independent glycosidic linkages. *J. Am. Chem. Soc.* **2012**, *134*, 18964−18972.

(19) Periole, X.; Marrink, S.-J. *Biomolecular Simulations*; Humana Press: Totowa, NJ, USA, Vol. *924*, pp 533−565, DOI: 10.1007/978-1-62703-017-5_20.

(20) Corradi, V.; Sejdiu, B. I.; Mesa-Galloso, H.; Abdizadeh, H.; Noskov, S. Y.; Marrink, S. J.; Tieleman, D. P. Emerging diversity in lipid−protein interactions. *Chem. Rev.* **2019**, *119*, 5775−5848.

(21) Tsai, Y.-H.; Liu, X.; Seeberger, P. H. Chemical biology of glycosylphosphatidylinositol anchors. *Angew. Chem., Int. Ed.* **2012**, *51*, 11438−11456.

(22) López, C. A.; Rzepiela, A. J.; De Vries, A. H.; Dijkhuizen, L.; Hunenberger, P. H.; Marrink, S. J. Martini coarse-grained force field: extension to carbohydrates. *J. Chem. Theory Comput.* **2009**, *5*, 3195−3210.

(23) Yesylevskyy, S. O.; Schäfer, L. V.; Sengupta, D.; Marrink, S. J. Polarizable water model for the coarse-grained MARTINI force field. *PLoS Comput. Biol.* **2010**, *6*, No. e1000810.

(24) Bennett, W. D.; Tieleman, D. P. Water defect and pore formation in atomistic and coarse-grained lipid membranes: pushing the limits of coarse graining. *J. Chem. Theory Comput.* **2011**, *7*, 2981−2988.

(25) Kociurzynski, R.; Pannuzzo, M.; Bockmann, R. A. Phase transition of glycolipid membranes studied by coarse-grained simulations. *Langmuir* **2015**, *31*, 9379−9387.

(26) Pluhackova, K.; Wassenaar, T. A.; Kirsch, S.; Bockmann, R. A. Spontaneous adsorption of coiled-coil model peptides K and E to a mixed lipid bilayer. *J. Phys. Chem. B* **2015**, *119*, 4396−4408.

(27) Gu, R.-X.; Ingolfsson, 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. *J. Phys. Chem. B* **2017**, *121*, 3262−3275.

(28) Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; Van Gunsteren, W. F. Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chem. Phys. Lett.* **1994**, *222*, 529−539.

(29) Mazzobre, M. F.; Román, M. V.; Mourelle, A. F.; Corti, H. R. Octanol−water partition coefficient of glucose, sucrose, and trehalose. *Carbohydr. Res.* **2005**, *340*, 1207−1211.

(30) Alessandri, R.; Souza, P. C.; Thallmair, S.; Melo, M. N.; De Vries, A. H.; Marrink, S. J. Pitfalls of the Martini Model. *J. Chem. Theory Comput.* **2019**, *15*, 5448−5460.

(31) Periole, X.; Cavalli, M.; Marrink, S.-J.; Ceruso, M. A. Combining an elastic network with a coarse-grained molecular force field: structure, dynamics, and intermolecular recognition. *J. Chem. Theory Comput.* **2009**, *5*, 2531−2543.

(32) Ormö, M.; Cubitt, A. B.; Kallio, K.; Gross, L. A.; Tsien, R. Y.; Remington, S. J. Crystal structure of the Aequorea victoria green fluorescent protein. *Science* **1996**, *273*, 1392−1395.

(33) 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. *J. Phys. Chem. B* **2007**, *111*, 7812−7824.

(34) Moiset, G.; Lopez, 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*, 16167−16175.

(35) Schmalhorst, P. S.; Deluweit, F.; Scherrers, R.; Heisenberg, C.-P.; Sikora, M. Overcoming the limitations of the MARTINI force field in simulations of polysaccharides. *J. Chem. Theory Comput.* **2017**, *13*, 5039−5053.

(36) Stark, A. C.; Andrews, C. T.; Elcock, A. H. Toward optimized potential functions for protein−protein interactions in aqueous solutions: osmotic second virial coefficient calculations using the martini coarse-grained force field. *J. Chem. Theory Comput.* **2013**, *9*, 4176−4185.

(37) Javanainen, M.; Martinez-Seara, H.; Vattulainen, I. Excessive aggregation of membrane proteins in the Martini model. *PLoS One* **2017**, *12*, No. e0187936.

(38) López, C. A.; Sovova, Z.; van Eerden, F. J.; de Vries, A. H.; Marrink, S. J. Martini force field parameters for glycolipids. *J. Chem. Theory Comput.* **2013**, *9*, 1694−1708.

(39) Petrov, D.; Zagrovic, B. Are current atomistic force fields accurate enough to study proteins in crowded environments? *PLoS Comput. Biol.* **2014**, *10*, No. e1003638.

(40) Best, R. B.; Zheng, W.; Mittal, J. Balanced protein−water interactions improve properties of disordered proteins and non-specific protein association. *J. Chem. Theory Comput.* **2014**, *10*, 5113−5124.
(41) Karunaweera, S.; Gee, M. B.; Weerasinghe, S.; Smith, P. E. Theory and simulation of multicomponent osmotic systems. *J. Chem. Theory Comput.* **2012**, *8*, 3493−3503.
(42) Ploetz, E. A.; Bentenitis, N.; Smith, P. E. Developing force fields from the microscopic structure of solutions. *Fluid Phase Equilib.* **2010**, *290*, 43−47.
(43) Ploetz, E. A.; Smith, P. E. A Kirkwood−Buff force field for the aromatic amino acids. *Phys. Chem. Chem. Phys.* **2011**, *13*, 18154−18167.
(44) Luo, Y.; Roux, B. Simulation of osmotic pressure in concentrated aqueous salt solutions. *J. Phys. Chem. Lett.* **2010**, *1*, 183−189.
(45) Sauter, J.; Grafmuller, A. Efficient Osmotic Pressure Calculations Using Coarse-Grained Molecular Simulations. *J. Chem. Theory Comput.* **2018**, *14*, 1171−1176.
(46) Miller, M. S.; Lay, W. K.; Li, S.; Hacker, W. C.; An, J.; Ren, J.; Elcock, A. H. Reparametrization of protein force field nonbonded interactions guided by osmotic coefficient measurements from molecular dynamics simulations. *J. Chem. Theory Comput.* **2017**, *13*, 1812−1826.
(47) Stigter, D. Interactions in aqueous solutions. II. Osmotic pressure and osmotic coefficient of sucrose and glucose solutions. *J. Phys. Chem.* **1960**, *64*, 118.
(48) Davis, D. J.; Burlak, C.; Nicholas, P. Osmotic pressure of fungal compatible osmolytes. *Mycol. Res.* **2000**, *104*, 800−804.
(49) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*, 19−25.
(50) 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.
(51) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(52) Dickson, C. J.; Madej, B. D.; Skjevik, Å. A.; Betz, R. M.; Teigen, K.; Gould, I. R.; Walker, R. C. Lipid14: the amber lipid force field. *J. Chem. Theory Comput.* **2014**, *10*, 865−879.
(53) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696−3713.
(54) Sorin, E. J.; Pande, V. S. Exploring the helix-coil transition via all-atom equilibrium ensemble simulations. *Biophys. J.* **2005**, *88*, 2472−2493.
(55) Marrink, S. J.; De Vries, A. H.; Mark, A. E. Coarse grained model for semiquantitative lipid simulations. *J. Phys. Chem. B* **2004**, *108*, 750−760.
(56) Wassenaar, T. A.; Ingólfsson, H. I.; Bockmann, R. A.; Tieleman, D. P.; Marrink, S. J. Computational lipidomics with insane: a versatile tool for generating custom membranes for molecular simulations. *J. Chem. Theory Comput.* **2015**, *11*, 2144−2155.
(57) Verlet, L. Computer" experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.* **1967**, *159*, 98.
(58) 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*, 8577−8593.
(59) Goga, N.; Rzepiela, A.; De Vries, A.; Marrink, S.; Berendsen, H. Efficient algorithms for Langevin and DPD dynamics. *J. Chem. Theory Comput.* **2012**, *8*, 3637−3649.
(60) Berendsen, H. J.; Postma, J. v.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(61) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
(62) Kim, H.; Afsari, H. S.; Cho, W. High-throughput fluorescence assay for membrane-protein interaction. *J. Lipid Res.* **2013**, *54*, 3531−3538.

https://dx.doi.org/10.1021/acs.jctc.0c00056
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

SUPPORTING INFORMATION

Coarse-grained molecular model for the glycosylphosphatidylinositol anchor with and without protein

Pallavi Banerjee,[†,‡] Reinhard Lipowsky,[†,‡] and Mark Santer[*,†]

†*Max Planck Institute of Colloids and Interfaces, Potsdam, Germany*
‡*University of Potsdam, Potsdam, Germany*

E-mail: mark.santer@mpikg.mpg.de

Parameterization of the EtNP linker

The bead definitions of the coarse-grained molecule containing the EtNP linker (see Fig.5b of the main article) are listed in Tab.S1. Coarse-grained simulations of the mapped molecule were also conducted for 200 ns. Fig.S1 and S2 show good overlap of the coarse-grained dihedral and angle distributions, respectively, with those of the all-atom system. The derived bonded parameters involving beads L1 and L2 were plugged in to the GFP-GPI coarse-grained system. The parameters are listed in Tab.S1.

Table S1: Bead names with corresponding bead types of coarse-grained EtNP linker, along with its bonded topology (bonds, angles, dihedrals). For the amino acid beads (BB, SC1, SCD), subscript 1 denotes GLY, 2 denotes ILE, and 3 denotes THR.

Bead name	Bead type	Bonds	r_0 (nm)	K_b (kJ/mol)	Angle	$\theta_0(°)$	K_a (kJ/mol)
T	PSN0	T-BB_1	0.3	20000	T-BB_1-BB_2	128	20
BB_1	PNda	BB_1-BB_2	0.39	20000	BB_1-BB_2-BB_3	132	30
BB_2	PP5	BB_2-$SC1_2$	0.23	13250	$SC1_2$-BB_2-BB_3	108	30
$SC1_2$	PC1	BB_2-BB_3	0.35	20000	BB_2-BB_3-$SC1_3$	96	80
BB_3	PNda	BB_3-$SC1_3$	0.26	94000	BB_2-BB_3-L1	116	30
$SC1_3$	PN0	BB_3-L1	0.34	10000	$SC1_3$-BB_3-L1	106	20
SCD_3	D	L1-L2	0.33	20000	BB_3-L1-L2	125	60
L1	GSNda	L2-C15	0.33	30000	L1-L2-C15	80	40
L2	GQa	C15-C14	0.33	20000	L2-C15-C13	140	20
C15	GNa	C15-C13	0.35	20000	L2-C15-C14	124	30
C14	GP3	C14-C13	0.28	40000	C15-C13-C10	65	450
C13	GSP1	C13-C10	0.36	20000	C14-C13-C10	95	100
C12	GP2	C12-C11	0.33	30000	Dihedral	$\phi_0(°)$	K_d
C11	GP3	C12-C10	0.35	30000	T-BB_1-BB_2-BB_3	-57.3	2
C10	GSN0	C11-C10	0.28	40000	BB_1-BB_2-BB_3-L1	57.3	3
		Constraints			BB_2-BB_3-L1-L2	180	3.5
					BB_3-L1-L2-C15	-34.4	5
		SCD_3-SCD_3	0.28		L2-C15-C13-C10 ($m = 2$)	-126	4

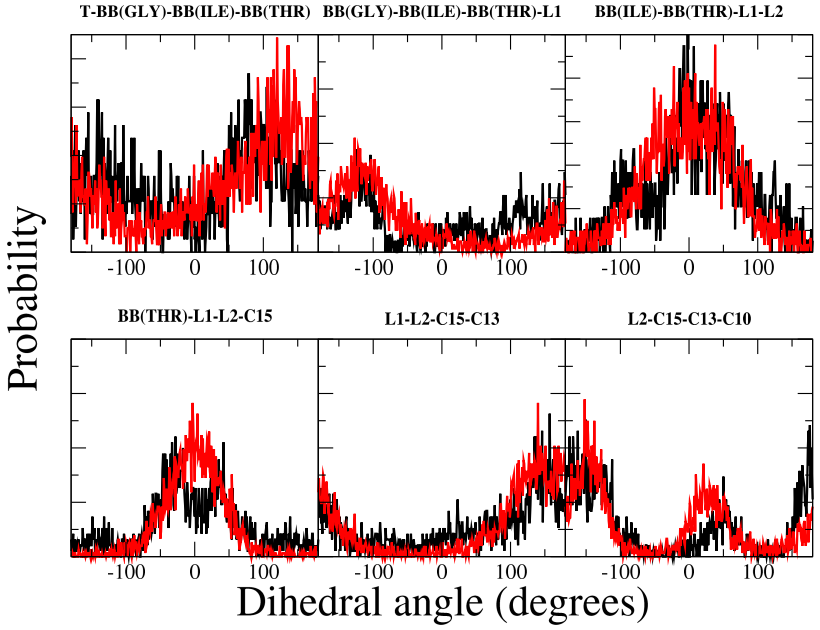


Figure S1: Comparing atomistic(black) and coarse-grained(red) dihedral angle profiles for the linker. The bead names are in correspondence with those shown in Fig.5b of the main article.

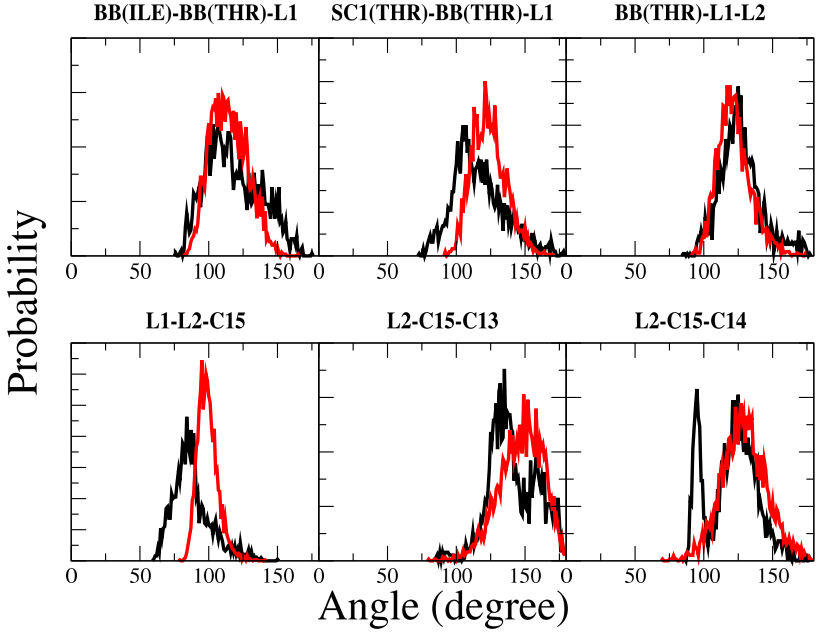


Figure S2: Comparing atomistic(black) and coarse-grained(red)angle profiles for the linker. The bead names are in correspondence with those shown in Fig.5b of the main article.

Parameterization of the chromophore inside GFP

Mapping scheme of the chromophore is shown in Fig.S3. Bonded parameters of the coarse-grained chromophore were derived from 1μs long all-atom trajectories of GFP in water. Bead types were assigned by following MARTINI's interaction matrix. Small (S) beads were used to represent the ring particles. The values are listed in Tab.S2.

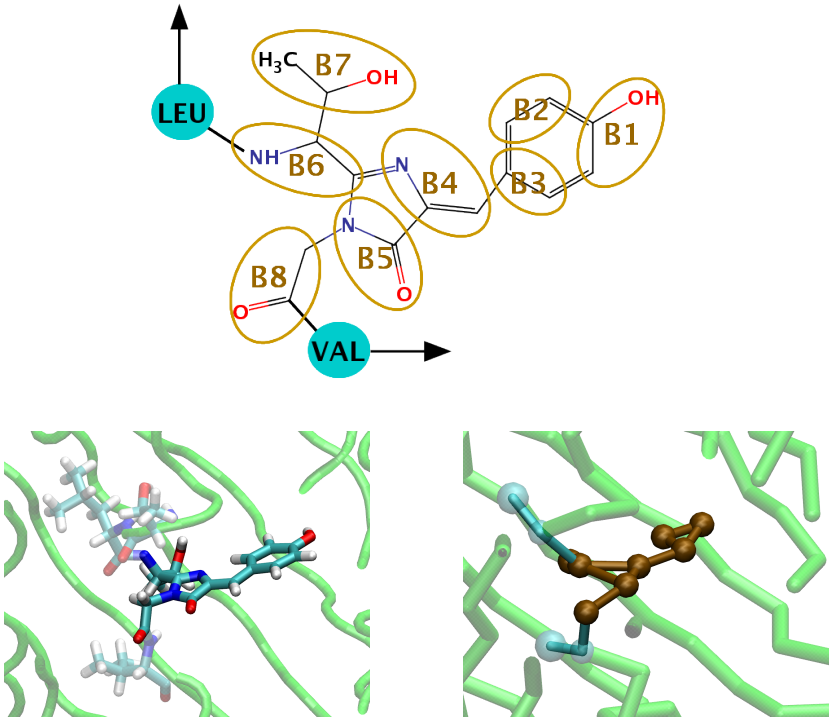


Figure S3: Mapping scheme of the chromophore is shown by superimposing coarse-grained beads with their bead names over the atomistic structure. The snapshots show the all-atom vs coarse-grained representation of the chromophore inside the barrel of GFP. In the coarse representation, the mapped beads that make up the chromophore are shown in brown. The translucent cyan beads represent the flanking LEU and VAL residues that the chromophore is bonded to. The colours in the coarse-grained snapshot are depicted in correspondence with those in the mapping scheme.

Table S2: Bead names and corresponding bead types of the coarse-grained chromophore. Bonded (bonds, angles, proper and improper torsions) parameters of the coarse-grained chromophore. MARTINI bead type terms have been prefixed with P.

Bead name	Bead type	Bonds	r_0 (nm)	K_b (kJ/mol)	Angle	$\theta_0(^\circ)$	K_a (kJ/mol)
B1	PSP1	B3-B4	0.28964	150000	B2-B3-B4	83	500
B2	PSC1	B4-B6	0.34846	150000	B3-B4-B5	140	450
B3	PSC1	B5-B6	0.34316	150000	B3-B4-B6	143	400
B4	PSNa	B5-B8	0.27216	150000	B4-B6-B7	112	400
B5	PSP3	B6-B7	0.23602	150000	B6-B5-B8	76	300
B6	PSNa	BB(LEU)-B6	0.28934	150000	BB(THR)-BB(LEU)-B6	80	80
B7	PSP2	BB(VAL)-B8	0.33769	150000	BB(LEU)-B6-B4	148	80
B8	PSP2				BB(VAL)-B8-B5	140	80
					BB(GLN)-BB(VAL)-B8	96	80
		Constraints			Proper Torsions	$\phi_0(^\circ)$	K_d
		B1-B2	0.24272		B1-B3-B4-B5	28.5	60
		B1-B3	0.28455		B7-B6-B5-B8	51.5	86
		B2-B3	0.22300		Improper Torsions	$\xi_0(^\circ)$	K_i
		B4-B5	0.22781		B4-B5-B6-B7	30	80
					B4-B5-B6-B8	155	100

Bonded interactions of GPI in water

Comparison of dihedrals and angles between atomistic and coarse-grained GPIs are shown in Fig.S4 and Fig.S5 respectively. The all-atom (black) and coarse-grained (red) profiles sufficiently overlap. The parameters that make up the link between inositol and phosphoglycerol, involving beads C1, C2, C3, PO4, GL1, and GL2, are crucial as they define the orientation of the glycan wrt the lipid head. These have been represented well, as can be seen in the profiles where even the double peaks have been reproduced at the coarse-grained level (see dihedral C2-C1-PO4-GL1 and C3-C1-PO4-GL1).

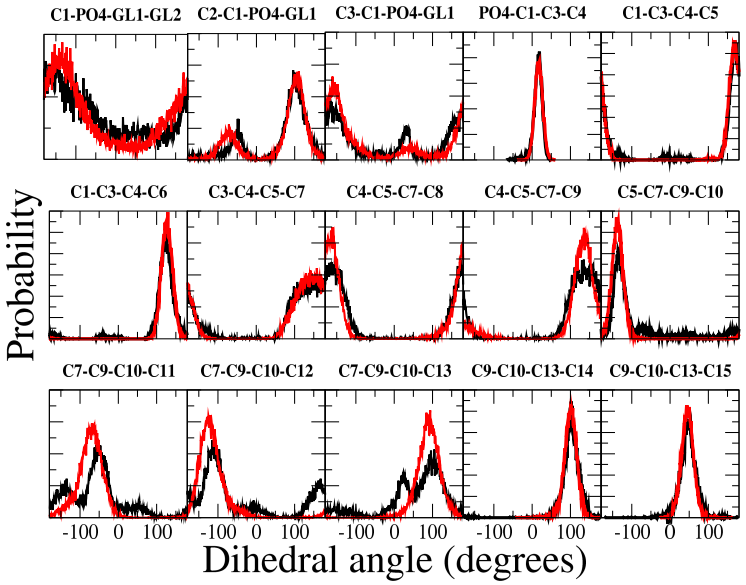


Figure S4: Comparing dihedral angles between atomistic(black) and coarse-grained(red) GPIs

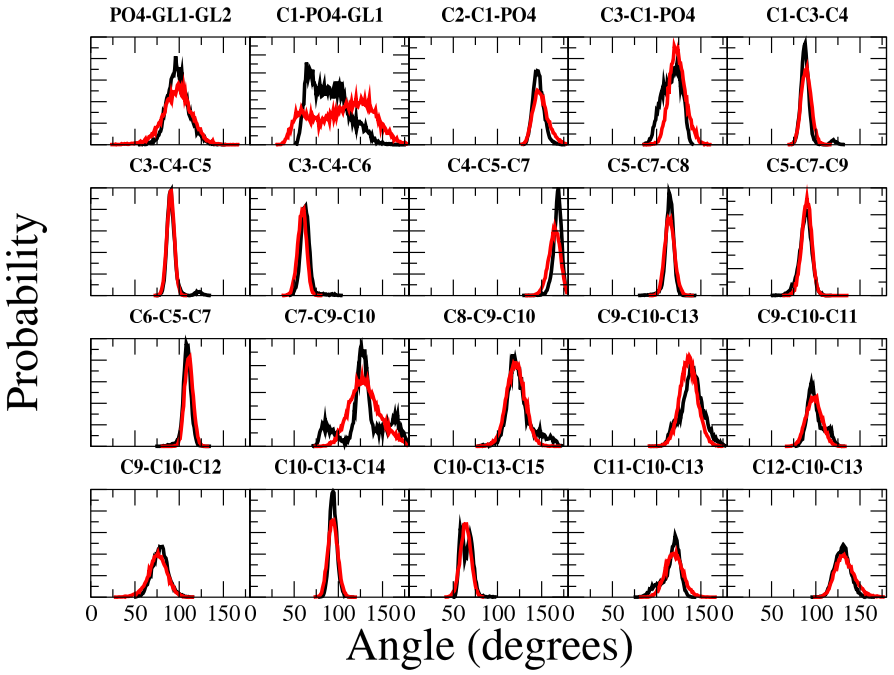


Figure S5: Comparing angles between atomistic(black) and coarse-grained(red) GPIs

Procedure for backmapping

To backmap a coarse-grained system from MARTINI to an all-atom structure file, the script *backward.py* devised by Wassenaar *et al.*[1] can be used. The script uses mapping files for each defined residue/molecule to transform a CG structure file to its atomistic counterpart. The mapping files for GPI and DMPC lipid have been provided as additional supporting information as gpi.amber.map and dmpc.lipid14.map respectively. The mapping files corresponding to the amino acid residues are already available on the database of MARTINI. The backmapped atomistic structure can then be subject to an atomistic simulation using the atomistic topology (gfp-gpi-aa.top) also provided as additional material. The following table gives an overview of the naming conventions used for the molecules and residues in our system.

Table S3: Residue names (res. name) of the coarse-grained (CG) DMPC lipid along with its atomistic (AA) residue names with corresponding description of the residues

CG res.name	AA res.name	AA residue
DMPC	MY	myristoyl tail
	PC	phosphocholine head

Table S4: Residue names (res. name) of the coarse-grained (CG) GPI along with its atomistic (AA) residue names with corresponding description of the residues

CG res.name	AA res.name	AA residue
GPIL	MY	myristoyl tail
	6DP	phosphoinositol-glycerol
	4YM	glucosamine
	6MA	mannose 1
	2MA	mannose 2
	6MA	mannose 3
	EG1	phosphoethanolamine

References

(1) Wassenaar, T. A.; Pluhackova, K.; Bockmann, R. A.; Marrink, S. J.; Tieleman, D. P. Going backward: a flexible geometric approach to reverse transformation from coarse

grained to atomistic models. *J. Chem. Theory Comput.* **2014**, *10*, 676–690.

Chapter 5

Discussion and Conclusions

In this thesis, we have modeled the GPI anchor for the purpose of molecular dynamics studies at two different levels of resolution – all-atom and coarse-grained. We developed the atomistic model by constructing an adapted bridging residue, phosphoinositolglycerol, in order to allow a smooth and continuous amalgamation of two different AMBER force-fields, GLYCAM06 and Lipid14, lying on either side of the juncture. The model was designed in such a manner that a modular framework is facilitated on both the glycan (GLYCAM06) and the lipid side (Lipid14). Considering that GPI anchors show extraordinary structural diversity (see Sec.1.2), in that the glycan can be variably branched and the lipid tails can be differently saturated, the modular feature of the model is of great importance. However, note that only GPIs carrying diacyl phospholipids can be studied with our model. Reparameterization would be required to model GPIs containing alkyl-acyl or ceramide lipids or even an additional lipid tail like that of *Plasmodium falciparum* (see Figure 1.3). In Chapter 2, we have described the method for construction of the hybrid unit in detail. From 1 μs long simulations, we observed that the GPI anchor core is a largely flexible structure adopting different conformations on the lipid surface as a result of the flexible glycosidic linkages, especially the $\alpha(1 \rightarrow 6)$ linkage connecting Man1 and Man2 (see Figure 1.2). Regardless of the internal flexibility of the glycan, GPI, largely flops down on the membrane, while making infrequent transitions to metastable erect conformations. Furthermore, by comparing the conformation of different fragments of the GPI core in lipid bilayers, it was realized that addition of the charged amino group of glucosamine brings about a marked difference to the dihedrals connecting inositol to the phosphoglycerol unit of the lipid. Interactions of this positively charged amino group with negatively charged phosphate of phosphoinositol contribute to the overall flop-down orientation of GPIs. This observation suggests a notable relevance of the unique presence of glucosamine in GPI to the overall structure, over N-acetylglucosamine or sulfated glucosamine that are more common in other naturally occurring glycans[20, 97].

In Chapter 3, we have utilized the modularity of the GPI model that is presented in Chapter2 to study three different parasitic GPIs – *Toxoplasma gondii*, *Toxoplasma gondii* LMW, *Trypanosoma congolense*

– and human GPI, all of which differ in their side chain decorations. For a systematic investigation, we retained the dimyristoyl lipid tail in all. An additional residue, phosphoethanolamine (EtNP) linker that is bonded to the last mannose, Man3, in all GPI forms (see Figure 1.2) was constructed using the GLYCAM protocol to ensure compatibility and modularity. Summing up the findings from Chapter2 and Chapter3, we report that the GPI glycan, irrespective of the level of branching, is a flexible structure, particularly at the Man2-$(1 \rightarrow 6)$-Man1 linkage that can be construed as a molecular hinge for the structure. With the addition of extra side chain residues (more than one), rotamer population was seen to significantly shift from the usually predominant *gg* conformer to *gt*. The protein-free GPI of *T.congolense* demonstrated remarkable conformational variation due to the presence of a $\beta(1 \rightarrow 6)$ linkage between Gal and GlcNac. In glycans, $(1 \rightarrow 6)$ glycosidic torsions are the most flexible because of the presence of an additional carbon[98]. Although the internal conformation of the different GPI variants is prone to variation, their overall orientation with respect to the bilayer, is more or less the same, that is, GPIs prefer to flop down on the bilayer surface and strongly interact with the lipid heads. This behavior is consistent with that of the basic GPI core from Chapter2. Through rigorous conformational analysis we noted that the side branch residues are consistently projected into the solvent, so much so that out of all the residues these are the most-solvent accessible. Here note that the terminal side branch residue Gal of *T.congolense* GPI is an outlier, as it was seen to be quite buried in the lipid heads. We theorize that the side branches could be potential epitopes for parasitic GPI recognition by antibodies. This observation concurs with experimental studies having reported that the side branches are a necessary feature to trigger an immune response. Using *in vitro* macrophage activation, Azzouz et al. demonstrated that the production of Tumor Necrosis Factor-α production was five-fold with protein-free *T.gondii* LMW GPI than with protein-attached *T.gondii* GPI[99]. In another study, synthetic GPIs of *T.gondii* were promoted as suitable diagnostic tools in differentiating between latent and acute stages of *Toxoplasmosis*[44]. Here too it was shown that *T.gondii* LMW GPI elicits a stronger immune response than the protein-bound form. They also reported that the minimal epitope for antibody binding was the side branch – Man1-GalNac-Glc, thus, agreeing qualitatively with our observations. We included human GPI anchor in our study, particularly to compare with *T.gondii* GPI as they are are structurally very similar. The only difference is the additional EtNP attached to O2 of Man1 in human GPI. Our conformational analysis revealed some difference between the two in the ψ values of Man2-Man1 linkage. Moreover, human GPI tended to assume an upright orientation more than *T.gondii* GPI. This observation provides a plausible explanation for the lack of an auto-immune response of antibodies in humans against their own GPIs. Morever, through the use of synthetically generated protein-free *T. gondii* and human GPIs, it was demonstrated that antibodies raised in infected human sera against *T.gondii* GPI variants did not bind with human GPI[44], thus proving the absence of cross-reactivity. In human GPI, we observed that the middle EtNP (mEtNP) is more exposed/accessible than the terminal EtNP (tEtNP), thus backing experimental reports wherein mEtNP was part of the minimal epitope for recognition by GPI transamidase, whereas tEtNP was not a crucial requirement[100]. Our findings indicate a putative structure-function relationship for the GPI anchor. The overall flop-down conformation of GPIs causes the side chain residues to stick out into the solvent. Had the erect/lollipop conformation been favored,

terminal mannoses of the core would have been the most exposed residues. Furthermore, the easy-accessibility of the side branches provides a plausible reason behind specificity in immune response against the different GPI forms[19]. As the GPIs differ from each other at the side branches, it is likely that the immune system is directed at these residues. The antigenic character of protein-free GPIs has been illustrated in experimental studies using both synthetic and isolated GPIs, suggesting that GPI anchors serve as promising vaccine candidates[101]. Our atomistic GPI model could, therefore, be used to design more judicious and targeted experiments to synthesize GPI-based vaccines and drugs against several parasitic diseases, for instance, by helping arrive at the minimal epitope for recognition or deducing the conformation in various environments which could have particular implications in drug-designing.

Interactions between GPI and its attached protein GFP (w.r.t. the present study), have been elucidated in Chapters 2 and 3. We noted that the protein is pulled close to the membrane surface, irrespective of the GPI type, however the orientation of the protein is subject to fluctuation which could largely be attributed to the flexibility of the EtNP linker. The orientation, and stability thereof, also depend on other factors – (a) the type of GPI, (b) the region of the protein in contact with GPI, (c) the lipid bilayer. Interestingly, although the degree of branching is the same in *T.gondii LMW* and *T.congolense*, the orientation of GFP was strikingly different between the two. GFP fluctuated a great deal when attached to the former, but predominantly reclined on the bilayer (barrel axis perpendicular to bilayer normal) when attached to the latter. This suggests that the use of GPI analogues, instead of actual GPIs, in experiments can lead to faulty interpretations of actual GPI-based phenomena.. Irrespective of the GPI type, GPI and GFP were seen to make significant number of contacts throughout the simulation time. Cluster analysis revealed that no matter the orientation of the protein, GPIs consistently flop down on the membrane. Besides, the flexibility of GPIs was considerably reduced upon attachment of protein. The flop-down orientation of GPI is consistent with a few existing experimental studies. For instance, through a FRET study on GPI-anchored alkaline phosphatase, it was concluded that the protein lies very close membrane to the extent of resting on it[26]. Such an orientation is possible for the protein if the GPI is close to the membrane too. Thus, our atomistic GPI model rules out any notion of a general erect conformation of GPI that would prevent interactions between the protein and lipid layer, as was presented by Paulick and group[60, 102]. This absolutely opposite outcome could be a consequence of using GPI analogs in the study, instead of actual GPIs, which as was discussed before could make a significant difference to the conformation. All GPI systems were simulated in bilayers of DMPC and POPC to investigate the effect of the lipid bilayer. And indeed, GPIs were seen to be more buried/embedded in POPC than DMPC. The reason for this discrepancy is most possibly the larger area per lipid in POPC than DMPC that allows more space for the glycan to explore. GPIs are highly prone to hydrogen bonding due to the large number of hydroxyl groups in their structure. By extensively interacting (including H-bonding) with both the lipid headgroup region and the protein, it appears that GPIs mediate communication and so, assume a space-filling role, between the two. This observation is particularly relevant to *T.congolense* GPI because in nature it is found attached to the coat protein Variant Surface Glycoprotein on the cell surface of Trypanosomes. Previously, experiments and early computational studies suggested the space

filling property of the GPI that confers protection to the parasite's membrane[78, 103]. Using circular dichroism spectra and MD simulations, a similar picture of the space-filling nature of GPI was presented by Barboni and coworkers[24]. Our results are in line with these reports. The importance of the side branch residues is further enforced in that on the one hand, they make large number of contacts with the protein, and on the other, are still the most hydrated even in the GPI-anchored GFP systems. Although GFP is not a natural GPI-anchored protein, recombinant forms of GPI-anchored GFP are in prevalent use for experimental investigations[55, 58, 88] in which our model can play an important part.

Next, we developed a coarse-grained model of the GPI anchor and GPI-anchored GFP using a modified form of the MARTINI force-field in an aqueous medium of polarizable water, that is presented in Chapter 4. Polarizable water as the solvent was essential to the study as interfacial interactions are known to be better represented here than with the standard MARTINI water[104]. As our aim was to study the mutual interactions among carbohydrates, protein, lipid heads and water, polarizable water was the natural choice. The model is based on mapping four heavy atoms to 1 bead/super-atom. A top-down approach was adopted to derive bonded parameters from the atomistic system, and a bottom-down strategy was employed to define non-bonded potentials based on octanol-water partitioning free energies and second virial coefficients of osmotic pressure. The modification was necessary to overcome unrealistic aggregation in sugars and proteins from overestimated attractive forces, a problem that is persistently seen to occur in the original MARTINI force-field[105, 106]. The parameterization protocol began with simple sugars such as glucose, sucrose and trehalose, because experimental observables for these molecules are readily available, but not for GPI. The coarse-grained model of GPI was then designed in a building-block approach based on the parameters of simple sugars and MARTINI's interaction matrix[107] which is basically a chart to assign bead types to groups of atoms based on their polarity, hydrogen-bonding character and charge. We arrived at a scaling factor of $\gamma = 0.85$ to impose on the Lennard-Jones ϵ_{ij}s constituting sugar-sugar interactions by optimizing the second virial coefficient of osmotic pressure of simple sugar solutions w.r.t. experimental values, which was then applied to GPI-GPI interactions too. There was good match of structure between the atomistic and coarse-grained GPIs, and aggregation among multiple GPIs was also overridden. Interestingly, when the same scaling factor was applied to GPI-GFP interaction potentials, they ended up making similar frequency of contacts as in the atomistic. Prior to scaling down the interactions, coarse-grained GPI and GFP exhibited excessive mutual stickiness in comparison with their atomistic counterparts. The orientation of both the protein-free and protein-attached GPI w.r.t. the bilayer showed great similarities between the AA and CG systems, with the GPI flopped-down and interacting with the lipid heads at both the resolutions. However, although the orientation of coarse-grained GFP w.r.t the bilayer was quite similar with atomistic within the same time-frame, enhanced conformational sampling revealed that the erect presentation of the GFP is revisited and fluctuations in the orientation increase beyond the time-scale of the atomistic simulation. This unsteadiness in the orientation of GFP could be imparted by (a) flexibility of the EtNP linker, (b) lack of specific adhesive/attractive interactions between GFP and phosphocholine lipid bilayers[108]. In a control simulation of just the anchorless GFP placed on lipid surface, we applied a biased force for 300 ns to force

the protein to locally interact with the bilayer surface, and subsequently released the force to observe that GFP detached from the surface after about 500 ns. This observation points towards the importance of coarse-graining in exploring conformational spaces inaccessible to atomistic models. Note that lipid-lipid, sugar-lipid and protein-lipid potentials were retained from the original MARTINI as these are known to work well in replicating cellular processes such lipid-mediated protein clustering, protein-mediated lipid flip-flop, cryogenic effect of sugars on lipids, to name a few[69, 109]. With the coarse-graining, a speed-up of an order of magnitude in the computational runtime was achieved. This faster sampling now allows us to investigate the dynamic behavior of GPIs when placed in membranes, that which we could not with the atomistic model, in order to understand their implications in plasma membrane organization like how do GPIs affect the diffusion of surrounding membrane components, how do they partition in different domains on the membrane (liquid-ordered and liquid-disordered), do they play a role in trafficking of attached/surrounding proteins. There have been long-standing and still on-going controversies regarding the involvement of GPIs in forming functionally active membrane microdomains termed as "lipid rafts" and the role of GPIs in protein trafficking (described in Sec.1.5), that can be attempted to solve using such reductive approaches through CG models. These questions can be readily explored by combining GPIs and GPI-APs with existing membrane-associated biomolecules including cholesterol, gangliosides, sphingolipids and various membrane proteins of the MARTINI database[5]. In a preliminary attempt to understand the partitioning behaviour of GPIs in membrane domains, we plugged in lipid tails of varying saturation levels to the coarse-grained GPI glycan. We found that the lipid-tail saturation has a considerable role to play in the localization of GPIs in either liquid-ordered or liquid-disordered domains or the domain boundaries. This study is as yet in its nascent stage and needs deeper investigation, yet it exemplifies the potential of the coarse-grained model.

To conclude, we hope to have designed viable models, atomistic and coarse-grained, of the GPI anchor and GPI-anchored GFP which can be employed to study many interesting questions regarding the diverse cellular functions of the GPIs. Through a conformational analysis of the atomistic GPIs, we have attempted to quell the controversy regarding the orientation and presentation of GPIs on membranes, and ventured ahead to establish a structure-function connection for parasitic GPIs. We believe that our modular atomistic model qualifies for computational investigations of many other parasitic GPIs to aid in vaccine development and drug designing. Our coarse-grained model can be used to conveniently traverse conformational landscapes of atomistic structures after mapping, and then back-mapping post enhanced sampling to return back to structural details of the all-atom model. Such a combined approach of using both atomistic and coarse-grained models will significantly contribute to conformational analysis especially for flexible molecules like GPIs that need plenty of sampling for sufficient convergence of properties. Lastly, our coarse-grained GPI model in conjunction with the MARTINI lipidome, can be used to explore several membrane-related phenomena that require large systems and long runtimes, and thereby contribute to modelling a realistic plasma membrane.

Bibliography

[1] Fiorella Malchiodi-Albedi, Silvia Paradisi, Andrea Matteucci, Claudio Frank, and Marco Diociaiuti. Amyloid oligomer neurotoxicity, calcium dysregulation, and lipid rafts. *International journal of Alzheimer's disease*, 2011, 2011.

[2] Michael AJ Ferguson, Taroh Kinoshita, and Gerald W Hart. Glycosylphosphatidylinositol anchors. In *Essentials of Glycobiology. 2nd edition.* Cold Spring Harbor Laboratory Press, 2009.

[3] Chengqian Yuan, Shukun Li, Qianli Zou, Ying Ren, and Xuehai Yan. Multiscale simulations for understanding the evolution and mechanism of hierarchical peptide self-assembly. *Physical Chemistry Chemical Physics*, 19(35):23614–23631, 2017.

[4] Jad Abbass and Jean-Christophe Nebel. Reduced fragment diversity for alpha and alpha-beta protein structure prediction using rosetta. *Protein and peptide letters*, 24(3):215–222, 2017.

[5] Siewert J Marrink, Valentina Corradi, Paulo CT Souza, Helgi I Ingólfsson, D Peter Tieleman, and Mark SP Sansom. Computational modeling of realistic cell membranes. *Chemical reviews*, 119(9): 6184–6226, 2019.

[6] Ryo Taguchi and Hiroh Ikezawa. Phosphatidyl inositol-specific phospholipase c from clostridium novyi type a. *Archives of biochemistry and biophysics*, 186(1):196–201, 1978.

[7] Martin G Low and Donald B Zilversmit. Role of phosphatidylinositol in attachment of alkaline phosphatase to membranes. *Biochemistry*, 19(17):3913–3918, 1980.

[8] David G Campbell, J Gagnon, KB Reid, and Alan F Williams. Rat brain thy-1 glycoprotein. the amino acid sequence, disulphide bonds and an unusual hydrophobic region. *Biochemical Journal*, 195(1):15–30, 1981.

[9] Anthony A Holder and George AM Cross. Glycopeptides from variant surface glycoproteins of trypanosoma brucei. c-terminal location of antigenically cross-reacting carbohydrate moieties. *Molecular and biochemical parasitology*, 2(3-4):135–150, 1981.

[10] Anthony A Holder. Carbohydrate is linked through ethanolamine to the c-terminal amino acid of trypanosoma brucei variant surface glycoprotein. *Biochemical Journal*, 209(1):261–262, 1983.

[11] M Lucia Cardoso de Almeida and Mervyn J Turner. The membrane form of variant surface glycoproteins of trypanosoma brucei. *Nature*, 302(5906):349, 1983.

[12] MA Ferguson and GA Cross. Myristylation of the membrane form of a trypanosoma brucei variant surface glycoprotein. *Journal of Biological Chemistry*, 259(5):3011–3015, 1984.

[13] MA Ferguson, Kasturi Haldar, and GA Cross. Trypanosoma brucei variant surface glycoprotein has a sn-1, 2-dimyristyl glycerol membrane anchor at its cooh terminus. *Journal of Biological Chemistry*, 260(8):4963–4968, 1985.

[14] MAl Ferguson, Martin G Low, and GA Cross. Glycosyl-sn-1, 2-dimyristylphosphatidylinositol is covalently linked to trypanosoma brucei variant surface glycoprotein. *Journal of Biological Chemistry*, 260(27):14547–14555, 1985.

[15] Anthony H Futerman, Martin G Low, Karen E Ackermann, William R Sherman, and Israel Silman. Identification of covalently bound inositol in the hydrophobic membrane-anchoring domain of torpedo acetylcholinesterase. *Biochemical and biophysical research communications*, 129(1):312–317, 1985.

[16] AG Tse, A Neil Barclay, Anthony Watts, and Alan F Williams. A glycophospholipid tail at the carboxyl terminus of the thy-1 glycoprotein of neurons and thymocytes. *Science*, 230(4729): 1003–1008, 1985.

[17] William L Roberts and Terrone L Rosenberry. Identification of covalently attached fatty acids in the hydrophobic membrane-binding domain of human erythrocyte acetylcholinesterase. *Biochemical and biophysical research communications*, 133(2):621–627, 1985.

[18] MA Ferguson, Steve W Homans, Raymond A Dwek, and Thomas W Rademacher. Glycosylphosphatidylinositol moiety that anchors trypanosoma brucei variant surface glycoprotein to the membrane. *Science*, 239(4841):753–759, 1988.

[19] Yu-Hsuan Tsai, Xinyu Liu, and Peter H Seeberger. Chemical biology of glycosylphosphatidylinositol anchors. *Angewandte Chemie International Edition*, 51(46):11438–11456, 2012.

[20] Alexander Adibekian, Pierre Stallforth, Marie-Lyn Hecht, Daniel B Werz, Pascal Gagneux, and Peter H Seeberger. Comparative bioinformatics analysis of the mammalian and bacterial glycomes. *Chemical Science*, 2(2):337–344, 2011.

[21] Sue Moody-Haupt, John H Patterson, David Mirelman, and Malcolm J McConville. The major surface antigens of entamoeba histolytica trophozoites are gpi-anchored proteophosphoglycans. *Journal of molecular biology*, 297(2):409–420, 2000.

[22] Olivier Nosjean. Mammalian gpi proteins: sorting, membrane resistance and functions. *Biochim Biophys Acta*, 1331:153–186, 1997.

[23] PETER BUTIKOFER, TATIANA MALHERBE, MONIKA BOSCHUNG, and ISABEL RODITI. Gpi-anchored proteins: now you see'em, now you don't. *The FASEB journal*, 15(2):545–548, 2001.

[24] E Barboni, B Pliego Rivero, AJ George, SR Martin, DV Renoup, EF Hounsell, PC Barber, and RJ Morris. The glycophosphatidylinositol anchor affects the conformation of thy-1 protein. *Journal of cell science*, 108(2):487–497, 1995.

[25] TW Rademacher, CJ Edge, and RA Dwek. Dropping anchor with the lipophosphoglycans. *Current Biology*, 1(1):41–42, 1991.

[26] Marty T Lehto and Frances J Sharom. Proximity of the protein moiety of a gpi-anchored protein to the membrane surface: a fret study. *Biochemistry*, 41(26):8368–8376, 2002.

[27] Peter J Robinson, Margaret Millrain, Jane Antoniou, Elizabeth Simpson, and Andrew L Mellor. A glycophospholipid anchor is required for qa-2-mediated t cell activation. *Nature*, 342(6245):85, 1989.

[28] David R Jones and Isabel Varela-Nieto. The role of glycosyl–phosphatidylinositol in signal transduction. *The international journal of biochemistry & cell biology*, 30(3):313–326, 1998.

[29] Kai Simons and Derek Toomre. Lipid rafts and signal transduction. *Nature reviews Molecular cell biology*, 1(1):31, 2000.

[30] Rui Tan, Iryna Leshchyns'ka, and Vladimir Sytnyk. Glycosylphosphatidylinositol-anchored immunoglobulin superfamily cell adhesion molecules and their role in neuronal development and synapse regulation. *Frontiers in molecular neuroscience*, 10:378, 2017.

[31] Michael P Lisanti, Ingrid W Caras, Michael A Davitz, and Enrique Rodriguez-Boulan. A glycophospholipid membrane anchor acts as an apical targeting signal in polarized epithelial cells. *The Journal of cell biology*, 109(5):2145–2156, 1989.

[32] Deborah A Brown and John K Rose. Sorting of gpi-anchored proteins to glycolipid-enriched membrane subdomains during transport to the apical cell surface. *Cell*, 68(3):533–544, 1992.

[33] Chiara Zurzolo, Michael P Lisanti, Ingrid W Caras, Lucio Nitsch, and Enrique Rodriguez-Boulan. Glycosylphosphatidylinositol-anchored proteins are preferentially targeted to the basolateral surface in fischer rat thyroid epithelial cells. *The Journal of Cell Biology*, 121(5):1031–1039, 1993.

[34] Simona Paladino, Daniela Sarnataro, Rudolf Pillich, Simona Tivodar, Lucio Nitsch, and Chiara Zurzolo. Protein oligomerization modulates raft partitioning and apical sorting of gpi-anchored proteins. *J Cell Biol*, 167(4):699–709, 2004.

[35] Sebastian Schuck and Kai Simons. Polarized sorting in epithelial cells: raft clustering and the biogenesis of the apical membrane. *Journal of cell science*, 117(25):5955–5964, 2004.

[36] David R Taylor and Nigel M Hooper. Gpi-anchored proteins in health and disease. In *Post-translational modifications in health and disease*, pages 39–55. Springer, 2011.

[37] Françoise Debierre-Grockiego, Nahid Azzouz, Jörg Schmidt, Jean-François Dubremetz, Hildegard Geyer, Rudolf Geyer, Ralf Weingart, Richard R Schmidt, and Ralph T Schwarz. Roles of glycosylphosphatidylinositols of toxoplasma gondii induction of tumor necrosis factor-α production in macrophages. *Journal of Biological Chemistry*, 278(35):32987–32993, 2003.

[38] Françoise Debierre-Grockiego, Marco A Campos, Nahid Azzouz, Jörg Schmidt, Ulrike Bieker, Marianne Garcia Resende, Daniel Santos Mansur, Ralf Weingart, Richard R Schmidt, Douglas T Golenbock, et al. Activation of tlr2 and tlr4 by glycosylphosphatidylinositols derived from toxoplasma gondii. *The journal of immunology*, 179(2):1129–1137, 2007.

[39] Françoise Debierre-Grockiego and Ralph T Schwarz. Immunological reactions in response to apicomplexan glycosylphosphatidylinositols. *Glycobiology*, 20(7):801–811, 2010.

[40] Igor C Almeida, Maristela M Camargo, Daniela O Procópio, Luiz S Silva, Angela Mehlert, Luiz R Travassos, Ricardo T Gazzinelli, and Michael AJ Ferguson. Highly purified glycosylphosphatidylinositols from trypanosoma cruzi are potent proinflammatory agents. *The EMBO Journal*, 19(7): 1476–1485, 2000.

[41] Shobhona Sharma and Sulabha Pathak. Malaria vaccine: a current perspective. *Journal of vector borne diseases*, 45(1):1, 2008.

[42] Louis Schofield, Michael C Hewitt, Krystal Evans, Mary-Anne Siomos, and Peter H Seeberger. Synthetic gpi as a candidate anti-toxic vaccine in a model of malaria. *Nature*, 418(6899):785, 2002.

[43] Faustin Kamena, Marco Tamborrini, Xinyu Liu, Yong-Uk Kwon, Fiona Thompson, Gerd Pluschke, and Peter H Seeberger. Synthetic gpi array to study antitoxic malaria response. *Nature chemical biology*, 4(4):238, 2008.

[44] Sebastian Götze, Nahid Azzouz, Yu-Hsuan Tsai, Uwe Groß, Anika Reinhardt, Chakkumkal Anish, Peter H Seeberger, and Daniel Varón Silva. Diagnosis of toxoplasmosis using a synthetic glycosylphosphatidylinositol glycan. *Angewandte Chemie International Edition*, 53(50):13701–13705, 2014.

[45] Stefan Magez, Benoît Stijlemans, Toya Baral, and Patrick De Baetselier. Vsg-gpi anchors of african trypanosomes: their role in macrophage activation and induction of infection-associated immunopathology. *Microbes and Infection*, 4(9):999–1006, 2002.

[46] S Jonathan Singer and Garth L Nicolson. The fluid mosaic model of the structure of cell membranes. *Science*, 175(4023):720–731, 1972.

[47] Elishalom Yechiel and Michael Edidin. Micrometer-scale domains in fibroblast plasma membranes. *The Journal of cell biology*, 105(2):755–760, 1987.

[48] Kai Simons and Gerrit Van Meer. Lipid sorting in epithelial cells. *Biochemistry*, 27(17):6197–6202, 1988.

[49] Dirk P Cerneus, Elles Ueffing, George Posthuma, GJ Strous, and A Van der Ende. Detergent insolubility of alkaline phosphatase during biosynthetic transport and endocytosis. role of cholesterol. *Journal of Biological Chemistry*, 268(5):3150–3155, 1993.

[50] Kai Simons and Elina Ikonen. Functional rafts in cell membranes. *Nature*, 387(6633):569, 1997.

[51] Satyajit Mayor, Karen G Rothberg, and Frederick R Maxfield. Sequestration of gpi-anchored proteins in caveolae triggered by cross-linking. *Science*, 264(5167):1948–1951, 1994.

[52] Satyajit Mayor and Frederick R Maxfield. Insolubility and redistribution of gpi-anchored proteins at the cell surface after detergent treatment. *Molecular biology of the cell*, 6(7):929–944, 1995.

[53] H Heerklotz. Triton promotes domain formation in lipid raft mixtures. *Biophysical journal*, 83(5): 2693–2701, 2002.

[54] Heiko Heerklotz, Halina Szadkowska, Thomas Anderson, and Joachim Seelig. The sensitivity of lipid domains to small perturbations demonstrated by the effect of triton. *Journal of molecular biology*, 329(4):793–799, 2003.

[55] Pranav Sharma, Rajat Varma, RC Sarasij, Karine Gousset, G Krishnamoorthy, Madan Rao, Satyajit Mayor, et al. Nanoscale organization of multiple gpi-anchored proteins in living cell membranes. *Cell*, 116(4):577–589, 2004.

[56] Debanjan Goswami, Kripa Gowrishankar, Sameera Bilgrami, Subhasri Ghosh, Riya Raghupathy, Rahul Chadda, Ram Vishwakarma, Madan Rao, and Satyajit Mayor. Nanoclusters of gpi-anchored proteins are formed by cortical actin-driven activity. *Cell*, 135(6):1085–1097, 2008.

[57] Thomas S van Zanten, Alessandra Cambi, Marjolein Koopman, Ben Joosten, Carl G Figdor, and Maria F Garcia-Parajo. Hotspots of gpi-anchored proteins and integrin nanoclusters function as nucleation sites for cell adhesion. *Proceedings of the National Academy of Sciences*, 106(44): 18557–18562, 2009.

[58] Eva Sevcsik, Mario Brameshuber, Martin Fölser, Julian Weghuber, Alf Honigmann, and Gerhard J Schütz. Gpi-anchored proteins do not reside in ordered domains in the live cell plasma membrane. *Nature communications*, 6:6969, 2015.

[59] Margot G Paulick and Carolyn R Bertozzi. The glycosylphosphatidylinositol anchor: a complex membrane-anchoring structure for proteins. *Biochemistry*, 47(27):6991–7000, 2008.

[60] Margot G Paulick, Amber R Wise, Martin B Forstner, Jay T Groves, and Carolyn R Bertozzi. Synthetic analogues of glycosylphosphatidylinositol-anchored proteins and their behavior in supported lipid bilayers. *Journal of the American Chemical Society*, 129(37):11543–11550, 2007.

[61] Ron O Dror, Robert M Dirks, JP Grossman, Huafeng Xu, and David E Shaw. Biomolecular simulation: a computational microscope for molecular biology. *Annual review of biophysics*, 41: 429–452, 2012.

[62] Matthieu Chavent, Anna L Duncan, and Mark SP Sansom. Molecular dynamics simulations of membrane proteins and their interactions: from nanoscale to mesoscale. *Current opinion in structural biology*, 40:8–16, 2016.

[63] Siewert J Marrink, H Jelger Risselada, Serge Yefimov, D Peter Tieleman, and Alex H De Vries. The martini force field: coarse grained model for biomolecular simulations. *The journal of physical chemistry B*, 111(27):7812–7824, 2007.

[64] Siewert-Jan Marrink and Alan E Mark. Molecular view of hexagonal phase formation in phospholipid membranes. *Biophysical journal*, 87(6):3894–3900, 2004.

[65] Roland Faller and Siewert-Jan Marrink. Simulation of domain formation in dlpc- dspc mixed bilayers. *Langmuir*, 20(18):7686–7693, 2004.

[66] Siewert J Marrink, Jelger Risselada, and Alan E Mark. Simulation of gel phase formation and melting in lipid bilayers using a coarse grained model. *Chemistry and physics of lipids*, 135(2): 223–244, 2005.

[67] WF Drew Bennett, Justin L MacCallum, Marlon J Hinner, Siewert J Marrink, and D Peter Tieleman. Molecular view of cholesterol flip-flop and chemical potential in different membrane environments. *Journal of the American Chemical Society*, 131(35):12714–12720, 2009.

[68] Xavier Periole, Adam M Knepp, Thomas P Sakmar, Siewert J Marrink, and Thomas Huber. Structural determinants of the supramolecular organization of g protein-coupled receptors in bilayers. *Journal of the American Chemical Society*, 134(26):10959–10965, 2012.

[69] Valentina Corradi, Besian I Sejdiu, Haydee Mesa-Galloso, Haleh Abdizadeh, Sergei Yu Noskov, Siewert J Marrink, and D Peter Tieleman. Emerging diversity in lipid–protein interactions. *Chemical reviews*, 119(9):5775–5848, 2019.

[70] Sergei Izvekov and Gregory A Voth. A multiscale coarse-graining method for biomolecular systems. *The Journal of Physical Chemistry B*, 109(7):2469–2473, 2005.

[71] Jian Zhou, Ian F Thorpe, Sergey Izvekov, and Gregory A Voth. Coarse-grained peptide modeling using a systematic multiscale approach. *Biophysical journal*, 92(12):4289–4303, 2007.

[72] Sergei Izvekov and Gregory A Voth. Multiscale coarse-graining of mixed phospholipid/cholesterol bilayers. *Journal of chemical theory and computation*, 2(3):637–648, 2006.

[73] Pu Liu, Sergei Izvekov, and Gregory A Voth. Multiscale coarse-graining of monosaccharides. *The Journal of Physical Chemistry B*, 111(39):11566–11575, 2007.

[74] Yanting Wang, Sergei Izvekov, Tianying Yan, and Gregory A Voth. Multiscale coarse-graining of ionic liquids. *The Journal of Physical Chemistry B*, 110(8):3564–3575, 2006.

[75] Christine Peter, Luigi Delle Site, and Kurt Kremer. Classical simulations from the atomistic to the mesoscale and back: coarse graining an azobenzene liquid crystal. *Soft Matter*, 4(4):859–869, 2008.

[76] Khongvit Prasitnok and Mark R Wilson. A coarse-grained model for polyethylene glycol in bulk water and at a water/air interface. *Physical Chemistry Chemical Physics*, 15(40):17093–17104, 2013.

[77] Alexander P Lyubartsev and Aatto Laaksonen. Calculation of effective interaction potentials from radial distribution functions: A reverse monte carlo approach. *Physical Review E*, 52(4):3730, 1995.

[78] SW Homans, CJ Edge, MAJ Ferguson, RA Dwek, and TW Rademacher. Solution structure of the glycosylphosphatidylinositol membrane anchor glycan of trypanosoma brucei variant surface glycoprotein. *Biochemistry*, 28(7):2881–2887, 1989.

[79] Johannes Zuegg and Jill E Gready. Molecular dynamics simulation of human prion protein including both n-linked oligosaccharides and the gpi anchor. *Glycobiology*, 10(10):959–974, 2000.

[80] Franck Chevalier, Javier Lopez-Prados, Serge Perez, Manuel Martín-Lomas, and Pedro M Nieto. Conformational study of gpi anchors: the common oligosaccharide gpi anchor backbone. *European journal of organic chemistry*, 2005(16):3489–3498, 2005.

[81] Franck Chevalier, Javier Lopez-Prados, Patrick Groves, Serge Perez, Manuel Martín-Lomas, and Pedro M Nieto. Structure and dynamics of the conserved protein gpi anchor core inserted into detergent micelles. *Glycobiology*, 16(10):969–980, 2006.

[82] Mari L DeMarco and Valerie Daggett. Characterization of cell-surface prion protein relative to its recombinant analogue: insights from molecular dynamics simulations of diglycosylated, membrane-bound human prion protein. *Journal of neurochemistry*, 109(1):60–73, 2009.

[83] Michael Levitt, Miriam Hirshberg, Ruth Sharon, and Valerie Daggett. Potential energy function and parameters for simulations of the molecular dynamics of proteins and nucleic acids in solution. *Computer physics communications*, 91(1):215–231, 1995.

[84] Chris Oostenbrink, Alessandra Villa, Alan E Mark, and Wilfred F Van Gunsteren. A biomolecular force field based on the free enthalpy of hydration and solvation: the gromos force-field parameter sets 53a5 and 53a6. *Journal of computational chemistry*, 25(13):1656–1676, 2004.

[85] Carla G Chiodi and Hugo Verli. Structural characterization of netnes glycopeptide from trypanosoma cruzi. *Carbohydrate research*, 373:28–34, 2013.

[86]

[87] Callum J Dickson, Benjamin D Madej, Åge A Skjevik, Robin M Betz, Knut Teigen, Ian R Gould, and Ross C Walker. Lipid14: the amber lipid force field. *Journal of chemical theory and computation*, 10(2):865–879, 2014.

[88] Ian A Prior, Cornelia Muncke, Robert G Parton, and John F Hancock. Direct visualization of ras proteins in spatially distinct cell surface microdomains. *The Journal of cell biology*, 160(2): 165–170, 2003.

[89] James A Maier, Carmenza Martinez, Koushik Kasavajhala, Lauren Wickstrom, Kevin E Hauser, and Carlos Simmerling. ff14sb: improving the accuracy of protein side chain and backbone parameters from ff99sb. *Journal of chemical theory and computation*, 11(8):3696–3713, 2015.

[90] Karl N Kirschner, Austin B Yongye, Sarah M Tschampel, Jorge González-Outeiriño, Charlisa R Daniels, B Lachele Foley, and Robert J Woods. Glycam06: a generalizable biomolecular force field. carbohydrates. *Journal of computational chemistry*, 29(4):622–655, 2008.

[91] Carlos J Bosques, Sarah M Tschampel, Robert J Woods, and Barbara Imperiali. Effects of glycosylation on peptide conformation: a synergistic experimental and computational study. *Journal of the American Chemical Society*, 126(27):8421–8425, 2004.

[92] Mari L DeMarco and Robert J Woods. Atomic-resolution conformational analysis of the gm3 ganglioside in a lipid bilayer and its implications for ganglioside–protein recognition at membrane surfaces. *Glycobiology*, 19(4):344–355, 2008.

[93] Marko Wehle, Ivan Vilotijevic, Reinhard Lipowsky, Peter H Seeberger, Daniel Varon Silva, and Mark Santer. Mechanical compressibility of the glycosylphosphatidylinositol (gpi) anchor backbone governed by independent glycosidic linkages. *Journal of the American Chemical Society*, 134(46): 18964–18972, 2012.

[94] Brian A Camley, Michael G Lerner, Richard W Pastor, and Frank LH Brown. Strong influence of periodic boundary conditions on lateral diffusion in lipid bilayer membranes. *The Journal of chemical physics*, 143(24):12B604_1, 2015.

[95] Danielle E Chandler, Johan Strümpfer, Melih Sener, Simon Scheuring, and Klaus Schulten. Light harvesting by lamellar chromatophores in rhodospirillum photometricum. *Biophysical journal*, 106 (11):2503–2510, 2014.

[96] Tsjerk A Wassenaar, Kristyna Pluhackova, Rainer A Böckmann, Siewert J Marrink, and D Peter Tieleman. Going backward: a flexible geometric approach to reverse transformation from coarse grained to atomistic models. *Journal of chemical theory and computation*, 10(2):676–690, 2014.

[97] Varma Saikam, Riya Raghupathy, Mahipal Yadav, Veeranjaneyulu Gannedi, Parvinder Pal Singh, Naveed A Qazi, Sanghapal D Sawant, and Ram A Vishwakarma. Synthesis of new fluorescently labeled glycosylphosphatidylinositol (gpi) anchors. *Tetrahedron letters*, 52(33):4277–4279, 2011.

[98] Barry J Hardy. The glycosidic linkage flexibility and time-scale similarity hypotheses. *Journal of Molecular Structure: THEOCHEM*, 395:187–200, 1997.

[99] Nahid Azzouz, Hosam Shams-Eldin, Sebastian Niehus, Françoise Debierre-Grockiego, Ulrike Bieker, Jörg Schmidt, Corinne Mercier, Marie-France Delauw, Jean-François Dubremetz, Terry K Smith, et al. Toxoplasma gondii grown in human cells uses galnac-containing glycosylphosphatidylinositol precursors to anchor surface antigens while the immunogenic glc–galnac-containing precursors remain free at the parasite cell surface. *The international journal of biochemistry & cell biology*, 38(11):1914–1925, 2006.

[100] Saulius Vainauskas and Anant K Menon. Ethanolamine phosphate linked to the first mannose residue of glycosylphosphatidylinositol (gpi) lipids is a major feature of the gpi structure that is recognized by human gpi transamidase. *Journal of Biological Chemistry*, 281(50):38358–38364, 2006.

[101] Sebastian Götze, Anika Reinhardt, Andreas Geissner, Nahid Azzouz, Yu-Hsuan Tsai, Reka Kurucz, Daniel Varón Silva, and Peter H Seeberger. Investigation of the protective properties of glycosylphosphatidylinositol-based vaccine candidates in a toxoplasma gondii mouse challenge model. *Glycobiology*, 25(9):984–991, 2015.

[102] Margot G Paulick, Martin B Forstner, Jay T Groves, and Carolyn R Bertozzi. A chemical approach to unraveling the biological function of the glycosylphosphatidylinositol anchor. *Proceedings of the National Academy of Sciences*, 104(51):20332–20337, 2007.

[103] Nicola G Jones, Daniel Nietlispach, Reuben Sharma, David F Burke, Isobel Eyres, Marsilius Mues, Helen R Mott, and Mark Carrington. Structure of a glycosylphosphatidylinositol-anchored domain from a trypanosome variant surface glycoprotein. *Journal of Biological Chemistry*, 283(6): 3584–3593, 2008.

[104] Kristyna Pluhackova, Tsjerk A Wassenaar, Sonja Kirsch, and Rainer A Böckmann. Spontaneous adsorption of coiled-coil model peptides k and e to a mixed lipid bilayer. *The Journal of Physical Chemistry B*, 119(12):4396–4408, 2015.

[105] Philipp S Schmalhorst, Felix Deluweit, Roger Scherrers, Carl-Philipp Heisenberg, and Mateusz Sikora. Overcoming the limitations of the martini force field in simulations of polysaccharides. *Journal of chemical theory and computation*, 13(10):5039–5053, 2017.

[106] Austin C Stark, Casey T Andrews, and Adrian H Elcock. Toward optimized potential functions for protein–protein interactions in aqueous solutions: osmotic second virial coefficient calculations using the martini coarse-grained force field. *Journal of chemical theory and computation*, 9(9): 4176–4185, 2013.

[107] Siewert J Marrink and D Peter Tieleman. Perspective on the martini model. *Chemical Society Reviews*, 42(16):6801–6822, 2013.

[108] Hyunjin Kim, Hamid Samareh Afsari, and Wonhwa Cho. High-throughput fluorescence assay for membrane-protein interaction. *Journal of lipid research*, 54(12):3531–3538, 2013.

[109] Cesar A López, Andrzej J Rzepiela, Alex H De Vries, Lubbert Dijkhuizen, Philippe H Hünenberger, and Siewert J Marrink. Martini coarse-grained force field: extension to carbohydrates. *Journal of Chemical Theory and Computation*, 5(12):3195–3210, 2009.