Why GPCRs behave differently in cubic and lamellar lipidic mesophases

George Khelashvili,*,[†] Pedro Blecua Carrillo Albornoz,[†] Niklaus Johner,[†] Sayan Mondal,[†] Martin Caffrey,[‡] and Harel Weinstein[†,§]

[†]Department of Physiology and Biophysics, Weill Cornell Medical College of Cornell University, New York, New York 10065, United States

[§]The HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, Weill Cornell Medical College, Cornell University, New York, New York 10065, United States

[‡]Membrane Structural and Functional Biology Group, School of Biochemistry and Immunology, and School of Medicine, Trinity College, Dublin, Ireland

S *Supporting Information*

ABSTRACT: Recent successes in the crystallographic determination of structures of transmembrane proteins in the G protein-coupled receptor (GPCR) family have established the lipidic cubic phase (LCP) environment as the medium of choice for growing structure-grade crystals by the method termed "in meso". The understanding of *in meso* crystallogenesis is currently at a descriptive level. To enable an eventual quantitative, energy-based description of the nucleation and crystallization mechanism, we have examined the properties of the lipidic cubic phase system and the dynamics of the GPCR rhodopsin reconstituted into the LCP with coarse-grained molecular dynamics simulations with the Martini force-field. Quantifying the differences in the hydrophobic/hydrophilic exposure of the GPCR to lipids in the cubic and lamellar phases, we found that the highly curved geometry of the cubic phase provides more efficient shielding of the protein from unfavorable hydrophobic exposure, which leads to a lesser hydrophobic mismatch and less unfavorable hydrophobic−hydrophilic interactions between the protein and lipid−water interface in the LCP, compared to the lamellar phase. Since hydrophobic mismatch is considered a driving force for oligomerization, the differences in exposure mismatch energies between the LCP and the lamellar structures suggest that the latter provide a more favorable setting in which GPCRs can oligomerize as a prelude to nucleation and crystal growth. These new findings lay the foundation for future investigations of *in meso* crystallization mechanisms related to the transition from the LCP to the lamellar phase and studies aimed at an improved rational approach for generating structure-quality crystals of membrane proteins.
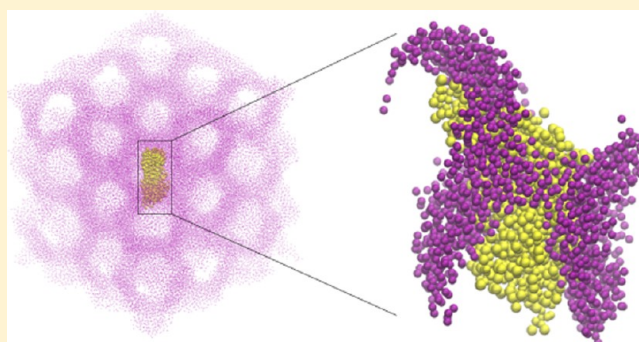
■ INTRODUCTION

Rapid advances in the crystallographic determination of transmembrane (TM) protein structures, and especially of G-protein coupled receptors (GPCRs), have established the lipidic cubic phase (LCP) as a suitable medium for growing high quality crystals of these proteins (see ref 1, and citations therein). The LCP consists of a highly curved lipid bilayer structure (see Figure 1B−C) that is continuous in three dimensions (3D) and separates in space two nonintersecting intertwined aqueous channels.[3−5] Because of the continuous nature of both the lipid and the water compartments, LCP has been described as a bicontinuous mesophase with its bilayer midplane tracing a triply periodic minimal surface (TPMS, characterized by zero mean curvature) (see refs 6−8 and citations therein). The particular kind of LCP utilized during *in meso* crystallization trials is of *Pn3m* symmetry type, where each

aqueous network of the LCP assumes a tetrahedral geometry (see Figure 2D−E).[5]

The rapid pace of new structural data acquisition[9] using the features of the *in meso* method underscores the usefulness and success of the technology. Thus, structural information obtained in this manner for several rhodopsin-like GPCRs (reviewed in ref 6) and a GPCR-G protein complex,[10] provided breakthrough insights about the structural basis for signal transduction through these 7-TM helical proteins. However, the present lack of a mechanistic understanding, at the molecular level, of the events that lead to *in meso* crystallization of membrane proteins, makes the success of the method somewhat enigmatic and results in the need for extensive trials to identify specific conditions, that is, the host and additive
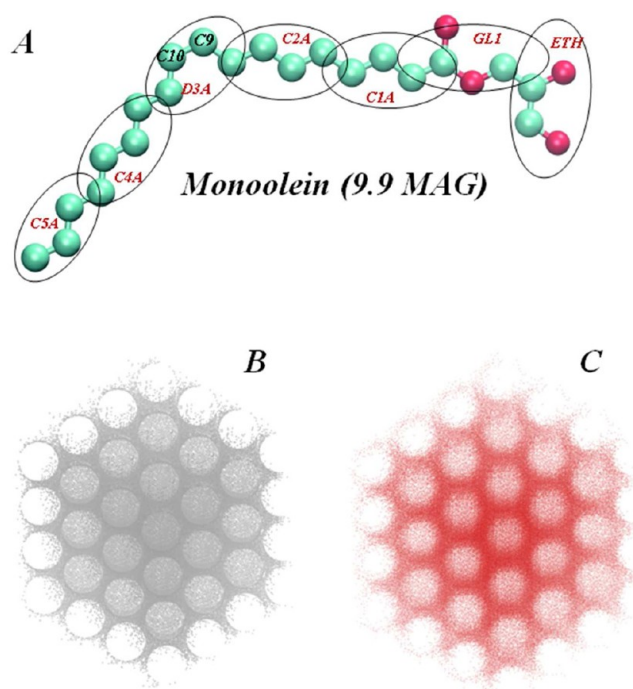
Figure 1. (A) Ball-and-stick representation of the Monoolein (9.9 MAG) lipid. Oxygen atoms are shown in red, carbon atoms are in cyan. Double bonded carbon atoms (C9 and C10) in the hydrocarbon chain of 9.9 MAG are highlighted. Hydrogens are omitted. The grouping of atoms in Martini coarse-grained beads (ETH, GL1, C1A, C2A, D3A, C4A, and C5A) is highlighted by ellipsoids. (B, C) Comparison of the triply periodic minimal surface (TPMS) of *Pn3m* symmetry calculated from the improved nodal approximation given in eqs 1−3 (B), to a snapshot after 6 *μ*s-long coarse-grained MD simulations of spontaneous self-assembly of 9.9 MAG/water complex at 40% (w/w) water concentration and at 20 °C temperature (C). The figure in C (see also Figures 2 and 3) shows the location of the C5A coarse-grained beads from the 9.9 MAG lipid (see A); note the cubic phase structure. The similarity evident in B and C underscores the close relation between the organization of the C5A beads (representing the density of terminal methylene and methyl groups in 9.9 MAG lipids) and the TPMS of the *Pn3m* cubic phase.

lipid, precipitant etc., that would yield high quality samples for crystallographic analysis.[5,11−18]

Such *in meso* trials begin with the target protein being reconstituted into the LCP, and a hypothetical descriptive mechanism based on the membrane curvature and hydrophobic mismatch has been proposed[13,19,20] to address the crucial step during *in meso* crystallization—the preferential partitioning of proteins from the LCP to the lamellar phase. In the original trials, the lipid used to form the LCP was monoolein (9.9 MAG, see Figure 1A), however recent experiments have utilized shorter chain analogs of monoolein as well as mixtures of monoolein with cholesterol (see refs 12, 18, and citations therein). The reconstituted proteins appear to be well-accommodated, functional, and mobile inside the LCP where they are believed to reside largely in a monomeric state.

The addition of a precipitant of suitable composition, which is the next step in the *in meso* trial, is proposed to result in local structural and compositional changes in the mesophase and likely in protein conformation as well,[8,13,19] which alter the interactions between the proteins and the cubic phase lipid bilayer; ultimately this drives the preferential partitioning of the proteins into the lamellar phase that presumably appears within

the LCP.[19] The critical aspect of the nucleation and subsequent crystal growth process that is triggered in this manner, and ultimately determines the quality of the grown crystals, is that once they have diffused from the LCP to the lamellar phase the proteins form tight 3D oligomeric arrays.[19]

The oligomerization of TM proteins in lipid bilayers has generally been suggested to involve the hydrophobic mismatch effects originating from the difference in the lengths of the hydrophobic core of the protein and the surrounding lipid hydrocarbon region.[22−26] Such mismatch effects have been shown from experiments to play a role not only in the spatial organization but also in the function of TM proteins, such as the GPCR Rhodopsin.[27,28] The hydrophobic mismatch is considered to drive the membrane to deform in order to alleviate the energetically unfavorable hydrophilic/hydrophobic exposure this mismatch entails. When such membrane deformation cannot achieve a complete hydrophobic adaptation, the "residual exposure" (or "residual mismatch") to unfavorable interactions gives rise to an unfavorable contribution to the free energy of protein/membrane interaction.[29] The mitigation of this energy term through oligomerization has been suggested as an important mechanistic contribution to membrane-driven oligomerization of multihelical TM proteins, such as GPCRs, in lipid bilayers.[28,30] Therefore, we investigated the potential role of differential hydrophobic exposure of such proteins in lipid bilayers of the cubic and lamellar phases as a possible determinant of the favorable outcome of *in meso* crystallization trials.

Here we present the results of a molecular dynamics-based study of a GPCR reconstituted into an LCP environment that reveal differences in the hydrophobic/hydrophilic exposure of the GPCR to lipid−water interface in the cubic and lamellar phases. The quantitative results reflect the drive for differential oligomerization behavior of these proteins in the two lipid environments that essentially define the end points along the *in meso* crystallization pathway. With the ability to connect quantitatively the difference in GPCR oligomerization behavior in the two different media, to the curvature and mismatch alleviation capabilities of the two environments, we are able to predict lipid types that would enhance/diminish the difference in the residual exposure for a GPCR in the cubic and lamellar phases.

Given the recognized importance of protein−membrane interaction for the functional properties of GPCRs, molecular dynamics simulation studies using all-atom[31,32] and coarse-grained representations[26] have been carried out for the prototypical GPCR, Rhodopsin, in lamellar membranes. To our knowledge the present work is the first molecular dynamics-based study of a GPCR reconstituted into the LCP. The results enable quantitative inferences about the energetically unfavorable hydrophobic−hydrophilic interactions that differentiate the behaviors of GPCRs in the two different lipid environments, LCP and lamellar membranes. We attain these insights from comparative microsecond time-scale coarse-grained simulations with the Martini force-field[33,34] of rhodopsin in monoolein-based cubic and lamellar lipid bilayers. Using our recently reported computational framework for quantifying residual mismatch energies from MD trajectories,[30] we evaluated the protein-dependent membrane deformations and their attendant energy cost, and identified specific TM regions and residues of rhodopsin that exhibit differential hydrophobic exposure in the LCP and lamellar bilayers. We show that the bilayer of the LCP more efficiently shields the

Table 1. Simulations Performed in Lipid Cubic phase (LCP)

number of monoolein lipids	water content (w/w %)	number of rhodopsin molecules	temperature,°C	number of simulations	duration of simulations[a]	simulation type
500	40	0	20	3	6 μs	self-assembly
13374	40	1[b]	20	1	4.8 μs	preformed LCP[c]

[a]Simulation times reported are effective times, taking into consideration typical for Martini force-field based coarse-grained simulations factor of 4. [b]Simulations with protein also included counterions to ensure electroneutrality of the system. [c]Simulations with protein were initiated from self-assembled lipid diamond cubic phase of 500 lipids by replicating the Monoolein/water system 27 times (3 times in positive and negative x, y, z directions) and inserting rhodopsin randomly into the expanded LCP phase.

protein from unfavorable interactions, and that the reduced level of hydrophobic mismatch in the LCP is attributable to the specific, highly curved geometry of the cubic phase. Better protection from unfavorable hydrophobic exposure of rhodopsin in the LCP phase is especially evident for residues in TM helix 1 (TMH1) and TMH5. The differences suggest that compared to the "protective" LCP environment, the planar bilayer of the lamellar phase provides a more favorable setting for GPCR oligomerization as a prelude to nucleation and crystal growth. Thus, our findings provide novel energy-based insights into driving elements of *in meso* crystallization mechanisms and lay the foundation for future quantitative exploration of rational approaches for the generation of structure-quality crystals of membrane proteins.

■ THEORY AND METHODS

Molecular Constructs and Coarse-grained Molecular Dynamics (MD) Simulations. Coarse-grained (CG) MD simulations of the lipid cubic phases (LCP) with or without rhodopsin were done with the Martini force field,[33,34] grouping atoms into coarse-grained beads (see Figure 1A), and the Gromacs 3.3.1 package[35] as listed for the various molecular constructs in Table 1. As a host lipid for the LCP, we used monoolein (9.9 MAG) which consists of an 18 carbon fatty acid with a *cis* double bond between carbons 9 and 10 in ester linkage to the primary hydroxyl of glycerol (Figure 1A).

Self-assembly Simulations of a Lipidic Diamond Cubic Phase. In the first phase of the study, we conducted extensive self-assembly simulations of randomly mixed CG 9.9 MAG lipids and water molecules into a lipidic diamond cubic phase. As listed in Table 1 and detailed in the Results, the diamond lipid phase was successfully achieved in multiple simulations conducted at 20 °C temperature and 40% (w/w) water weight percent.

All the self-assembly simulations (see Table 1) were initiated from a random placement of 500 CG 9.9 MAGs (a number that is within the experimentally determined range of 9.9 MAG molecules in the unit cell of the diamond cubic phase), and a corresponding number (determined by the desired percentage) of CG water beads in a cubic box. Each system was first equilibrated for a short period of time with the isotropic pressure coupling scheme and the Lennard-Jones (LJ) parameters of all the atoms set to those of the beads representing water molecules in the Martini force-field. This setup, implemented as well in earlier CG self-assembly simulations of the LCP,[36] allowed efficient adaptation of the volume of the simulation box to the number of CG molecules in the system, and ensured complete mixing of lipids and waters prior to self-assembly simulation.

After this initial equilibration phase, the LJ parameters for 9.9 MAG lipids were reset to their proper values (see details of force-field parametrization, below), and 6 μs-long self-assembly MD simulations were performed (the simulation times reported throughout are effective times,[34] taking into consideration the factor of 4 typical for Martini force-field-based coarse-grained simulations), and using isotropic pressure coupling (with 3.0 ps and 3e^{-5} bar^{-1} time constant and compressibility, respectively), and a 30 fs time step.

Simulations of Rhodopsin in Lipid Diamond Cubic Phase. To simulate rhodopsin in the diamond phase, we constructed the system by replicating 27 times (along positive and negative x, y, z

directions) one of the cubic phase structures obtained from the self-assembly simulations conducted at 20 °C and at 40% (w/w) water. The diamond cubic phase in this enlarged construct remained stable in a microsecond-long CG MD simulation (data not shown). A rhodopsin molecule was inserted at a random position in this large cubic phase (at ~122:1 9.9MAG/rhodopsin molar ratio, corresponding to 4 mg/mL concentration, see Figure S1 in the Supporting Information). After removing overlapping waters and 9.9 MAG molecules and adding counterions for electroneutrality, a 4.8 μs-long CG MD simulation was carried out at 20 °C. The total number of CG beads in this large construct was 139 979, which corresponds to ~1 400 000 actual atoms. The simulation utilized the isotropic pressure coupling scheme as above, and a 30 fs time step.

Simulations of Rhodopsin in Lamellar Bilayers. Rhodopsin was simulated in 9.9 MAG lipid lamellar bilayers with the CG Martini force-field. The CG model of rhodopsin (see below) was inserted into a pre-equilibrated CG 9.9 MAG lipid membrane containing 1772 lipids, and after solvation and ionization steps, a 3.2 μs-long MD simulation was carried out at 20 °C temperature using a 40 fs integration time step and the semi-isotropic pressure coupling scheme.

Force-field Parameters. The details of the Martini CG force fields and the parametrization scheme employed for 9.9 MAG lipid and Rhodopsin can be found in the Supporting Methods section of Supporting Information.

Quantification of the Residual Exposure of Rhodopsin Residues in Lipid Bilayers. The residual exposure energy is the energy cost of the hydrophobic mismatch that persists in the equilibrium state of the membrane-protein interaction, even after membrane remodeling around multi-TM proteins has taken place to reduce the hydrophobic mismatch.[27,29,30] This result of the anisotropy of the protein−membrane boundary when multi-TM proteins are inserted in the bilayer has been discussed in detail.[29,30] For rhodopsin, the residual exposure of residues in the TM segments was quantified with the computational protocol described in ref 30 in both the lipid cubic and lamellar phases, using solvent accessible surface area (SASA) calculations with the NACCESS software[37] and a probe radius of 1.4 Å. As described previously,[30] SASA values were obtained either with the solute comprising the protein only (SA$_{prot}$), or with the solute containing the protein and the hydrophobic core of the lipid bilayer (SA$_{mem}$)—the latter being defined as the 9.9 MAG bilayer region consisting of the GL1-C1A-C2A-D3A-C4A-C5A CG beads (see above for the CG bead definition, depicted in Figure 1A). The residual exposure SA$_{res}$ was then calculated as SA$_{res}$ = SA$_{prot}$ − SA$_{mem}$ for polar residues and as SA$_{res}$ = SA$_{mem}$ for hydrophobic residues (see ref 30 for details).

These accessible surface values were used to obtain the residual mismatch energy penalty ΔG_{res} associated with a particular residue as $\Delta G_{res} = \sigma_{res} SA_{res}$, where σ_{res} = 0.028 kcal/(mol·Å^2) is a constant of proportionality related to the free energy of transfer of amino acids between aqueous and lipid environments.[38,39] According to a well-accepted protocol,[30] interfacial Trp, Arg, and Lys residues were not penalized for residual mismatch due to their ability to adapt favorably to hydrophobic/hydrophilic interfaces.

Quantifying Lipid Bilayer Shapes in Computationally Derived Bicontinuous Structures. *Evaluation of the Lipid Bilayer Shape in Protein-free LCP.* For a quantitative assessment of the spatial organization of the bicontinuous lipid/water structures assembled computationally, we utilized an analytical fit of the data obtained from

the simulations. To obtain this fit we represented the location of the lipid bilayer midplane of the bicontinuous cubic phase by the triply periodic minimal surface (TPMS, surface with zero mean curvature):

$$f(x, y, z, \lambda) = 0 \tag{1}$$

where $f(x, y, z, \lambda)$ is the improved nodal approximation given by:

$$
\begin{aligned}
f(x, y, z, \lambda) = \; & (Eccc + Ecss + Escs + Essc)\{1, 1, 1\} \\
& + 0.1407(Eccc + Ecss + Escs + Essc + Occc \\
& + Ocss + Oscs + Ossc)\{1, 1, 1\} \\
& + 0.020(Eccc + Ecss - Escs + Essc + Occc \\
& + Ocss - Oscs - Ossc)\{3, 1, 1\} \\
& - 0.0138(Eccc - Ecss + Escs - Essc + Occc \\
& - Ocss + Oscs + Ossc)\{3, 1, 3\} \\
& - 0.0028(Eccc + Ecss + Escs + Essc + Occc \\
& + Ocss + Oscs + Ossc)\{1, 1, 5\} \\
& - 0.0021(Eccc + Ecss + Escs + Essc + Occc \\
& + Ocss + Oscs + Ossc)\{3, 3, 3\} \\
& + 0.001(Eccc + Ecss - Escs - Essc + Occc \\
& + Ocss - Oscs - Ossc)\{3, 1, 5\}
\end{aligned}
\tag{2}
$$

with

$$
\begin{aligned}
Opqr \; &= p(hX)q(kZ)r(lY) + p(hY)q(kX)r(lZ) + p(hZ)q(kY) \\
& \quad r(lX) \\
Epqr \; &= p(hX)q(kY)r(lZ) + p(hY)q(kZ)r(lX) + p(hZ)q(kX) \\
& \quad r(lY)
\end{aligned}
\tag{3}
$$

and $X = \lambda\pi x$, $Y = \lambda\pi y$, $Z = \lambda\pi z$, and $p,q,r = c,s$ (c and s stand for Cosine and Sine, respectively). The $\{h,k,l\}$ in the above equations are the crystallographic Miller indices, and λ is a fitting parameter such that $1/\lambda$ represents the linear size of the primitive unit cell. This surface was shown to provide an accurate representation of the surface in the $Pn3m$ symmetry group (see, for example, ref 3 and citations therein).

For each 9.9MAG/water trajectory listed in Table 1, a fit to the analytical solution was performed on 20 separate frames spaced 350 ns apart. To this end we extracted the coordinates $\{\vec{x}_i\}$ of N C5A coarse-grained beads representing the location of the terminal methylene and methyl groups in the hydrocarbon tail of 9.9 MAG lipids for each frame, and the fit was done by optimizing $f(\vec{x}_i, \lambda) = C_i$ through a search for the translational and rotational transformations that minimized $\sigma^2 = (1/N)\sum_{i=1}^{N}C_i^2$. The quality of the fit was assessed from the distribution of $\{C_i\}$ values obtained from the minimization procedure; this distribution is centered around the theoretical mean, zero (i.e., the analytical solution of eq 1). Such a distribution was collected for each MD trajectory by merging the $\{C_i\}$ coefficients obtained from fitting all 20 separate frames, and the quality was assessed as described above (see distribution plots and captions in Figure 4 and in Figure S3 in the Supporting Information).

Quantification of the Lipid Bilayer Shape in Protein-containing LCP. To compare quantitatively the structural arrangement of the LCP bilayer near the protein, to that in the bulk cubic phase, we employed two complementary approaches. In the first, we obtained an analytical solution for the lipid bilayer midplane around the protein by fitting the simulation data (coordinates of C5A beads) to $f(x, y, z, \lambda) = C_i$, as described above in the TPMS representation, except that for this purpose the $\{C_i\}$ coefficients were also considered as variable parameters that must be self-consistently minimized (see above). In this manner, the extent of the deviation of the mean value of the $\{C_i\}$ distribution obtained from the solution for the $Pn3m$ phase ($\langle C_i \rangle \geq 0$,

see above), measures the distortion of the diamond cubic phase due to the protein.

The alternative approach we developed for quantifying the perturbation of lipid bilayer shape induced by the GPCR departs from the symmetry group definition and focuses on the lipid–water boundary. In this approach, different trajectory frames from the simulation of rhodopsin in LCP are aligned onto snapshots from the simulations of the large protein-depleted 9.9 MAG/water system. Using a locally defined scoring function (see below), we assess how well the lipid bilayers of the protein-containing and protein-free systems align (the large protein-free system was constructed by replicating 27 times the equilibrated system composed of 500 9.9 MAG lipids/6800 water molecules, and subjecting it to microsecond long CG MD simulation during which the $Pn3m$ phase was maintained; data not shown).

To align the protein-free and protein-containing structures, we used water densities to first align 10 frames from the last 100 ns of the protein-free simulation, which served as a reference on which we aligned 10 frames from the last 100 ns of the protein-containing trajectory. Only water beads at least 50 Å away from the protein were considered, in order to avoid degradation of the alignment quality due to deformations of the lipid cubic phase near the protein.

The alignments utilized the electron density $\rho(\vec{y})$ at any point y of a set of beads with coordinates $\vec{r}_i$; the density function was constructed as a superposition of Gaussians, so that for a particular trajectory frame

$$\rho(\vec{y}) = \sum_{i=1}^{N} C \cdot \exp(-k \cdot \| \vec{r}_i - \vec{y} \|^2) \tag{4}$$

where C and k are the amplitude and width of the Gaussians, respectively, chosen to produce a 10 Å resolution density map.[40]

The quality of the superposition of two frames was estimated by measuring the overlap of the water-containing and water-free regions using:

$$S(X) = \sum_{j=1}^{N} s_j \rho_w(\vec{x}_j) \tag{5}$$

where $X = \{\vec{x}_j\}$ ($j = 1,2,...,M_1, M_1+1,...,N$) are coordinates of beads in the fitted frame which contains M_1 water beads and $N - M_1$ lipid beads; ρ_w is the density of water in the reference frame obtained using eq 5; $s_j = 1$ for $j \leq M_1$, and $s_j = -1$ for $j > M_1$. The best alignment was calculated by maximizing S with respect to the rotational and translational transformations applied to X.

The local quality of the alignment was assessed by a normalized scoring function defined as:

$$S_i(X) = \frac{\rho(\vec{x}_i)}{\sum_{j=1}^{M}\rho(\vec{x}_j)} \tag{6}$$

where $\rho(\vec{x}_j)$ represents the electron density of the beads in the reference frame.

RESULTS

Self-Assembly of a 9.9 MAG/Water System into a Lipidic Diamond Cubic Phase. The self-assembly of the 9.9 MAG/water complex with 40% (w/w) water composition was probed at 20 °C, which are conditions similar to those used in the original *in meso* crystallization trials of GPCR proteins (see ref 14 and citations therein). Figures 2 and 3 show several views of 9.9 MAG/water complexes obtained under these conditions produced by the 6 μs-long self-assembly simulation of 500 9.9 MAG lipids. In Figure 2, panels A−C offer views of the unit simulation cell, whereas Figure 3 depicts various snapshots of the same system replicated 27 times (3 times in each orthogonal direction). Typically, in our simulations the initially randomly mixed 9.9 MAG/water system became organized as
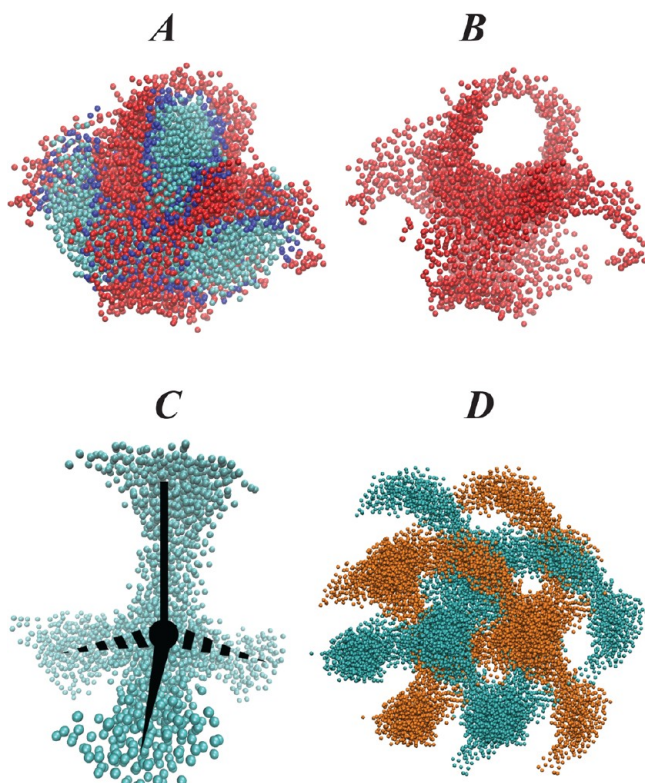
Figure 2. (A, B) Views of the unit simulation cell containing 9.9 MAG/water complex at 40% (w/w) water concentration and at 20 °C temperature, after 6 μs of self-assembly simulation. (A) Entire system where different components are colored as following: 9.9 MAG lipid headgroup beads, blue; 9.9 MAG lipid hydrocarbon tail beads, red; and water beads, cyan. (B) Separately 9.9 MAG hydrocarbon chain beads from the same perspective as in A. (C) View of the tetrahedral arrangement of the water channel from A. The four arms of the tetrahedral geometry are visible. (D) Snapshot of two nonintersecting continuous intertwined water channels (in gold and cyan, respectively). The representations are for the "expanded" system obtained by replicating the simulated system 9 times (in positive x, y, and z directions). Note the tetrahedral geometry formed by each of the aqueous networks.

Figure 3. Views of the 9.9 MAG/water complex from Figure 2 replicated 27 times (3 times in positive and negative x, y, and z directions) reveal bicontinuous diamond cubic phase nature of the self-assembled structure. Different panels show: the entire system (A), 9.9 MAG headgroup and water CG beads (B), 9.9 MAG hydrocarbon chain beads (C), only water beads (D). Color code is from Figure 2A.

illustrated in Figures 2–3 within the first 350 ns of simulation, and remained stable for the remainder of the 6 μs trajectory.

The bicontinuous cubic phase nature of the self-assembled structure is apparent from inspection of the spatial organization of both lipid and water components. Indeed, the hydrophobic core of the 9.9 MAG lipids (shown in red in Figures 2–3) traces a single highly curved continuous surface in 3D, surrounded by 9.9 MAG headgroups (dark blue shades in Figures 2–3) adjacent to aqueous compartments (cyan colors in Figures 2A,C and 3). The aqueous part of the system consists of two nonintersecting three-dimensional continuous networks of water channels (Figure 2D), which makes the assembled structure a bicontinuous cubic phase.

Figure 2C shows that in each of the water compartments four aqueous channels meet in a tetrahedral arrangement. A similar arrangement can be observed as well for the surface traced by the hydrocarbon chains of 9.9 MAG lipids (see Figures 2B and 3C). Thus, the assembled structure resembles closely the doubly bicontinuous lipidic diamond cubic phase. One general feature of such mesophases is that the surface traced by the midplane of the lipid bilayer, which separates the two aqueous compartments, can be approximated by the triply periodic
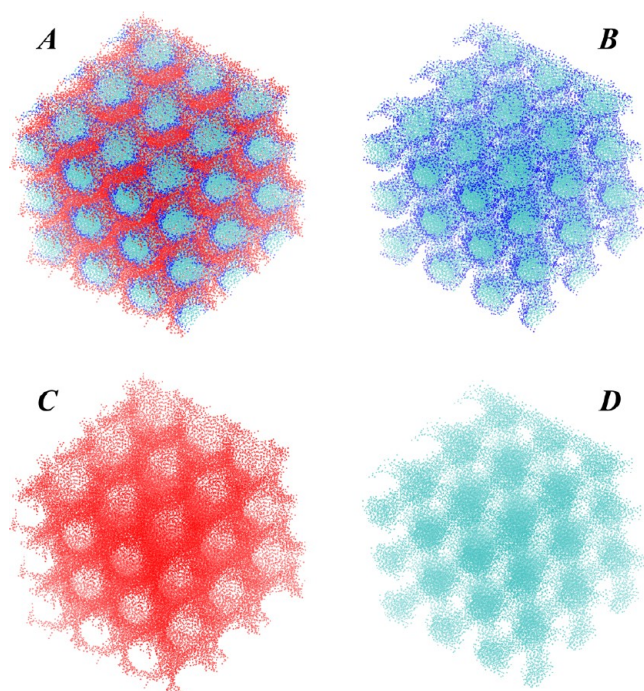
minimal surface (TPMS) that has a special property: it is characterized at every point by zero mean curvature. Consequently, the TPMS is saddle-shaped (Figure S2, Supporting Information). This saddle-shaped geometry of the 9.9 MAG/water assembly resulting from our calculations can be appreciated, for example, by tracing the surface formed by the 9.9 MAG hydrophobic core (see Figures 2–3).

For a quantitative assessment of the adherence of the bicontinuous structure from our simulations to the organization of the $Pn3m$ diamond cubic phase, we performed a fit of the simulation data to the analytical solution, as described in Methods. The quality of the fit was assessed quantitatively from the distribution of the fitting coefficients $\{C_i\}$ (see Methods) shown in Figure 4. Clearly, the distribution is centered around zero mean, and is relatively narrow (with a standard deviation of ~0.4), suggesting that the surface traced by the C5A beads is consistent with the TPMS of the $Pn3m$ cubic phase (see eqs 1–3).

As indicated in Table 1, the self-assembly of 9.9 MAG/water mixture was investigated in two additional simulations performed at T = 20 °C and 40% (w/w) water content. In all cases 9.9 MAG and water molecules aggregated in doubly bicontinuous diamond lipid cubic phases, similar to that described in Figures 2–3 (see Figure S3 in the Supporting Information).

Different Residual Exposure Patterns for Rhodopsin in Bilayers of the Cubic and Lamellar Phases. To explore the organization of a GPCR inside the cubic mesophase we simulated the prototypical class-A GPCR, rhodopsin, in 9.9 MAG-containing lipidic cubic phase in a large system constructed as described in Methods. The simulation results were used as described below to compare quantitatively the
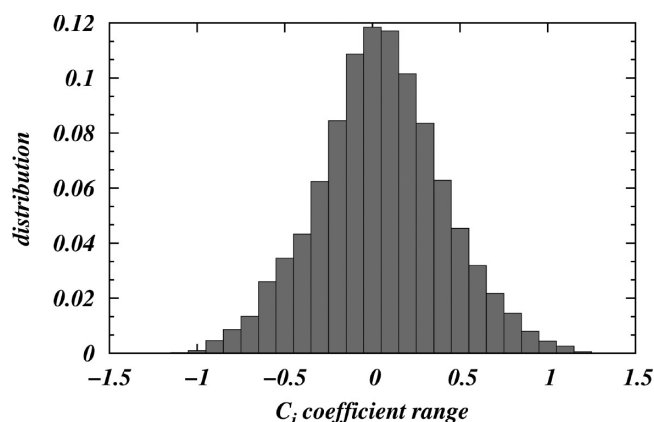
Figure 4. Distribution of $\{C_i\}$ coefficients from the fitting of the C5A bead positions from the simulations to the analytical TPMS approximation (see Methods). The standard deviation of the distribution is ~0.4. The χ^2 test established the significance of the fit with probability $p < 0.05$.

interactions of rhodopsin with bilayers of the cubic and lamellar phases, as a basis for interpreting the mechanistic role of the hydrophobic mismatch in triggering the nucleation process during the *in-meso* crystallization.[19]

Residues in the rhodopsin TM bundle that participate in unfavorable interfacial hydrophobic/hydrophilic interactions with the lipid bilayers of the cubic and lamellar phase were identified, as described in Methods, with a SASA-based approach.[30] The last 1.5 μs intervals from the respective CG MD trajectories for rhodopsin in the 9.9 MAG LCP (4.8 μs), and in the 9.9 MAG lipid lamellar phase (for 3.2 μs) were used for the residual exposure analysis (see Methods for definitions), and the corresponding residual exposure energies were evaluated as described.[30] Table 2 lists rhodopsin TM residues

Table 2. Residual Exposure Penalties (in k_BT)[a] for Rhodopsin TM Residues That Exhibit Different Hydrophobic Mismatch in the Lipid Bilayers of the Cubic (ΔG_{LCP}) Compared to the Lamellar ($\Delta G_{LAMELLAR}$) Phases

rhodopsin residue[b]	ΔG_{LCP}	$\Delta G_{LAMELLAR}$	$\Delta G_{LCP} - \Delta G_{LAMELLAR}$
Pro 34 (1.29)	0	4.7	−4.7
Gln 36 (1.31)	3.4	1.4	2.0
Phe 228 (5.63)	0	1.8	−1.8

[a]Analysis was carried out on the last 1.5 μs time interval of the respective trajectories where SASA values and corresponding residual exposure energies remained, within fluctuations, unchanged. [b]Numbering shown in parentheses corresponds to the generic residue numbering scheme for GPCRs defined by Ballesteros and Weinstein.[41]

that were found to exhibit different hydrophobic exposures in the LCP and lamellar phase simulations, alongside their respective residual exposure energy penalties (ΔG_{res}). All other residues in the TM bundle experienced a residual penalty of <1 k_BT (k_B denoting the Boltzmann constant) in both LPC and lamellar phases.

Different residual exposure in the two lipid environments became evident for residues in the extracellular (EC) end of TMH1, and the intracellular side (IC) of TMH5 (Table 2). Specifically, the Pro34[1.29]/Gln36[1.31] pair on TMH1 (superscript numbers identify the residues by the Ballesteros and Weinstein generic residue numbering scheme for GPCRs[41]), and Phe228[5.63] in TMH5 were found to generate a lower residual

exposure energy penalty in the LCP than in the lamellar phase. Especially remarkable is the difference in ΔG_{res} for the hydrophobic Pro34[1.29] residue, that is, 4.7 k_BT in the lamellar bilayers, but <1 k_BT in the cubic phase. This difference results from Pro34[1.29] being largely exposed to the lipid polar headgroups and/or water environment in the planar membrane, but shielded from such unfavorable interactions in the LCP.

The large residual exposure predicted for Pro34[1.29] in the lamellar 9.9 MAG bilayers is consistent with findings from our earlier all-atom MD simulations of rhodopsin in planar membranes composed of diC$_{n:1}$PC (n = 14, 16, 18, 20) lipids; the largest residual exposure for Pro34[1.29] was found in the bilayers with the thinnest hydrophobic core, that is, diC$_{14:1}$PC and diC$_{16:1}$PC membranes. Due to the substantial difference between the hydrophobic length of TMH1 in rhodopsin (~37 Å, measured along the membrane normal) and the bulk hydrophobic thickness of the thinner bilayers (~27.2 and 30.5 Å, respectively for diC$_{14:1}$PC and diC$_{16:1}$PC membranes), the mismatch was not alleviated by membrane deformations around the protein (see ref 30). Here, the hydrophobic thickness of the 9.9 MAG lamellar bilayers, calculated from the average distance between the lipid GL1 backbone beads on the two leaflets, is 31.5 Å, so that the hydrophobic mismatch between TMH1 and the bilayer is again substantial, and can be alleviated only partially by membrane deformation, as shown below.

The average thickness of the 9.9 MAG lamellar bilayer around rhodopsin, calculated from the last 1.5 μs of the MD trajectory and the membrane thinning around TMH2—the helix adjacent to TMH1—is evident in Figure 5. Interestingly, we find that the bilayer deforms near the polar Ser98[2.65] residue (see Figure 5D), which is apparently shielded from the unfavorable exposure to the hydrophobic lipid core by the local thinning of the lipid membrane. Usually, the membrane-facing −OH of Ser residues in a TM bundle can avoid hydrophobic contact by forming hydrogen bonds with the helix backbone. When this is not feasible energetically, the bilayer thins around this residue. In our simulations we find that in order to shield Ser98[2.65] from unfavorable hydrophobic exposure, the thinning at TMH2 also constrains the lipids at the adjacent TMH1 (see ref 30 for a discussion of this type of constraint), so that the membrane cannot deform significantly and thus leaves the Pro34[1.29] residue highly exposed (see Figure 5D and Table 2).

Figure 5 shows that the 9.9 MAG planar bilayer deforms around other TM helices as well. Specifically, the membrane thins around TMH4, alleviating the residual mismatch for this helix, and around TMH7 where the deformation efficiently accommodates the juxtaposed amphipathic helix 8 at the hydrophobic/hydrophilic interface. Overall, the heterogeneous pattern of membrane deformations we observe around rhodopsin in the current simulations agrees with earlier findings from all-atom MD simulations showing the pattern of membrane remodeling around rhodopsin and the homologous serotonin 5-HT$_{2A}$ GPCR.[30,31,42] Consistent with those studies, we find here that the deformations of the lamellar 9.9 MAG bilayer do not completely alleviate the hydrophobic mismatch with the receptor (Table 2). The different organization of the two lipid phases around the GPCR, as presented below, explains why the residual exposure was larger for rhodopsin in the lamellar 9.9 MAG bilayer than in the 9.9 MAG LCP.
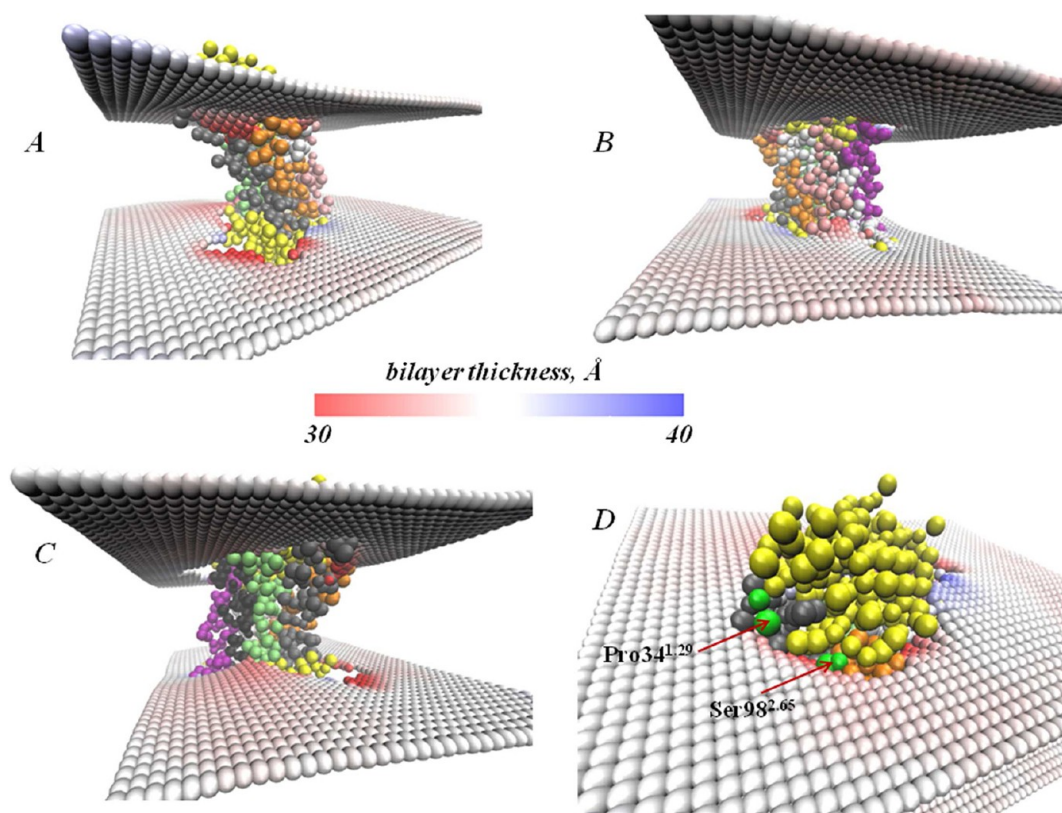
Figure 5. (A−C) Views of membrane deformation patterns around the rhodopsin immersed into 9.9 MAG lamellar bilayers. The average positioning of membrane leaflets is identified by the two surfaces colored according to the local thickness of the bilayer (see the bar for color definition). The GPCR is shown in van der Waals representation with the TM helices colored as follows: TM1 in gray, TM2 in orange, TM3 in white, TM4 in pink, TM5 in purple, TM6 in black, and TM7 in lime The rest of the protein (loops, etc.) is rendered in yellow. The membrane representation is focused on the region near the protein. (D) Magnified view of the region containing residues Pro34[1.29] and Ser98[2.65] (both rendered in green) showing the membrane deformations in their vicinity.

The Intrinsically Curved Geometry of the Lipidic Cubic Phase Is Key to the Low Residual Exposure Experienced by Rhodopsin in the Cubic Mesophase. Figure 6 shows the organization of the lipid cubic phase around rhodopsin in the initial configuration where the protein was randomly inserted in the LCP (see also Figure S1 in the Supporting Information), compared to the organization after 4.8 μs of CG MD simulations. Thus, comparison of A−C with D−F in Figure 6 shows the substantial rearrangement of the lipids and solvent around rhodopsin during the long MD trajectory. Specifically, it appears that the LCP tends to minimize hydrophobic/hydrophilic contacts with the GPCR as the water and lipid headgroup beads organize around polar regions of the protein (yellow colors in Figure 6), whereas the lipid hydrocarbon chain beads surround the hydrophobic core of the protein (see caption of Figure 6 for TM definitions).

The snapshots in Figure 6 also show that rhodopsin equilibrates, as expected, around the saddle-point inside the LCP[5] so that the TM bundle perpendicularly traverses a narrow but relatively flattened region of the LCP bilayer, much like the GPCR protein does in the lamellar bilayers. However, due to continuous and intrinsically saddle-shaped geometry of the diamond cubic phase, the LCP bilayer is dramatically curved near TMH1 and TMH5, which effectively confines the protein to a small region of the LCP. This special arrangement of the cubic mesophase around rhodopsin shields the GPCR in the LCP from the kind of unfavorable hydrophobic/hydrophilic interactions with the lipid bilayer seen in the lamellar

membranes. Thus, as illustrated in Figure S4 (in the Supporting Information) and reflected in the residual energy penalties (in Table 2), residues Pro34[1.29] and Phe228[5.63] are in contact with the hydrophobic core of the LCP bilayer and therefore do not incur a residual mismatch penalty, although Gln36[1.31] experiences somewhat higher residual exposure in the LCP compared to the lamellar phase (see Figure S4 and Table 2). Overall, we find the difference in energy $\Delta G_{\text{LCP}} - \Delta G_{\text{LAMELLAR}}$ to be -2.7 $k_{\text{B}}T$ for the Pro34[1.29]/Gln36[1.31] pair, and -1.8 $k_{\text{B}}T$ for Phe228[5.63] (Table 2). Notably, the shielding of Ser98[2.65] that faced the 9.9 MAG lipid bilayer in the lamellar phase simulations, occurs in the LCP simulations through interaction with the protein backbone.

To relate the calculated residual mismatch energies to the organization of the lipidic cubic mesophase around the GPCR, we quantitatively compared the structural arrangement of the LCP bilayer near the protein, to that in the bulk cubic phase. As described in Methods, two different approaches were used: In the first, we obtained an analytical fit of the lipid bilayer midplane shape around the protein by fitting the coordinates of 9.9 MAG lipid C5A methyl terminus beads (Figure 1A) from the simulations to $f(x, y, z, \lambda) = C_i$, considering $\{C_i\}$ coefficients as variable parameters that must be minimized during the fitting (see Methods). The application of this procedure to spherical lipid shells situated at different distances from rhodopsin (see Figure 7A) revealed that the organization of the LCP close to the GPCR is different from that expected for the *Pn3m* lipid bilayer ($\langle C_i \rangle \neq 0$ in Figure 7A). However, already at distances
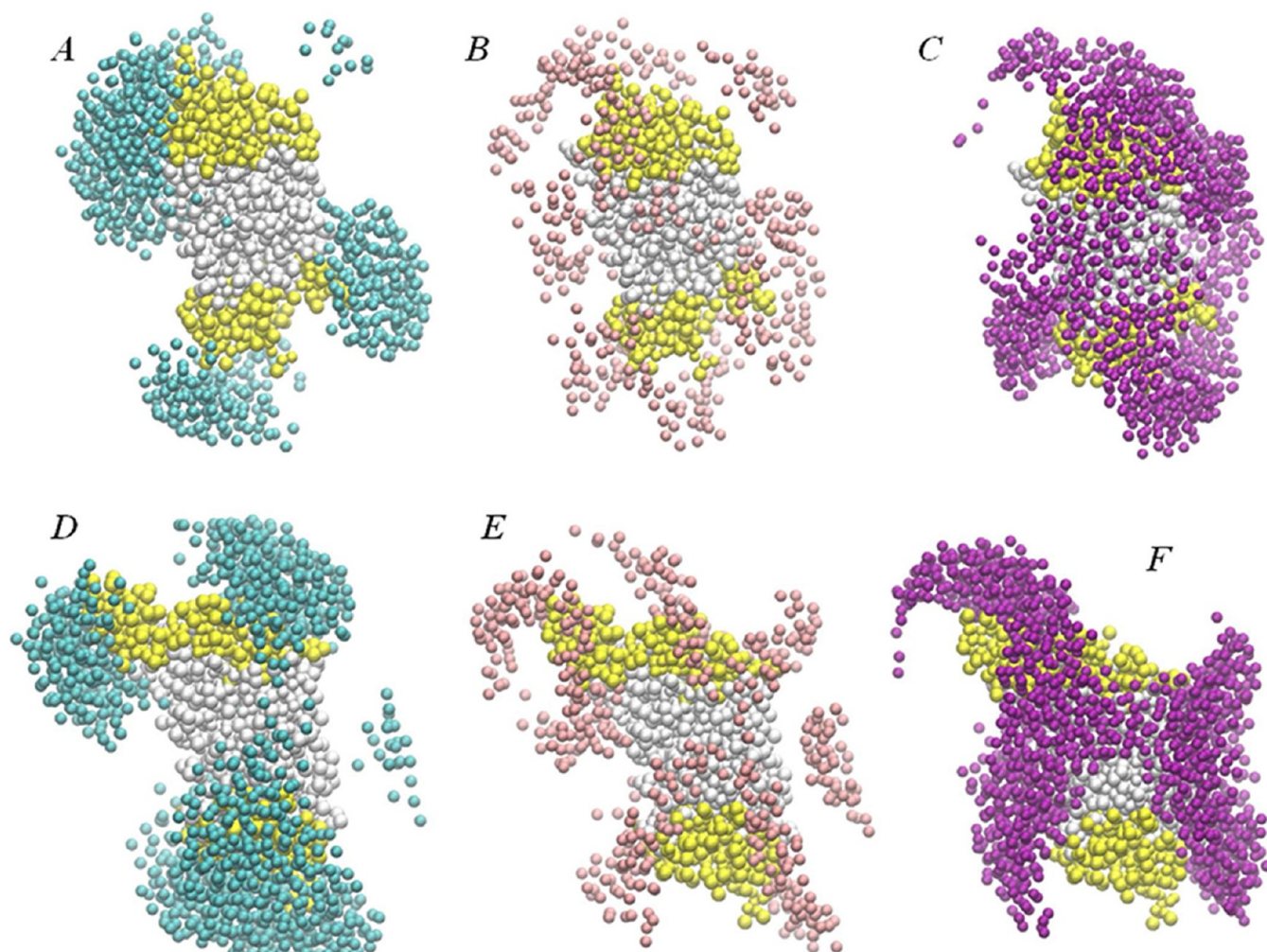
Figure 6. Snapshots illustrating initial (A−C), and final (after 4.8 μs simulations) positioning (D−F) of rhodopsin (in yellow/white) inside the cubic phase. A and D show the organization of water beads (in cyan), B and E illustrate the organization of 9.9 MAG lipid headgroup beads (in pink), and C and F show the organization of 9.9 MAG lipid hydrophobic core beads (in purple). In all panels rhodopsin transmembrane (TM) helices are colored in white and the rest of the protein in yellow. Shown are only water and 9.9 MAG lipids that are within 20 Å of rhodopsin; rhodopsin is oriented so that its intracellular and extracellular ends (as defined in the lipid bilayer) point toward the bottom and top of the panels, respectively.

∼50 Å from the center of rhodopsin (compare to ∼35−40 Å linear dimension of a rhodopsin-like GPCR in bilayer $x-y$ plane directions), the simulated structure relaxes to the organization of the diamond cubic phase ($\langle C_i \rangle = 0$, Figure 7A). The second approach confirms that the distorted LCP bilayer near the protein tends to relax toward the *Pn3m* phase arrangement at larger distances (Figure 7B). In this alternative approach we locally quantify the shape of the lipid bilayer around the GPCR by aligning trajectory frames from the simulation of rhodopsin in LCP onto snapshots from the simulations of the pure LCP system, so as to optimize the overlap of the water-containing and water-free regions (see Methods). A scoring function defined locally (see eq 5) is used to assess the quality of the alignment. As seen from Figure 7B, this procedure revealed that the fit between the densities of C5A lipid beads in protein-containing and protein-free systems progressively improves with the radial distance from the GPCR (increasing S in Figure 7B). Indeed, consistent with the result in Figure 7A, the scoring function reaches its plateau ∼50 Å away from the protein, indicating that at these distances the LCP arrangement becomes similar to that expected for the *Pn3m* phase.

Taken together, our results establish a quantitative link between the residual mismatch penalty and the lipid bilayer deformations around the GPCR in the cubic phase. Thus, the areas where the largest perturbations from the diamond cubic phase were observed included residues Pro34[1.29] and Phe228[5.63] (see Figure S5 in the Supporting Material), which are also the residues with the largest value of $\Delta G_{LCP} - \Delta G_{LAMELLAR}$ (Table 2).

■ DISCUSSION

The main findings from this study offer unprecedented insight into molecular level processes related to *in meso* crystallization, which are likely leading to protein nucleation and eventually to crystal growth. Thus, we found that the difference in the hydrophobic/hydrophilic exposure of the protein to the lipid bilayers of 9.9 MAG LCP and lamellar phases is responsible for a lower residual mismatch (and its corresponding energy penalty) in the cubic mesophase compared to the lamellar membranes. This indicates why GPCRs reconstituted into the LCP in the initial stages of the *in meso* trial are well accommodated inside the cubic phase, where they are believed to reside mostly in the monomeric form. Indeed, from our
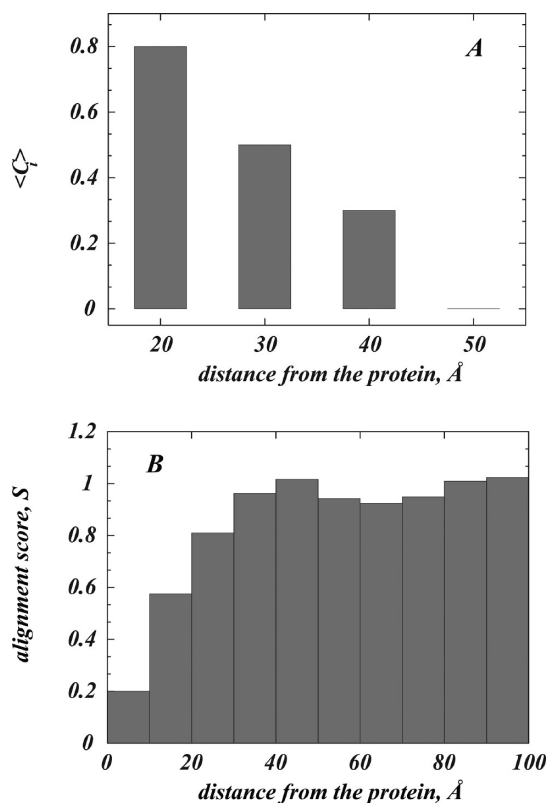
Figure 7. Distance dependence of the relaxation of membrane deformation near rhodopsin in the cubic phase. (A) Analytical fit of the density of C5A 9.9 MAG lipid beads from CG MD simulation of rhodopsin in the LCP phase to the surface defined by the $f(x, y, z, \lambda) = C_i$ equation (see Methods). The mean values of the $\{C_i\}$ fitting coefficient distribution as a function of r, are shown for 10 Å spherical lipid shells located at $r = 20$, 30, 40, and 50 Å away the protein center-of-mass. (B) Reasults from the same measurements performed with an alternative approach based on the alignment of the lipid bilayer midplane in protein-containing and protein-free cubic phases (see Methods). The quality of the fit upon alignment was determined for 10 Å spherical lipid shells located at distances r from the protein center-of-mass, and the panel shows the dependence of the scoring function S on r. Note that the deformed cubic phase near rhodopsin relaxes to the arrangement that is characteristic of the $Pn3m$ phase ~50 Å away from the protein.

findings there is insignificant drive toward rhodopsin oligomerization in the LCP. We showed, however, why this drive is increased in the planar bilayer environment of the lamellar phase where there are several critical mechanistic factors that determine the probability for proteins to oligomerize. One important energy component that has been established from both experiments[28,29] and computations,[26,30] involves the hydrophobic mismatch. In particular, a plausible mechanism to relieve the energy penalty due to residual exposure is protein oligomerization, whereby the TM segments incurring the highest energy penalty from the residual exposure come together in the lipid bilayer.[29,30]

For such hydrophobic-mismatch driven association to occur, the residual exposure penalty alleviated by the protein association must counteract other critical factors that could favor proteins in the monomeric state. One such important consideration is the mobility of proteins, because constrained diffusion reduces the chance for protein−protein encounter. As shown in Figure 8, the spatial restriction encountered by the
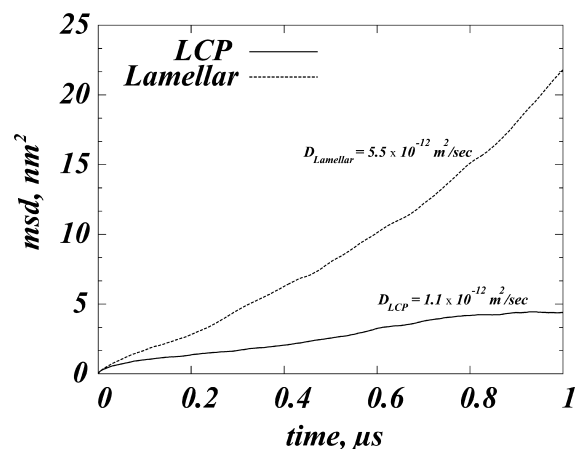

Figure 8. Mean-square-displacement (msd) as a function of time calculated for rhodopsin in the 9.9 MAG lipidic cubic phase (LCP) and in the 9.9 MAG lamellar membrane (Lamellar). Corresponding diffusion coefficient values, calculated from the linear fit performed in $[0\ \mu s; 1\ \mu s]$ time interval, are $D_{LCP} = 1.1 \times 10^{-12}\ m^2/sec$ and $D_{LAMELLAR} = 5.5 \times 10^{-12}\ m^2/sec$.

GPCR inside the LCP (see Figure 6) significantly affects protein mobility, with the diffusion coefficient of the GPCR in the cubic phase being smaller ($D_{LCP} = 1.1 \times 10^{-12}\ m^2/sec$) than that in the lamellar bilayers ($D_{LAMELLAR} = 5.5 \times 10^{-12}\ m^2/sec$). The restricted diffusion mode for rhodopsin in the LCP environment is consistent with the suggested high energy cost in the LCP for GPCR-sized proteins to cross between different regions of the mesophase.[8] This energy penalty was quantified from phenomenological principles to be as high as ~12−20 $k_B T$ in the $Pn3m$ phase with lattice parameter values in the 75−110 Å range.[8] This estimated energy cost of protein diffusion in the LCP is significantly higher than the residual mismatch penalty we calculated for rhodopsin in the cubic phase (Table 2). When considered together, the relatively low residual exposure energy for the GPCR in the LCP is not likely to be sufficient to compensate for the high energy barrier for protein mobility in the LCP and drive protein association. The prediction is, therefore, that in the cubic mesophase GPCR proteins will remain largely in the monomeric state.

In contrast to the LCP, the lamellar membrane environment is more favorable for GPCR oligomerization: The residual energy penalty for the rhodopsin TM bundle in the lamellar bilayers is ~8 $k_B T$ (Table 2) and is significantly higher compared to that in the LCP, indicating a stronger drive for GPCR oligomerization in the lamellar membranes. A higher propensity for association in the planar bilayers is also supported by our calculations of the protein mobility in the two lipid environments, whereby we find a 5-fold larger diffusion coefficient for the GPCR in the lamellar bilayers (Figure 8).

Taken together, results from our simulations suggest that planar 9.9 MAG lipid membranes provide a more suitable setting for oligomerization of the GPCR proteins. Notably, this prediction is based on results for oligomerization in a plane of the lipid bilayer. This 2D nucleation process together with the formation of protein-enriched lamellar stacks in 3D[19] should ultimately drive the emergence of a bulk crystal.

The question remains regarding the mechanism that drives the proteins from the LCP to the lamellar membranes and the potential role of the differential residual interactions between

the protein and the bilayers of the two lipidic phases in the nucleation process. We note that the nucleation is triggered by the addition of the precipitant which is known to cause transient water depletion from the LCP interior,[19] resulting in structural changes in the cubic phase bilayer. Our current study has not attempted to quantify the effects of the precipitant, but it is reasonable to speculate that dehydration of the cubic phase will only increase the unfavorable residual interactions between the GPCR and the LCP bilayer.[19] This is likely to increase the drive for protein oligomerization inside the LCP upon precipitant addition, which could lead to the formation of locally flattened lamellar bilayers as a prelude to protein crystallization. In order to address quantitatively the structural perturbations due to precipitant, such as changes in curvature of the LCP bilayer, it is critical to calculate the corresponding deformation energies in the presence of the protein. However, the representation of the complex geometry of the LCP around the insertion (Figures 6–7) in the numerical approach developed in the current work for quantifying the LCP shape (i.e., fitting the MD data to the analytical solution) is not sufficiently refined to serve in the calculation of reliable energies. We will be addressing these numerical challenges in future work.

We note, however, that recent experimental studies[12,18,21] showed that, compared to 9.9 MAG, in meso trials conducted on the GPCR-Gs protein complex with a shorter chain MAG analog, 7.7 MAG (acyl chain 14 carbon atoms long with the cis double bond between carbon atoms 7 and 8) resulted in better quality crystals. These results support the mechanism we describe based on the residual mismatch. Thus, results from our current studies as well as from recently reported all-atom simulations of GPCRs in membranes of different thicknesses,[30] explain why 7.7 MAG would provide a better platform in which GPCRs can aggregate more extensively: based on our findings we can predict that, due to the expected ∼7−8 Å difference in the hydrophobic thickness of planar 9.9 MAG and 7.7 MAG membranes, the residual exposure of the GPCR will be substantially higher in the thinner 7.7 MAG lamellar bilayers. As an example, we recently reported that the rhodopsin TM bundle will exhibit a 12 k_BT higher residual energy penalty in $diC_{14:1}PC$ than in $diC_{18:1}PC$ bilayers.[30] A more complete quantitative elucidation of the GPCR nucleation process in relation to the residual mismatch energy component requires further studies that would take into consideration the effect of the different precipitants on various lipids.

The structurally specific predictions of the regions where the residual mismatch with lipid bilayers of cubic and lamellar phases differed for the rhodopsin TM bundle made it possible to probe the resulting predictions and general character of the findings by examining crystallographic data. In particular, we showed here that the intracellular (IC) end of TMH1 (Pro34[1.29]), and the extracellular (EC) part of the TMH5 including Phe228[5.63] have higher residual exposure in lamellar membranes, which leads to the prediction (e.g., see Shan et al.[42]) that in the planar bilayers rhodopsin will exhibit strong hydrophobic-mismatch driven tendency for oligomerization through TMH1 and TMH5. To assess this hypothesis in the context of structural information available for GPCRs, we examined crystallographic contact interfaces for 12 different structures of rhodopsin-like GPCRs obtained by means of the in meso technology (β_2 receptor − PBD codes: 2RH1, 3PDS, 3SN6; A2A receptor − PDB codes: 3EML, 3QAK; Chemokine CXCR4 receptor − PDB codes: 3ODU, 3OE0, 3OE6, 3OE8,

3OE9; Dopamine D3 receptor − PDB code: 3PBL, and Histamine H1 receptor − PDB code: 3RZE). For all these GPCR structures, we analyzed the content of the unit crystallographic cell and quantified for each TM helix (1) the frequency of its occurrence at the crystallographic contact interface and (2) the number of residue interactions each TM helix forms at the contact interface. In this analysis only "canonical" interfaces were considered, where the crystallographic contacts were formed exclusively through TM−TM interactions between monomers in parallel orientations.

Remarkably, we found that TMH1 and TMH5, the two helices implicated in the largest residual mismatch in our simulations of rhodopsin with 9.9 MAG lamellar bilayers, contribute to the most common contact interfaces in the crystallographic structures of the homologous GPCRs. More specifically, as illustrated in Figure S6A (see Supporting Information), all but 4 of the 14 distinct intermolecular TM−TM interfaces identified in this analysis, involve TMH1 and/or TMH5. Furthermore, examining the residues involved in the interactions at the contact interfaces revealed that the TMH1 stretch of residues in positions 1.29−1.34 (that would include Pro34[1.29] in the homologous rhodopsin GPCR) contribute significantly to the formation of the crystallographic interface.

During the final preparation of this manuscript, two new crystal structures of rhodopsin-like GPCRs, the κ-opioid receptor (KOR) and the μ-opioid receptor (MOR) were reported to be obtained by means of in meso crystallization.[43,44] Consistent with the data presented above, the crystallographic interface of KOR consists of TM1, TM2 and H8, with the N-terminal end of TM1 (region harboring the residue homologous to Pro34[1.29] in rhodopsin) forming extensive dimeric contacts.[43] For MOR,[44] two crystallographic interfaces were reported: the more prominent interface consisted of TM5/TM6 helices with the I256[5.62] residue (aligning with the residue next to Phe228[5.63] in rhodopsin) residing at the dimer interface, and more limited contacts formed by TM1, TM2 and H8.

Taken together, our findings support the mechanistic inferences that (1) the reduced level of hydrophobic mismatch in the LCP is attributable to the specific highly curved geometry of the cubic phase that provides for more coverage from unfavorable hydrophobic exposure; (2) compared to the LCP, lamellar structures provide a more favorable setting in which GPCRs can oligomerize as a prelude to nucleation and crystal growth; and (3) the extent of the residual mismatch penalty is likely one of the critical mechanistic factors that determines not only the drive for GPCR proteins to oligomerize during in meso crystallization, but also the structural elements that are likely to participate in contact interfaces.

These findings provide novel energy-based insights into in meso crystallization mechanisms and lay a foundation for future computational explorations involving other class-A GPCRs toward designing rational approaches for generation of structure-quality crystals of membrane proteins.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Supporting methods and six figures describing organization of rhodopsin in the lipid cubic and lamellar mesophases. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author
*gek2009@med.cornell.edu

Notes
The authors declare no competing financial interest.

ACKNOWLEDGMENTS

REFERENCES

(1) Cherezov, V. *Curr. Opin. Struct. Biol.* **2011**, *21*, 559−66.
(2) Caffrey, M.; Li, D.; Dukkipati, A. *Biochemistry* **2012**, DOI: 10.1021/bi300010w.
(3) Schwarz, U. S.; Gompper, G. *Phys. Rev. E: Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics* **1999**, *59*, 5528−41.
(4) Landau, E. M.; Rosenbusch, J. P. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 14532−5.
(5) Caffrey, M. *Annu. Rev. Biophys.* **2009**, *38*, 29−51.
(6) Cherezov, V. *Curr. Opin. Struct. Biol.* **2011**, *21*, 559−66.
(7) Schnering, H.; Nesper, R. *Z. Z. Phys. B* **1991**, *83*, 407−12.
(8) Grabe, M.; Neu, J.; Oster, G.; Nollert, P. *Biophys. J.* **2003**, *84*, 854−68.
(9) Katritch, V.; Cherezov, V.; Stevens, R. C. *Trends Pharmacol. Sci.* **2011**, *33*, 17−27.
(10) Rasmussen, S. G.; DeVree, B. T.; Zou, Y.; Kruse, A. C.; Chung, K. Y.; Kobilka, T. S.; Thian, F. S.; Chae, P. S.; Pardon, E.; Calinski, D.; Mathiesen, J. M.; Shah, S. T.; Lyons, J. A.; Caffrey, M.; Gellman, S. H.; Steyaert, J.; Skiniotis, G.; Weis, W. I.; Sunahara, R. K.; Kobilka, B. K. *Nature* **2011**, *477*, 549−55.
(11) Cherezov, V.; Clogston, J.; Papiz, M. Z.; Caffrey, M. *J. Mol. Biol.* **2006**, *357*, 1605−18.
(12) Li, D.; Lee, J.; Caffrey, M. *Cryst. Growth Des.* **2011**, *11*, 530−7.
(13) Nollert, P.; Qiu, H.; Caffrey, M.; Rosenbusch, J. P.; Landau, E. M. *FEBS Lett.* **2001**, *504*, 179−86.
(14) Misquitta, L. V.; Misquitta, Y.; Cherezov, V.; Slattery, O.; Mohan, J. M.; Hart, D.; Zhalnina, M.; Cramer, W. A.; Caffrey, M. *Structure* **2004**, *12*, 2113−24.
(15) Misquitta, Y.; Caffrey, M. *Biophys. J.* **2003**, *85*, 3084−96.
(16) Misquitta, Y.; Cherezov, V.; Havas, F.; Patterson, S.; Mohan, J. M.; Wells, A. J.; Hart, D. J.; Caffrey, M. *J. Struct. Biol.* **2004**, *148*, 169−75.
(17) Cherezov, V.; Fersi, H.; Caffrey, M. *Biophys. J.* **2001**, *81*, 225−42.
(18) Cherezov, V.; Clogston, J.; Misquitta, Y.; Abdel-Gawad, W.; Caffrey, M. *Biophys. J.* **2002**, *83*, 3393−407.
(19) Caffrey, M. *Cryst. Growth Des.* **2008**, *8*, 4244−54.
(20) Caffrey, M. *Curr. Opin. Struct. Biol.* **2000**, *10*, 486−97.
(21) Lyons, J. A.; Aragao, D.; Slattery, O.; Pisliakov, A. V.; Soulimane, T.; Caffrey, M. *Nature* **2012**, DOI: 10.1038/nature11182.
(22) Goforth, R. L.; Chi, A. K.; Greathouse, D. V.; Providence, L. L.; Koeppe, R. E., 2nd; Andersen, O. S. *J. Gen. Physiol.* **2003**, *121*, 477−93.
(23) Harroun, T. A.; Heller, W. T.; Weiss, T. M.; Yang, L.; Huang, H. W. *Biophys. J.* **1999**, *76*, 937−45.
(24) Huang, H. W. *Biophys. J.* **1986**, *50*, 1061−70.
(25) Nielsen, C.; Goulian, M.; Andersen, O. S. *Biophys. J.* **1998**, *74*, 1966−83.
(26) Periole, X.; Huber, T.; Marrink, S. J.; Sakmar, T. P. *J. Am. Chem. Soc.* **2007**, *129*, 10126−32.
(27) Botelho, A. V.; Gibson, N. J.; Thurmond, R. L.; Wang, Y.; Brown, M. F. *Biochemistry* **2002**, *41*, 6354−68.
(28) Botelho, A. V.; Huber, T.; Sakmar, T. P.; Brown, M. F. *Biophys. J.* **2006**, *91*, 4464−77.
(29) Marsh, D. *Biophys. J.* **2008**, *94*, 3996−4013.
(30) Mondal, S.; Khelashvili, G.; Shan, J.; Andersen, O. S.; Weinstein, H. *Biophys. J.* **2011**, *101*, 2092−101.
(31) Huber, T.; Botelho, A. V.; Beyer, K.; Brown, M. F. *Biophys. J.* **2004**, *86*, 2078−100.
(32) Grossfield, A.; Feller, S. E.; Pitman, M. C. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 4888−93.
(33) Marrink, S. J.; de Vries, A. H.; Mark, A. E. *J. Phys. Chem. B* **2004**, *108*, 750−60.
(34) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. *J. Phys. Chem. B* **2007**, *111*, 7812−24.
(35) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. *J. Comput. Chem.* **2005**, *26*, 1701−18.
(36) Fuhrmans, M.; Knecht, V.; Marrink, S. J. *J. Am. Chem. Soc.* **2009**, *131*, 9166−7.
(37) Hubbard, S. J.; Thornton, J. M. *NACCESS Software*; University College London: London, 1993.
(38) Choe, S.; Hecht, K. A.; Grabe, M. *J. Gen. Physiol.* **2008**, *131*, 563−73.
(39) Ben-Tal, N.; Ben-Shaul, A.; Nicholls, A.; Honig, B. *Biophys. J.* **1996**, *70*, 1803−12.
(40) DiMaio, F.; Tyka, M. D.; Baker, M. L.; Chiu, W.; Baker, D. *J. Mol. Biol.* **2009**, *392*, 181−90.
(41) Ballesteros, J. A.; Weinstein, H. *Methods Neurosci.* **1995**, *40*, 366−428.
(42) Shan, J.; Khelashvili, G.; Mondal, S.; Mehler, E. L.; Weinstein, H. *PLoS Comput. Biol.* **2012**, *8*, e1002473.
(43) Wu, H.; Wacker, D.; Mileni, M.; Katritch, V.; Han, G. W.; Vardy, E.; Liu, W.; Thompson, A. A.; Huang, X.-P.; Carroll, F. I.; Mascarella, S. W.; Westkaemper, R. B.; Mosier, P. D.; Roth, B. L.; Cherezov, V.; Stevens, R. C. *Nature* **2012**, DOI: 10.1038/nature10939.
(44) Manglik, A.; Kruse, A. C.; Kobilka, T. S.; Thian, F. S.; Mathiesen, J. M.; Sunahara, R. K.; Pardo, L.; Weis, W. I.; Kobilka, B. K.; Granier, S. *Nature* **2012**, DOI: 10.1038/nature10954.

SUPPORTING INFORMATION

Why GPCRs behave differently in cubic and lamellar lipidic mesophases

George Khelashvili[1], Pedro Blecua Carrillo Albornoz[1], Niklaus Johner[1], Sayan Mondal[1], Martin Caffrey[3], Harel Weinstein[1,2]

[1]Weill Cornell Medical College of Cornell University, New York, NY, 10065

[2]The HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, Weill Cornell Medical College, Cornell University, New York, NY 10065

[3]Membrane Structural and Functional Biology Group, School of Biochemistry and Immunology, and School of Medicine, Trinity College, Dublin

Supporting Methods

Coarse-grained parameters for 9.9 MAG lipid

The parameterization of 9.9 MAG compatible with the CG Martini lipid force-field version 2.0 was achieved by dividing the 9.9 MAG all-atom structure into 7 standard coarse-grained beads (Fig. 1A): one bead for the headgroup (ETH), one for the ester linkage (GL1), and 5 beads for the hydrocarbon tail (C1A, C2A, D3A, C4A, C5A) [1]. For the glycerol and the 5 tail beads, the Martini types were set, respectively, to the standard "Na" type, and the "C1-C1-C3-C1-C1" sequence of bead types [1]. The headgroup ETH bead was designated as "P4" type, similar to that found in the Martini parameter set for 1,2-dioleoyl-*sn*-glycero-3-phospho-(1'-*rac*-glycerol) (DOPG) lipid [2] which also possesses an OH-C-C-OH headgroup (Fig. 1A). However, for 9.9 MAG, the ETH bead/water LJ interaction parameter was increased here from 5.0 to 5.6 kJ/mol [3] in order to take into account the special properties of 9.9 MAG, as described below. The parameterization of the bonded interactions was performed following the standard set of bonded parameters for lipids in the Martini version 2.0 [1] with a single exception: the equilibrium angle for the GL1-C1A-C2A triplet was set to 140^0 (instead of the default value of 180^0). The special structure of 9.9 MAG made it necessary to introduce the two adjustments described above, implemented equally in both the cubic and the lamellar phase simulations, in order to account for the different curvature propensity expected of a single-tail 9.9 MAG compared to the double-tail phospholipids, such as DOPG.

The set of CG force-field parameters was tested first on lamellar bilayers formed by 9.9 MAG, in simulations of membranes containing 472 9.9 MAG lipids and 15104 water molecules. These simulations, conducted at 27^0C for 40ns under semi-isotropic pressure coupling conditions and with a 40fs time step, yielded structural characteristics of the 9.9 MAG bilayer in good agreement with those reported from all-atom MD simulations [4], as well as from experimental measurements on 9.9 MAG [38,39]. Specifically, in our CG simulations we found the area per 9.9 MAG head-group to be 37.5 0.5 $Å^2$, a value similar to experimentally measured 37.9 $Å^2$ [4]. Furthermore, the calculated thickness of the 9.9 MAG bilayer (measured from the peak-to-peak distance on the density plot of ETH/GL1 pair of beads) was 34.7 1 Å, consistent with the 33-37 Å range for the bilayer thickness (depending on the extent of

solvation) reported from X-ray diffraction experiments [5] and all-atom MD simulations [4]. In addition, measured from the peak-to-peak distance on the density plot of the D3A double bond bead, the average separation between the C9=C10 unsaturated bonds on the two leaflets of the bilayer is 12.8 1 Å, in good agreement with results from atomistic MD simulations [4]. The close match between the structural properties of CG 9.9 MAG bilayer obtained from our CG simulations and those obtained experimentally or with all-atom MD simulations supports the CG parametrization we developed for the exploration of the cubic phases formed by 9.9 MAG.

Coarse-grained representation of rhodopsin

The structure of rhodopsin was obtained from PBD 1U19, represented at the CG level as described before [6], and using the recent 2.1 version of the Martini force-field [7]. Similar to the protocols used in earlier CG studies of rhodopsin [6]: 1)-the protein secondary structure elements were maintained unchanged throughout the CG simulations through appropriate bond interaction parameters; 2)-the disulfide bridge between C110 and C187 was introduced in the force-field through covalent bonding; 3)-the protonated forms of Asp83, Glu122, and Glu181 in the CG model were represented by Asp and Glu by Asn and Gln, respectively; and 4)-palmitoyl chains were not included in the CG model.

References

(1)   Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. *J Phys Chem B* **2007**, *111*, 7812-24.

(2)   Baoukina, S.; Monticelli, L.; Risselada, H. J.; Marrink, S. J.; Tieleman, D. P. *Proc Natl Acad Sci U S A* **2008**, *105*, 10803-8.

(3)   Marrink, S. J.; Mark, A. E. *Biophys J* **2004**, *87*, 3894-900.

(4)   Wilson, M. A.; Pohorille, A. *J Am Chem Soc* **1994**, *116*, 1490-501.

(5)   Pezron, I.; Pezron, E.; Bergenstahl, B. A.; Claesson, P. M. *Journal of Physical Chemistry* **1990**, *94*, 8255-8261.

(6)   Periole, X.; Huber, T.; Marrink, S. J.; Sakmar, T. P. *J Am Chem Soc* **2007**, *129*, 10126-32.

(7)   Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S. J. *Journal of Chemical Theory and Computation* **2008**, *4*, 819–834.

Figure Legends

Figure S1: Two views of Rhodopsin (in *yellow*) inserted randomly in the lipid cubic phase for the start of the simulation. The lipid system as shown was obtained by magnifying 27 times (i.e., 3 times in positive and negative *x, y,* and *z* directions) the last snapshot from the 9.9 MAG/water complex simulation (Figure 2 of the main text). The cubic phase is shown by depicting water beads (cyan) and the 9.9 MAG lipid headgroup beads (*pink*).

Figure S2: TPMS approximation to the cubic phase of *Pn3m* symmetry (*continuous analytical surface)* constructed using Eqs. (1-3). The locations of the C5A beads (in *green*) from the simulations of 9.9 MAG/water complex at 40 % w/w water and 20^{0}C are obtained from the alignment of 20 trajectory frames using a fitting procedure described in Methods.

Figure S3: Views of the 9.9 MAG/water system from simulations at 40 % (w/w) water concentration and 20^{0}C temperature (see Table 1 of the main text), after independent 6 µs of self-assembly simulations (the third simulation at 40 % (w/w) water and at 20^{0}C is discussed in detail in the main text). For each simulated system, panels *A* and *B* show the organization of 9.9 MAG head-group and water beads, respectively, in representations of the of the simulation unit cell magnified 27 times (i.e., 3 times in positive and negative *x, y,* and *z* directions). Panels *C-E* offer various views of the two non-intersecting continuous water channels (shown in cyan and gold colors) in the respective simulations. Panels *F* in each Figure show the distributions of $\{C_i\}$ coefficients as derived from the fitting of the simulated structures to the TPMS of the *Pn3m* phase (the standard deviation of each distribution is ~0.4). The χ^2 test established the significance of the fit with probability p<0.05.

Figure S4: Views of Rhodopsin TMH1 and TMH5 helices (in *yellow*) equilibrated in the lamellar mesophase showing the mismatched positions of residues Pro34[1.29] (in *red*), Gln36[1.31] (*green*), and Phe228[5.63] (*blue*). The lipid environment is represented by the location of 9.9 MAG lipid hydrophobic core beads (*in purple*). The snapshots were generated from the last frame of a 3.2 µs-long MD trajectory.

Figure S5: Classification of rhodopsin residues according to their proximity to the LCP region that is most perturbed from the *Pn3m* organization (according to the *S* score defined in Eq. (5) in the

Main text). The residues that contact the lipid bilayer where it is most deformed from the diamond cubic phase, are shown as blue spheres; the rest of the protein is depicted as red spheres. The lipid bilayer within 25Å of the protein is represented by the location of the C1A 9.9 MAG lipid backbone beads (*gray spheres*). Residues Pro34$^{1.29}$ and Phe228$^{5.63}$ are highlighted in yellow.

Figure S6: : Identification of contact interfaces in the crystal structures of various GPCRs obtained from *in meso* crystallization. The 12 rhodopsin-like GPCRs used in this analysis are: β$_2$ adrenergic receptor – PBD codes: 2RH1, 3PDS, 3SN6; A2A receptor – PDB codes: 3EML, 3QAK; Chemokine CXCR4 receptor – PDB codes: 3ODU, 3OE0, 3OE6, 3OE8, 3OE9; Dopamine D3 receptor – PDB code: 3PBL, and Histamine H1 receptor – PDB code: 3RZE). Only interfaces where the crystallographic contacts were formed exclusively through inter-monomer TM-TM interactions were considered. (*A*) Shows the fraction of GPCR structures that have at least one specific TM-TM contact interface (*red bars*), and the average number of contacts between TM pairs throughout the structural library (*blue bars*). (*B*) Shows the fraction of GPCR structures that have specific residue on TMH1 at least once at the contact interface throughout the structural library (*red bars*), and the average number of contacts particular residue on TMH1 makes throughout the structural library (*blue bars*). (*C*) Same as in panel *B* only for TMH5 residues. Residue numbering in panels *B-C* are given in Ballesteros & Weinstein numbering.
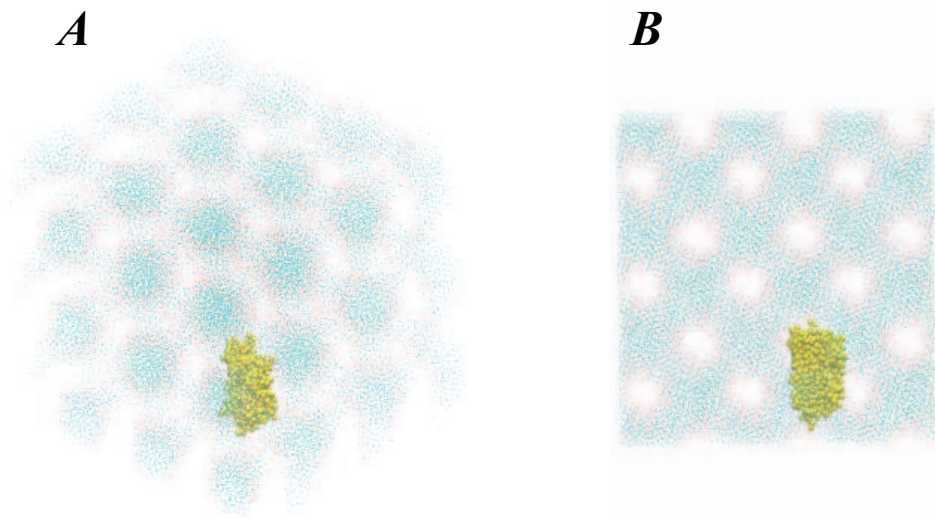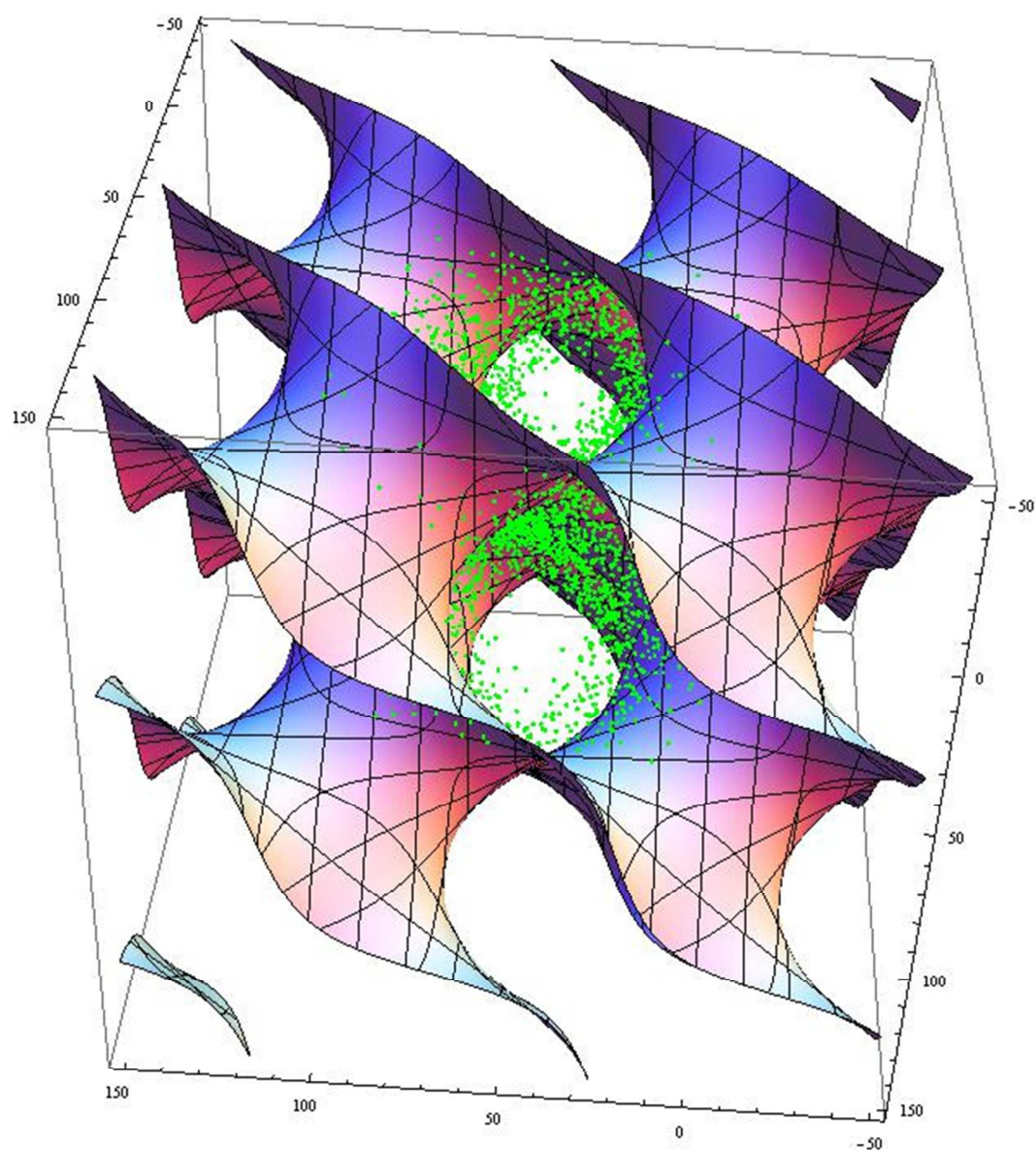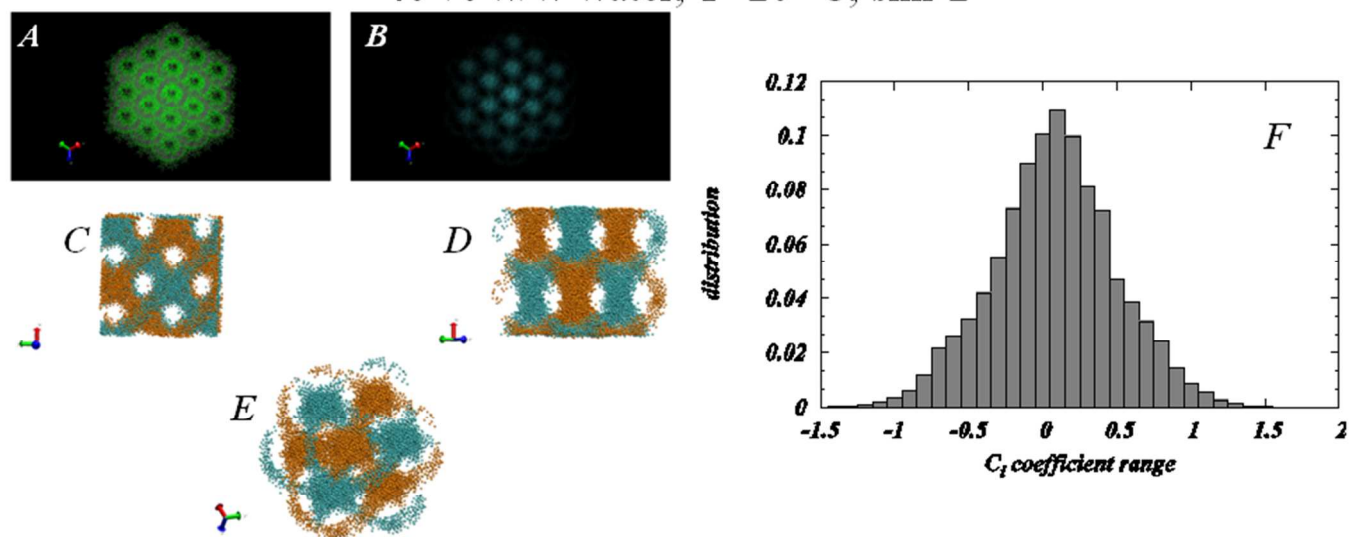
A

B

Figure S1

Figure S2

40 % w/w water, T=20°C, sim 2

40 % w/w water, T=20°C, sim 3
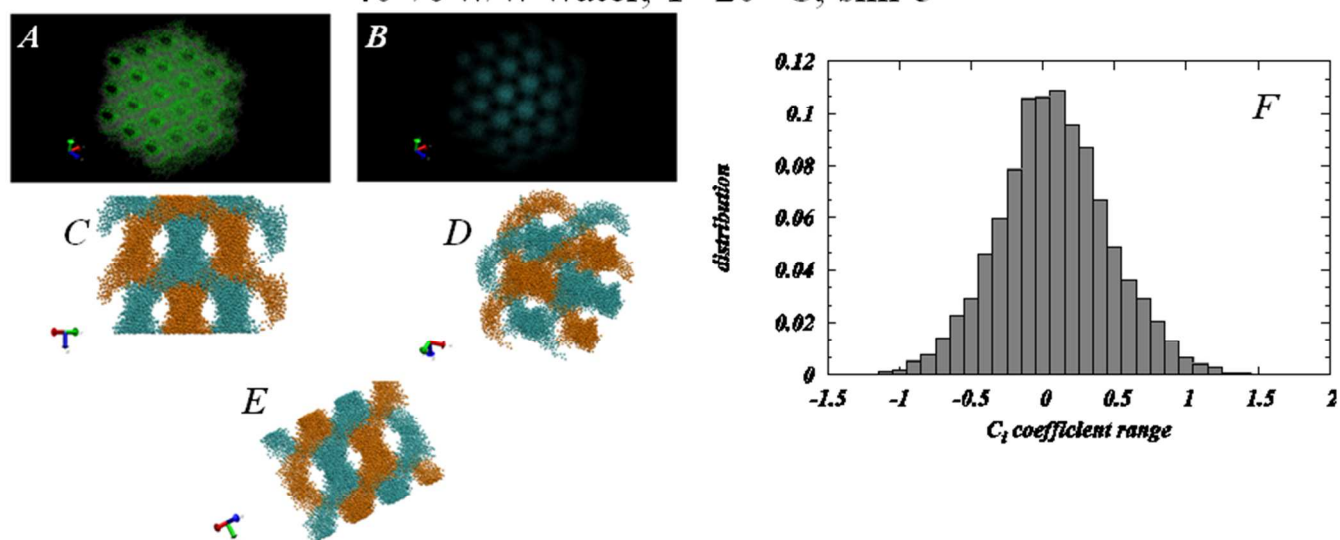
Figure S3

A
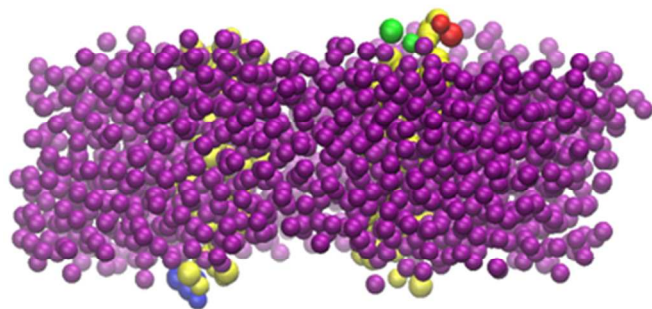
B
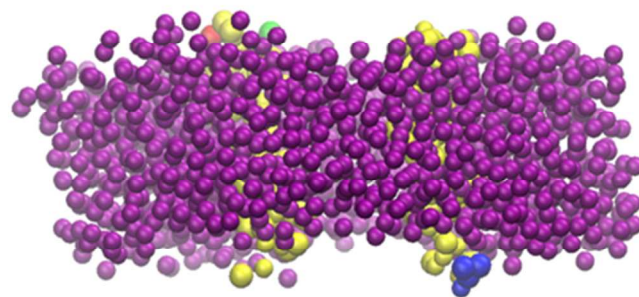
Pro34[1.29] – red
Gln36[1.31] – green
Phe228[5.63] – blue

Figure S4

Pro34$^{1.29}$

Phe228$^{5.63}$

Figure S5

Figure S6