Protein and Lipid Interactions Driving Molecular Mechanisms of *in meso* Crystallization

Niklaus Johner,[†] Sayan Mondal,[†] Giulia Morra,[†,‡] Martin Caffrey,[§] Harel Weinstein,[†] and George Khelashvili*,[†]

[†]Weill Cornell Medical College of Cornell University, New York, New York, 10065, United States

[‡]Istituto di Chimica del Riconoscimento Molecolare, Consiglio Nazionale delle Ricerche, 20131, Milano, Italy

[§]Membrane Structural and Functional Biology Group, School of Biochemistry and Immunology, and School of Medicine Trinity College, Dublin 2, Ireland

S *Supporting Information*

ABSTRACT: The recent advances in the *in meso* crystallization technique for the structural characterization of G-protein coupled receptor (GPCR) proteins have established the usefulness of the lipidic-cubic phases (LCPs) in the field of crystallography of membrane proteins. It is surprising that despite the success of the approach, the molecular mechanisms of the *in meso* method are still not well understood. Therefore, the approach must rely on extensive screening for a suitable protein construct, for host and additive lipids, and for the appropriate precipitants and temperature. To shed light on the *in meso* crystallization mechanisms, we used extensive coarse-grained molecular dynamics simulations to study, in molecular detail, LCPs under different conditions (compositions and temperatures relevant to crystallogenesis) and their interactions with different types of GPCR constructs. The results presented show how the modulation of the lattice constant of the LCP (triggered by the addition of precipitant during the *in meso* assay), or of the host lipid type, can destabilize monomeric proteins in the bilayer of the LCP and thus drive their aggregation into the stacked lamellae, where the residual hydrophobic mismatch between the protein and the membrane can drive the formation of lateral contacts leading to nucleation and crystal growth. Moreover, we demonstrate how particular protein designs (such as transmembrane proteins engineered to contain large polar regions) can promote protein stacking interactions in the third, out-of-plane, dimension. The insights provided by the new aspects of the specific molecular mechanisms responsible for protein−protein interactions inside the cubic phase presented here should be helpful in guiding the rational design of future *in meso* trials with successful outcomes.
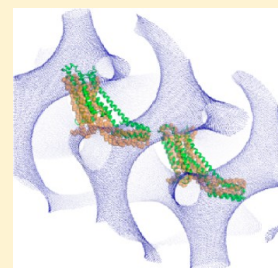
■ INTRODUCTION

The use of nonlamellar lipidic phases[1] has taken an increasing importance in structural studies of membrane proteins.[2−4] Preeminent among those is the lipid cubic phase (LCP), which consists of a highly curved lipid bilayer structure that is continuous in three dimensions (3D) and separates two interpenetrating, but nonintersecting, aqueous channels.[5] Both the aqueous and bilayer compartments are continuous in space, and the midplane of the lipid bilayer traces a triply periodic minimal surface (characterized by zero mean curvature) with cubic symmetry.[6,7] Of particular interest is the specific type of cubic phase that belongs to the *Pn3m* symmetry group, where each aqueous channel has a tetrahedral geometry.[5] The *Pn3m* mesophase (also termed the diamond cubic phase) has been extensively utilized in the crystallography of transmembrane (TM) proteins (*in meso* crystallization), and its suitability for supporting growth of high quality crystals for structure determination has been firmly established by the rapid pace of new structural data acquired with the *in meso* method in the class A, and more recently class C, G-protein coupled receptor (GPCR) families. Indeed, structures of a number of GPCR proteins in complex with a variety of ligands (reviewed in ref 8)

and even bound to the G-protein[9] have been solved with high resolution using the *in meso* approach.

Despite the unquestioned success of the *in meso* method, several important mechanistic aspects remain unclear,[5,8,10−12] leaving the approach to rely on extensive trials which include screening for a suitable host and additive lipid, precipitants, and temperature. There is evidence that during *in meso* crystallization, the membrane proteins partition into a lamellar phase formed of stacked bilayers and connected to the LCP.[10] This organization is reflected in the layered crystal packings seen for all proteins crystallized *in meso*, suggesting the formation of lateral contacts in the membrane plane and stacking interactions in the third dimension, perpendicular to the membrane planes. But several fundamental questions about the key mechanistic steps leading to crystallization remain unanswered: One concerns the reason for partitioning of the proteins, previously well-accommodated in the LCP, into the lamellar phase (stacked bilayers that form in the LCP during crystallization) upon addition of precipitants.[10,13−15] Understanding this essential step for the spatial organization of the

crystal lattice is key to progress of *in meso* crystallization. The second unanswered question is what drives the proteins to form tight aggregates once they are in the lamellar phase.

In our previous work,[15] we have used large-scale Martini coarse-grained molecular dynamics (CGMD) simulations to study the GPCR rhodopsin embedded in an LCP and lamellar bilayers of 9.9 monoolein lipids (9.9 MAG). We found that in the bilayer of the LCP the protein is more efficiently shielded from unfavorable hydrophobic–hydrophilic interactions with its environment, than in a planar membrane; the reduced level of hydrophobic mismatch in the LCP is attributable to its specific, highly curved geometry.[15] Because hydrophobic mismatch has previously been shown to be a drive for protein oligomerization in the membrane plane,[16,17] the results from our earlier studies suggested a plausible molecular mechanism for the observed protein aggregation in the lamellar phase and not in the LCP.

Here, we extend our previous computational studies of the cubic phases and embedded GPCRs to address the two specific open mechanistic questions regarding *in meso* crystallization process, discussed above, by investigating: (i) the drive toward partitioning of the integral proteins from the LCP into the lamellar phase; (ii) the formation of stacking interactions between the proteins; (iii) how the stacking interactions complement protein oligomerization in the lipid bilayer plane to support the formation of a bulk crystal by *in meso* crystallization;[10] and (iv) the role of precipitants and host lipid types in the nucleation and crystal growth process. To this end, we studied with extensive Martini CGMD simulations the dynamics of a GPCR protein (adenosine A_{2A} receptor, $A_{2A}R$) in LCPs of different lattice sizes and in lamellar bilayers, comprised of monoacylglycerols (MAGs or monoolein) with different hydrocarbon chain lengths: 9.9 (18 carbon acyl chain), 7.9 (16 carbon acyl chain), and 7.7 (14 carbon acyl chain) MAGs (see Figure S1). We took advantage of the high-resolution structure of the $A_{2A}R$ (PDB ID: 4EIY), recently obtained by *in meso* crystallization,[18] to study this protein in the absence ($A_{2A}R$ construct) or presence ($A_{2A}R$-BRIL construct) of the covalently attached thermostabilized protein $b_{562}IL$ used during the crystallization experiments,[18] and in the different lipidic environments described above, allowing for direct comparison of our computational results with the crystal packing of this structure.

We found that the $A_{2A}R$-BRIL complex is organized within the LCP in a manner that favors strong stacking interactions with neighboring $A_{2A}R$-BRIL structures, when inserted in the 9.9 MAG cubic phase with a lattice constant of ~102 Å. This lattice size is achieved in the crystallization trials[18] at room temperature prior to any precipitant addition. The simulations revealed that the formation of these intermolecular contacts is mediated through the water channels of the cubic phase which accommodate the BRIL extensions by bringing the BRIL from one protein into close proximity of the extracellular loops from an adjacent $A_{2A}R$-BRIL complex. This alone produces an arrangement of the proteins that is surprisingly similar to the one seen from the crystal packing (PDB ID: 4EIY) of $A_{2A}R$-BRIL obtained from *in meso* crystallyzation.[18] But under the same lattice size conditions, and contrary to the $A_{2A}R$-BRIL complex, we find that the constructs of the GPCR alone ($A_{2A}R$) are fully solvated by the cubic phase environment, precluding the stacking interactions in the absence of precipitants.

The geometric parameters (for example, curvature and bilayer thickness) of the LCP with lattice constant of 102 Å

computed from the simulations revealed that both $A_{2A}R$-BRIL and $A_{2A}R$ constructs could be accommodated inside the cubic phase with only minimal perturbations to the mesophase. But remarkably, we found that modulation of the lattice constant in a range that is expected from the addition of different precipitants,[13,14,19] i.e. either increasing the unit cell dimensions to 113 Å or reducing them to 82 Å, resulted in strong destabilization of the LCP matrix around the inserted proteins, leading to unfavorable hydrophobic–hydrophilic interactions between the insertions and the cubic phase environment. This suggests that such energetically costly protein–LCP interactions (mainly due to hydrophobic mismatch) are the molecular mechanism driving the formation of, and/or the segregation of the proteins into, a lipid lamellar environment.[15–17] Finally, hypothesizing that the formation of the crystal contacts in the plane of the membrane occurs in this lamellar environment and is driven by the unfavorable protein–membrane interactions, we quantified the cost of such hydrophobic mismatch for the $A_{2A}R$ protein in monoolein lamellar bilayers using a previously established protocol[15–17,20] and found that the TMs with the highest drive for oligomerization in the 9.9 MAG membrane correspond to the specific interfaces seen in the crystal packing of the $A_{2A}R$-BRIL complexes (PDB ID: 4EIY).

The novel quantitative and mechanistic results presented here provide, for the first time, a molecular level explanation for the manner in which the modulation of the lattice constant of the LPC (triggered by the addition of precipitant during the *in meso* crystallization assay) can destabilize monomeric proteins inside the LCP, therefore driving their aggregation into the stacked lamellar bilayers. We also demonstrate how the specific protein constructs can influence the formation of stacking interactions, hence the possible formation of crystals during *in meso* experiments, which offers valuable insights for the choice of precipitants and constructs leading to successful crystallogenesis.

■ METHODS

Molecular Constructs and Force Field Parameters. For all the molecular dynamics (MD) simulations we used the Martini force field representations[21,22] that combine atoms into coarse-grained (CG) beads. The grouping is performed by replacing, on average, four heavy atoms with a single CG bead, as illustrated in Figure S1 for 9.9, 7.9, and 7.7 monoolein (MAG) lipids.

In all CGMD runs, parameters compatible with the Martini lipid force field (version 2.0) were used for the MAG lipids[21] (see Supporting Information), and the Elnedyn force field implemented in Martini was used for the proteins.[23] The structure from the PDB ID: 4EIY[18] was used with ($A_{2A}R$-BRIL) or without (full-length human $A_{2A}R$) the BRIL sequence covalently attached to the intracellular loop 3 (ICL3) of the GPCR, as starting structure for all the simulations. We note that the engineered GPCR-BRIL construct studied here was implemented in recent *in meso* crystallography trials of the $A_{2A}R$[18] and other GPCRs,[24,25] yielding high-resolution structures. Throughout the text, we used the $A_{2A}R$ residue numbering according to the sequence of the human $A_{2A}R$ (UNIPROT: P29274) with superscripts indicating the Ballesteros–Weinstein generic numbering[26] (for example, the most conserved residue in TM1: $N24^{1.50}$). The 45 crystallographically resolved water molecules in the core of the receptor (from PDB ID: 4EIY) were retained in the CG representations by transforming them into 12 Martini water beads, to conserve the correct ratio between CG and all atom particles (Martini combines on average 4 heavy atoms into single CG beads). The optimal positions of the 12 CG water beads were found by using a K-means clustering algorithm and the R package[27] applied to the set of coordinates of crystal water oxygen

atoms. This procedure decomposes the coordinate set into a predetermined number of clusters (set to 12 in this case). The coordinates of the center of each cluster were then used as starting positions for the CG water beads, which were further optimized by a cycle of 1000 steps of steepest-descent minimization.

Simulations in LCPs. Several sets of CGMD simulations were carried out in the LCP environment in the presence or absence of proteins. First, self-assembly simulations of the Pn3m cubic phase were conducted for 9.9, 7.9, and 7.7 MAGs at different lipid-to-water ratios and at various temperatures. In the next stage, the $A_{2A}R$ and $A_{2A}R$-BRIL constructs were studied in specifically selected self-assembled Pn3m systems. These computational experiments are described in detail below.

Self-Assembly Simulations of Pn3m LCP. Self-assembly simulations of Pn3m LCP were performed for 9.9, 7.9, and 7.7 MAG lipids under the lipid-to-water ratio (% w/w) and temperature (T) conditions listed in Tables S1–S3. For a given concentration, lipid and water components were first ideally mixed by performing short CGMD simulations at $T = 50$ °C 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 studies from our group[15] and others,[28] allows for efficient adaptation of the volume of the simulation box to the number of CG molecules in the system and ensures complete mixing of lipids and waters prior to the self-assembly simulation.

After this initial equilibration phase, 3 μs-long self-assembly simulations were carried out at the desired temperature (Tables S1–S3), again with the isotropic pressure coupling but with the LJ parameters for MAG lipids reset to their proper values, as prescribed in Martini (see Supporting Information for the detailed parameters used). These MD simulations implemented periodic boundary conditions and were run using a 30 fs integration time step. As detailed in the Results, at all the conditions probed, lipid/water mixtures spontaneously assembled into structures representing the crystallographic unit cells of the Pn3m cubic phases with different lattice constants (unit cell size) a presented in Tables S1–S3, ranging from 82 to 120 Å.

Simulations of GPCR Protein Constructs in Pn3m LCPs. To study GPCR proteins in LCPs of different lattice constants, we used specific self-assembled Pn3m setups from the conditions probed in the self-assembly simulations (shaded in gray in Table S1). These selected systems were first replicated 27 times in positive and negative x, y, and z directions, and a single GPCR construct (either $A_{2A}R$ or $A_{2A}R$-BRIL) was inserted randomly in the enlarged mesophase (resulting in a 14743:1 to 32281:1 range of lipid-to-protein number ratios equivalent to protein concentrations of 2–3 mg/mL up to 30 times lower than the concentrations used during crystallization). Specifically the $A_{2A}R$ was studied in the 82, 86, and 102 Å LCPs and $A_{2A}R$-BRIL in the 102 and 113 Å LCPs. After removing waters and lipid molecules overlapping with the protein and adding counterions for electroneutrality, 2 μs-long CGMD trajectories were accumulated (six 1 μs-long replicas with different initial conditions were simulated for the $A_{2A}R$-BRIL in the 113 Å LCP, see Table S1). These MD runs used a 10 fs integration step, to ensure the stability of the protein under Elnedyn, and were conducted with the isotropic pressure coupling algorithm.

Simulations of GPCR Proteins in Lamellar Bilayers. The $A_{2A}R$ protein was inserted in pre-equilibrated 9.9, 7.9, or 7.7 MAG lamellar membranes containing 3104 lipids and at a hydration level of ~40 waters per lipid. Following a previously established protocol,[15] after an initial equilibration phase with the protein backbone atoms restrained, a 2 μs-long MD simulation was carried out with a 10 fs integration step at a temperature $T = 25$ °C and with the semi-isotropic pressure coupling scheme.

Quantifying Structural Parameters of Computationally Derived Cubic Phases. *Trajectory Alignment.* Trajectories from the LCP simulations were aligned using the same approach as described in ref 15. Briefly, the density of the water beads was generated for a reference frame in the trajectory as a sum of Gaussians:

$$\rho_w(\vec{y}) = \sum_{i=1}^{N} C \cdot \exp(-k \cdot \| \vec{x}_i^w - \vec{y} \|^2) \tag{1}$$

where $\vec{x}_i^w$ represent the positions of the water beads. This density was then smoothed using a low-pass filter (see Figure S2a). The quality of the superposition of a frame with the reference frame was estimated by measuring the overlap of the water-containing and water-free regions with this density using:

$$S(X) = \sum_{i=1}^{M1} \rho_w(\vec{x}_i^w) - \sum_{i=1}^{N-M1} \rho_w(\vec{x}_i^l) \tag{2}$$

where $X = \{\vec{x}_i^w, \vec{x}_j^l\}$ $(i = 1, 2, ..., M_1; j = 1, ..., N - M_1)$ are the coordinates of the beads in the fitted frame which contains M_1 water beads and $N - M_1$ lipid beads. The best alignment was obtained by maximizing S with respect to the rotational and translational transformations applied to X.

For the systems containing a protein, two different alignments were done: one on the whole system (comprising 27 unit cells) and another on a square box of size 1.5 times the lattice constant centered on the protein and excluding the waters and lipids within 20 Å of any backbone bead of the protein. Focusing the alignment on the region around the protein allows a better determination of the geometrical properties of the LCP close to the protein and therefore was used for all the assessments of the perturbations of the LCP around the protein. The global alignment was only used for illustration purposes to represent the surfaces further than one unit cell away from the protein. Since, after the initial equilibration phase, the protein remained confined to a particular region of the LCP (with standard deviations of its center of mass position and backbone root-mean square deviations (RMSDs) relative to the average structure typically of the order of 4–9 Å, when measured after aligning the trajectory as described above), the aligned trajectories were suitable for determining different structural properties of the LCP around the protein as detailed hereafter.

Surface Determination and Analysis. Two surfaces, the bilayer midplane and the interface between the waters and lipids, were determined from the aligned trajectories using average densities over ~300 frames from the second half of each simulation (last 1.5 μs for simulations without proteins and 1 μs for simulations containing a protein, except for the $A_{2A}R$-BRIL complex in the LCP of largest lattice constant, i.e., the 113 Å LCP, where the last 0.5 μs were used). The surfaces were determined on a 3D grid with a 1 Å mesh size, except for the $A_{2A}R$-BRIL in the 113 Å LCP where, for practical reasons, a 2 Å grid was used (a 1 Å grid was nevertheless used for this simulation when performing analysis only on the unit cell around the protein). Periodic boundary conditions were accounted for in all the analyses except when we focused on the alignment and analysis on the unit cell around the protein (see Results).

The interface between waters and lipids was determined as the surface where the density of waters was equal to the density of lipids.[29] This procedure is illustrated in Figure S2b and in Supporting movie 1 and Supporting files showing the different densities and the extracted surface for one chosen system. To represent the membrane midplane, we first generated densities of the last bead in the lipid tail and determined the set of points that corresponded to the location of a maximum in density along any of the x, y, or z directions. To further refine the midplane surface definition, for each point P on the surface, we determined the best fit plane to the set of points within 15 Å of P and constructed the normal vector at P to that surface. Then, only points that were at a maximum density along the direction of the normal vector were kept. This refinement allowed to obtain a smooth and well-defined surface for the midplane (see Results).

To compute the curvatures and areas at the midplane and lipid/water interface surfaces, we defined normal vectors for each point P on the respective surface (using a best fit plane to a 15 Å neighborhood around each point P). Then, for each point P, an optimization protocol (described in ref 30) was used to obtain the principal directions $e1$ and $e2$ and the associated principal curvatures k_1 and k_2 best describing the local geometry of the surface around P (within 15 Å of P). As the density of points on the surface is not constant (the surfaces being
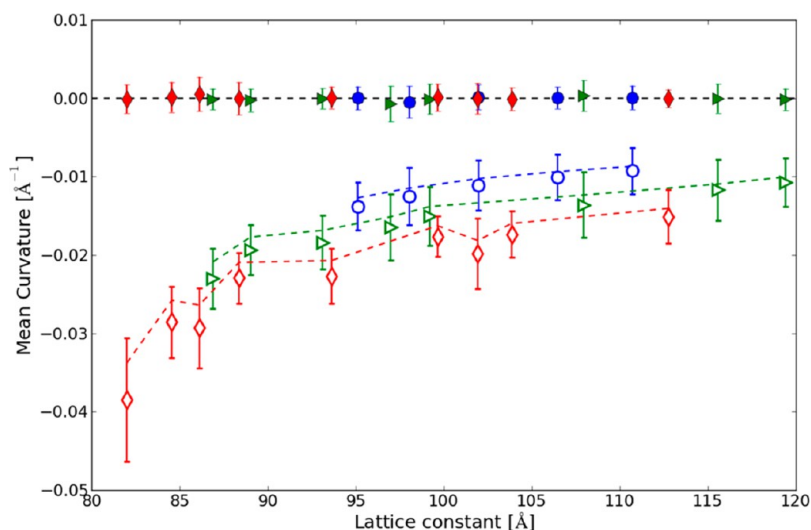
Figure 1. Mean curvature of the membrane midplane (full symbols) and interface (empty symbols), averaged over the unit cell, are shown as a function of the lattice constant for different lipids (blue circles, 7.7 MAG; green right triangles, 7.9 MAG; and red diamonds, 9.9 MAG). The dashed lines are the theoretical estimates from eq 7 as described in the text. The error bars measure the width of the distributions of the mean curvatures.

represented by points on a 3D grid), the surface element associated with each point P was calculated by dividing the area of an 8 Å disc by the number of points of the surface in such a disc around P. We have tested this method for convergence and accuracy by varying the size of the disc (see Figure S3). The total surface areas (SAs) for the midplane and interface were determined by summing the areas of the surface elements associated with each point on the surface.

Averages of the geometrical properties such as mean and Gaussian curvatures were calculated as discrete integrals over the surface, taking into consideration the elemental surface associated with each point (similar to the averages defined in eqs 4, 6 and 7 below). Prior to averaging, the data were truncated at the 1st and 99th percentiles (p_1 and p_{99}) to eliminate some very rare outliers (for example points not lying on the midplane but not eliminated by the refinement procedure described above could subsequently result in curvatures several orders of magnitude higher than points on the surface). The error bars in the figures discussed in the Results show the width of the distributions measured as $p_{75} - p_{25}$.

All analyses and visualizations were done using OpenStructure.[31,32]

Quantification of the Residual Hydrophobic Mismatch. MD simulations have shown[15−17] that, even after the membrane remodeling around multi-TM proteins has taken place in order to reduce the hydrophobic mismatch, it is often impossible to eliminate completely the exposure of some of the TM residues to energetically unfavorable environments. The relation of the residual hydrophobic mismatch (RHM) to the structural properties of multi-TM proteins has been discussed in detail.[17,20] For the $A_{2A}R$ in lamellar bilayers, the RHM of residues in the TM segments was quantified with the computational protocol described in refs 15−17 and 20, and the energy cost was calculated using solvent accessible surface area (SASA) values obtained with the NACCESS software[33] and a probe radius of 5.2 Å, as done for Martini simulations.[34] Only residues with energy penalties larger than 1 k_BT are considered to contribute to the drive for oligomerization.[20] More details about the protocol can be found in ref 15.

Representations of the CG Proteins. In all representations, except for Figure S8 which presents a particular snapshot of a trajectory, we show the average structure of the CG protein, obtained by averaging the positions of the beads over the aligned (as described above) trajectories. For clarity, an all-atom model is overlaid onto the average protein structure by superposition of the $C\alpha$ atoms of the all-atom model onto the backbone beads of the CG protein.

RESULTS

Geometrical Characteristics of the Self-Assembled *Pn3m* LCPs. The molecular organization of the *Pn3m* lipidic cubic phases formed by different MAG lipid types has not yet been characterized in a systematic way. Indeed, existing analytical formulations approximately describe the average geometric features (curvatures, SAs) of the mesophase,[8,35,36] but little is known about the way in which these properties vary from one region of the LCP to another. Yet the ability to obtain such information is critical in order to quantify the effects of inserted proteins on the local mesophase structure. To gain such insights and, at the same time, to verify the robustness of our computational models and of the analysis tools (see Methods), we first carried out an in-depth characterization of the geometry of the lipidic cubic phases formed in self-assembly CGMD simulations (see Methods) by different lipids, i.e. 9.9, 7.9, and 7.7 MAGs, under different composition and temperature conditions (see Tables S1−S3). For the computationally modeled mesophases, we determined the midplane surfaces and lipid/water interfaces and calculated their SAs as well as local geometrical characteristics of interest, such as the mean and the Gaussian curvatures, membrane thickness, and water channel width.

The results, shown in Figures 1−3, reveal remarkable overall agreement between the geometric properties measured for the self-assembled systems and those obtained from the corresponding theoretical estimates for the *Pn3m* mesophase. Specifically, the bilayer midplane of the *Pn3m* LCP must trace a so-called "minimal surface" where the mean curvature H_0 is, by definition, 0 everywhere, and for which the following empirical relationships are expected[8] to hold between the surface area A_0, the average Gaussian curvature $\langle K \rangle_0$, and the size of the unit cell a (i.e., the lattice constant):

$$A_0 = 1.919a^2 \qquad (3)$$

$$\langle K \rangle_0 = -2\pi\chi/A_0 \qquad (4)$$

Here $\langle \cdots \rangle_0 = 1/A_0 \int_{A_0} \cdots dA_0$ is the average taken over the midplane surface and the topological parameter $\chi = 2$ for the *Pn3m* LCP. As a matter of fact, we find that the mean curvature
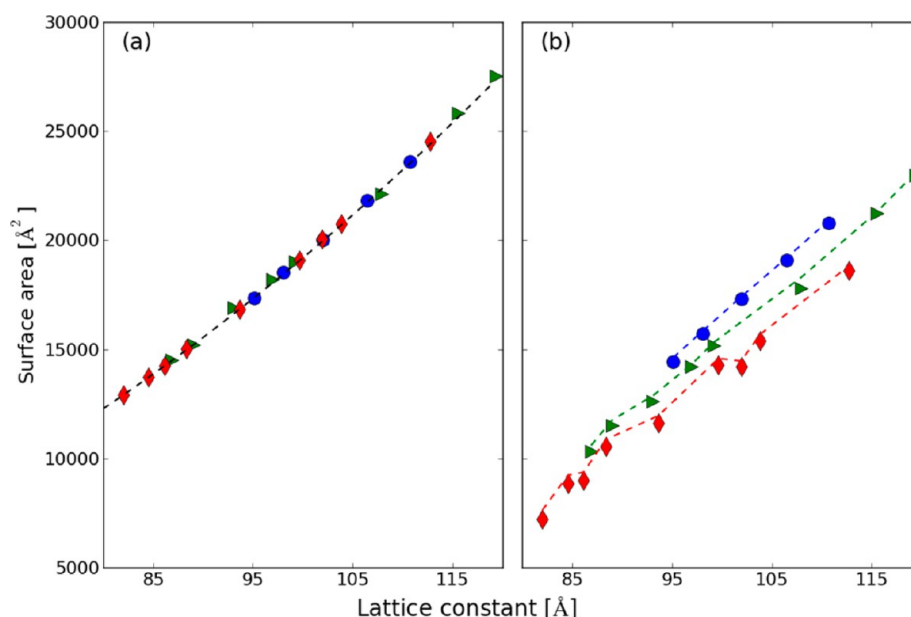
Figure 2. Surface area of the membrane midplane (a) and water−lipid interface (b) in the lipidic cubic phase for different lipids (blue circles, 7.7 MAG; green right triangles, 7.9 MAG; and red diamonds, 9.9 MAG) and compositions, shown as a function of the lattice constant. The dashed lines are the theoretical values from eqs 3 and 5 as described in the text. In panel (b), half of the total interfacial surface is shown, see derivation of the equation in Supporting Information for more information.
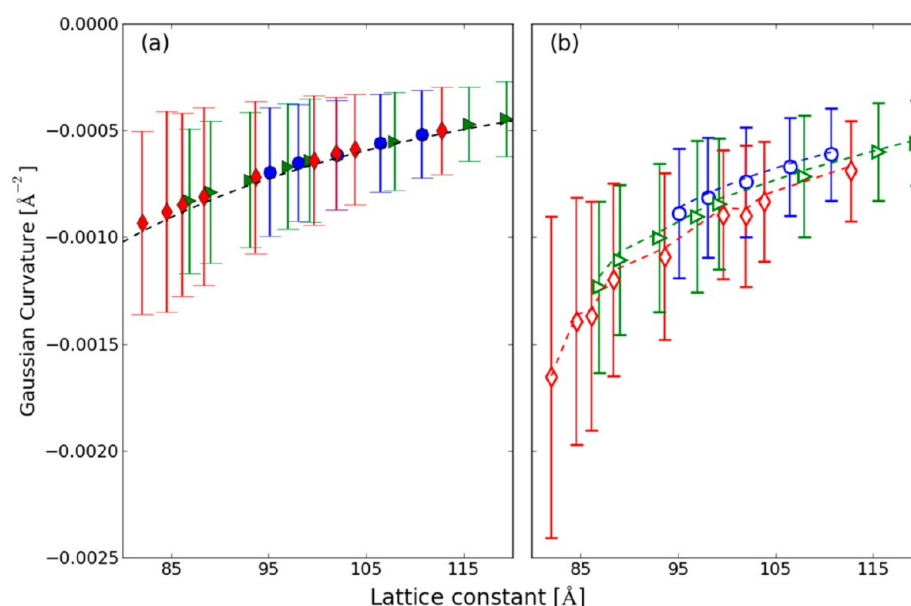

Figure 3. Gaussian curvature of the membrane midplane (a) and interface (b), averaged over the unit cell, are shown as a function of the lattice constant for different lipids (blue circles, 7.7 MAG; green right triangles, 7.9 MAG; and red diamonds, 9.9 MAG). The error bars measure the width of the distributions of Gaussian curvatures. Lines are theoretical values from eqs 4 and 6 as described in the text.

$\langle H \rangle_0$ averaged over the LCP bilayer midplane is ∼0 for all the self-assembled mesophases (Figure 1 and Figure S4 for the principal curvatures), and the trends in A_0 (Figure 2a) and $\langle K \rangle_0$ (Figure 3a) calculated from the different computationally modeled LCPs are nearly identical to their theoretical estimates (eqs 3 and 4). This suggests that the bilayer midplane of each simulated system traces the minimal surface of $Pn3m$ symmetry.

Assuming a constant monolayer thickness (l), the analytical formulations also relate the surface area (A_l) and the mean ($\langle H \rangle_l$) and Gaussian ($\langle K \rangle_l$) curvatures measured on the lipid/water interface of the $Pn3m$ phase to those computed at the LCP bilayer midplane:[35,36]

$$A_l = (1 + l^2 \langle K \rangle_0) A_0 \tag{5}$$

$$\langle K \rangle_l = \frac{\langle K \rangle_0}{1 + l^2 \langle K \rangle_0} \tag{6}$$

$$\langle H \rangle_l = l \langle K \rangle_l \tag{7}$$

In the above, $\langle \cdots \rangle_l$ is the average taken over a surface at a distance l from, and parallel to, the LCP midplane (the derivation of these equations is given in Supporting Information).
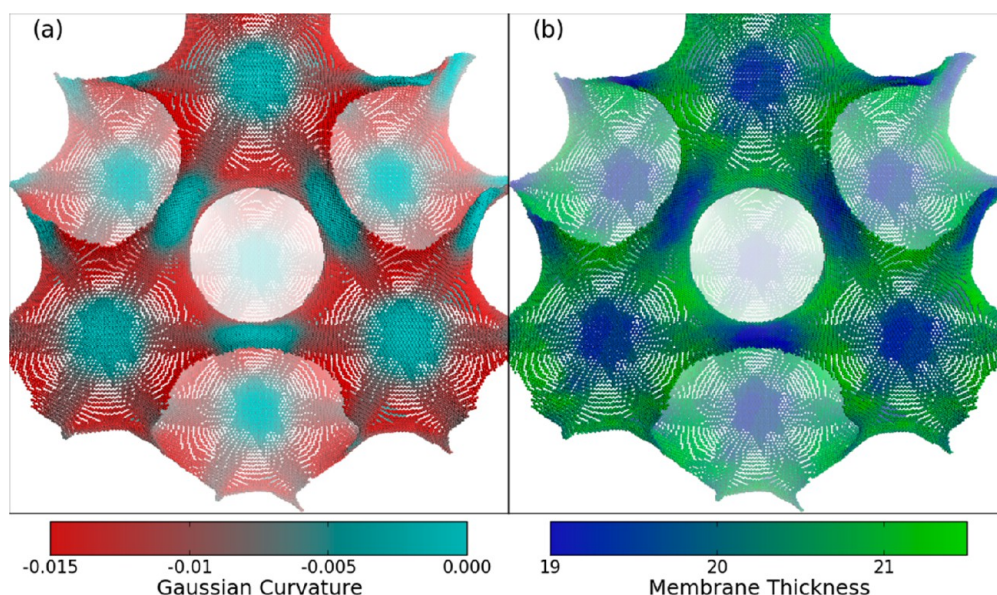
Figure 4. Midplane surface for the 9.9 MAG LCP with lattice constant a = 82 Å colored according to the Gaussian curvature (a) and according to the membrane monolayer thickness (b). Note, that the regions with Gaussian curvature $K \approx 0$ also have thinner membrane.

As illustrated in Figures 1, 2b, and 3b, the values for A_l, $\langle H \rangle_l$ and $\langle K \rangle_l$ computed for the self-assembled LCPs using the bilayer thickness values determined for the respective systems (shown in Figure S5) are also in excellent agreement with the analytical predictions (eqs 5−7). Not surprisingly, the calculations on the lipid/water interfacial surfaces suggest that the geometric properties of these surfaces are strongly dependent on the thickness l of the LCP monolayer. This is evidenced by the curves for different MAG lipids in Figures 1, 2b, and 3b, which can no longer be described by a common functional relationship (unlike the measurements on the midplanes shown in Figures 1, 2a, and 3a). In particular, for a fixed lattice constant, longer-tail lipids lead to larger (in magnitude) curvatures but smaller SAs on the interfaces, as illustrated for example in Figures 1, 2b, and 3b for 9.9 MAG and 7.7 MAG LCPs, both at $a \approx 102$ Å. Similarly, the data show the general trend that, compared to the LCP bilayer midplane, the lipid/water interface of the $Pn3m$ mesophase is characterized by a smaller SA but larger (in magnitude) mean and Gaussian curvatures.

Importantly, we find that the quantities reported in Figures 1−3 and S4−S5 are not constant throughout the unit cell of the LCP but rather vary from one location of the mesophase to another. The extent of the variations is quantified by the size of the error bars on the graphs in Figures 1−3 and Figures S4−S5 (see Methods) and is illustrated in Figure 4a where we plot the distribution of the Gaussian curvature on the bilayer midplane of the computationally assembled $Pn3m$ LCP with a lattice constant of 82 Å. In addition, Figure 4b shows the changes in the local thickness of the LCP bilayer in the same system. These plots reveal that the membrane is thinnest at the saddle points of the cubic phase, where the Gaussian curvature is minimal (∼0). Interestingly, these are the regions of the LCP where the reconstituted membrane proteins are thought to reside[14,37] (see below). Conversely the areas of larger Gaussian curvatures correspond to the regions of the LCP where the membrane is thicker (see Figure 4). This correspondence can also be seen on the lipid/water interfacial surface (see Figure S6), where the Gaussian and mean curvatures are minimal in

the saddle point regions of the mesophase and increase in regions more distal from those points. Moreover the regions of small Gaussian curvature on the membrane midplane correspond to the locations where the Gaussian curvature on the interface is low as well (Figure S7). Such molecular organization of the LCP implies thickening of the membrane in the regions of higher curvature due to the packing frustration of the lipids in these regions of the cubic phase, as discussed in previous theoretical works (see for example ref 38).

The analysis of the geometrical characteristics presented above validates the computationally self-assembled MAG/water systems by showing that it exhibits all the expected structural characteristics of the $Pn3m$ mesophase; it also demonstrates the numerical accuracy of our methodological tools and the ability to quantify various geometric properties of the LCP locally inside the cubic phase unit cell.

Protein Dynamics in the LCPs. To characterize, with computational modeling, the dynamics and arrangement of GPCRs in environments relevant to *in meso* crystallization trials, we simulated different GPCR protein constructs inserted in $Pn3m$ phases of various lattice sizes. To this end, we reconstituted (as described in Methods) the $A_{2A}R$ protein in LCPs of lattice constants a = 102, 86, and 82 Å, and the $A_{2A}R$-BRIL construct in LCPs with a = 102 and 113 Å (see Table S1 for details on the simulation setups and Methods for the protein constructs), corresponding to specific conditions encountered by each of these proteins during the crystallization assays. These protein/LCP complexes were then studied in microsecond long CGMD simulations (see Methods for details). For both the $A_{2A}R$ and $A_{2A}R$-BRIL constructs, the 102 Å $Pn3m$ phase is representative of the starting experimental setup for *in meso* trials of the LCP (at room temperature and without precipitant). The other simulations, of the $A_{2A}R$ in the 86 and 82 Å cubic phases and of the $A_{2A}R$-BRIL construct in the 113 Å LCP, mimic the effects of various precipitants used during *in meso* trials to initiate crystallization (see Discussion).

$A_{2A}R$ in 9.9 MAG LCPs. Figure 5 shows the equilibrium positioning of the $A_{2A}R$ in 9.9 MAG LCPs and reveals that the local environment created by the LCP around the inserted
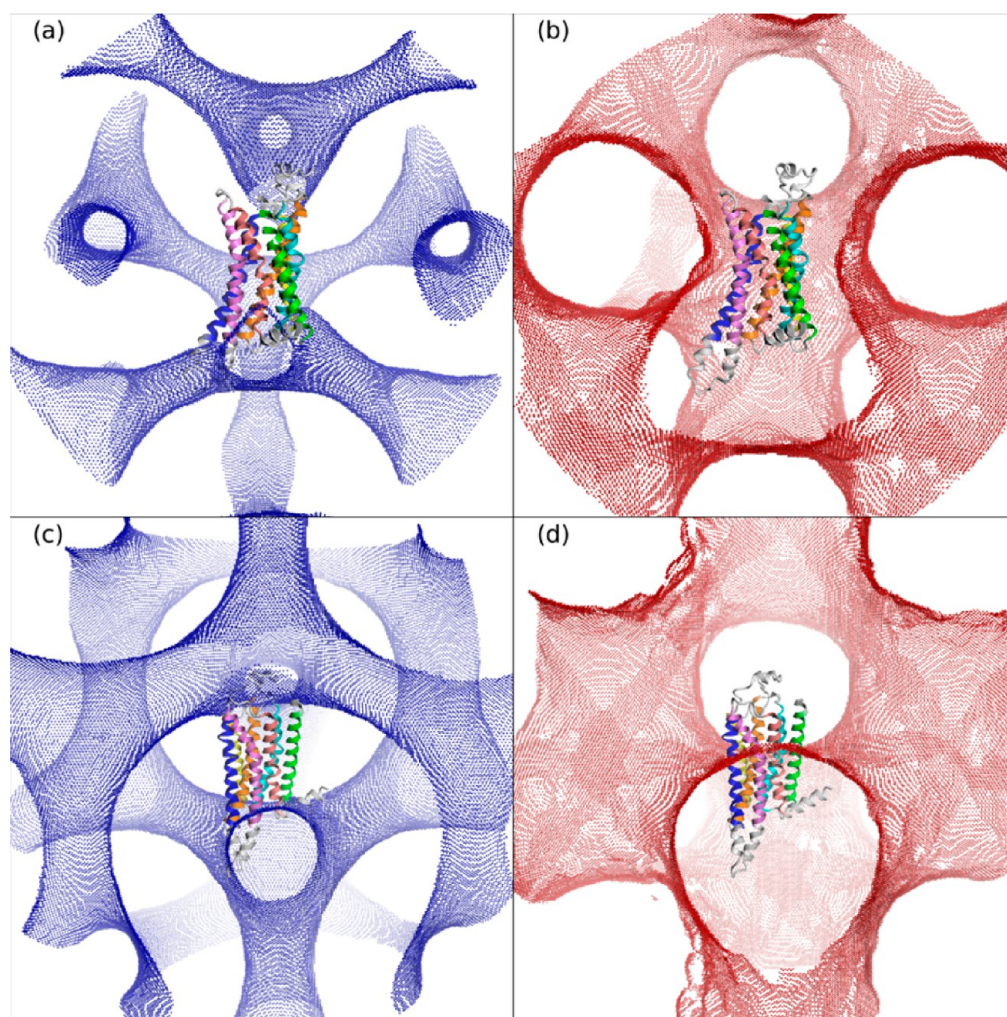
Figure 5. Arrangements of the $A_{2A}R$ (cartoon) in LCPs of 82 Å (panels a and b) and 102 Å (panels c and d) lattice constants. The midplane is shown in red and the interface in blue. In the 102 Å LCP, the protein is positioned in a "canonical" manner where it traverses the membrane midplane with its intracellular and extracellular sides protruding into the nodes where 4 water channels meet. In contrast, in the smaller 82 Å LCP, the GPCR breaks the water network disrupting one of the arms of the water channel tetrahedron and has the cubic phase bilayer midplane wrapped around it. Surfaces shown are from the analyses using the global alignments (all 27 unit cells) of the trajectories, and the cartoons are an all-atom model of the $A_{2A}R$ superposed onto the structure of the CG protein average over the trajectory (see Methods, colored as: TM1 in green, TM2 in cyan, TM3 in orange, TM4 in yellow, TM5 in blue, TM6 in light purple, and TM7 in salmon and white for the loops and helix 8.

protein is different in the LCPs of different lattice constants studied. Indeed, in the cubic phase with $a = 102$ Å (Figure 5c,d), the protein is situated in a "canonical" monkey saddle point,[14] where the Gaussian curvature of the midplane surface is minimal. In this position, the protein is in an environment resembling a flat lamellar membrane where the midplane of the lipid bilayer perpendicularly traverses the hydrophobic core of the protein (Figure 5d), and the polar loop regions of $A_{2A}R$ are exposed to the water compartments of the LCP (Figure 5c). Interestingly, the calculations of the local structural properties of the cubic phase around the GPCR reveal that $A_{2A}R$ can be accommodated at the monkey saddle point of the 102 Å LCP with minimal perturbation to the cubic phase matrix. This is illustrated in Figure 6c, which shows that the mean curvature on the LCP midplane around the inserted protein is near zero, as in the unperturbed LCP (Figure 1). The positioning of the GPCR in the cubic phase of $a = 86$ Å is similar to that in $a = 102$ Å, although with slightly higher perturbations to the structure of the LCP around the protein (Figure 6b,e). However, our simulations reveal that in the smallest LCP,

with a lattice spacing of 82 Å, the $A_{2A}R$ adopts a dramatically different position (Figure 5a,b). Indeed, the GPCR in this small cubic phase no longer traverses the membrane midplane (compare Figures 5b and 5d). Instead it breaks the water network of the LCP by displacing one of the arms of the water channel tetrahedron (compare Figures 5a and 5c). This unusual arrangement was confirmed in a repeat trajectory (initiated with a different random seed) as well as in CGMD simulations of another multihelical TM protein, the Leucine transporter (LeuT), inserted in the same LCP ($a = 82$ Å, data not shown). The characteristic of this arrangement is that the bilayer midplane is wrapped around the TM protein, with the lipids oriented nearly perpendicular to the protein's long axis and partly organized as a monolayer. For the $A_{2A}R$, lipid tails cover the hydrophobic surfaces of the TMs 2, 5, and 6, while the rest of the $A_{2A}R$ faces lipids forming a bilayer, with the headgroups facing the hydrophobic region of TMs 1, 4, and 7 (Figure S8 and Supporting movie 2). This arrangement results in unfavorable hydrophobic−hydrophilic interactions between the protein and the lipids. This is substantiated by the
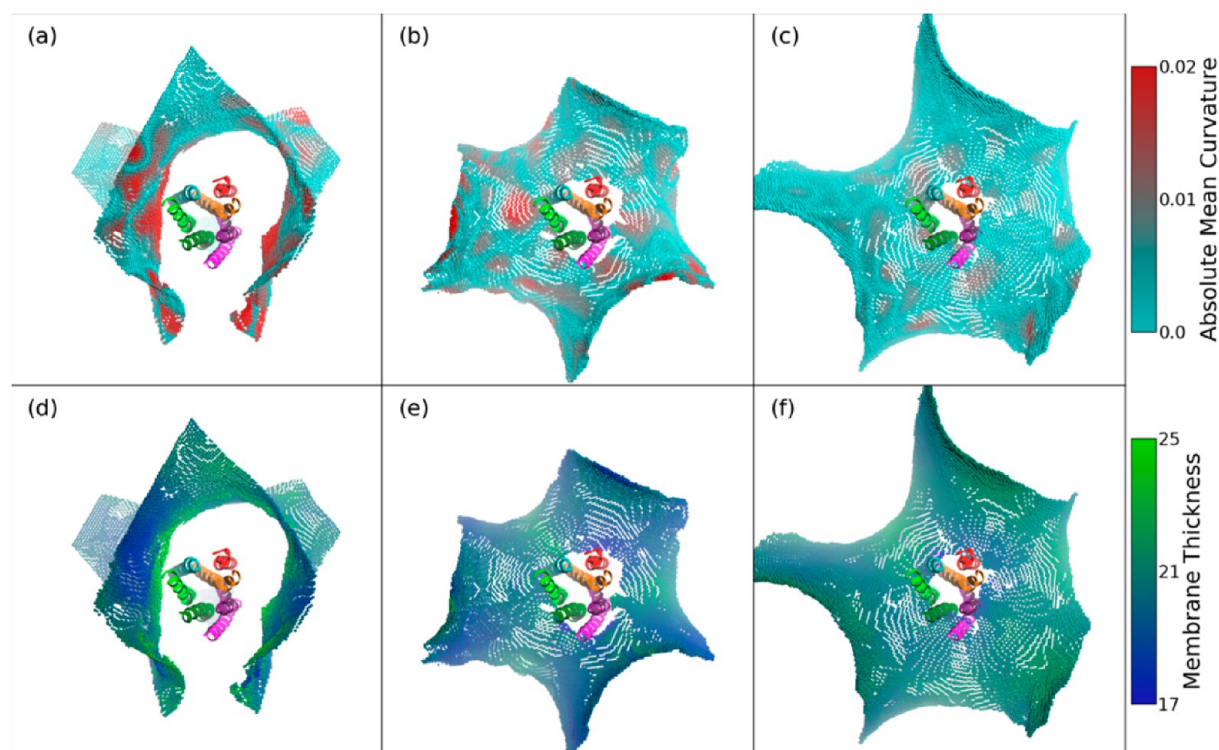
Figure 6. Membrane midplane around the $A_{2A}R$ for the LCP with lattice constant of 82 Å (a,d), 86 Å (b,e), and 102 Å (c,f) respectively. (a−c) Surface colored according to the absolute value of the mean curvature. (d−f) Surface colored according to the membrane monolayer thickness. Cartoons of the receptor are average structures as described in Figure 5 and Methods and colored as: TM1 in green, TM2 in cyan, TM3 in orange, TM4 in yellow, TM5 in blue, TM6 in light purple, and TM7 in salmon.

calculated SAs of the hydrophobic cores of TMs 1, 4, and 7 accessible to the water and lipid headgroup beads in the 82 Å LCP, compared to those in the larger cubic phases and in CGMD simulations of the $A_{2A}R$ in a 9.9 MAG lamellar lipid bilayer. Results in Table 1 show that the SA values for the TM1, TM4, and TM7 segments calculated in the small LCP are indeed high and significantly larger than those obtained in the 86 and 102 Å cubic phases and in the lamellar membrane, suggesting strong unfavorable interactions between the GPCR and its environment in the 82 Å LCP.

Table 1. SASA (in Å²) of Hydrophobic Residues in the Core of TM1 (Tyr9$^{1.35}$-Trp29$^{1.55}$), TM4 (Ala121$^{4.42}$-Pro139$^{4.60}$), and TM7 (Met270$^{7.35}$ to Tyr288$^{7.53}$) in $A_{2A}R^a$

	SASA (in Å²) of $A_{2A}R$			
unit cell	82 Å	86 Å	102 Å	lamellar bilayer
TM1	70	5	15	8
TM4	96	4	4	1
TM7	60	11	6	15

^{a}SASAs were calculated using MSMS[39] with a probe radius of 2.35 Å with the protein and lipid tails (excluding first ETH bead, Figure S1) as solute.

As shown in Figure 6a,d, we find that the above-described specific mode of interactions between the $A_{2A}R$ and the lipids in the 82 Å LCP is accompanied by important local deformations of the LCP bilayer near the inserted protein. Indeed, in striking contrast with the larger LCPs where the GPCR was accommodated with minimal perturbation to the mesophase (see Figure 6b,c), the lipid bilayer of the 82 Å LCP becomes significantly curved (Figure 6a) and thicker by 3−4 Å

(measured as the monolayer thickness close to the protein. The structural changes are especially significant around TMs 2, 5 and 6, corresponding to the regions of the protein that interact with the lipids forming a monolayer. Consequently, in order to accommodate the $A_{2A}R$, the membrane in this region of the 82 Å LCP would have to both bend and thicken, to enable the hydrophobic tails to cover TMs 2, 5, and 6 in order to alleviate the unfavorable exposure of these TM regions of the receptor to the otherwise hydrophilic environment. Interestingly, although the mesophase is less perturbed around TMs 1, 4, and 7 of the $A_{2A}R$, where the lipids form a bilayer, the protein hydrophobic surface at these TMs remains unfavorably exposed to the solvent and lipid headgroups (Table 1), suggesting that the 82 Å LCP is unable to adapt sufficiently to properly accommodate the receptor. Overall, the above analyses suggest that it is energetically less favorable for $A_{2A}R$ monomers to reside in the LCP of the smallest lattice constant (82 Å) compared to the larger ones (86 and 102 Å), where they are well accommodated.

$A_{2A}R$-BRIL in 9.9 MAG LCPs. The $A_{2A}R$-BRIL construct was studied with CGMD simulations in LCPs with lattice constants of 102 and 113 Å. In the smaller LCP (a = 102 Å), the complex equilibrates in the canonical monkey saddle point, similar to $A_{2A}R$ construct in the same LCP, with its BRIL attachment protruding into one of the water channels (see Figures 7a,b). This arrangement of the $A_{2A}R$-BRIL is well tolerated by the lipidic environment leading to minimal structural perturbations of the mesophase (see Figure 7c).

However, our simulations of $A_{2A}R$-BRIL in the larger LCP (a = 113 Å) revealed several sources of instability for this protein complex bound to an enlarged cubic phase. Specifically, out of four initial simulations of the $A_{2A}R$-BRIL in the 113 Å LCP
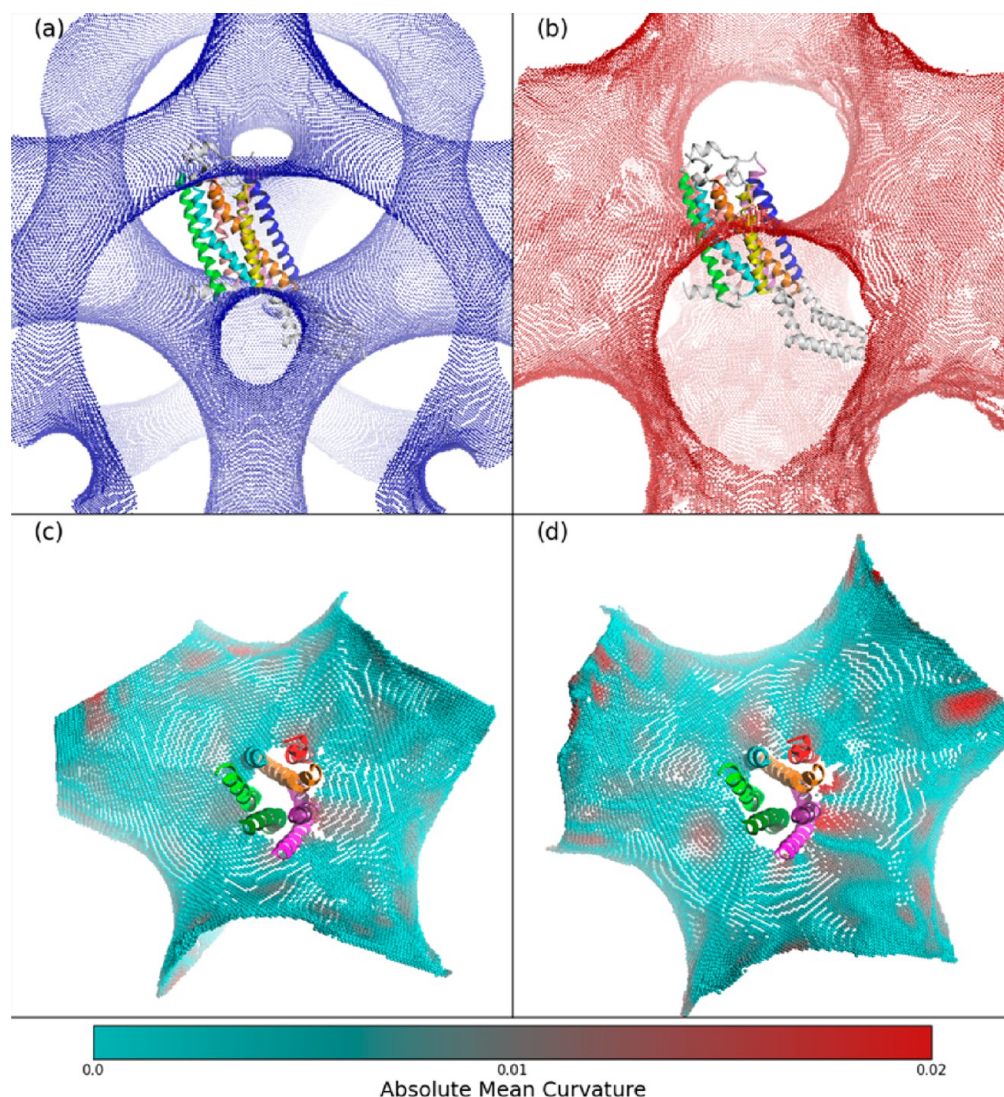
Figure 7. (a,b) Water−lipid interface and the membrane midplane, respectively, around the A_{2A}R-BRIL in the 102 Å LCP. (c,d) Midplane surface of the 102 and 113 Å LCPs, respectively, colored according to the absolute value of the mean curvature. Cartoons are an all-atom model of the A_{2A}R-BRIL superposed onto the structure of the CG protein averaged over the trajectory (see Methods) and colored as: TM1 in green, TM2 in cyan, TM3 in orange, TM4 in yellow, TM5 in blue, TM6 in light purple, and TM7 in salmon and loops, HX8, and the BRIL extension in white.

(initiated from different random seeds and placements of the protein complex in the cubic phase), three converged (after 1 μs CGMD) to a positioning of the protein in the LCP very similar to that described above for the A_{2A}R in the 82 Å LCP. In those simulations TM4 was substantially exposed to the solvent and the lipid polar headgroups, with SAs of 200−300 Å^2 (compare to SA data for A_{2A}R in Table 1), and in two of them we observed very large deformations of the protein, notably in the positioning of the BRIL relative to the rest of the GPCR (backbone root-mean-square deviation (RMSD) in the range of 8.4−12.6 Å relative to the initial structure when aligned on the whole protein). The fourth simulation did not converge to a stable position during the 1 μs simulation and also showed substantial exposure to solvent of the TM regions (total of 80 Å^2 for TMs 2, 4, and 7 compared to ∼20 Å^2 for the simulation of the A_{2A}R-BRIL or A_{2A}R in the 102 Å LCP).

To further examine the behavior of A_{2A}R-BRIL in the large cubic phase, we carried out two additional CGMD simulations of A_{2A}R-BRIL in the 113 Å LCP (differing in initial random seed), where in the starting configuration the A_{2A}R-BRIL

construct was ideally positioned in the monkey saddle point, with its BRIL attachment hydrated in one of the water channels (Figure 7a,b). Interestingly, in one of these two simulations, the protein again deformed substantially, with its BRIL extension completely changing its position relative to the rest of the protein (backbone RMSD of 10.6 Å relative to the initial structure). But in the other simulation, the protein was more stable (RMSD of 7.2 Å) and remained positioned in the monkey saddle point of the mesophase during the microsecond-long simulation. We therefore used this trajectory for further structural analyses of the A_{2A}R-BRIL complex in the LCP with lattice constant of 113 Å. But even in this case we found (Figure 7c,d) that the deformations of the LCP's bilayer midplane close to the protein were larger than in the smaller LCP (a = 102 Å). Taken together, these results suggest that the A_{2A}R-BRIL monomer is much better accommodated in the LCP with a = 102 Å than in the larger one with a = 113 Å.

Residual Hydrophobic Mismatch of A_{2A}R in Lamellar Membranes of MAG Lipids. So far our results have suggested that significant deviations in the size of the protein-

embedded LCP from that measured at room temperature (a = 102 Å) result in unfavorable interactions between the mesophase and the embedded proteins. These unfavorable interactions arise from perturbations of the matrix around the insertions, from deformations of the protein, and/or from exposure of the hydrophobic surface of the protein to the hydrophilic environment. Such protein−LCP interactions are energetically costly and could, in principle, drive the segregation of the proteins into planar lipid bilayers (see Discussion), where the crystallization is thought to take place. Since unfavorable hydrophobic−hydrophilic interactions between membrane proteins and their environment (hydrophobic mismatch) have been shown to be an important drive for the oligomerization of the proteins in lamellar bilayers[16,20] we quantified the cost of such hydrophobic mismatch from CGMD simulations of the $A_{2A}R$ in lamellar membranes of 9.9 MAG lipids and using the previously established CTMD (continuum MD) protocol[15−17] (see Methods). The results, reported in Table 2, show that the overall residual hydrophobic mismatch

Table 2. Residual Mismatch Energies (in k_BT units) for Each TM Segment and the Entire TM Bundle of $A_{2A}R$ in 9.9 MAG, 7.9 MAG, and 7.7 MAG Lamellar Membranes

	9.9 MAG	7.9 MAG	7.7 MAG
TM1	1.3	2	7.6
TM2	0	0	0
TM3	0	0	0
TM4	0	2.1	0
TM5	1.4	3	1.4
TM6	0	2.8	2.2
TM7	1.2	1.6	1.6
total	3.9	11.5	12.8

(RHM) energy for the TM bundle of the $A_{2A}R$ is ∼4 k_BT, consistent with the value previously reported for another class-A GPCR, rhodopsin, in 9.9 MAG planar membranes.[15] Here, we found that the $A_{2A}R$ TM segments with the highest residual mismatch penalties were TM1 (energy penalty of 1.3 k_BT at Pro2$^{1.28}$), TM5 (1.4 k_BT at Asn175$^{5.36}$), and TM7 (1.2 k_BT at Leu267$^{7.32}$). No other residue in the TM region of the protein showed an RHM energy >1 k_BT.

Since the hydrophobic mismatch effects have been shown to depend on the hydrophobic thickness of the embedding planar lipid membrane,[16] and with the goal of testing the generality of the interfaces determined above in different bilayers, we also simulated with CGMD the $A_{2A}R$ in the shorter tail 7.9 and 7.7 MAG planar lipid membranes. Comparison of the mismatch energy penalties for this GPCR in different MAG lipid membranes showed that the RHM energy in 7.7 and 7.9 MAG membranes is ∼12 k_BT, i.e. ∼8 k_BT higher than in 9.9 MAG bilayers (see Table 2). The drive for oligomerization at TMs 1, 5 and 7, although present in all three membranes, is much stronger at TM1 in 7.7 MAG (∼7.5 k_BT). Differences between the shorter tail lipids and 9.9 MAG lipids were notable at TM6, where a residual mismatch energy of ∼2.5 k_BT was calculated in both 7.9 and 7.7 MAG membranes, and at TM4, where the energy penalty was ∼2 k_BT in the 7.9 MAG bilayer. Overall these results suggest larger energy costs of the RHM for the $A_{2A}R$ in lamellar bilayers of shorter tail MAG lipids and a consistent (for all MAG lipids) drive for oligomerization at TM1, TM5, and TM7.

■ DISCUSSION

The results presented here from the large-scale CGMD simulations of $A_{2A}R$ and $A_{2A}R$-BRIL in 9.9 MAG LCPs of various sizes, as well as of $A_{2A}R$ in planar lamellar membranes, offer novel insights into the organization of these protein constructs in different lipidic environments relevant to the *in meso* crystallization process for membrane proteins. Importantly, as we discuss below, these results illuminate key molecular mechanisms that make the *in meso* method so successful for the crystallogenesis of GPCRs. On this basis, our current understanding is that the addition of precipitants, which is of fundamental importance in initiating the formation of protein crystals during *in meso* trials, will change the lattice constant of the LCP, which drives the formation of, or the partitioning of the proteins into, stacked lamellar bilayers connected to the LCP, where the crystals are formed.

Crystallization of protein constructs with specifically engineered bulky polar regions, such as the $A_{2A}R$-BRIL studied here, generally appears to require the action of precipitants that substantially (>25%) enlarge the LCP. However, for TM proteins with more compact water-soluble segments (for example, bacteriorhodopsin, a 7-TM protein structurally similar to class-A GPCRs) the *in meso* crystallization succeeds with precipitants that shrink (by as much as 20%) the unit cell of the LCP. These experimental setups were explored here by considering two model protein constructs, the $A_{2A}R$ and the $A_{2A}R$-BRIL complex, in LCPs with different unit cell sizes. With these conditions, we sought to identify from the analysis of the simulations, the molecular mechanisms that would oppose the presence of proteins inside the cubic phase as monomers and thus support protein−protein interactions.

The generation of high-quality crystals during *in meso* experiments is strongly determined by the formation of two complementary types of interprotein interactions:[10] 2D lateral protein aggregation in the lipid bilayer plane and 3D stacking interactions between the proteins. Therefore, our specific goal was to identify from the simulations specific conditions that relate to these steps. As discussed below, the specific results obtained from the simulations of the various constructs studied here reveal the conditions and processes that produce these effects, which are key to crystallogenesis *in meso* and pertain to the rational design of future membrane protein crystallization experiments.

■ MOLECULAR MECHANISM OF THE FORMATION OF PROTEIN STACKS

$A_{2A}R$-BRIL Constructs. Our computational results showed that in the LCP, at room temperature and prior to addition of precipitant (a = 102 Å), both the $A_{2A}R$ and the $A_{2A}R$-BRIL constructs are well accommodated in the region of the mesophase with the minimal Gaussian curvature (monkey saddle point), where the lipidic environment around the inserted proteins resembles a lamellar bilayer. For the $A_{2A}R$-BRIL construct, a key structural feature of this specific arrangement is the complete exposure of the soluble BRIL extension into one of the water channels of the cubic phase. We hypothesized that, in the presence of sufficiently high concentration of proteins (*in meso* trials are generally conducted at protein concentrations at least 10 times higher than those simulated here, which would lead to close to 1 protein per unit cell), the observed protrusion into the water compartment of the BRIL from one $A_{2A}R$-BRIL protein could
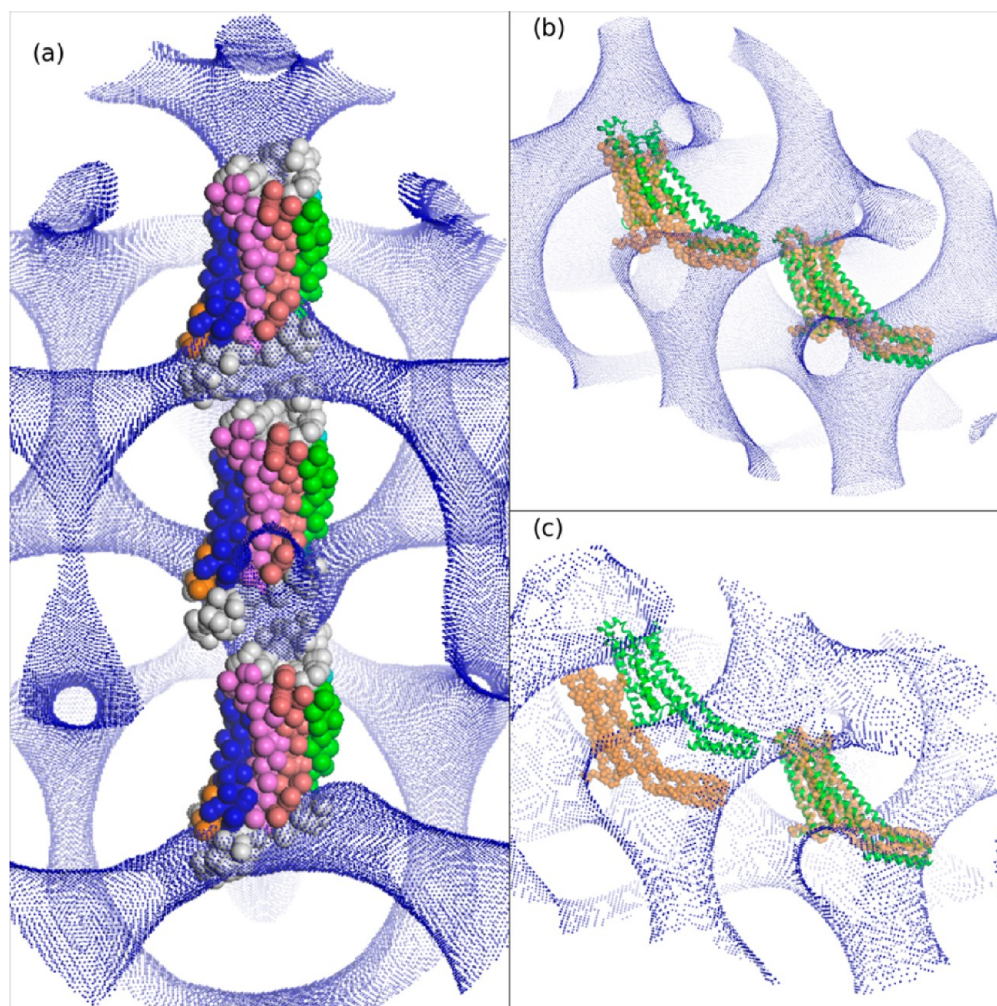
Figure 8. Illustration of contacts that could be formed by neighboring proteins in the LCP. Interfacial surfaces (determined from the trajectories aligned on the whole simulation box, i.e., the 27 unit cells) are shown in blue. (a) Structure of the $A_{2A}R$ (averaged over the last 1 μs of the simulations) in the 82 Å LCP (first protein form the top of the image) and two images displaced by half a unit cell (protein in the middle) and a whole unit cell (bottom protein), respectively. The bottom protein shows where a protein in a neighboring unit cell would be in the 82 Å LCP, whereas the central protein would be occupying the canonical binding site (monkey saddle point, where the protein resides in larger LCPs). (b,c) Structure of the protein (averaged over the last 1 μs of the simulations) in the LCP with $a = 103$ and 113 Å, respectively, and one copy of this structure displaced by one unit cell vector. In green are two proteins from the $A_{2A}R$ crystal packing (PDB ID: 4EIY), superposed on one of the CG proteins.

lead to strong intermolecular interactions with the polar regions of the neighboring $A_{2A}R$-BRIL complex en route to the formation of 3D protein stacks.

Indeed, examination of the crystallographic unit cell of the $A_{2A}R$-BRIL (PDB ID: 4EIY)[18] reveals crystal contacts between the stacks of $A_{2A}R$-BRIL proteins achieved through intermolecular interactions that involve the BRIL of one protein and the extracellular loop 2 of another (see Figure 8b). The feasibility of the formation of such protein–protein contacts from the specific arrangement of $A_{2A}R$-BRIL seen in our simulations is demonstrated in Figure 8b. This figure illustrates the striking similarity of both the relative positions of the protein and in the interprotein contacts between the $A_{2A}R$-BRIL crystal and the computationally predicted arrangement of $A_{2A}R$-BRILs.

Importantly, our results showed that in the larger LCP ($a = 113$ Å) the $A_{2A}R$-BRIL is also stable in the monkey saddle point, but the neighboring proteins in such an enlarged mesophase will not be sufficiently close in space to engage in the type of stacking interactions seen in the $A_{2A}R$-BRIL crystal

(see Figure 8c). As this lack of contact in the 113 Å LCP emerges both from the large unit cell size and a different orientation of the GPCR relative to the normal of the membrane plane it traverses (cf. Figures 8b,c and S9), this analysis suggests that the 3D stacking interactions between $A_{2A}R$-BRIL constructs formed during *in meso* experiments are likely to originate at the stage of the reconstitution of the proteins in the cubic phase at room temperature, prior to any precipitant action. In fact, the analysis indicates that the addition of precipitants that significantly enlarge the cubic phase will oppose these stacking interactions.

Importantly, this key finding rationalizes the size, shape, and arrangement of a hydrophilic insertion best-suited to promote crystallization in the LCP. Indeed, we find that the stacking contacts form in the LCP environment and are mediated by the soluble regions of the protein that extend into the water channels (toward a neighboring protein in an adjacent unit cell of the simulated LCP). These results suggest that the engineered soluble extension must extend somewhat diagonally to allow contacts with the protein in the neighboring monkey
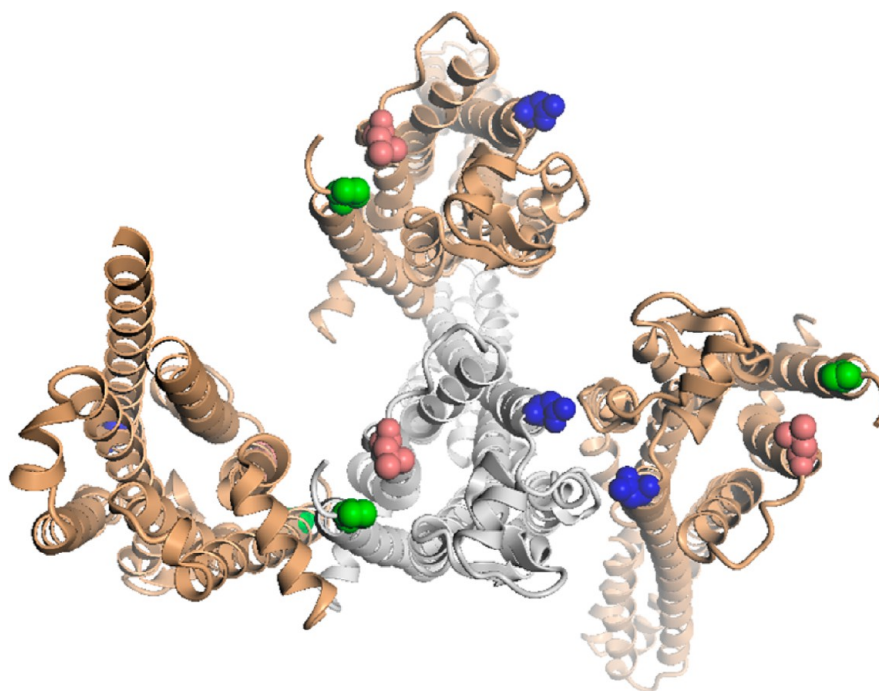
Figure 9. Intermolecular packing in the A$_{2A}$R-BRIL crystal structure (PDB ID: 4EIY).[18] Residues identified in the CGMD as having significant hydrophobic mismatch are shown in space-filling representations: Pro2$^{1.28}$ in TM1 is shown in green, Asn175$^{5.36}$ in TM5 in blue, and Leu276$^{7.32}$ in TM7 in red. Leu276$^{7.32}$ is buried in a symmetrical TM4-TM5 interface, Pro2$^{1.28}$ in a symmetrical TM1-TM7 interface, showing a remarkable agreement between the locations of the hydrophobic mismatch and the oligomeric interfaces in the crystal structure.

saddle point, because the simulations of the A$_{2A}$R-BRIL construct reveal that neighboring proteins are not lined up vertically (along the main axis of the membrane protein, see Figure 8). We note that for nearly all the GPCR constructs that led to successful crystallization in the enlarged LCPs (apart from BRIL, most notably the T4 Lysozyme fused to the IL3),[40] the structural features satisfy this condition.

A$_{2A}$R Constructs. The results for the A$_{2A}$R in the LCP with a = 102 Å show that under room temperature conditions and without precipitants, these proteins would not form stacking interactions, even though they reside in the monkey saddle regions of the cubic phase (Figure 5c,d). From the mechanistic picture discussed above, it appears that this can be attributed to the fact that A$_{2A}$R itself (without the BRIL) does not possess large polar segments that could protrude sufficiently deep into the water channels separating neighboring monkey saddle points of the cubic phase. We reasoned, therefore, that the crystal-like stacking contacts between the A$_{2A}$R proteins must originate from the precipitant's effects. Since A$_{2A}$Rs structurally resemble bacteriorhodopsin[41] (i.e., both share a 7-TM architecture and possess compact hydrophilic loop regions), which was successfully crystallized using additives that shrink the cubic phase lattice,[13,42] we probed this conjecture in the simulations comparing the arrangement of the A$_{2A}$Rs in LCPs with 102 and 82 Å lattice constants. Indeed, the A$_{2A}$R in the small (82 Å) cubic phase equilibrates in an arrangement which is dramatically different from that found in the room temperature *Pn3m* mesophase: instead of residing in the monkey saddle point, the GPCR assumes a position that would be occupied by one of the arms of the water channel tetrahedron. This finding suggests an intriguing hypothesis for the emergence of the stacking interactions between A$_{2A}$Rs in an LCP under the influence of additives that decrease its lattice constant: as the precipitant mediates the shrinkage of the LCP

and the preferred binding position for the GPCR shifts from the monkey saddle point to the water channel, the two poses will coexist for the membrane protein in the LCP. At sufficiently high protein concentrations, this would give rise to juxtapositions of A$_{2A}$R molecules arranged in the monkey saddle points (as seen in the 102 Å LCP) and the water channels (as found in the 82 Å LCP). This hypothetical scenario, illustrated in Figure 8a, would directly lead to stacking of GPCRs in the direction perpendicular to the membrane plane of the GPCR sitting in the monkey saddle point. As seen in the figure, the GPCR constructs without any large soluble extension on both the extracellular and intracellular side would be just at the right distance from each other to form contacts if both binding poses were to coexist at some transitional state between the 102 and 82 Å LCPs.

■ MECHANISM OF PROTEIN OLIGOMERIZATION IN THE MEMBRANE PLANE

Segregation of the Proteins into Stacked Planar Lipid Bilayers. Along with 3D stacking interactions, protein aggregation in the 2D lipid bilayer plane is a requirement for producing high-quality crystals for structure determination. Indeed, crystallographic unit cells of GPCR protein constructs[18,41] obtained with the *in meso* method reveal strong intermolecular interactions between TM segments of neighboring proteins in the membrane plane. Our simulations in small (a = 82 Å) and enlarged (a = 113 Å) *Pn3m* mesophases have identified several molecular mechanisms that could lead to destabilization of monomeric proteins in these specific cubic phase environments, notably large deformations of the lipid matrix (Figures 6a and 7d) accompanied by unfavorable hydrophobic−hydrophilic interactions between the protein and the LCP (Table 1) and large deformations of the proteins (in most simulation of the A$_{2A}$R-BRIL in the 113 Å LCP).

3282

dx.doi.org/10.1021/ja4129839 | *J. Am. Chem. Soc.* 2014, 136, 3271−3284

Moreover, for the $A_{2A}R$-BRIL construct, increase of the lattice constant is likely to break the stacking contacts previously formed in the LCP under the initial experimental conditions (prior to addition of the precipitant). This destabilization of the protein−LCP complex could lead to the formation of, and/or the segregation of the proteins into, stacked lipid bilayers, where the proteins could aggregate and form in-plane contacts.

Formation of crystal contacts in the membrane plane. Since it has been previously shown by us[16,20] and others[43] that unfavorable hydrophobic−hydrophilic protein−membrane interactions play an important mechanistic role in driving oligomerization of membrane proteins in the lipid bilayer plane, calculations of the corresponding energy cost for the $A_{2A}R$ in lamellar bilayers of 9.9, 7.9, and 7.7 MAG lipids can be used to identify preferred oligomerization interfaces.[16,20] Assuming that the in-plane oligomerization takes place in the lamellar bilayer environment during crystallization, we compared the interfaces predicted from our residual mismatch analysis to the interfaces observed in the packing of the $A_{2A}R$-BRIL crystal structure (Figure 9). The most significant residual mismatch calculated in the 9.9 MAG membrane, which was used in the crystallization of the $A_{2A}R$-BRIL (PDB ID: 4EIY)[18] occurred at TM1 (Pro2[1.28]), TM5 (Asn175[5.36]), and TM7 (Leu267[7.32]) of $A_{2A}R$ (Table 2). These results are in remarkable agreement with the lateral contacts between proteins in the $A_{2A}R$-BRIL crystal structure (PDB ID: 4EIY, see Figure 9). The findings provide support for the mechanistic hypothesis that the in-plane crystal packing takes place in the context of the stacked planar bilayers as well as for the concept that oligomerization is largely driven by the hydrophobic mismatch between the reconstituted proteins and lipid environment.[16,17,20]

CONCLUDING REMARKS

Regarding the central mechanistic question of the use of precipitants in the crystallization process to either increase or decrease the lattice constant of the LCP, the detailed findings and interpretation of the results from this work suggest two specific mechanistic paths: (i) When the precipitants increase the lattice constant, stacking contacts are formed between proteins in neighboring unit cells of the LCP prior to the addition of the precipitant, through a water-soluble extension, these contacts are maintained during the rest of the crystallization process. (ii) When precipitants are used that shrink the LCP unit cell, the stacking contacts are formed after addition of the precipitant and are realized by interactions between proteins coexisting in the monkey saddle points and in their alternative location previously occupied by the water channel.

In both of these mechanistic paths, the protein is well accommodated in the LCP under the initial conditions but less so after the addition of the precipitant; this leads to the formation and/or segregation of the proteins into stacked bilayers as discussed above. The lateral contacts are then formed in the bilayer environment and are mainly driven by the residual mismatch between the membrane and the protein, which is alleviated upon formation of the oligomers.[16] Importantly these results suggest that precipitants decreasing the lattice constant could be more practical for *in meso* crystallization of proteins with small soluble regions, whereas precipitants swelling the LCP could be more appropriate for proteins with larger soluble extensions. Moreover, the optimal size and specific configuration of these extensions would allow the formation of stacking contacts in the LCP under initial conditions (prior to action of any additives). These predictions, obtained here for the first time from a detailed quantitative analysis of the underlying mechanisms, can guide the rational choice of experimental conditions for successful outcomes from *in meso* crystallization trials.

ASSOCIATED CONTENT

Ⓢ Supporting Information

Figures show quantifications of various properties of the cubic phases. Two movies and three pdb visualization files illustrate the procedures used for determination of surfaces. Additional methodological details and Martini force field parameters for 9.9, 7.9, and 7.7 MAG lipids. 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.

REFERENCES

(1) Lindblom, G.; Rilfors, L. *Adv. Colloid Interface Sci.* **1992**, *41*, 101.
(2) Dürr, U. H. N.; Gildenberg, M.; Ramamoorthy, A. *Chem. Rev.* **2012**, *112*, 6054.
(3) Caffrey, M. *J. Struct. Biol.* **2003**, *142*, 108.
(4) Wang, H.; Elferich, J.; Gouaux, E. *Nat. Struct. Mol. Biol.* **2012**, *19*, 212.
(5) Caffrey, M. *Annu. Rev. Biophys.* **2009**, *38*, 29.
(6) Cherezov, V. *Curr. Opin. Struct. Biol.* **2011**, *21*, 559.
(7) Katritch, V.; Cherezov, V.; Stevens, R. C. *Trends Pharmacol. Sci.* **2012**, *33*, 17.
(8) Caffrey, M.; Li, D.; Dukkipati, A. *Biochemistry (Moscow)* **2012**, *51*, 6266.
(9) Rasmussen, S. G. F.; 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. A.; 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.
(10) Caffrey, M. *Cryst. Growth Des.* **2008**, *8*, 4244.
(11) Caffrey, M. *Curr. Opin. Struct. Biol.* **2000**, *10*, 486.
(12) Aherne, M.; Lyons, J. A.; Caffrey, M. *J. Appl. Crystallogr.* **2012**, *45*, 1330.
(13) Nollert, P.; Qiu, H.; Caffrey, M.; Rosenbusch, J. P.; Landau, E. M. *FEBS Lett.* **2001**, *504*, 179.
(14) Grabe, M.; Neu, J.; Oster, G.; Nollert, P. *Biophys. J.* **2003**, *84*, 854.
(15) Khelashvili, G.; Albornoz, P. B. C.; Johner, N.; Mondal, S.; Caffrey, M.; Weinstein, H. *J. Am. Chem. Soc.* **2012**, *134*, 15858.
(16) Mondal, S.; Khelashvili, G.; Shan, J.; Andersen, O. S.; Weinstein, H. *Biophys. J.* **2011**, *101*, 2092.
(17) Mondal, S.; Khelashvili, G.; Shi, L.; Weinstein, H. *Chem. Phys. Lipids* **2013**, *169*, 27.

(18) Liu, W.; Chun, E.; Thompson, A. A.; Chubukov, P.; Xu, F.; Katritch, V.; Han, G. W.; Roth, C. B.; Heitman, L. H.; IJzerman, A. P.; Cherezov, V.; Stevens, R. C. *Science* **2012**, *337*, 232.

(19) Cherezov, V.; Clogston, J.; Papiz, M. Z.; Caffrey, M. *J. Mol. Biol.* **2006**, *357*, 1605.

(20) Mondal, S.; Johnston, J.; Wang, H.; Khelashvili, G.; Filizola, M.; Weinstein, H. *Nat. Sci. Rep.* **2013**, *3*, 2909.

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

(22) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S.-J. *J. Chem. Theory Comput.* **2008**, *4*, 819.

(23) Periole, X.; Cavalli, M.; Marrink, S.-J.; Ceruso, M. A. *J. Chem. Theory Comput.* **2009**, *5*, 2531.

(24) Wang, C.; Jiang, Y.; Ma, J.; Wu, H.; Wacker, D.; Katritch, V.; Han, G. W.; Liu, W.; Huang, X.-P.; Vardy, E.; McCorvy, J. D.; Gao, X.; Zhou, X. E.; Melcher, K.; Zhang, C.; Bai, F.; Yang, H.; Yang, L.; Jiang, H.; Roth, B. L.; Cherezov, V.; Stevens, R. C.; Xu, H. E. *Science* **2013**, *340*, 610.

(25) Wang, C.; Wu, H.; Katritch, V.; Han, G. W.; Huang, X.-P.; Liu, W.; Siu, F. Y.; Roth, B. L.; Cherezov, V.; Stevens, R. C. *Nature* **2013**, *497*, 338.

(26) Ballesteros, J. A.; Weinstein, H. In *Methods in Neurosciences*; Sealfon, S. C., Ed.; Receptor Molecular Biology; Academic Press, 1995; Vol. 25, pp 366–428.

(27) R Development Core Team. *R: A language and environment for statistical computing*; R Foundation for Statistical Computing: Vienna, Austria, 2011.

(28) Fuhrmans, M.; Knecht, V.; Marrink, S. J. *J. Am. Chem. Soc.* **2009**, *131*, 9166.

(29) Marrink, S.-J.; Tieleman, D. P. *J. Am. Chem. Soc.* **2001**, *123*, 12383.

(30) Zhang, X.; Li, H.; Cheng, Z.Proccedings from ASIAGRAPH 2008, Shanghai, China, June 27–July 1, 2008.

(31) Biasini, M.; Schmidt, T.; Bienert, S.; Mariani, V.; Studer, G.; Haas, J.; Johner, N.; Schenk, A. D.; Philippsen, A.; Schwede, T. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2013**, *69*, 701.

(32) Biasini, M.; Mariani, V.; Haas, J.; Scheuber, S.; Schenk, A. D.; Schwede, T.; Philippsen, A. *Bioinformatics* **2010**, *26*, 2626.

(33) Hubbard, S. J.; Thornton, J. M. *Naccess*; Department of Biochemistry and Molecular Biology, University College: London, 1993.

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

(35) Anderson, D. M.; Gruner, S. M.; Leibler, S. *Proc. Natl. Acad. Sci.* **1988**, *85*, 5364.

(36) Shearman, G. C.; Ces, O.; Templer, R. H.; Seddon, J. M. *J. Phys.: Condens. Matter* **2006**, *18*, S1105.

(37) Hofer, N.; Aragao, D.; Lyons, J. A.; Caffrey, M. *Cryst. Growth Des.* **2011**, *11*, 1182.

(38) Seddon, J. M.; Templer, R. H. *Philos. Trans. R. Soc. London* **1993**, *344*, 377.

(39) Sanner, M. F.; Olson, A. J.; Spehner, J.-C. *Biopolymers* **1996**, *38*, 305.

(40) Audet, M.; Bouvier, M. *Cell* **2012**, *151*, 14.

(41) Schobert, B.; Cupp-Vickery, J.; Hornak, V.; Smith, S. O.; Lanyi, J. K. *J. Mol. Biol.* **2002**, *321*, 715.

(42) Landau, E. M.; Rosenbusch, J. P. *Proc. Natl. Acad. Sci.* **1996**, *93*, 14532.

(43) Marsh, D. *Biophys. J.* **2008**, *94*, 3996.