Molecular driving forces defining lipid positions around aquaporin-0

Camilo Aponte-Santamaría[a], Rodolfo Briones[a], Andreas D. Schenk[b], Thomas Walz[b,c], and Bert L. de Groot[a,1]

[a]Department of Theoretical and Computational Biophysics, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany; [b]Department of Cell Biology, and [c]Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115

Edited by Wolfgang Baumeister, Max-Planck-Institute of Biochemistry, Martinsried, Germany, and approved May 4, 2012 (received for review December 22, 2011)

Lipid–protein interactions play pivotal roles in biological membranes. Electron crystallographic studies of the lens-specific water channel aquaporin-0 (AQP0) revealed atomistic views of such interactions, by providing high-resolution structures of annular lipids surrounding AQP0. It remained unclear, however, whether these lipid structures are representative of the positions of unconstrained lipids surrounding an individual protein, and what molecular determinants define the lipid positions around AQP0. We addressed these questions by using molecular dynamics simulations and crystallographic refinement, and calculated time-averaged densities of dimyristoyl-phosphatidylcholine lipids around AQP0. Our simulations demonstrate that, although the experimentally determined crystallographic lipid positions are constrained by the crystal packing, they appropriately describe the behavior of unconstrained lipids around an individual AQP0 tetramer, and thus likely represent physiologically relevant lipid positions.While the acyl chains were well localized, the lipid head groups were not. Furthermore, in silico mutations showed that electrostatic inter actions do not play a major role attracting these phospholipids towards AQP0. Instead, the mobility of the protein crucially modulates the lipid localization and explains the difference in lipid density between extracellular and cytoplasmic leaflets. Moreover, our simulations support a general mechanism in which membrane proteins laterally diffuse accompanied by several layers of localized lipids, with the positions of the annular lipids being influenced the most by the protein surface. We conclude that the acyl chains rather than the head groups define the positions of dimyristoyl-phosphatidylcholine lipids around AQP0. Lipid localization is largely determined by the mobility of the protein surface, whereas hydrogen bonds play an important but secondary role.

electron crystallography | lipd bilayer | atomistic simulations

Lipids and membrane proteins form biological membranes that constitute the boundary of cells and their intracellular compartments. Lipids arrange in a bilayer conformation that serve as a 2D fluid for membrane proteins. The lipid bilayer, however, is more than a passive fluid and influences many aspects of membrane proteins, including their insertion into the membrane (1, 2), assembly into complexes (3–5), and activity (6, 7). Conversely, membrane proteins alter the conformational properties of lipid bilayers, mediating for instance pore formation (8), fusogenicity (9), and membrane bending (10, 11). Detailed knowledge of how lipids and membrane proteins interact with each other is therefore crucial to understand the molecular machinery of biological membranes.

To date, spectroscopic methods have contributed most to our understanding of lipid–protein interactions, providing insight into the dynamics of such interactions (1, 12). Atomistic views were obtained by structures of membrane proteins either with few specifically bound lipids or surrounded by a complete ring of lipids, determined by X-ray (13–17) and electron crystallography (18–20). Furthermore, molecular dynamics (MD) and coarse-grained simulations have added a wealth of dynamic and energetic information creating a better understanding of the principles underlying lipid–protein interactions (for comprehensive reviews see refs. 21 and 22).

Electron crystallographic studies of the lens-specific water channel aquaporin-0 (AQP0) have emerged as a promising approach for systematic structural studies of lipid–protein interactions (19, 20, 23–27). Electron crystallography uses 2D crystals of membrane proteins reconstituted into artificial lipid bilayers and thus allows the structure of membrane proteins to be determined in their native environment (24, 25, 28). The 1.9-Å structure of AQP0 crystallized in dimyristoyl-phosphatidylcholine (DMPC) revealed not only the protein, but also the first shell of lipids, called the annular lipids, surrounding the AQP0 tetramers (19). The structure of the complete ring of annular lipids defined the preferred lipid positions around the protein and provided insights into the nature of nonspecific lipid–protein interactions. Moreover, the annular lipids were also observed in the recent 2.5-Å structure of AQP0 crystallized in *Escherichia coli* polar lipids (20), demonstrating that high-quality 2D crystals of AQP0 can be produced with different lipids.

The electron crystallographic structures of AQP0 raised several questions: Are the observed crystallographic lipid structures, which correspond to lipids sandwiched in between two tetramers in the 2D crystals, representative of the positions adopted by unconstrained lipids surrounding a single AQP0 tetramer? What are the molecular driving forces stabilizing the observed lipid positions around AQP0? How does AQP0 affect lipids beyond the first annular layer? We addressed these questions by using MD simulations and crystallographic refinement. We calculated time-averaged density maps of DMPC bilayers either surrounding an individual AQP0 tetramer or constrained by four AQP0 tetramers simulating the situation in a 2D crystal.

Results

Lipid Arrangement Around a Single AQP0 Tetramer. We first performed 100-ns MD simulations of a single AQP0 tetramer embedded in a DMPC bilayer (Fig. 1*A*, *Left*, and *SI Appendix*, Fig. S1) and calculated a time-averaged lipid-density map ρ around the tetramer (in the following called MD map). Because each monomer in the AQP0 tetramer has identical lipid interfaces, composed of surfaces S1 and S2 (Fig. 1*A*), we fourfold symmetrized the map to produce the average lipid density around a single AQP0 monomer, which we could compare with the lipids seen in the electron crystallographic structure of AQP0 (19) (Fig. 1*B*).

For most of the crystallographic lipids (labeled PC1–PC7), portions of their tails fall into high-density regions of the MD map. In particular, almost the entire tails of lipid PC1 in the

Author contributions: C.A.S., R.B., A.D.S., T.W., and B.L.d.G. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: bgroot@gwdg.de.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1121054109/-/DCSupplemental.

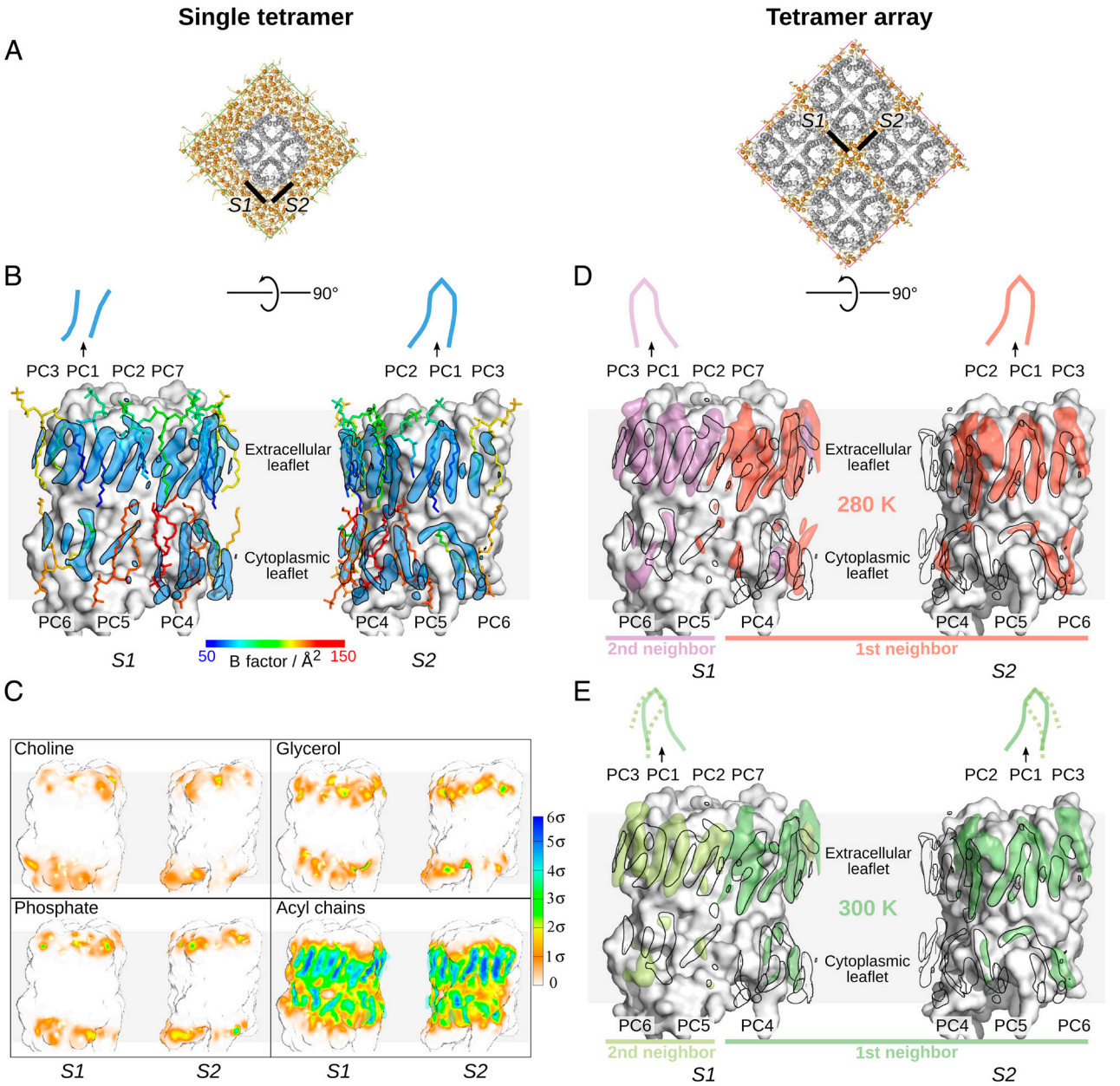

Fig. 1. Time-averaged lipid-density maps around AQP0 obtained from MD simulations. (*A*) Top views (perpendicular to the membrane plane) of a single AQP0 tetramer embedded in a DMPC lipid bilayer (*Left*) and an array of four AQP0 tetramers with sandwiched DMPC molecules (*Right*). AQP0 tetramers are shown in white and DMPC lipids in orange and yellow. S1 and S2 indicate the two lipid-facing surfaces of an AQP0 monomer. *B*–*E* show side views of surfaces S1 and S2 indicated in *A*. (*B*) Comparison between the DMPC molecules seen in the electron crystallographic structure of AQP0 (19) and the lipid-density map obtained from an MD simulation of an AQP0 tetramer inserted into an equilibrated DMPC patch (*A*, *Left*). The crystallographic lipids, labeled PC1 to PC7, are shown in stick representation and color-coded according to their B factor (color scale at the bottom). The MD-derived density map is contoured at 4σ and shown as blue surface enclosed by black lines. (*C*) Contributions of different lipid groups to the MD-derived lipid-density map shown in *B*. The density map at a distance of 5.6 Å from the protein is depicted according to the scale at the right side. (*D*–*E*) Lipid-density maps obtained from MD simulations of a system containing an array of four AQP0 tetramers in a 2D crystal arrangement (*A*, *Right*), at 280 (*D*) and 300 K (*E*). Maps were contoured at 3.2σ. For comparison, the density map obtained from the simulation with a single tetramer (*B*) is shown as black contours. The sketches on top represent the major conformations sampled during the simulations by the lipids at the position of crystallographic lipid PC1.

extracellular leaflet at surface S2 are represented by high density in the computed map. The middle part of lipid PC3 at S1 and one of the tail ends of lipid PC3 at S2 also coincide with high-density regions in the MD map, which also captures the separation of the tails of lipid PC6 at S1. In addition, the MD map shows a favorable degree of correlation with the crystallographic B factors of the lipids: Portions of the lipids with low B factors (less positional uncertainty) match with high-density regions, and portions of the lipids with high B factors (more positional uncertainty) correspond to regions of weak density. Remarkably, two independent simulations of a single tetramer, either including the crystallographic lipids or inserted into an equilibrated lipid patch (*SI Appendix*, Fig. S1*A*), produced similar density maps (*SI Appendix*, Fig. S2). Moreover, shortening the production runs to only 50% of the simulation length (50 ns) did not induce substantial changes in the density maps (*SI Appendix*, Fig. S3). These two results indicate convergence of the lipid positions on the simulated timescale. Unless stated otherwise, the MD map obtained with a single AQP0 tetramer inserted into a DMPC patch (without the crystallographic lipids) was used for all further analysis.

To further analyze the MD map, we inspected the contributions made by the different parts of the lipids (Fig. 1*C*). High-density values ($\rho > 4\sigma$), corresponding to more localized positions during the simulations, are mainly observed for the acyl chains, whereas the densities representing the head groups (choline, phosphate, and glycerol) are found to be weak ($\rho < 3\sigma$).

In addition, the MD map revealed an asymmetry between the two leaflets regarding the number of well-defined tail positions (Fig. 1*B*). Eleven well-defined tail positions are observed in the extracellular leaflet and only six in the cytoplasmic leaflet. Moreover, the density is much weaker in the cytoplasmic leaflet than in the extracellular leaflet. The asymmetry is also reflected in the computed deuterium parameters, which show different average tilting angles (with respect to the axis normal to the membrane) for the acyl chains of annular lipids in the extracellular leaflet compared to those in the cytoplasmic leaflet (*SI Appendix*, Fig. S4 *A* and *B*).

Effect of Crystal Packing on the Lipid Positions. To investigate the effect of the dense protein packing in the 2D crystals on the positioning of the annular lipids surrounding AQP0, we performed MD simulations with lipids sandwiched between four tetramers in the crystal packing arrangement (Fig. 1*A*, *Right*, and *SI Appendix*, Fig. S1*B*) at two different temperatures, 280 and 300 K.

To calculate the time-averaged lipid-density map, each lipid molecule was assigned to its closest AQP0 monomer (labeled first neighbor in Fig. 1 *D* and *E*). In the resulting density maps, high-density regions are mainly located near surface S2 (Fig. 1 *D* and *E*). High-density regions in these tetramer-array maps colocalize with those in the single-tetramer map on surface S2, with the best-defined position observed at the place of crystallographic lipid PC1. Nevertheless, the density at the position of crystallographic lipid PC3 is stronger in the tetramer-array maps than in the single-tetramer map (see also differences in *SI Appendix*, Fig. S5), indicating that this lipid is more strongly localized in the context of a crystalline array than when associated only with a single AQP0 tetramer. In addition, the deuterium parameters were found to be in a broader range of values for the acyl chains of lipids sandwiched in between AQP0 tetramers than for the acyl chains of lipids surrounding an individual tetramer (compare *SI Appendix*, Fig. S4 *C* and *D* with Fig. S4 *A* and *B*).

Interestingly, the tetramer-array maps revealed that lipids at crystallographic position PC1 adopt primarily one conformation in the simulation at 280 K (Fig. 1*D*) and alternate between two conformations in the simulation at 300 K (Fig. 1*E*). The single conformation in the 280 K map matches the conformation seen on surface S2 in the single-tetramer map, but deviates from the conformation seen on surface S1, as observed when the tetramer-array map is projected onto surface S1 by assigning the lipid positions to the second-closest (second neighbor) AQP0 monomer (compare Fig. 1*D* with Fig. 1*B*). In contrast, the 300 K map displays two conformations: The first one matches the one seen on surface S1 in the single-tetramer map, and the second corresponds to the one seen on surface S2 in the single-tetramer map (compare Fig. 1*E* with Fig. 1*B*).

Refinement of Lipid Structures Based on Lipid-Density Maps Derived from MD Simulations. Although the MD map shows many similarities with the electron crystallographic structure, there are also numerous differences in lipid conformations. To assess whether these differences constitute inconsistencies between the MD-derived and the crystallographic datasets or represent alternative lipid conformations, we modeled lipids into the MD density and then refined them against the electron crystallographic data (19) (Fig. 2). Refinement did not affect the positions of the acyl chains and glycerol backbone of lipid PC1, but parts of the remaining lipids moved to slightly different positions (Fig. 2, *Left*). The refined structure includes five complete lipids in the extracellular leaflet and four in the cytoplasmic leaflet, and both leaflets also contain a single acyl chain near the fourfold axis. In comparison to the original structure, the MD map allowed identification of an additional acyl chain in the extracellular leaflet (PC11) and an additional full lipid in the cytoplasmic leaflet (PC10).

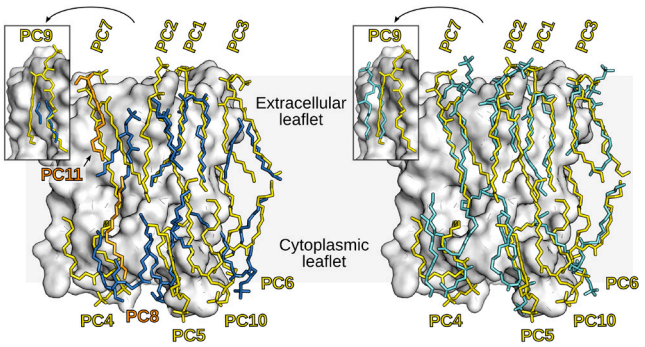

Fig. 2. Refinement of lipid structures around AQP0 by using MD and electron crystallography data. The lipid moieties modeled into the density map derived from the MD simulation of a single tetramer (presented in Fig. 1*B*) are shown in blue. The lipids after refinement against the electron crystallography data (19) are presented in yellow (complete lipids) and orange (individual acyl chains). The original crystallographic lipids (19) are shown in cyan for comparison. The AQP0 monomer is shown in surface representation. For clarity, lipid PC9 is depicted separately in the insets, and the individual acyl chains PC8 and PC11 are not shown in the right figure.

The lipids in the refined structure based on the MD map show many similarities with those in the original crystallographic structure, especially those in the extracellular leaflet, PC1, PC2, PC3, and PC7 (Fig. 2, *Right*), but there are also some local differences. For example, one acyl chain of lipid PC1, which shows the best-defined density in all datasets, has a bent conformation in the MD map but has a more straight conformation in the crystallographic structure. Moreover, lipid PC9 is not in direct contact with the protein surface in the original crystallographic structure, but it is in contact in the MD map. The refined structure also exhibits a different orientation for the two hydrogen bond-forming protein residues Arg113 and Arg196 (*SI Appendix*, Fig. S6*D*).

In addition to this converged MD map, we also used a nonconverged MD simulation (100-ps length) as a control. The lipids modeled into the nonconverged density map matched neither the lipid positions in the converged MD map nor in the original crystal structure (*SI Appendix*, Fig. S6 *E* and *F*). However, upon refinement against the crystallographic data, the lipids modeled into the nonconverged map exhibited big shifts (especially in the first refinement steps) and finally approached the same positions as those seen in the converged model after refinement (compare *SI Appendix*, Fig. S6*C* with Fig. S6*G*, and see further details in the *SI Appendix*).

Strong Protein-Lipid Interaction Sites and in Silico Mutations. We monitored the potential energy between lipids and AQP0 during the MD simulations (Fig. 3*A*), and thereby identified seven residues at the AQP0 surface that strongly interact with lipids: R5, Y105, R113, R196, and K238 through electrostatic interactions, and W10 and W202 through van der Waals interactions (see also *SI Appendix*, Fig. S7*A*). Correlation between the Coulomb interaction energy and the presence of hydrogen bonds suggests that the electrostatic residue-lipid interactions are mediated by hydrogen bonds (*SI Appendix*, Fig. S7*B*). In the extracellular leaflet, the side chain of R196 forms up to four hydrogen bonds, for more than half of the simulation time, mainly with the carboxyl and phosphatidyl-ester oxygens (*SI Appendix*, Fig. S7*C*) of lipids at positions PC1, PC2, and PC7 (*SI Appendix*, Fig. S7*D*). The neighboring residues Y105 and R113 form hydrogen bonds with lipids at positions PC1 and PC3, but less frequently than R196. Similar

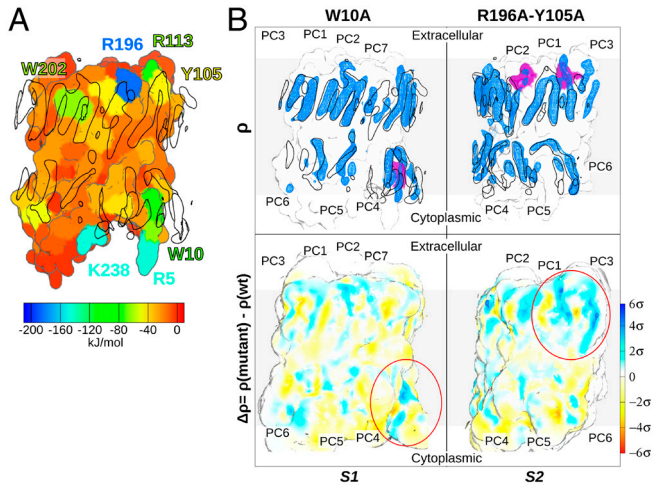

Fig. 3. Interaction energy between the lipids and AQP0 computed from MD simulations. (*A*) Time-averaged potential interaction energy derived from a simulation of a single AQP0 tetramer embedded in a DMPC lipid bilayer. The energy is depicted in color representation according to the scale at the bottom. Labels correspond to AQP0 residues strongly interacting with lipids, either through electrostatic (R5, Y105, R113, R196, and K238) or van der Waals (W10 and W202) interactions. High lipid-density regions (presented in Fig. 1*B*) contoured at 4σ are outlined in black. (*B*) Effects of mutating AQP0 residues that strongly interact with lipids on the resulting MD-derived lipid-density maps. Two representative maps, obtained with the W10A and the R196A-Y105A mutations, are shown (see *SI Appendix*, Figs. S8–S10 for all 12 maps). The *Top* panels show the mutated residues in purple and the resulting lipid-density maps contoured at 4σ as blue mesh. For comparison, the lipid-density map obtained with wild-type AQP0 contoured at 4σ is displayed as black contours. The lower panels depict differences between the lipid-density maps obtained with mutant and wild-type AQP0 (see scale for color coding). The red circles indicate regions with an increase in lipid-density near the mutated residues.

to R196 in the extracellular leaflet, the side chains of R5 and K238 in the cytoplasmic leaflet also make intermittent hydrogen bonds mainly with carboxyl and ester oxygens of the annular lipids.

To test whether these strongly interacting residues constitute phospholipid-binding sites, we performed MD simulations in which we mutated them to alanine: 10 simulations of 100 ns each, including single, double, and triple mutations. Changes in the lipid-density around AQP0 mutants are in the range from 1σ to 2σ. Larger changes in the density map primarily correspond to lipid atoms occupying space freed by deletion of the side-chains of the mutated residues (Fig. 3*B*, *Right*, and *SI Appendix*, Figs. S8 and S9). Compared to wild-type AQP0, the mutations did not substantially modify the lipid positions, and the well-defined tail positions were always observed throughout all mutant simulations.

The Effect of Protein Mobility on the Lipid Positions. Calculation of the rms fluctuations (RMSF) in the MD trajectories revealed that the protein surface displays a wide range of flexibility (Fig. 4*A*). Residues in transmembrane helices (especially those close to the extracellular surface of the protein) are mostly rigid, whereas the N and C termini are highly flexible. Despite their strong electrostatic interactions, residues R5, R196, and K238 are also very mobile. Overall, we found that the RMSF are positively correlated with the crystallographic B factors of the protein (19) (*SI Appendix*, Fig. S11).

To analyze the relation between protein mobility and lipid density, we defined F as the fraction of a cylinder (of radius 7.5 Å and height 4 Å) that is occupied by high lipid-density points ($\rho > 4\sigma$), and plotted this quantity as a function of the RMSF of the protein atom centered in the cylinder (Fig. 4*B*). Near high-RMSF atoms, such as the ones in the flexible N and C termini, F invariably takes

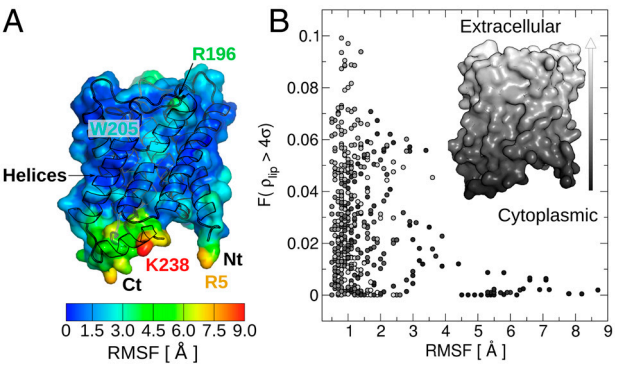

Fig. 4. Effect of AQP0 mobility on lipid positions. (*A*) RMSF derived from MD simulations are shown as a measure of the flexibility of the protein surface. The RMSF are color-coded according to the scale at the bottom. (*B*) F defined as the fraction of a cylinder (of radius 7.5 Å and height 4 Å) occupied by high lipid-density points ($\rho > 4\sigma$) plotted as a function of the RMSF of the surface atom centered in the cylinder. The gray scale represents the position of the atom along the coordinate normal to the membrane (as illustrated in the inset).

on small values. In contrast, in the vicinity of low-RMSF atoms, F displays a broader range of values, thus allowing highly localized lipid positions.

Lipid Behavior Distant from AQP0. An analysis of the behavior of lipids distant from the protein (Fig. 5 and *SI Appendix*, Fig. S12) revealed that only the annular lipids display strong density regions, whereas with increasing distance from the protein the lipid density rapidly weakens and becomes "bulk-like." However, at intermediate distances (a few lipid shells away from AQP0), the lipids still show a moderately localized behavior. Very different density patterns are observed for the two leaflets. For instance, at the height of the tips of the acyl chains, the map presents stronger densities in the extracellular leaflet than in the cytoplasmic leaflet. In contrast, at the height of the head groups (both glycerol and phosphate), the density is predominantly higher in the cyto-

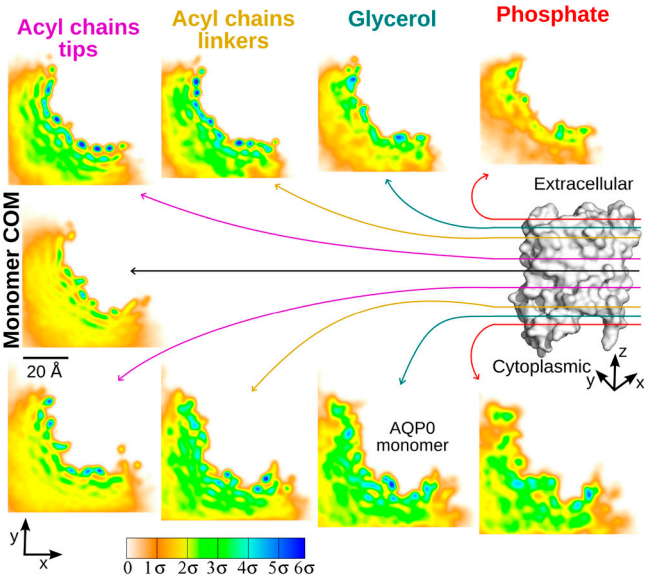

Fig. 5. Lipid density around an AQP0 monomer beyond the annular lipid shell, recovered from the simulation of a single tetramer embedded in a DMPC lipid bilayer without the crystallographic lipids (see *SI Appendix*, Fig. S12 for simulations with and without the crystallographic lipids). The color maps represent lateral projections (onto the *xy* membrane plane) of the lipid density, at the different *z* positions indicated on the AQP0 monomer (white). Projections were taken at the average *z* positions of the center of masses (COM) of the indicated lipid groups (for both leaflets, upper and lower maps) and the AQP0 monomer (middle map).

plasmic leaflet compared to that in the extracellular leaflet. This finding indicates that the effect of leaflet asymmetry (noted before for annular lipids) extends to lipids distant from the protein.

Discussion

Lipids in AQP0 2D Crystals Behave Similarly to Lipids Surrounding a Single AQP0 Tetramer. Our results demonstrate that the majority of lipid structures and their B factors seen by electron crystallography of AQP0 2D crystals are representative of the average positions adopted by unconstrained lipids surrounding an individual AQP0 tetramer seen in the MD map (Fig. 1*B*). Nevertheless, lipids sandwiched in between AQP0 tetramers have a stronger degree of alignment than lipids only in contact with a single AQP0 tetramer, as reflected in the calculation of the deuterium order parameters (*SI Appendix*, Fig. S4).

The favorable agreement between the computed lipid-density maps and the crystallographic structures also provides an independent validation of the used lipid and protein simulation parameters (force field). Moreover, the similarity of the lipid-density maps obtained in two independent simulations (*SI Appendix*, Fig. S2) indicates that lipids adopt stable positions around AQP0 in a timescale of tens of nanoseconds, and independent of their initial positions.

Our simulations reveal that the acyl chains are the most localized parts of annular lipids, generating strong density in the time-averaged map, whereas the lipid head groups are less localized and thus create only weak density (Fig. 1*C*). This result supports the hypothesis deduced from electron crystallographic structures, that acyl chains are mainly responsible for stabilizing the saturated DMPC lipids around AQP0, whereas the head groups make only a secondary contribution to lipid localization (20). Electron crystallography also revealed that acyl chains of unsaturated *Escherichia coli* polar lipids occupy similar positions as those of the saturated DMPC lipids (20), thus suggesting that acyl chains play an important stabilizing role not only for saturated but also for unsaturated lipids.

Our simulations also show that the lipids in the 2D crystal accommodate closer to protein surface S2 of a tetramer than to S1 of their adjacent tetramer, suggesting that S2 mostly defines the lipid positions in the crystal (Figs. 1 *D* and *E*). This effect can be attributed to the observed strong electrostatic interactions between the lipids and residues R196, Y105, and R113 (extracellular leaflet), and R5 (cytoplasmic leaflet), all of which are located at surface S2 (Fig. 3*A*). The lack of electrostatic interactions with residues at S1 may also explain why high-density contours in the MD map (resulting from stable lipid positions) and the positions of the crystallographic lipids do not match as well at S1 as they do at S2.

Lipids at the crystallographic position PC1 were found to adopt two conformations at 300 K, when located in the crystal environment. Each conformation matches one of the two conformations adopted by lipids at position PC1 around an isolated AQP0 tetramer facing either surface S1 or S2 of the monomer (Fig. 1*E*). This result thus suggests that the lipid at this position in the 2D crystal, sandwiched between surfaces S1 and S2, alternates between two conformations at 300 K, and also illustrates how the lipid conformations may be influenced by the surface of the protein. When the temperature was decreased to 280 K, only one of the two conformations was sampled (Fig. 1*D*). At the very low temperature at which the electron crystallographic data were collected, the lipids would be expected to only adopt this conformation, which was indeed the case.

Refinement of Crystallographic Lipid Positions Validates the MD Data. When lipids were built into the MD map and subsequently refined against the electron crystallographic data, annular lipid PC1 in the extracellular leaflet retained its position, and the refined structures matched the original crystallographic lipid structures (Fig. 2 and *SI Appendix*, Fig. S6). When, as a control, the lipids were deliberately built into a nonconverged MD map, refinement against the electron crystallographic data moved the lipids close to the crystallographic positions (*SI Appendix*, Fig. S6). These results underscore the validity of the MD-derived density maps and show that the refined structures are not strongly biased by the initial model. In addition, they reveal that the lipids adopt preferred positions around AQP0 not only in the context of a 2D crystal but that they use similar positions when surrounding an individual AQP0 tetramer.

AQP0 Does Not Have Specific Phospholipid-Binding Sites. Strong electrostatic protein–lipid interactions—mediated by hydrogen bonds—suggested the possibility that Y105, R113, and R196 at the extracellular leaflet, and R5 and K238, at the cytoplasmic leaflet, correspond to phospholipid-binding sites that drive the lipids into the positions observed in both simulations and experiments (Fig. 3*A*). However, these residues did not form stable hydrogen bonds with a specific lipid, but rather transient bonds with all the lipids in their vicinity. The transient nature of the hydrogen bonds is also reflected in the different conformations of the involved protein residues in the refined structures (*SI Appendix*, Fig. S6*D*) and also in their high mobility (Fig. 4*A*). Moreover, in silico mutations of these residues to alanine did not appreciably change the well-defined positions of the lipid tails, in 10 independent simulations with AQP0 mutants spanning a total time of 1.0 μs (Fig. 3*B*, *Right*, and *SI Appendix*, Figs. S8 and S9). Our simulations thus dispose of electrostatic interactions as the main cause that defines the positions of phospholipids around AQP0, and corroborate the conclusion from the electron crystallographic AQP0 structures (20) that residues R196 and Y105 are not part of a phospholipid-binding site as defined by Palsdottir and Hunte (29).

Protein Mobility Interferes with the Localization of Lipids. Our simulations revealed that AQP0 is not a rigid entity but displays a broad range of flexibility. Transmembrane-helix residues at the protein surfaces facing the lipid environment are the most rigid parts, while the N and C termini are highly mobile (Fig. 4*A*). This result is in perfect agreement with the experimental B factors of the protein (19) (*SI Appendix*, Fig. S11). Interestingly, lipid density near the flexible termini was found to be more diffuse compared to the lipid density close to the rigid parts, and only a minor fraction of high-density points was found to reside near highly mobile protein atoms (Fig. 4*B*). Our results thus suggest that mobile segments of the protein interferes with the localization of lipids.

Protein mobility may also provide an explanation for the observed asymmetry in lipid density between the two leaflets (Figs. 1*B* and 5). This asymmetry does not appear to result from lipid immobilization due to protein contacts between the two layers in the double-layered 2D crystals. Instead, it appears that rigid AQP0 residues allow lipids in the extracellular leaflet to be localized, while the flexible termini interfere with localization of lipids in the cytoplasmic leaflet.

Irregularities in the Shape of the Protein Surface Modulate the Lipid Density. Our simulations with AQP0, which has an uneven surface, showed highly localized positions of individual lipid tails for the annular lipids, whereas simulations with transmembrane helices, which have smoother surfaces, did not (30). This result is consistent with the hypothesis by Niemelä et al. that lipid positions in the annular shell are modulated by irregularities in the protein surface (31). Moreover, our MD maps obtained with alanine substitution mutants showed increased lipid density in the space originally occupied by the side chains of the mutated residues (Fig. 3*B* and *SI Appendix*, Figs. S8–S10), illustrating the effect of the shape of the protein surface on lipid arrangement.

An analysis of the protein surface curvature allowed us to distinguish between low curved convex regions (bumps), and highly curved concave areas (clefts) (*SI Appendix*, Fig. S13). Surprisingly, the lipid density did not show a strong correlation with either type of concavity, and high lipid-density points were observed near both concave and convex surface regions. Our curvature calculations therefore support the notion that lipids adapt to the roughness of the exterior surface (bumps or clefts) to form a tight seal around the protein that prevents leakage of solutes across the membrane (for a detailed analysis of the protein surface curvature and concavity see the *SI Appendix*).

AQP0 Influences Lipid Behavior Beyond the First Lipid "Solvation" Shell. In our simulations, lipids only gradually recover their bulk properties with increasing distance from AQP0 (Fig. 5). The protein thus influences not only the localization of the first lipid shell, the annular lipids, but also the following lipid shells. Similar lipid-immobilization patterns have been observed in previous simulation studies with transmembrane helices (30) and ion channels (31). Our results therefore support the model proposed by Niemelä et al. (31), in which the protein forms an obstacle for lateral lipid diffusion perpendicular to the protein surface, thereby influencing the localization of several lipids shells around the protein. Our results, together with these computational studies (30, 31), thus suggest a general mechanism in which membrane proteins laterally diffuse with a highly coordinated lipid solvation shell that consists of several lipid layers, with the positions of the annular lipids being influenced the most by the protein surface.

Conclusions

We used MD simulations and crystallographic refinement to study the localization of DMPC lipids around AQP0. We found that the positions of the constrained lipids in the 2D crystals determined by electron crystallography together with their B factors are representative of the behavior of unconstrained lipids surrounding individual AQP0 tetramers. We conclude that positions of DMPC lipids around AQP0 are defined by the acyl chains rather than the head groups. Furthermore, we observed that the positions of these lipids are largely influenced by the local mobility of the protein, whereas specific hydrogen bonds play a secondary role. Finally, our results are consistent with a general mechanism in which membrane proteins laterally diffuse associated with several layers of lipids, with the positions of the lipids in the first solvation shell being also modulated by irregularities in the protein surface. It will be interesting to investigate if these features are specific for DMPC lipids surrounding AQP0, or rather represent general principles underlying lipid–protein interactions.

Materials and Methods

MD simulations were carried out using the GROMACS 4.0 simulation package (32, 33). Two different systems were simulated (Fig. 1*A* and *SI Appendix*, Fig. S1). The first system consisted of a single AQP0 tetramer embedded in a fully solvated DMPC lipid bilayer, simulating a membrane at low protein concentration. The second system included four densely packed AQP0 tetramers in the 2D crystal arrangement, with DMPC molecules filling the gaps between the tetramers and surrounded by explicit water molecules. The production runs were 100 ns in length and the first 10 ns were excluded to account for equilibration time. Additional simulations with AQP0 mutants (12 in total), in which residues of interest were substituted by alanine, were carried out following the same simulation scheme as for the single-tetramer system. The lipid density around a single AQP0 monomer was time-averaged over a concatenated trajectory consisting of fitted trajectories of individual AQP0 monomers (four in the single-tetramer and 16 in the four-tetramer system) together with their closest surrounding lipids. Additional simulation details, the methods used to calculate the lipid-density maps and other observables from the simulations, and the structure refinement procedure are described in the *SI Appendix*.

ACKNOWLEDGMENTS. We thank Ulrike Gerischer and Dirk Matthes for careful reading of the manuscript. This work was supported by grants from the Max Planck Society (C.A.S. and B.L.d.G.), the European Commission (Marie Curie Research Training Network MRTN-CT-2006-035995 to C.A.S. and B.L.d.G.), the Deutsche Forschungsgesellschaft (Sonderforschungsbereich 803 to R.B. and B.L.d.G.), and the National Institutes of Health (Grant EY015107 to T.W.). A.D.S. is supported by a Swiss National Science Foundation fellowship. T.W. is a Howard Hughes Medical Institute investigator.

1. Marsh D (2008) Protein modulation of lipids, and vice-versa, in membranes. *Biochim Biophys Acta* 1778:1545–1575.
2. Dowhan W, Bogdanov M (2009) Lipid-dependent membrane protein topogenesis. *Annu Rev Biochem* 78:515–540.
3. Dalbey R, Wang P, Kuhn A (2011) Assembly of bacterial inner membrane proteins. *Annu Rev Biochem* 80:161–187.
4. Raja M (2011) The potassium channel KcsA: A model protein in studying membrane protein oligomerization and stability of oligomeric assembly? *Arch Biochem Biophys* 510(1):1–10.
5. Vitrac H, Bogdanov M, Heacock P, Dowhan W (2011) Lipids and topological rules of membrane protein assembly: Balance between long and short range lipid–protein interactions. *J Biol Chem* 286:15182–15194.
6. Phillips R, Ursell T, Wiggins P, Sens P (2009) Emerging roles for lipids in shaping membrane-protein function. *Nature* 459:379–385.
7. Lee AG (2011) Biological membranes: The importance of molecular detail. *Trends Biochem Sci* 36:493–500.
8. Brogden KA (2005) Antimicrobial peptides: Pore formers or metabolic inhibitors in bacteria? *Nat Rev Microbiol* 3:238–250.
9. Jahn R, Scheller RH (2006) SNAREs-engines for membrane fusion. *Nat Rev Mol Cell Biol* 7:631–643.
10. Graham TR, Kozlov MM (2010) Interplay of proteins and lipids in generating membrane curvature. *Curr Opin Cell Biol* 22:430–436.
11. Qualmann B, Koch D, Kessels MM (2011) Let's go bananas: Revisiting the endocytic BAR code. *EMBO J* 30:3501–3515.
12. Marsh D (2010) Electron spin resonance in membrane research: Protein-lipid interactions from challenging beginnings to state of the art. *Eur Biophys J* 39:513–525.
13. Lange C, Nett JH, Trumpower BL, Hunte C (2001) Specific roles of protein-phospholipid interactions in the yeast cytochrome bc1 complex structure. *EMBO J* 20:6591–6600.
14. Murata T, Yamato I, Kakinuma Y, Leslie AGW, Walker JE (2005) Structure of the rotor of the V-Type Na^+-ATPase from Enterococcus hirae. *Science* 308:654–659.
15. Long SB, Tao X, Campbell EB, MacKinnon R (2007) Atomic structure of a voltage-dependent K^+ channel in a lipid membrane-like environment. *Nature* 450:376–382.
16. Shinzawa-Itoh K, et al. (2007) Structures and physiological roles of 13 integral lipids of bovine heart cytochrome c oxidase. *EMBO J* 26:1713–1725.
17. Hunte C, Richers S (2008) Lipids and membrane protein structures. *Curr Opin Struct Biol* 18:406–411.
18. Mitsuoka K, et al. (1999) The structure of bacteriorhodopsin at 3.0 Å resolution based on electron crystallography: Implication of the charge distribution. *J Mol Biol* 286:861–882.
19. Gonen T, et al. (2005) Lipid–protein interactions in double-layered two-dimensional AQP0 crystals. *Nature* 438:633–638.
20. Hite RK, Li Z, Walz T (2010) Principles of membrane protein interactions with annular lipids deduced from aquaporin-0 2D crystals. *EMBO J* 29:1652–1658.
21. Lindahl E, Sansom MSP (2008) Membrane proteins: Molecular dynamics simulations. *Curr Opin Struct Biol* 18:425–431.
22. Marrink SJ, de Vries AH, Tieleman DP (2009) Lipids on the move: Simulations of membrane pores, domains, stalks and curves. *Biochim Biophys Acta* 1788:149–168.
23. Gonen T, Sliz P, Kistler J, Cheng Y, Walz T (2004) Aquaporin-0 membrane junctions reveal the structure of a closed water pore. *Nature* 429:193–197.
24. Hite RK, Raunser S, Walz T (2007) Revival of electron crystallography. *Curr Opin Struct Biol* 17:389–395.
25. Raunser S, Walz T (2009) Electron crystallography as a technique to study the structure on membrane proteins in a lipidic environment. *Annu Rev Biophys* 38:89–105.
26. Schenk AD, Hite RK, Engel A, Fujiyoshi Y, Walz T (2010) Electron crystallography and aquaporins. *Methods Enzymol* 483:91–119.
27. Hite R, Gonen T, Harrison S, Walz T (2008) Interactions of lipids with aquaporin-0 and other membrane proteins. *Pflügers Arch* 456:651–661.
28. Fujiyoshi Y, Unwin N (2008) Electron crystallography of proteins in membranes. *Curr Opin Struct Biol* 18:587–592.
29. Palsdottir H, Hunte C (2004) Lipids in membrane protein structures. *Biochim Biophys Acta Biomembr* 1666:2–18.
30. Ash WL (2009) Helix-helix interactions in membrane proteins probed with computer simulations. PhD thesis, http://hdl.handle.net/1880/48166 (University of Calgary, Alberta, Canada), pp 164–170.
31. Niemelä PS, et al. (2010) Membrane proteins diffuse as dynamic complexes with lipids. *J Am Chem Soc* 132:7574–7575.
32. Spoel DVD, et al. (2005) GROMACS: Fast, flexible, and free. *J Comput Chem* 26:1701–1718.
33. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput* 4:435–447.

Supplementary information for

Molecular driving forces defining lipid positions around aquaporin-0

Camilo Aponte-Santamaría[a], Rodolfo Briones[a], Andreas D. Schenk[b], Thomas Walz[b,c] and Bert L. de Groot[a]

[a]Department of Theoretical and Computational Biophysics, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany

[b]Department of Cell Biology and [c]Howard Hughes Medical Institute, Harvard Medical School, Boston, USA

1 Supplementary figures

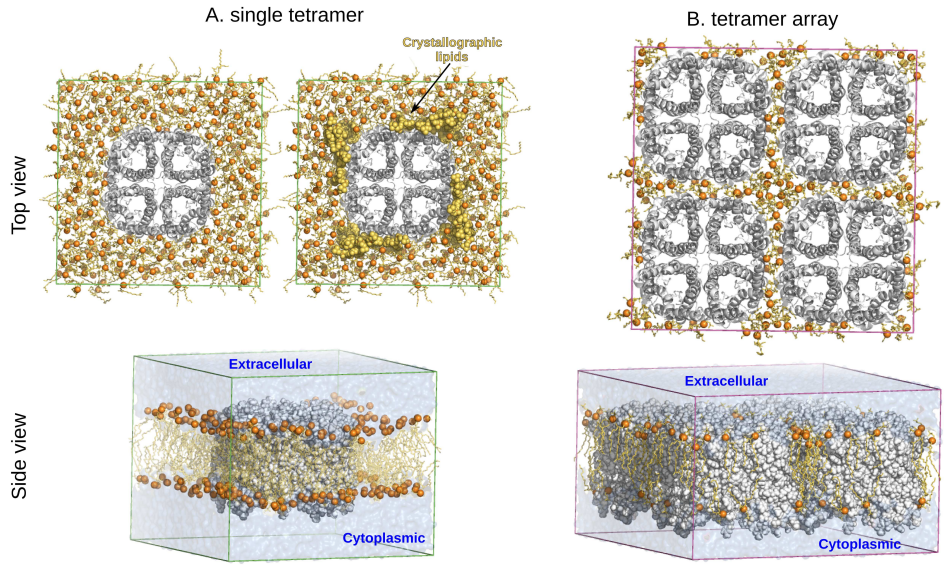

Figure S1. The two systems used for MD simulations of AQP0 in a DMPC lipid bilayer. A. A single tetramer embedded in a DMPC lipid bilayer. Simulations were carried out without (top view, left) or with (top view, right) the lipids seen in the electron crystallographic structure of AQP0 (1), facing the indicated surface of each AQP0 monomer. **B.** An array of four AQP0 tetramers in the 2D crystal arrangement. Simulations were carried out either at a temperature of 280 K or 300 K. The protein is shown in cartoon (top views) or sphere (side views) representation. Lipids are shown in yellow with their phosphor atom in orange. In both simulation systems, the lipid bilayer was fully solvated by explicit water molecules (blue in side views).

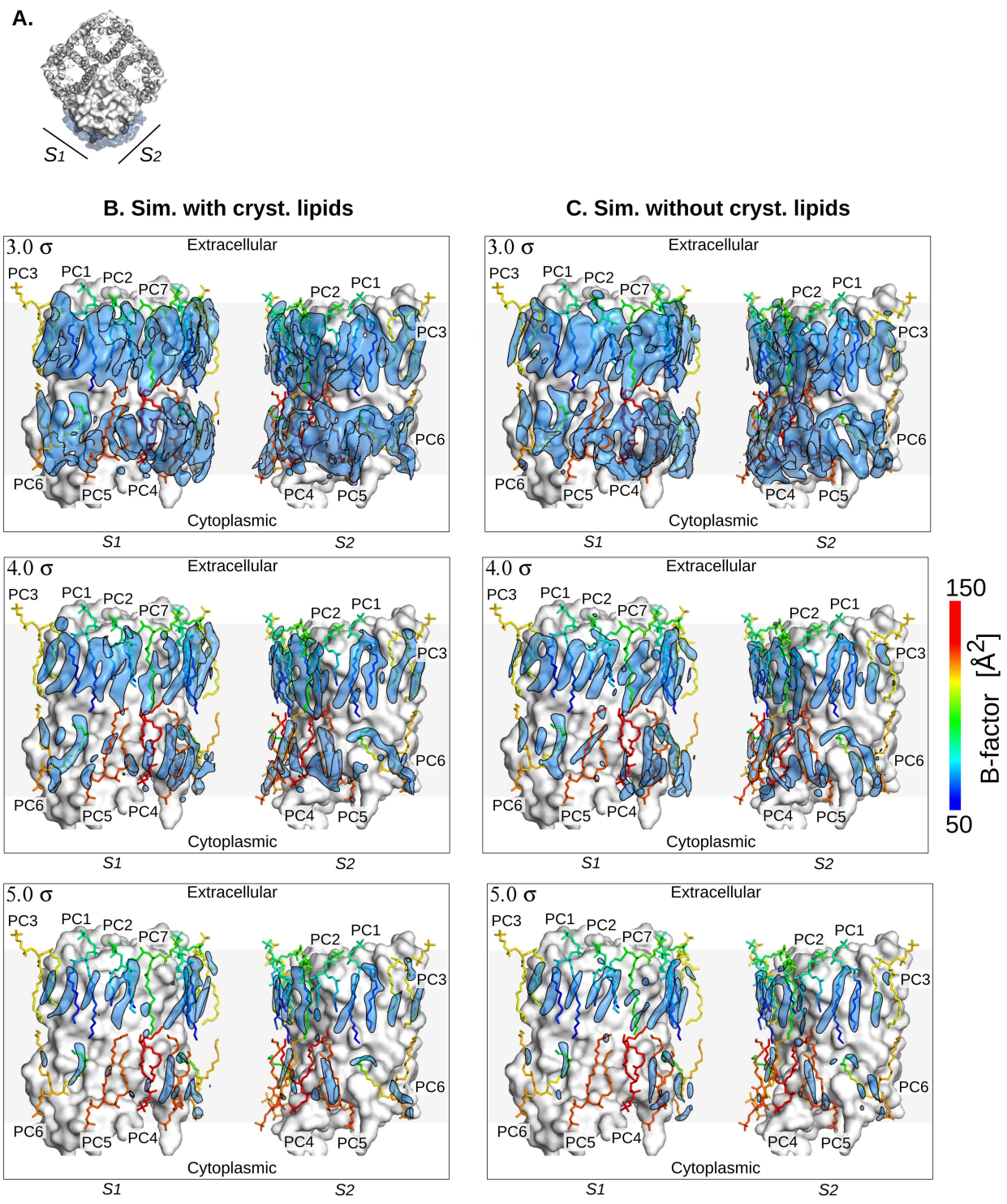

Figure S2. Comparison of density maps obtained in simulations with a single AQP0 tetramer with or without the crystallographic lipids in the starting configuration. A. Top view (perpendicular to the membrane plane) of an AQP0 tetramer, with S1 and S2 indicating the two lipid-facing surfaces of an AQP0 monomer. **B, C.** Side views of an AQP0 monomer showing the lipid density on surfaces S1 and S2 obtained in the simulation with (B) and without (C) crystallographic lipids. The three panels show the lipid densities (blue surface enclosed by black lines) contoured at different sigma values. For comparison, the crystallographic lipids seen in the electron crystallographic structure of AQP0 (1) are shown in stick representation (color-coded according to the B-factor; scale at the right side).

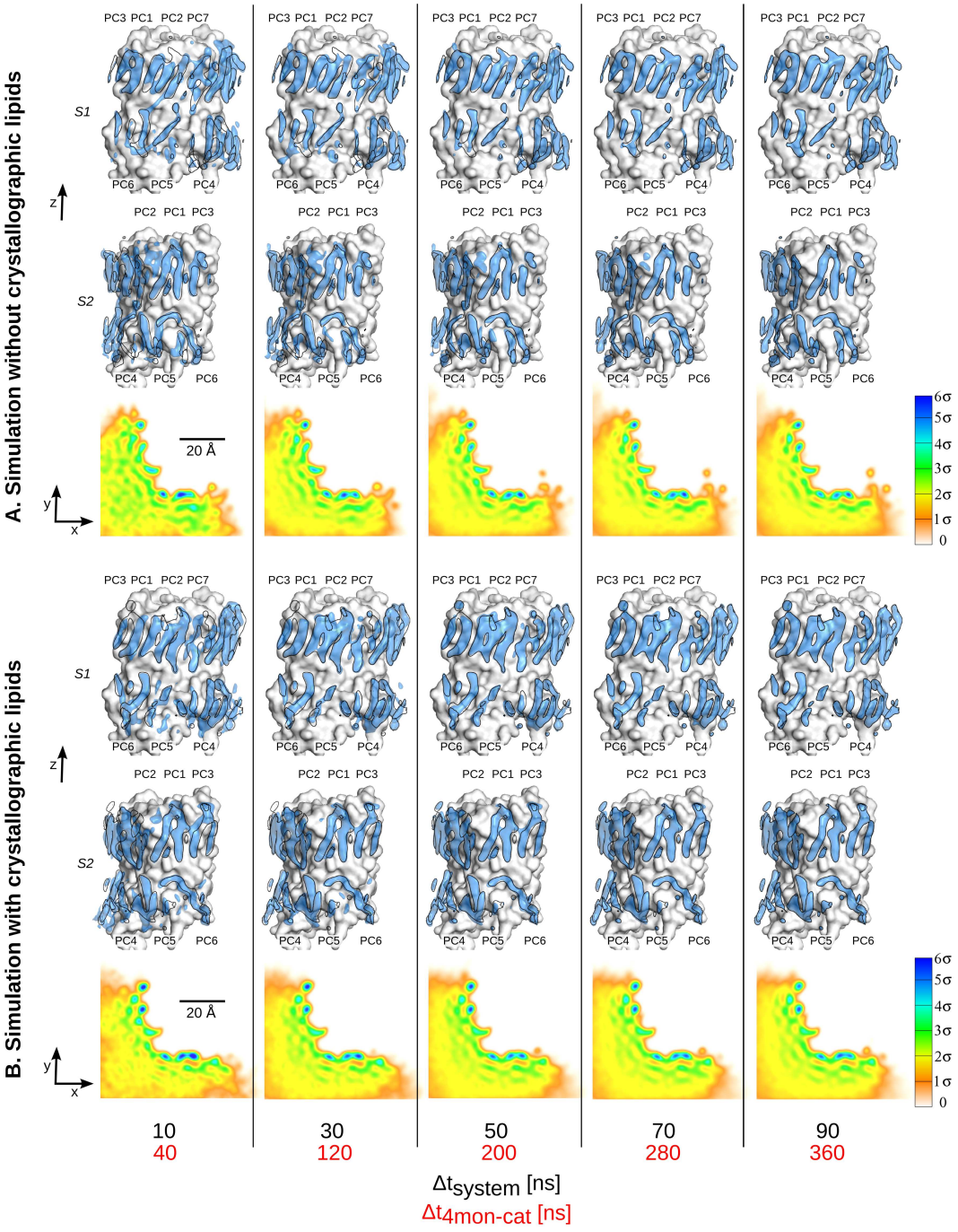

Figure S3. Density maps as a function of the length of trajectory used for their calculation. A, B. Lipid density maps obtained in the simulation with (A) and without (B) crystallographic lipids, by using a simulation of Δt_{system} in length. From each simulation, four independent trajectories of individual monomers (together with their closest surrounding lipids) were generated, yielding four-fold longer production runs for the calculation of the lipid density (lengths indicated with $\Delta t_{4mon-cat}$). The first two rows (both in A and B) show side views of the density maps projected onto protein surfaces S1 and S2. The maps are contoured at 4σ and shown as blue surfaces. For comparison, the density map obtained from the complete simulation ($\Delta t_{system} = 90$ ns) is shown as black contours. The color maps presented in the third row (both in A and B) represent lateral projections (onto the xy membrane plane) of the lipid density, at the average z position of the center of masses of the acyl-chain tips in the cytoplasmic leaflet.

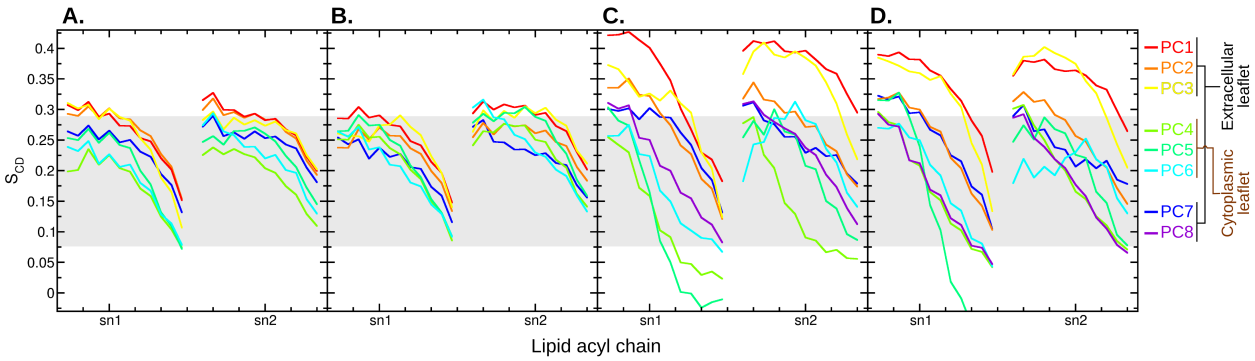

Figure S4. Deuterium order parameters, S_{CD}, for the lipid acyl chains derived from MD simulations. **A.** Lipids surrounding a single AQP0 tetramer including the crystallographic lipids. **B.** Lipids surrounding a single AQP0 tetramer without the crystallographic lipids. **C.** Lipids in the AQP0 array at 280 K. **D.** Lipids in the AQP0 array at 300 K. Order parameters were calculated for the carbon atoms of the two acyl chains (labeled sn1 and sn2). Lipids are labeled as in the electron crystallographic structure of AQP0 (1). For comparison, the gray bar shows the range of the order parameters derived from MD simulations of a pure DMPC bilayer.

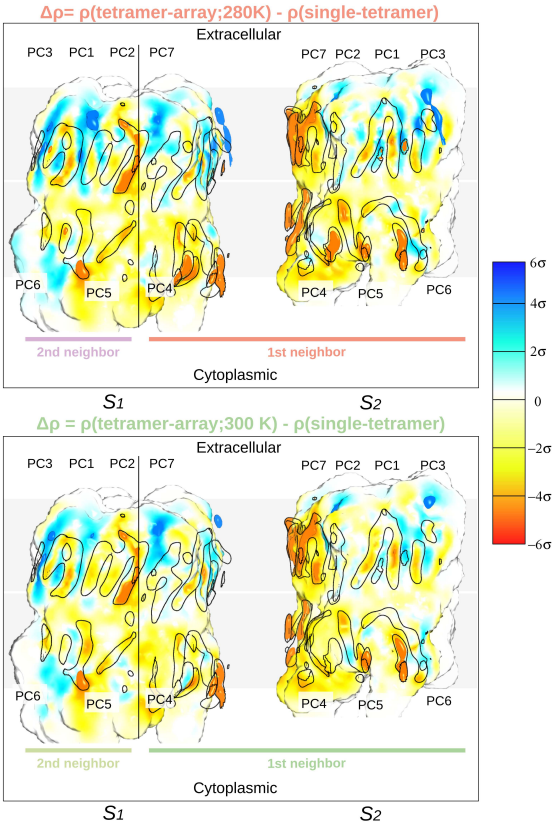

Figure S5. Difference of the lipid density computed from the simulation of the tetramer array (2D crystal) minus the lipid density computed from the simulation of a single tetramer. The difference is shown for the tetramer-array simulation at 280 K (upper panel) and 300 K (lower panel). The difference is displayed at a distance of 5.6 Å from the protein and color-coded according to the scale at the right. Before subtraction, in the tetramer-array system, the density was computed at surface S2 by assigning the lipids to their closest AQP0 monomer (labeled 1st neighbor) and at surface S1 by assigning the lipids to their second closest AQP0 monomer (labeled 2nd neighbor).

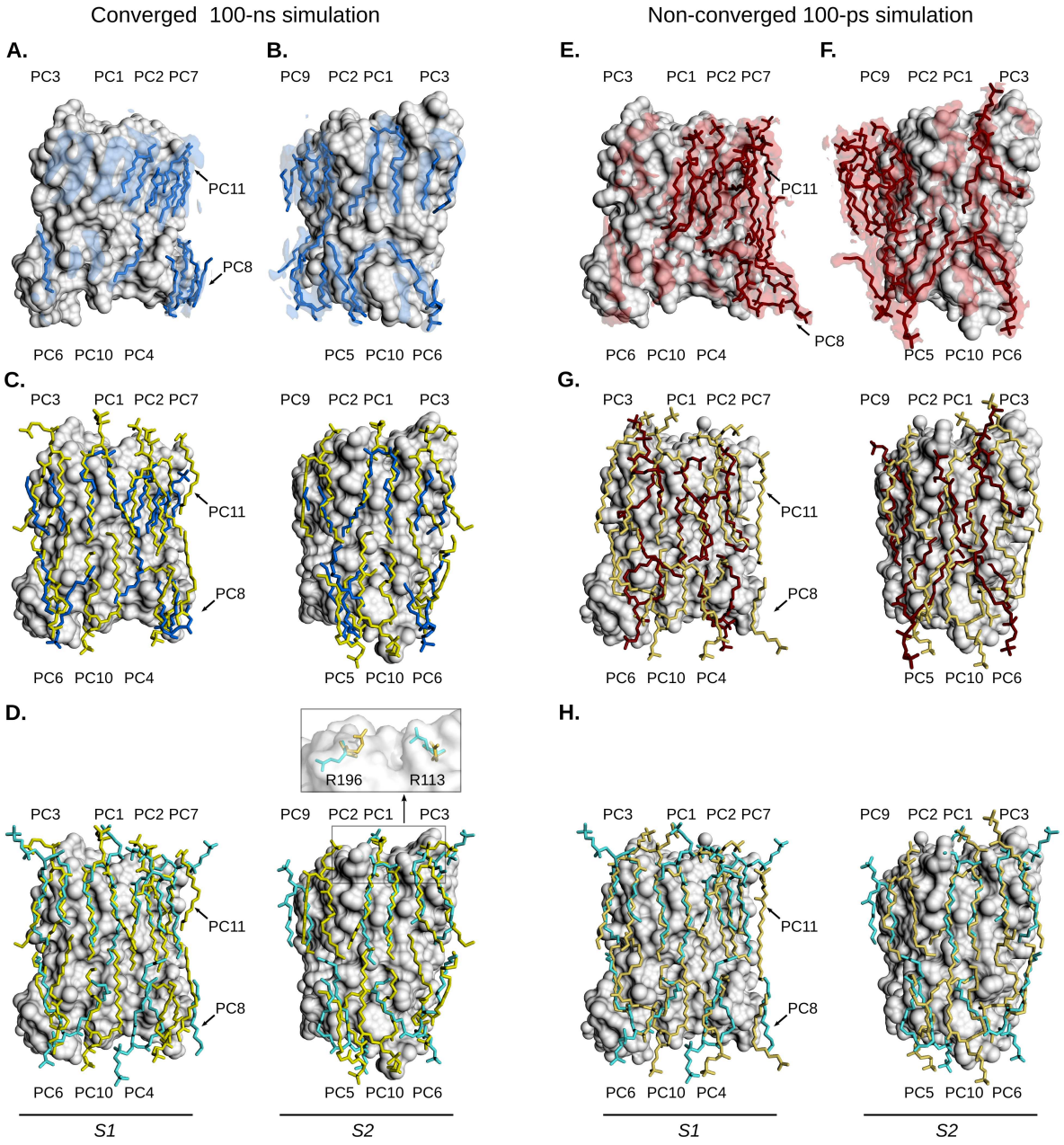

Figure S6. Refinement of lipids modeled into density maps derived from MD simulations, against the electron crystallographic data. A, B. Lipids (blue sticks) were first modeled into the density map derived from a converged 100-ns MD simulation of a single AQP0 tetramer that started from a configuration that did not include the crystallographic lipids. Panel A shows lipids modeled into density at surface S1, and panel B lipids modeled into density at surface S2. **C.** Comparison of the lipids before (blue) and after refinement against the electron crystallographic data (yellow). **D.** Comparison of the refined lipids (yellow) with the original lipids in the electron crystallographic structure of AQP0 (cyan). The inset shows the orientations adopted by the side-chains of residues R113 and R196 located at the AQP0 surface,in the refined structure (yellow) compared to the original structure (cyan). **E, F.** As a control, lipids (red) were modeled into the density map derived from a non-converged 100-ps MD simulation. Panel E shows lipids modeled into density at surface S1, and panel F shows lipids modeled into density at surface S2. **G.** Comparison of the lipids before (red) and after refinement against the electron crystallographic data (gold) for the non-converged data set. **H.** Comparison of the refined lipids (gold) with the original lipids in the electron crystallographic structure of AQP0 (cyan) for the non-converged data set.

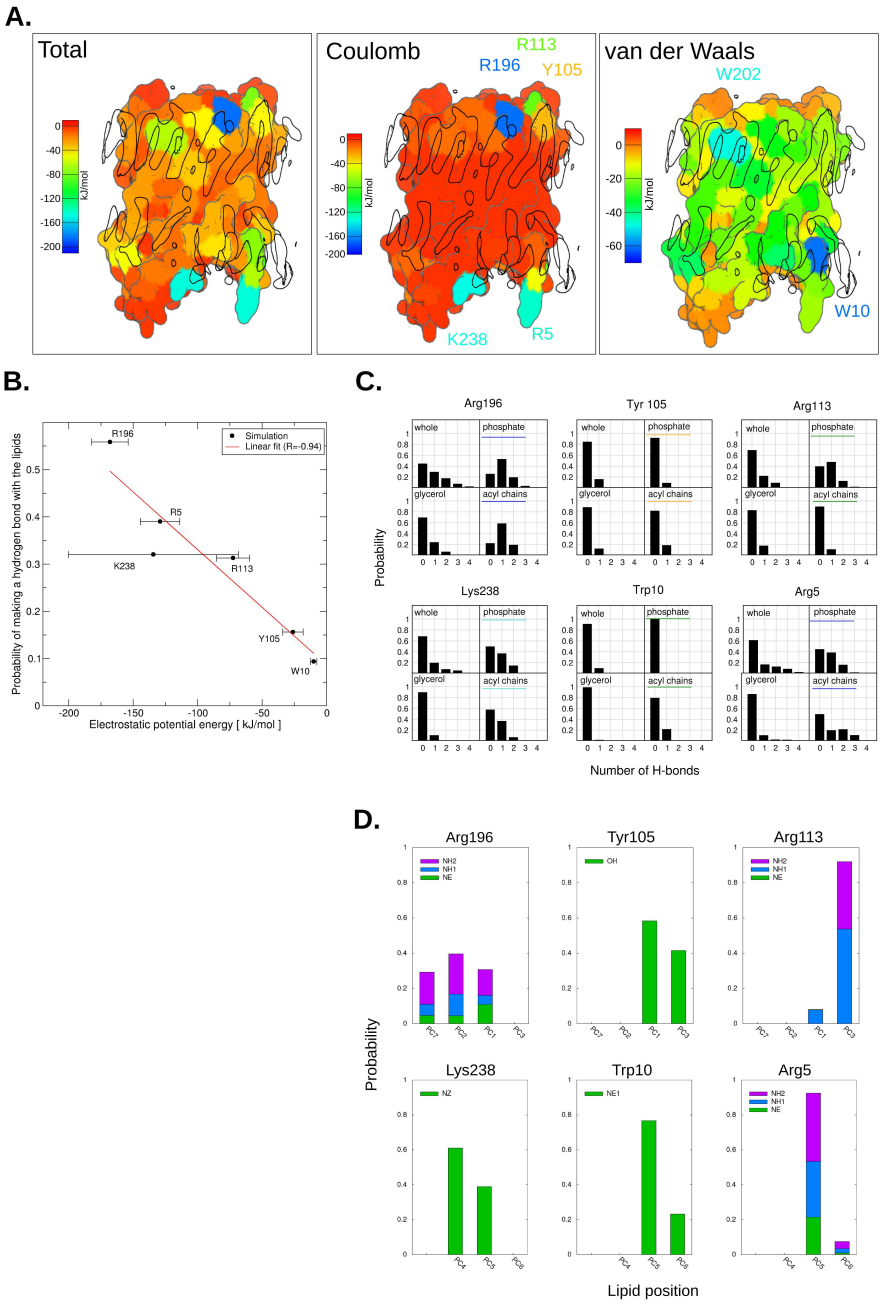

Figure S7. Interaction energy between lipids and AQP0 derived from MD simulations. A. The left panel shows the total time-averaged potential energy. The middle and right panels show the contributions of electrostatic Coulomb and short-range van der Waals interactions, respectively, to the total potential energy. The colors represent the energies according to the scales in the individual panels. Labels indicate AQP0 residues that interact strongly with lipids. Black outlines represent high lipid-density regions contoured at 4σ. **B.** Probability that the strongly interacting AQP0 residues form a hydrogen bond with a lipid as a function of the electrostatic potential energy. **C.** Probability that the strongly-interacting AQP0 residues form a certain number of hydrogen bonds either with any part of a lipid or specifically with the phosphate, glycerol or acyl chains of a lipid. **D.** Probability that the strongly-interacting AQP0 residues form a hydrogen bond with lipids located at different crystallographic positions (labeled PC1 to PC7). Potential energies (A) and probability distributions (B-D) were obtained from the simulation of a single AQP0 tetramers surrounded by a DMPC bilayer without the crystallographic lipids (system shown in Figure S1A, left).

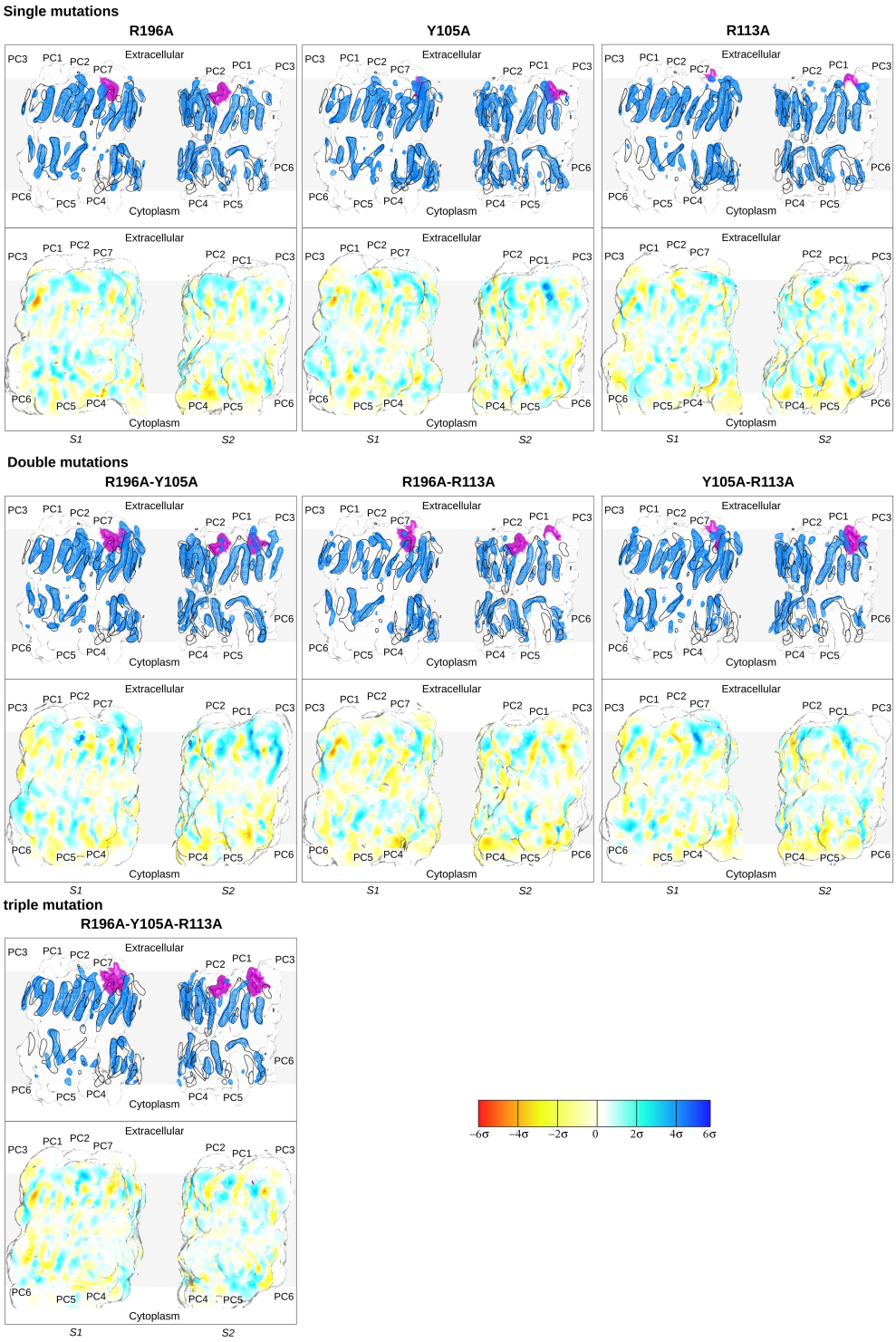

Figure S8. Effect of mutations of AQP0 surface residues on the lipid-density maps derived from MD simulations: mutations of residues involved in strong electrostatic interactions with lipids in the extracellular leaflet. Each panel corresponds to one of the introduced mutations. The upper figures in each panel show the mutated residue(s) in purple and the resulting lipid-density maps contoured at 4σ as blue mesh. For comparison, the lipid-density map obtained with wild-type AQP0 contoured at 4σ is shown as black contours. The lower figures in each panel show the differences between the lipid densities in maps obtained with mutant and wild-type AQP0 (see scale for color coding).

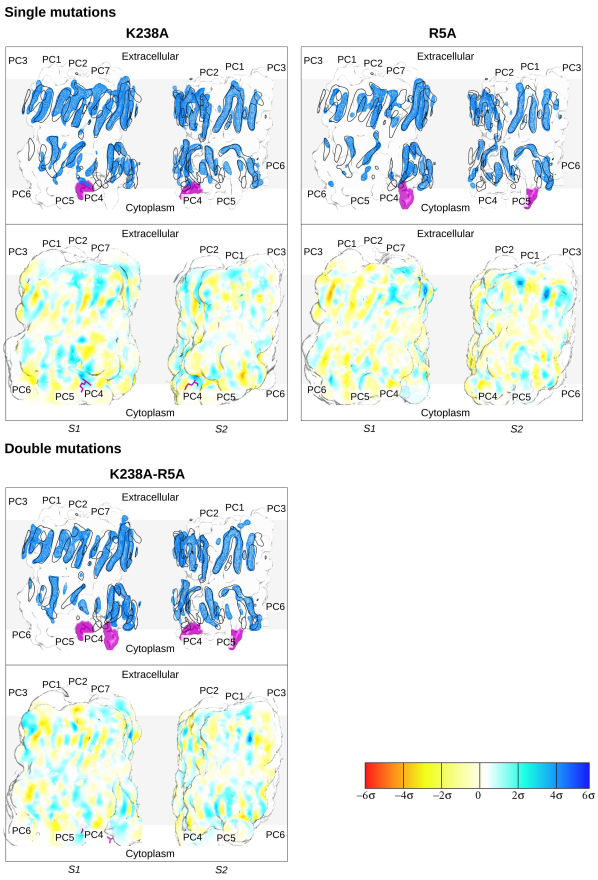

Figure S9. Effect of mutations of AQP0 surface residues on the lipid-density maps derived from MD simulations: mutations of surface residues involved in strong electrostatic interactions with lipids in the cytoplasmic leaflet. The same format is used as in Figure S8.

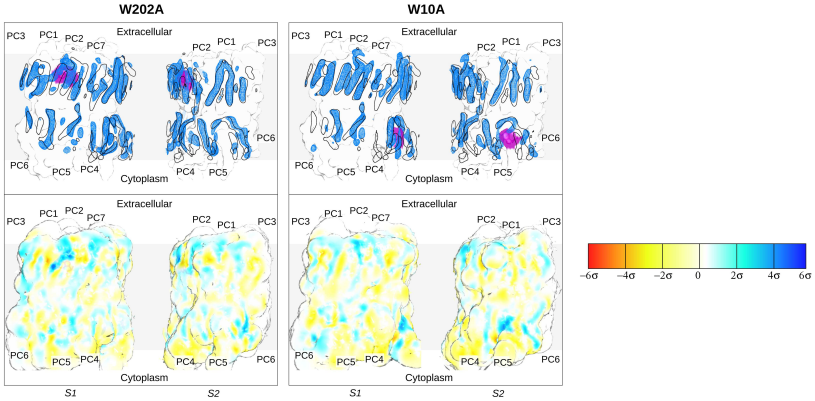

Figure S10. Effect of mutations of AQP0 surface residues on the lipid-density maps derived from MD simulations: mutations of surface residues involved in strong van der Waals interactions with lipids. The same format is used as in Figure S8.

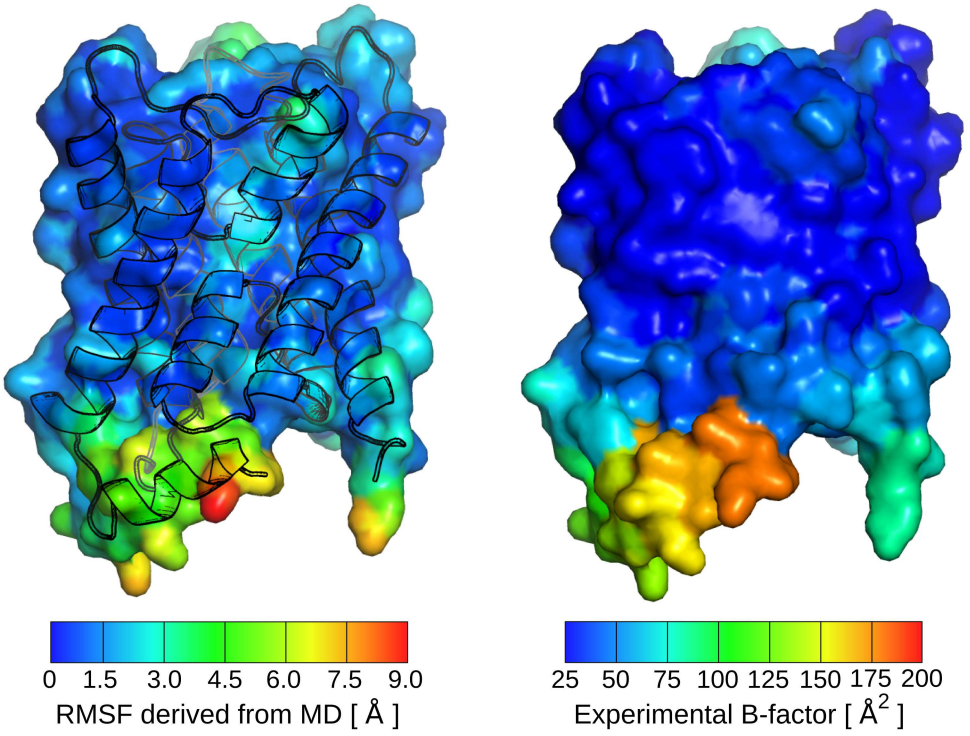

Figure S11. Comparison of the RMSF of AQP0 atoms derived from MD simulations with their B-factors from the electron crystallographic structure of AQP0 (1). The RMSF values and B-factors are color-coded according to the scales at the bottom.

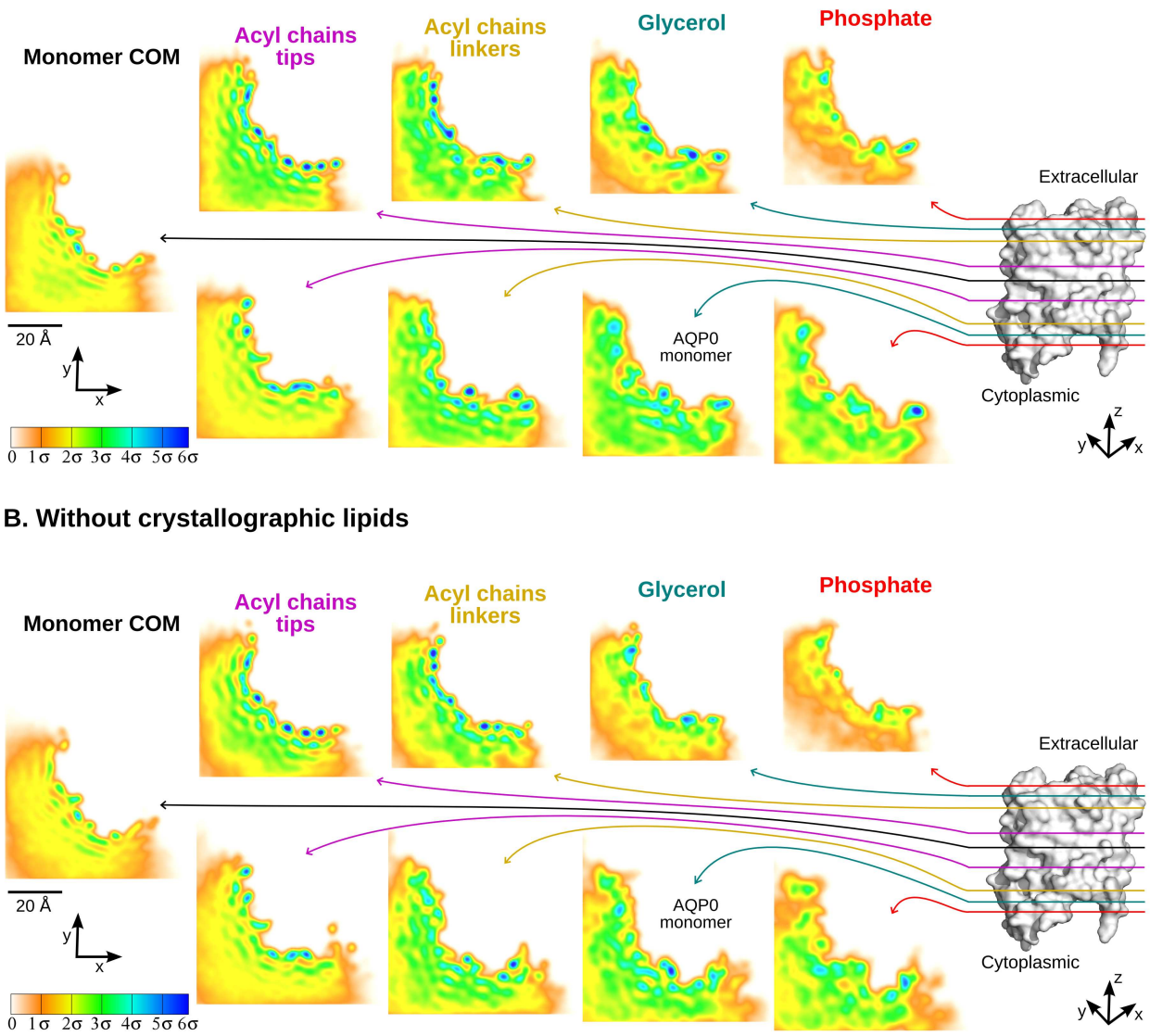

Figure S12. Lipid density beyond the annular shell of lipids, recovered from the MD simulations of a single AQP0 tetramer with (A) or without (B) the crystallographic lipids. The color maps at the left represent lateral projections (onto the xy membrane plane) of the lipid density, at the different z positions indicated in the right figure. Projections were taken at the average z positions of the center of masses (COM) of the indicated lipid groups (for both leaflets) and the AQP0 monomer.

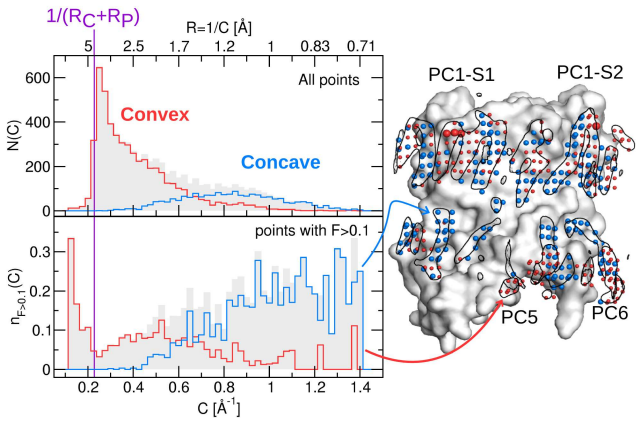

Figure S13. Protein surface curvature, C, and concavity computed from MD simulations. Left, top panel: Histogram of the curvature values $N(C)$ (gray), separated into contributions by convex (red) and concave (blue) surfaces. Left, bottom panel: density of curvature points with a high lipid density in their vicinity ($F > 0.1$), $n_{F>0.1}(C)$. As a reference, the purple vertical line shows the curvature value equal to $1/(R_C + R_p)$, with $R_C = 1.88$ Å and $R_p = 2.5$ Å being the carbon and probe radii, respectively. Curvature points at high lipid-density regions ($F > 0.1$) are illustrated on the AQP0 monomer (right). To guide the eye, the black contours show high lipid-density regions ($\rho > 4\sigma$) on the AQP0 monomer.

2 Structure refinement

Comparison of the crystallographic lipids with the density map derived from MD simulations

To assess the similarities and differences between the lipid conformations in the 2D crystal structure and the conformation they adopt in the MD simulation of a single AQP0 tetramer, the lipid conformations in the crystal were compared to the lipid densities computed from MD simulations. The crystal data exhibits a fourfold rotational symmetry within each protein-lipid layer, and the two lipid-facing surfaces of the protein (S1 and S2) thus inherently show identical lipid conformations. In contrast, the MD simulation is not restrained by any symmetry, and lipids at surfaces S1 and S2 can thus adopt different conformations. As a result, the comparison has to treat the two lipid-facing surfaces independently, even though the conformations of the crystallographic lipids are identical.

The MD-derived density map shows high-density regions for the hydrophobic part of the lipid bilayer (Figs. 1 and S2). Comparison with the lipids in the electron crystallographic structure of AQP0 (PDB code 2B6O) (1) shows that some of the high-density regions coincide with parts of the acyl chains of the crystallographic lipids. In the extracellular leaflet, density in the MD map shows good agreement with both chains of PC1 on S2 and a fragment of a chain on S1, as well as with acyl chain fragments of PC3 on S1 and S2. In the cytoplasmic leaflet, the MD density overlaps with a fragment of an acyl chain of PC6 on S1 and on S2. As a general trend, the MD density correlated best with the acyl chains with the lowest B-factors.

The presence of overlapping positions of the MD-derived density map with fragments of the crystallographic lipids suggests that these positions are not only occupied by lipids in the context of an AQP0 2D crystal but also when lipids surround an individual AQP0 tetramer.

The most striking difference between the MD density and the electron crystallographic lipids is the almost complete absence of density for the lipid head-groups. Additional differences can be seen for crystallographic lipids PC2, PC4, PC5 and PC7 and the two bulk lipids PC8 and PC9, for which either strong density is absent in the MD map or at a different position. To assess whether these differences constitute inconsistencies between the two datasets or represent alternative lipid conformations, which can also be seen in the crystallographic density map, we modeled lipids into the MD density (Fig. S6A, B) and then refined them against the electron crystallographic data.

Comparison of the lipids modeled into the MD density map with those after refinement against the electron crystallographic data

Due to the lack of symmetry restrictions in the MD map, each lipid could be modeled either into the density at surface S1 (Fig. S6A) or S2 (Fig. S6B). Unless it conflicted with the position of the protein side-chains of the crystallographic structure, the lipids were modeled into the stronger of the two densities.

After refining the initial lipid models against the electron crystallographic data, the refined structure included five complete lipids in the extracellular leaflet and four in the cytoplasmic leaflet. Both leaflets in addition contained a single acyl chain near the fourfold axis. The four symmetry-related acyl chain densities at the fourfold axis presumably represent two bulk lipids that do not follow the overall fourfold symmetry of the AQP0 2D crystal.

Comparison of the final refined model with the initial model shows that some lipids retained their position, whereas others changed their position completely (Fig. S6C). In the extracellular leaflet, PC1 remained essentially unchanged, and the two stretches of acyl chain for PC9 that were initially modeled into the MD density also stayed in the same position. PC3 represents a special case, because it is positioned very close to the twofold symmetry axis of the 2D crystal. It could not be modeled directly into the MD density, as the MD density for this lipid extends over the twofold axis (Fig. S6A, B). Of the two acyl chain stretches that were initially modeled into the MD density, one stayed in place while the other one rotated by approximately 45°, resulting in a different conformation. The refined structure of PC2 only shared a very small fragment of one of its acyl chains with the initial structure.

In the cytoplasmic leaflet, PC5 retained the overall conformation of one of its acyl chains, even though it is shifted by approximately 2.5 Å. PC4 shared the position of a fragment of one acyl chain with its initial position. PC9 and PC11 in the extracellular leaflet and PC6, PC8, and PC10 in the cytoplasmic leaflet showed no similar position in the initial and refined structures.

The conservation of lipids from the model based on the MD density in the refined model shows that these extracellular lipid positions are not only preferred in a crystalline environment, but also when the lipids surround an individual AQP0 tetramer. The lack of packing and symmetry constraints in the MD simulation of a single tetramer allow the lipids, however, to adopt a larger variety of conformations that have no equivalent in the crystal structure.

Comparison of the refined lipids initially modeled into the MD density map with the lipids in the electron crystallographic structure of AQP0

The information provided by the MD simulation allowed identification of an additional acyl chain in the extracellular leaflet and an additional full lipid in the cytoplasmic leaflet that were not modeled in the original electron crystallographic structure (Fig. S6D). Comparison of the lipids modeled into the MD density map and refined against the electron crystallographic data with the lipids in the original electron crystallographic structure of AQP0, shows that all four annular lipids in the extracellular leaflet (PC1, PC2, PC3 and PC7) retained their original position (Fig. S6D). The conserved position of these four lipids strongly indicates that these lipid positions are indeed dominant in the crystalline environment.

In addition, one of the acyl chains of at least three more lipids in the refined structure coincides with an acyl chain of the original crystallographic lipids. PC6 in the refined structure shares one acyl chain with PC6 of the crystallographic structure, and one acyl chain of each PC9 and PC11 in the refined structure occupy the space originally filled by the two acyl chains of PC9 in the crystallographic structure.

There are, however, some local differences between the lipids in the refined structure and their direct counterparts in the original crystallographic structure. For example, one acyl chain of lipid PC1, which shows the best-defined density in all the datasets, exhibits a different conformation in the two structures. The lower part of the acyl chain (atoms C19-C24) is oriented at an angle of 49° versus its upper part, while it shows a straighter conformation in the crystallographic structure. In addition, the conformations of lipids PC3 and PC7 are straighter in the refined structure than in the crystallographic structure, in which the glycerol backbone and the first few carbon atoms of the acyl chains form a distinct bend.

The two hydrogen bond-forming protein residues Arg113 and Arg196 also have different orientations in the two structures, indicating that they are mobile. The mobility of these residues strongly suggests that the hydrogen bonds they form with lipids are of a transient nature.

Comparison of the lipids modeled into the non-converged MD density map with those after refinement against the electron crystallographic data

To address the question whether the initial choice of lipid positions and conformations biases the final refined lipid structures, a second set of lipid starting conditions was chosen to repeat the refinement against the experimental crystallographic data. For this purpose, an MD simulation was stopped after 100 ps and used to calculate a time-averaged lipid-density map. This map corresponds to a non-converged situation that is dominated by the starting coordinates of the lipids, that have not equilibrated with respect to the protein surface. Lipids were then modeled into this non-converged density map to provide an alternative starting model for refinement against the electron crystallographic data (Fig. S6E, F). As before, lipids were modeled into the stronger density at S1 or S2, unless it conflicted with the position of a protein side-chain or the crystal symmetry.

The initial refinement steps for lipids modeled into the non-converged density map exhibited

big shifts for those lipids that were at positions different from where they were in the converged density map or the original 2D crystal structure (Fig. S6G, H). PC2 is the only lipid that retained the same position for both acyl chains before and after refinement, but even this lipid showed a considerable shift in the z-height of its glycerol backbone. For lipid PC10, the position of one acyl chain overlapped in the initial and refined structures, and PC4 retained the position of its glycerol backbone. The conformation of all other lipids changed with refinement.

In terms of acyl chain and glycerol backbone, the positions of the lipids in both leaflets gradually converged during refinement to the positions previously seen for the lipids modeled into the converged MD density map and refined against the electron crystallographic data (compare Figs. S6C and S6G). The positions and conformations of the lipids in the refined structure therefore do not appear to be noticeably biased by the chosen starting conditions. This result proves that potentially wrong lipid positions and conformations used as initial model are not propagated to the final refined lipid structures, and that the crystallographic data are sufficiently strong to allow determination of the correct lipid positions during refinement even if the chosen starting model is imprecise.

3 Surface curvature and concavity

To analyze the effect of protein surface curvature, C, on the lipid positions, we first plotted a histogram, $N(C)$, of the curvature values of the protein surface that is in contact with lipids (Fig. S13, left upper panel). The curvature values range from $0.1\,\text{Å}^{-1}$ to $1.4\,\text{Å}^{-1}$ with a main peak close to $C = 0.228\,\text{Å}^{-1}$. This value corresponds to the inverse of the sum of the carbon atom radius ($R_C = 1.88\,\text{Å}$) plus the probe radius ($R_C = 2.5\,\text{Å}$), $C = 1/(R_C + R_p)$. Classifying the points according to their concavity revealed that surfaces with low curvature were mostly convex, whereas surfaces with large curvatures were mostly concave.

We then calculated a reduced histogram, in which we only considered surface points surrounded by high lipid density ($F > 0.1$; in this case we used a cylinder of radius 1.5 Å and height 2 Å centered at each curvature point). Approximately 8% of the convex and 21% of the concave surface regions were found to be close to high lipid-density regions. Subsequently, this histogram was divided by $N(C)$ to obtain the density of curvature points surrounded by high lipid density, $n_{F>0.1}(C) = N_{F>0.1}(C)/N(C)$ (Fig. S13, left lower panel). Up to 30% ($n_{F>0.1} = 0.3$) of surface points with curvature values smaller than $0.228\,\text{Å}^{-1}$ or larger than $0.8\,\text{Å}^{-1}$ were close to high lipid-density points, while only about 10% of surface points with intermediate curvature values (0.228-$0.8\,\text{Å}^{-1}$) were close to high lipid-density points. Accordingly, marking curvature points with $F > 0.1$ on the AQP0 monomer (Fig. S13, right figure) allowed the lipid-accessible protein surface to be characterized according to its concavity. Lipids accommodated to both concave (blue) and convex (red) surfaces of AQP0, as illustrated by the highly localized lipid positions of PC1 at S2 and PC6 on concave AQP0 surfaces, and PC1 at S1 on a convex AQP0 surface.

Our calculations capture the essential features of the AQP0 protein surface, allowing us to distinguish between low curved convex regions (bumps), and highly curved concave areas (clefts).

Accordingly, the lipid density did not show a strong preference for either type of concavity, and high lipid-density points were observed near to both concave and convex surface regions, such as PC1 sitting in a cleft in AQP0 surface S2 or the same PC1 residing on a bump on surface S1. Our curvature calculations therefore support the notion that lipids adapt to the roughness of the protein surface (clefts or bumps) to form a tight seal around the protein that prevents leakage of solutes across the membrane.

4 Force field consistency

The following results indicate that the OPLS protein force field, the Berger lipid parameters and the TIP4P water model are consistent and can be combined to study the localization of lipids around AQP0. We used a simulation of a pure lipid bilayer of 98 DMPC Berger lipids solvated by 3528 TIP4P water molecules to compute the area per lipid and the membrane thickness, two critical structural properties of a lipid bilayer. The computed area per lipid was $0.610\,\mathrm{nm}^2$ ($0.008\,nm^2$ standard deviation), which is in excellent agreement with the experimental value of $0.606\pm0.005\,\mathrm{nm}^2$ (2). The membrane thickness (measured as the phosphate-to-phosphate distance between the two leaflets) was found to be 3.60 nm (0.07 nm standard deviation), which is also very close to the experimental value of 3.53 nm (2). The ability of the simulation to reproduce these two structural parameters underscores the validity of the used force field to study the localization of lipids around AQP0.

In recent molecular dynamics simulations (3), Berger parameters for lipids and the TIP4P model for the waters were used to study the partition properties of DMPC lipid bilayers (among six other types of lipid bilayers). The energetic cost to move several solutes (ammonia, ethanol, nitric oxide, benzene, propane, and neopentane) from bulk water to the inner part of the lipid bilayer was found to be in good agreement with the energetic cost to move such solutes from water to hexadecane, $\Delta G = -K_B T \ln K_{hex}$. Here, K_B is the Boltzmann constant, T is the temperature (300 K), and K_{hex} is the hexadecane/water partition coefficient. This result indicates that the use of Berger parameters for lipids together with the TIP4P model for waters correctly reproduces the partition properties of DMPC lipid bilayers.

The favorable agreement between the computed lipid-density maps and the crystallographic structures (that we obtained) provides an independent validation of the used force field parameters.

Finally, our simulations revealed similar immobilization patterns when lipids were close to AQP0, as the ones predicted in a computational study of an ion channel (4). In the latter, the OPLS force field was used for the protein, Berger parameters for the lipids, and the SPC model for the water. The fact that two independent computational studies yielded similar results suggests that the lipid localization around membrane proteins (and ultimately the mechanisms underlying lipid-protein interactions) is properly described by the use of the OPLS force field in combination with the Berger lipid parameters, and is not severely affected by the choice of water models (SPC or TIP4P).

5 Materials and methods

MD simulations

Two different systems were simulated (Fig. S1). The first system consisted of a single AQP0 tetramer embedded in a DMPC lipid bilayer, simulating a membrane at low protein concentration. The second system included four densely packed AQP0 tetramers in the 2D crystal arrangement with 128 DMPC molecules filling the gaps in between the tetramers. For the single-tetramer system, one simulation was performed with the AQP0 tetramer inserted into an equlibrated patch of DMPC molecules, and a second one also included the lipids seen in the crystallographic structure of AQP0 (1). The resulting number of DMPC lipids around the AQP0 tetramer was 278 and 288 for the simulation with and without the crystallographic lipids, respectively. For the four-tetramer system, simulations were carried out at temperatures of 280 K and 300 K.

In both systems the membrane was solvated by around 23000 explicit water molecules. The initial structure of AQP0 was taken from the Protein Data Bank (PDB ID code 2B6O (1)). Water molecules observed in the crystallographic structure were also included, and the system was neutralized by adding chloride ions. In the simulations with a single tetramer, the tetramer was inserted into the lipid bilayer by using the g_membed software (5). In the simulations with four tetramers, the initial arrangement of the four tetramers was generated by applying the crystallographic symmetry operations provided in the pdb file of the electron crystallographic AQP0 structure.

The OPLS-AA all-atom force field (6, 7) was used for the protein, Berger parameters (8) for the lipids, and the TIP4P model (9) for water molecules. The simulations were carried out using the GROMACS 4.0 simulation package (10–12). Equations of motion were numerically integrated by using the leap frog algorithm (13). Bond lengths and angles of water molecules were constrained by using the Settle algorithm (14). The remaining bonds were constrained with Lincs (15), and angular vibrations involving hydrogen atoms were removed by using the virtual interaction-sites algorithm (16). The production runs were 100 ns in length, and the integration time step was 4 fs. Electrostatic interactions were calculated with the particle-mesh Ewald method (17, 18). Short-range non-bonded interactions were considered by a Lennard-Jones potential, within a cut-off of 1.0 nm. To maintain the temperature constant, the system was coupled to a velocity-rescaling thermostat (19, 20). The reference temperature for the simulations with the single-tetramer system was 300 K, while it was 280 K and 300 K for the four-tetramer system. In both cases the coupling constant t was 0.1 ps. The pressure was kept constant at 1 bar by employing the semiisotropic Berendsen barostat (19), with a coupling constant of $t = 1.0$ ps. In all simulations, a 4 ns equilibration step preceded the production run, in which the coordinates of the protein were harmonically restrained, with a harmonic force constant of 1000 $\mathrm{kJmol^{-1}nm^{-2}}$. The first 10 ns of the production runs were removed to account for equilibration time.

Additional simulations with AQP0 mutants, in which residues of interest were substituted by alanine, were carried out following the same simulation scheme described for the single-tetramer system without the crystallographic lipids.

To validate the consistency of the Berger lipid parameters with the TIP4P water model, an additional 240-ns MD-simulation of a pure lipid bilayer of 98 DMPC Berger lipids solvated by 3528 TIP4P water molecules was carried out. The same algorithms and simulation parameters described for the single-tetramer simulation were used, but with a coupling constant of 1 ps for the thermostat and 5 ps for the barostat.

Lipid-density maps derived from MD simulations

The density map (defined in real space) is correlated with the atomic scattering factors (defined in reciprocal space) (21). The atomic scattering factor is defined as:

$$f(k = 2\sin\theta/\lambda) = \sum_{i=1}^{4} a_i \exp\left\{-(b_i + B)k^2/4\right\} + c_1 \exp\left\{-Bk^2/4\right\}. \quad (1)$$

Here, $f(k)$ is fitted to four Gaussian functions, with the fitting parameters a_i, b_i and c_1 depending on the atom type. a_i, b_i and c_1 were taken from Hirai et al. (21). The B-factor is explicitly taken into account. The lipid-density map $\rho(r)$ is calculated by Fourier transformation of $f(k)$ (21):

$$\rho(r) = \sum_{i=1}^{4} \frac{a_i\sqrt{\pi}}{\sqrt{b_i + B}} \exp\left\{-\frac{4\pi^2 r^2}{b_i + B}\right\} + \frac{c_1\sqrt{\pi}}{\sqrt{B}} \exp\left\{-\frac{4\pi^2 r^2}{B}\right\}. \quad (2)$$

where r is the lipid atom coordinate. Due to the fourfold symmetry of the AQP0 tetramer, each monomer has identical lipid interfaces, and $\rho(r)$ was thus calculated for a single AQP0 monomer. The trajectories of the monomers (four in the single-tetramer and 16 in the four-tetramer system) together with their closest surrounding lipids were concatenated, after fitting the monomer to the reference crystallographic structure. The average lipid density, $\langle\rho\rangle$, was calculated by time-averaging the instantaneous lipid density, $\rho(t)$, over the concatenated trajectories (360 ns for the single-tetramer system and 1440 ns for the four-tetramer system) over time. $\rho(t)$ was calculated at every time step t, in a 3D grid of 70 Å × 70 Å × 60 Å, with a resolution of 0.4 Å, and centered at the center of mass of the AQP0 monomer. The density at the $i-$th point of the grid was estimated by summing up the ρ contributions (given by equation 2) of the atoms nearby the grid point:

$$\rho_i(t) = \sum_{j \in \text{cut off}} \rho(|\mathbf{R}_i(t) - \mathbf{r}_j(t)|), \quad (3)$$

where $R_i(t)$ and $r_j(t)$ are the coordinates of the $i-$th grid point and $j-$th lipid atom, respectively. Only atoms within a cut-off distance of 0.3 Å to the $i-$th grid point were included in the summation. ρ remained practically unchanged for $B-$factors of 1 Å^2 and 20 Å^2, and a $B-$factor of 20 Å^2 was thus chosen for the calculations. The maps were displayed and analyzed with the PyMOL software (22).

Model building and refinement

The topology files for the DMPC lipids were generated using the ProDRG server (http://davapc1.bioch.dundee.ac.uk/prodrg/ (23), followed by manual renaming of the atom labels to adhere to the conventions used in PDB file 2B6O. As initial model for refinement the protein coordinates of PDB 2B6O (without the C-terminal helix) were used. Lipids or lipid fragments were modeled into the densities of the converged or non-converged MD-derived maps using COOT (24). For each lipid the stronger of the two possible densities at S1 or S2 was chosen, unless the stronger density conflicted with the positions of protein side chains of the electron crystallographic AQP0 structure. Lipid densities that conflicted with the *p422* symmetry of the AQP0 crystal were only partially modeled with lipid fragments to avoid symmetry clashes. The model was refined using CNS version 1.3 (25). After each round of simulated annealing, atoms represented by strong densities in $2F_o - F_c$, $F_o - F_c$ or composite-omit maps were added to the acyl chains. The refinement was continued iteratively until further cycles did not result in additional lipid density. The refinement was also guided by the use of MolProbity (26) to check for potential clashes.

Structures were visualized with DNG of the OpenStructure framework (27), and molecular surfaces were generated with msms (28). The MD-derived density maps, especially for the non-converged case, contained additional layers of bulk lipids in addition to the annual lipids. Because these lipid positions are forbidden in the context of a crystal (i.e., they would overlap with protein from neighboring unit cells), these lipids were removed by masking the MD density maps to only retain the densities that were either within 6 Å of any atom of the protein or within 3 Å of any atom of any of the lipids.

Lipid-protein interaction energy

The short-range non-bonded potential interaction energy of lipids with individual AQP0-surface residues was extracted from the simulations. This energy was then separated into electrostatic and van der Waals contributions and averaged over the entire simulation. The maximum standard deviations were 67.5 kJ/mol for the total interaction energy, and 5.5 kJ/mol and 65.8 kJ/mol for the total van der Waals and electrostatic terms, respectively.

Surface curvature and concavity

The surface curvature C was estimated as $C = \theta/d$, where θ is the angle formed by two normal vectors at two points on the protein surface separated by a distance d. Concavity values were assumed -1 (convex) if both angles formed by such normal vectors with the line connecting the surface points were larger than 90°, and +1 (concave) if these angles were smaller than 90°. Surface points were generated by rolling a sphere with a radius of 2.5 Å on the protein-surface atoms, as described by the Connolly algorithm (29). Subsequently, triplets of adjacent points were grouped to form triangles. For every triangle, a normal vector was calculated and placed at the center of the triangle. Curvature and concavity were then computed for every pair of adjacent triangles, by

evaluating their normal vectors and their separation. C was calculated over the entire surface, and stored in a 3D grid of 120 Å x 120 Å x 100 Å, with a resolution of 2 Å, and centered at the center of mass of the AQP0 monomer. The final curvature value at the $i-$th point of the grid corresponds to the time-average over the trajectories of the four AQP0 monomers in the two independent single-tetramer simulations. Grid points that were not assigned with a curvature value for at least 10% of the simulation time were discarded. Concavity was stored in an identical 3D grid and time-averaged over the entire trajectory. Grid points with average negative concavity values (ranging between -1 and 0) were assumed to be convex, whereas points with positive concavity values (ranging between 0 and +1) were assumed to be concave.

Order parameters

The deuterium-order parameter, S_{CD}, of the $i-$th carbon atom of the lipid acyl chains (C_i) was calculated according to the following formula (30):

$$S_{CD} = \frac{2}{3}S_{xx} + \frac{1}{3}S_{yy}, \tag{4}$$

where S_{xx} and S_{yy} are defined as follows:

$$\begin{aligned} S_{xx} &= \frac{1}{2}\langle 3\cos^2\theta - 1\rangle \quad , \\ S_{yy} &= \frac{1}{2}\langle 3\cos^2\alpha - 1\rangle. \end{aligned} \tag{5}$$

Here, θ is the angle between the vector normal to the membrane plane ($\vec{z}$) and the vector normal to the plane defined by C_{i-1}, C_i and C_{i+1}. α is the angle between $\vec{z}$ and the vector defined in the plane through C_{i-1}, C_i and C_{i+1} but perpendicular to the vector connecting C_{i-1} to C_{i+1}. Order parameters, time-averaged over the entire MD trajectory, were calculated for both acyl chains of the lipids at the different crystallographic positions (labeled PC1 to PC8) around AQP0, by using the GROMACS (10–12) analysis tools.

6. Supplementary information references

[1] Gonen T, Cheng Y, Sliz P, Hiroaki Y, Fujiyoshi Y, Harrison SC, Walz T (2005) Lipid-protein interactions in double-layered two-dimensional AQP0 crystals *Nature* **438**:633–638.

[2] Kucerka N, Liu Y, Chu N, Petrache HI, Tristram-Nagle S, Nagle JF (2005) Structure of fully hydrated fluid phase dmpc and dlpc lipid bilayers using x-ray scattering from oriented multilamellar arrays and from unilamellar vesicles *Biophys J* **88**:2626–2637.

[3] Wennberg CL, van der Spoel D, Hub JS (2012) Large influence of cholesterol on solute partitioning into lipid membranes *J Am Chem Soc* **DOI**:10.1021/ja211929h.

[4] Niemelä PS, Miettinen MS, Monticelli L, Hammaren H, Bjelkmar P, Murtola T, Lindahl E, Vattulainen I (2010) Membrane proteins diffuse as dynamic complexes with lipids. *J Am Chem Soc* **132**:7574–7575.

[5] Wolf MG, Hoefling M, Aponte-Santamaría C, Grubmüller H, Groenhof G (2010) g_membed: Efficient insertion of a membrane protein into an equilibrated lipid bilayer with minimal perturbation. *J Comput Chem* **31**:2169–2174.

[6] Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) Development and testing of the OPLS All-Atom force field on conformational energetics and properties of organic liquids *J Am Chem Soc* **118**:11225–11236.

[7] Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on Peptides *J Chem Phys B* **105**:6474–6487.

[8] Berger O, Edholm O, Jähnig F (1997) Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature. *Biophys J* **72**:2002–2013.

[9] Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water *J Chem Phys* **79**:926–935.

[10] Spoel DVD, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC (2005) GROMACS: fast, flexible, and free. *J Comput Chem* **26**:1701–1718.

[11] Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation *J Chem Theory Comput* **4**:435–447.

[12] van der Spoel D, Lindahl E, Hess B, van Buuren A, Apol E, Meulenhoff P, Tieleman PD, Sijbers AM, Feenstra A, van Drunen R, Berendsen HJC (2005) *Gromacs User Manual version 4.0.* (www.gromacs.org).

[13] Hockney RW, Eastwood JW (1988) *Computer simulation using particles* (Bristol: Hilger).

[14] Miyamoto S, Kollman PA (1992) Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models *J Comput Chem* **13**:952–962.

[15] Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: A linear constraint solver for molecular simulations *J Comput Chem* **18**:1463–1472.

[16] Feenstra KA, Hess B, Berendsen HJC (1999) Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems *J Comput Chem* **20**:786–798.

[17] Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems *J Chem Phys* **98**:10089–10092.

[18] Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995) A smooth particle mesh Ewald method *J Chem Phys* **103**:8577–8593.

[19] Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath *J Chem Phys* **81**:3684–3690.

[20] Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling *J Chem Phys* **126**:14101.

[21] Hirai T, Mitsuoka K, Kidera A, Fujiyoshi Y (2007) Simulation of charge effects on density maps obtained by high-resolution electron crystallography *J Electron Micr* **56**:131–140.

[22] Schrödinger, LLC (2010) The PyMOL Molecular Graphics System, Version˜1.3r1.

[23] Schüttelkopf AW, van Aalten DMF (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes *Acta Crystallogr D* **60**:1355–1363.

[24] Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot *Acta Crystallogr D* **66**:486–501.

[25] Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination *Acta Crystallogr D* **54**:905–921.

[26] Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WBr, Snoeyink J, Richardson JS, Richardson DC (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids *Nucleic Acids Res* **35**:W375–383.

[27] Biasini M, Mariani V, Haas J, Scheuber S, Schenk AD, Schwede T, Philippsen A (2010) OpenStructure: a flexible software framework for computational structural biology *Bioinformatics* **26**:2626–2628.

[28] Sanner MF, Olson AJ, Spehner JC (1996) Reduced surface: an efficient way to compute molecular surfaces *Biopolymers* **38**:305–320.

[29] Connolly ML (1983) Analytical molecular surface calculation *J Appl Crystallogr* **16**:548–558.

[30] Egberts E, Berendsen HJC (1988) Molecular dynamics simulation of a smectic liquid crystal with atomic detail *J Chem Phys* **89**:3718–3732.