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
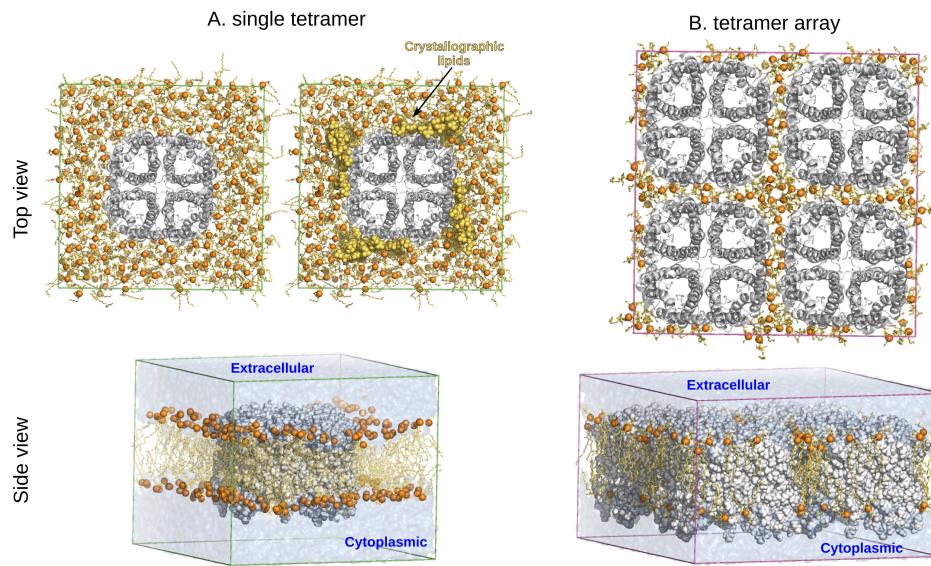
Figure S1. The two systems used for MD simulations of AQP0 in a DMPC lipid bilayer. A. A single tetramer embedded in a DMPC lipid bilayer. Simulations were carried out without (top view, left) or with (top view, right) the lipids seen in the electron crystallographic structure of AQP0 (1), facing the 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).
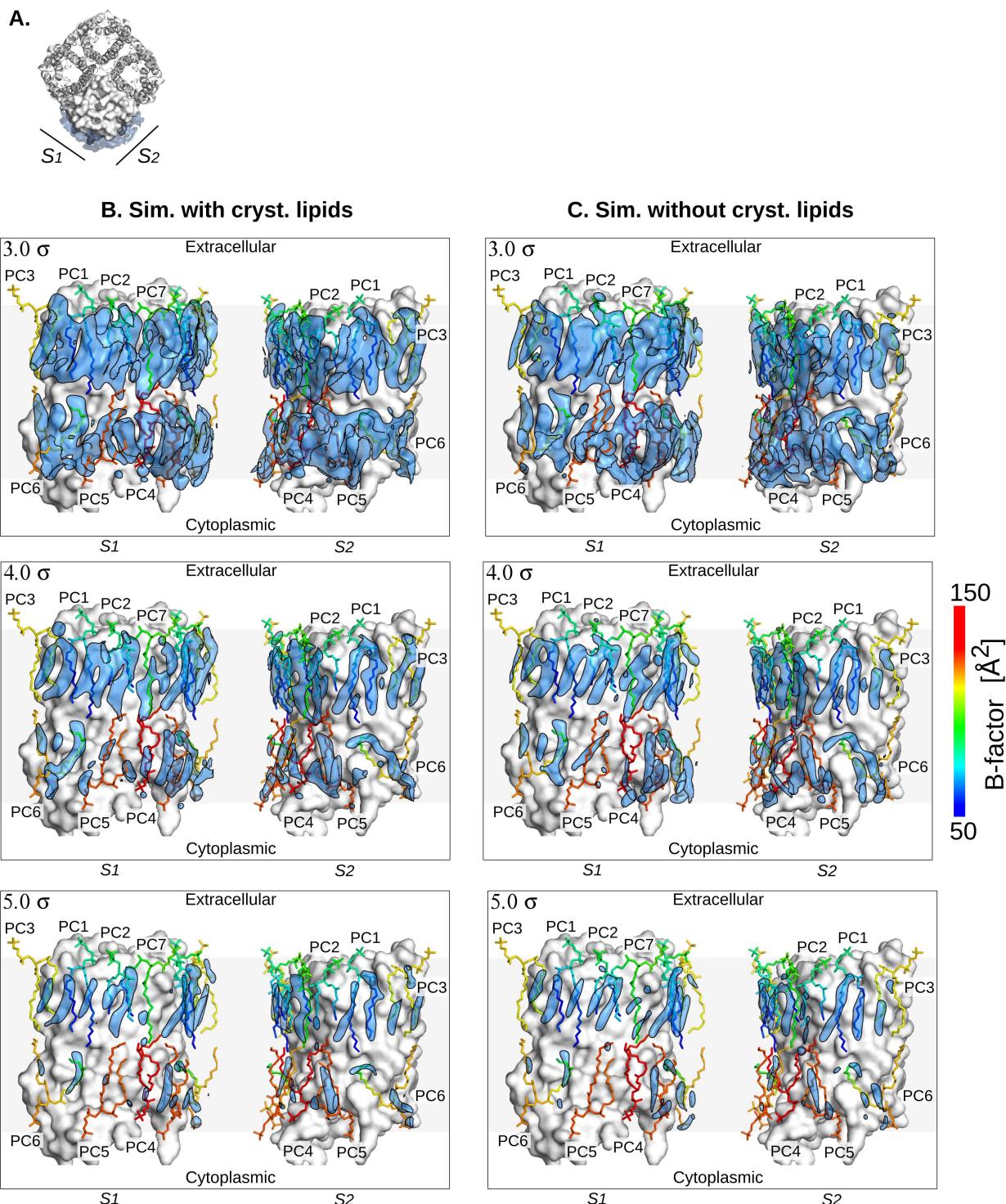
Figure S2. Comparison of density maps obtained in simulations with a single AQP0 tetramer with or without the crystallographic lipids in the starting configuration. A. Top view (perpendicular to the membrane plane) of an AQP0 tetramer, with S1 and S2 indicating the two lipid-facing surfaces of an AQP0 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).
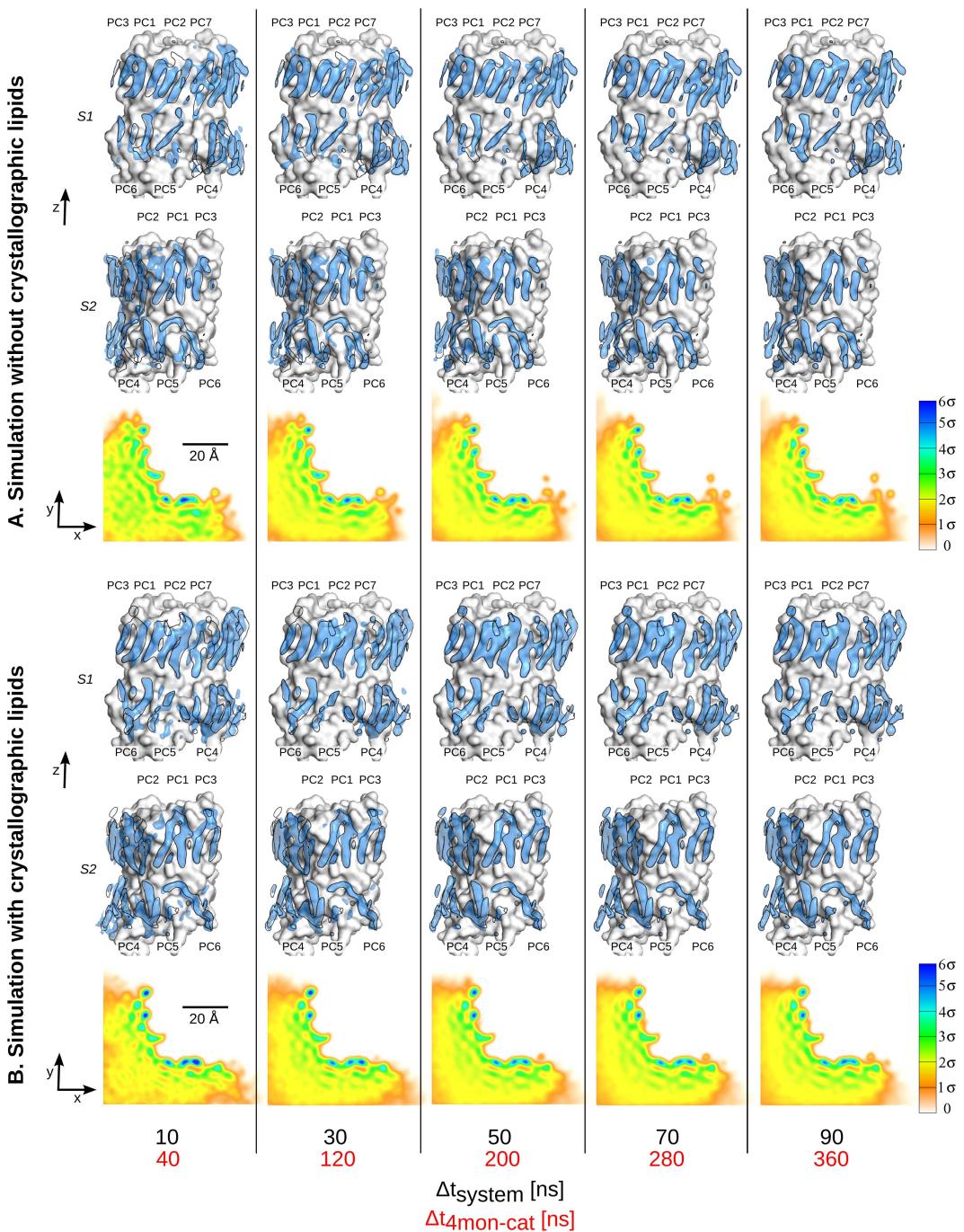
Figure S3. Density maps as a function of the length of trajectory used for their calculation. A, B. Lipid density maps obtained in the simulation with (A) and without (B) crystallographic lipids, by using a simulation of Δt_{system} in length. From each simulation, four independent trajectories of individual 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.

3
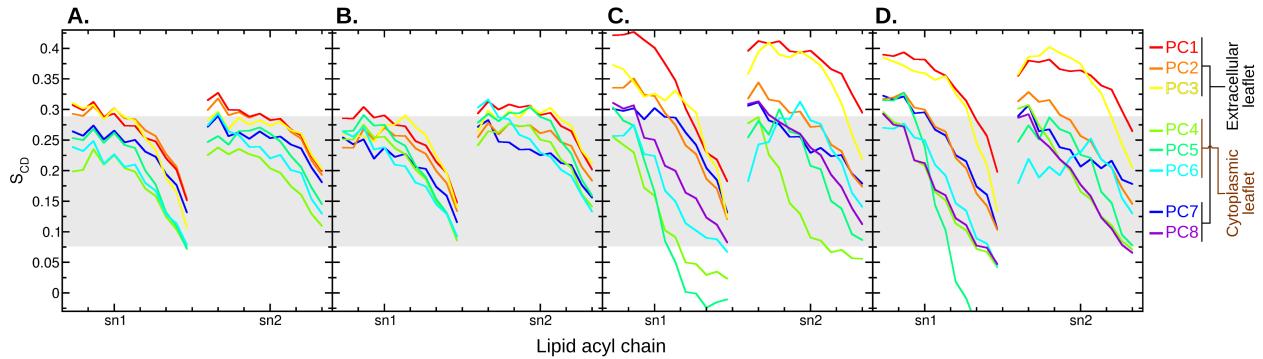
Figure S4. Deuterium order parameters, S_{CD}, for the lipid acyl chains derived from MD simulations. **A.** Lipids surrounding a single AQP0 tetramer including the crystallographic lipids. **B.** Lipids surrounding a single AQP0 tetramer without the crystallographic lipids. **C.** Lipids in the AQP0 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.
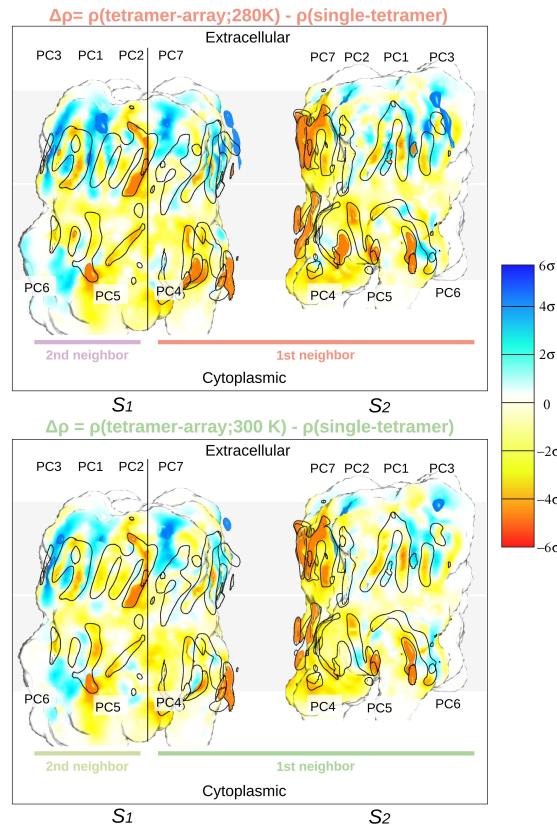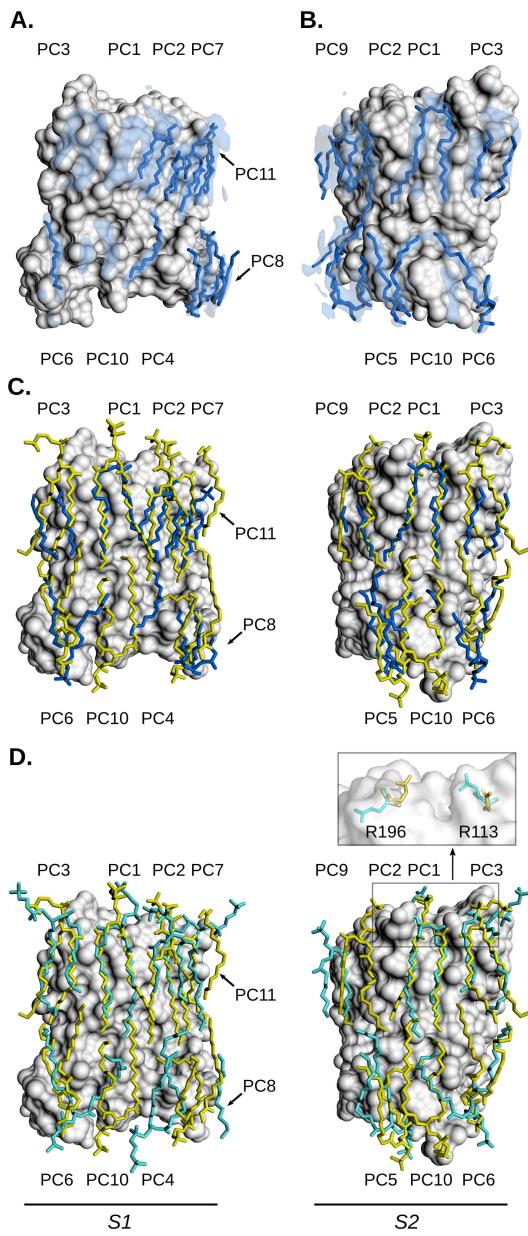

Figure S5. Difference of the lipid density computed from the simulation of the tetramer array (2D crystal) minus the lipid density computed from the simulation of a single tetramer. The difference is shown for the tetramer-array simulation at 280 K (upper panel) and 300 K (lower panel). The difference 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).

4
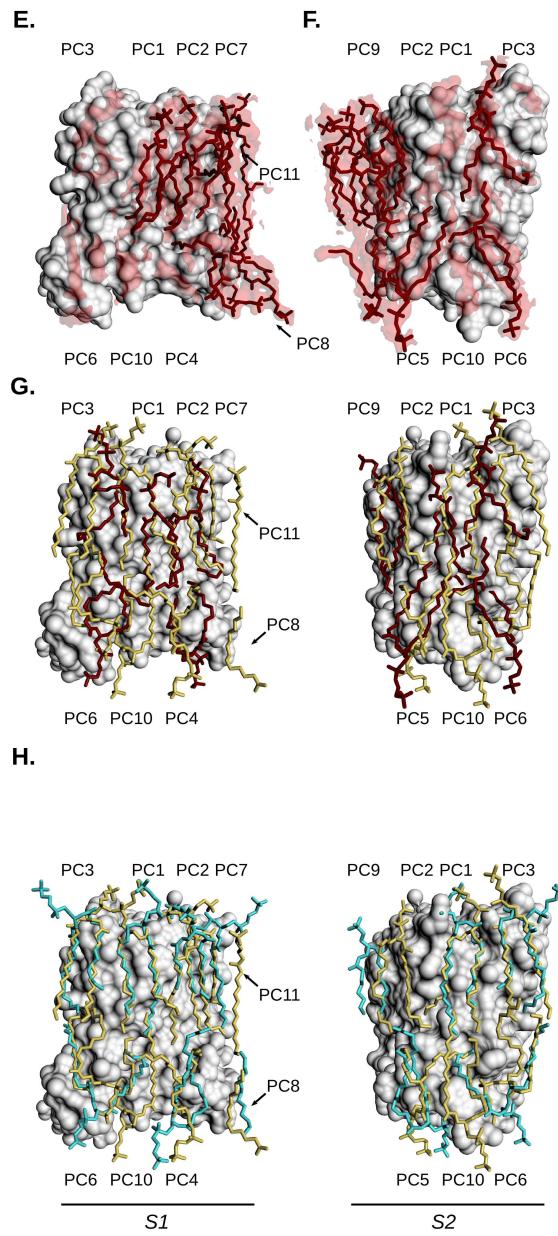
Figure S6. Refinement of lipids modeled into density maps derived from MD simulations, against the electron crystallographic data. A, B. Lipids (blue sticks) were first modeled into the density map derived from a converged 100-ns MD simulation of a single AQP0 tetramer that started from a configuration 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.
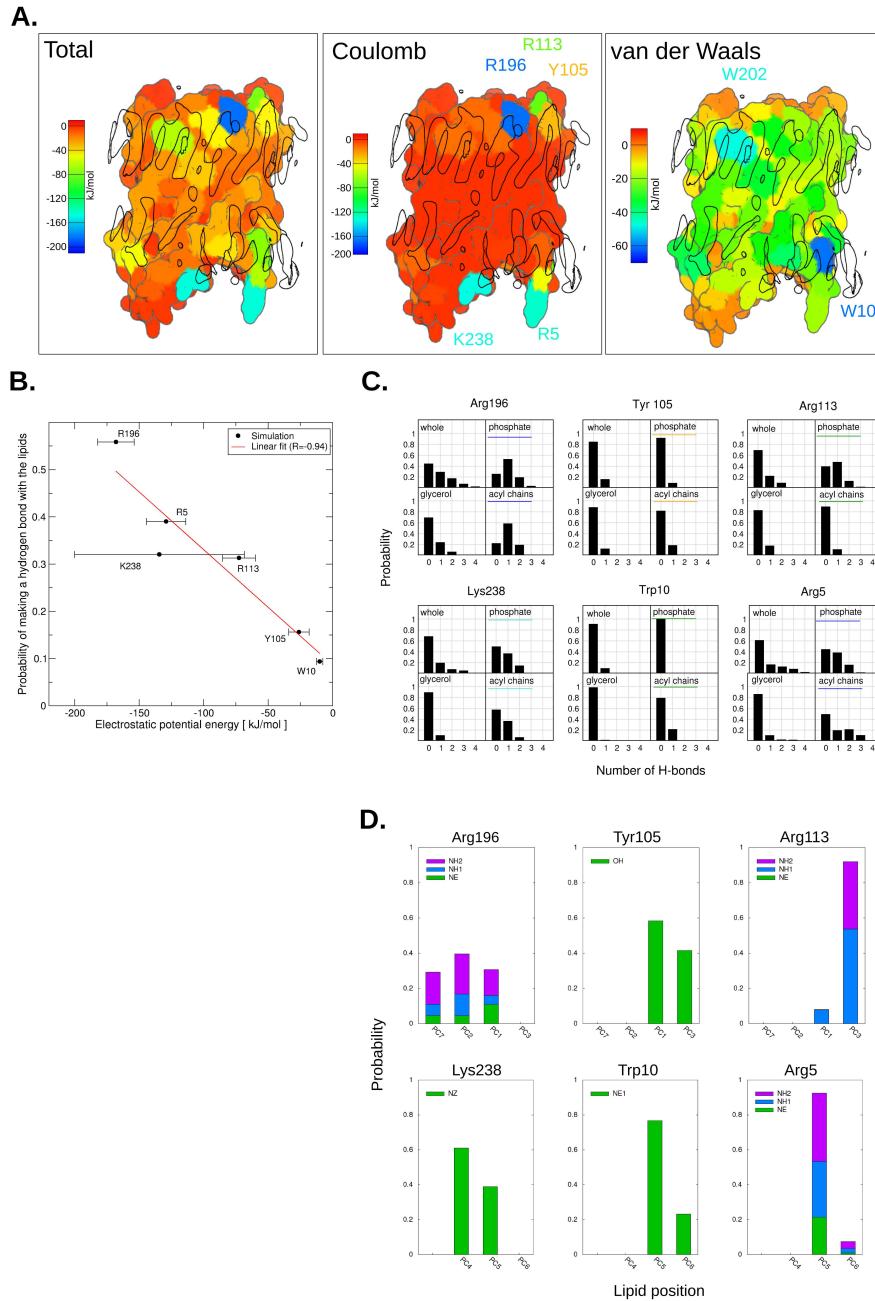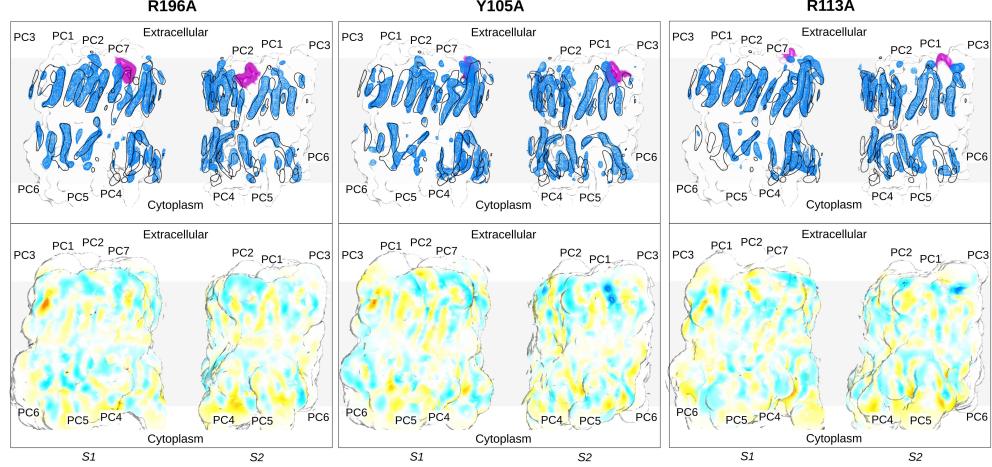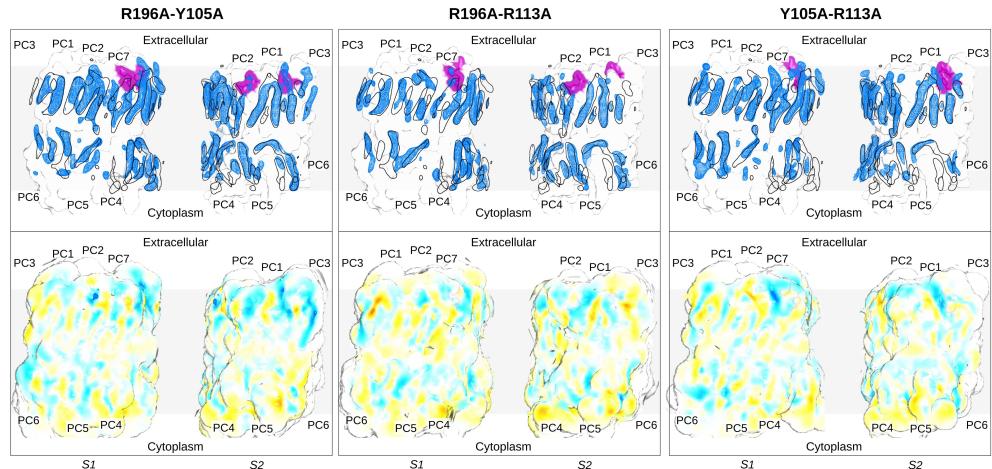
Figure S7. Interaction energy between lipids and AQP0 derived from MD simulations. A. The left panel shows the total time-averaged potential energy. The middle and right panels show the contributions of electrostatic Coulomb and short-range van der Waals interactions, respectively, to the total potential 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).

6
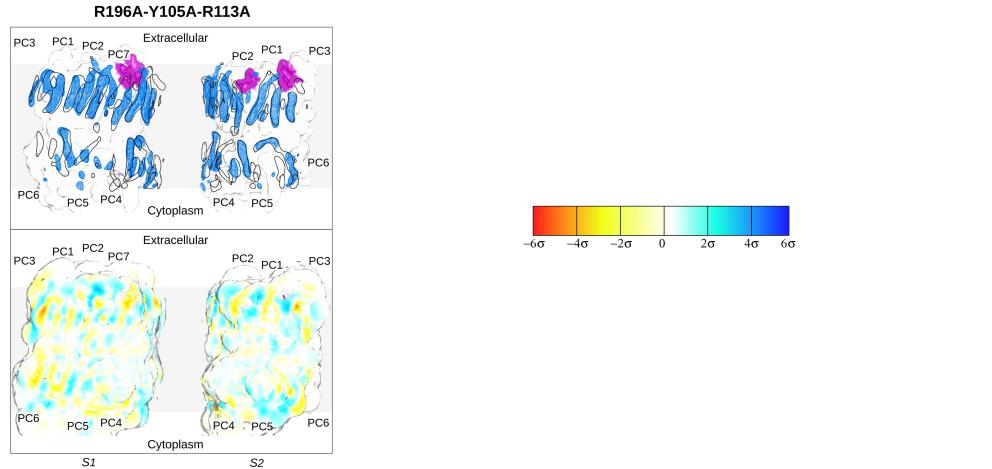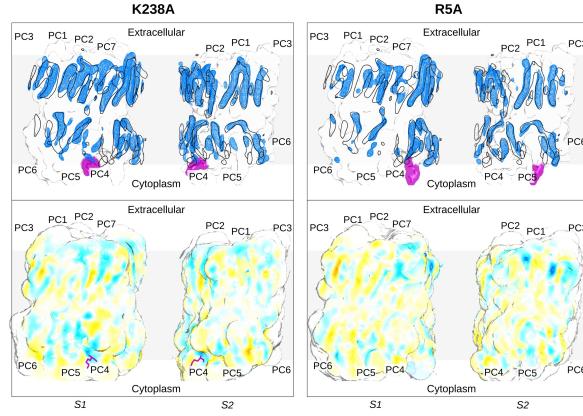
Figure S8. Effect of mutations of AQP0 surface residues on the lipid-density maps derived from MD simulations: mutations of residues involved in strong electrostatic interactions with lipids in the extracellular leaflet. Each panel corresponds to one of the introduced mutations. The upper figures in 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).
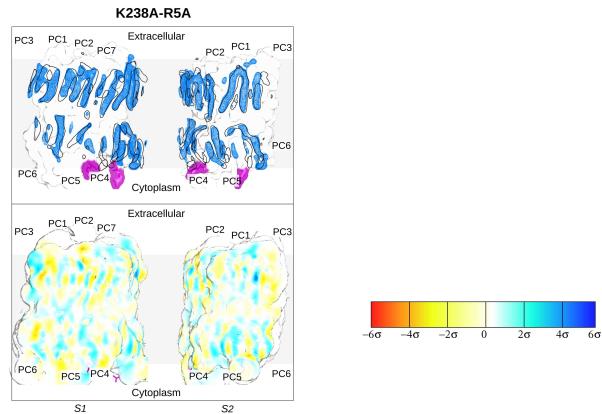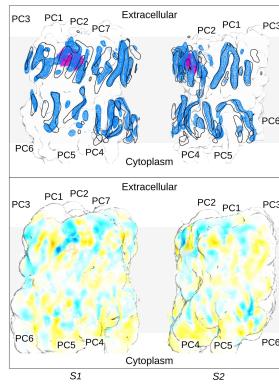
Figure S9. Effect of mutations of AQP0 surface residues on the lipid-density maps derived from MD simulations: mutations of surface residues involved in strong electrostatic interactions with lipids in the cytoplasmic leaflet. The same format is used as in Figure S8.
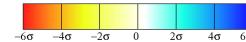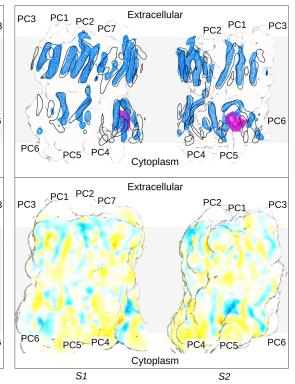

Figure S10. Effect of mutations of AQP0 surface residues on the lipid-density maps derived from MD simulations: mutations of surface residues involved in strong van der Waals interactions with lipids. The same format is used as in Figure S8.
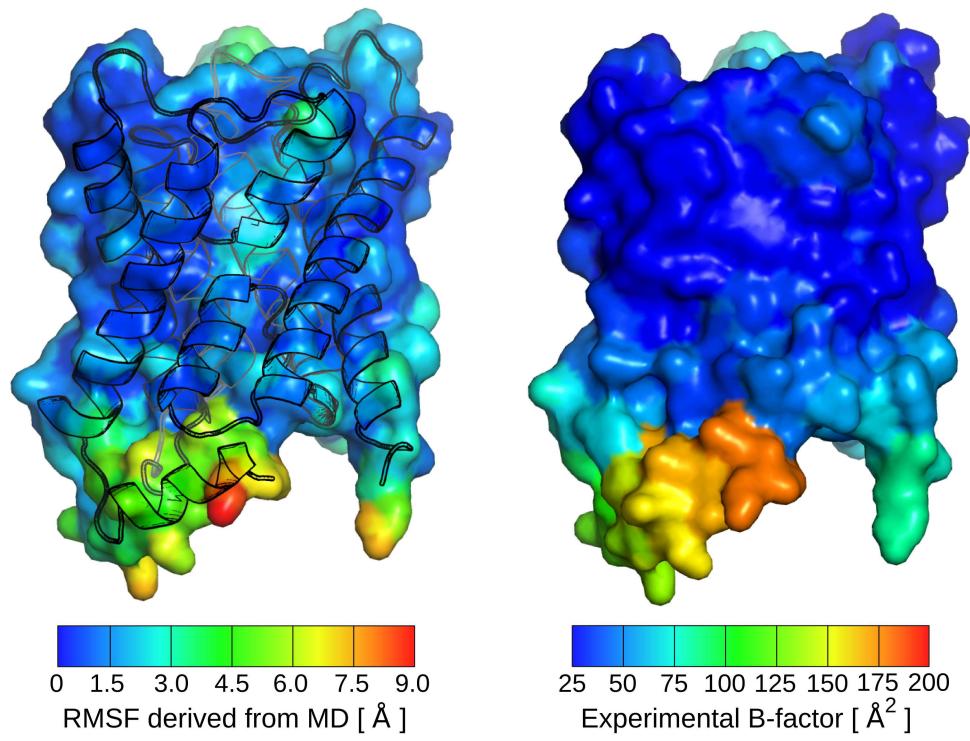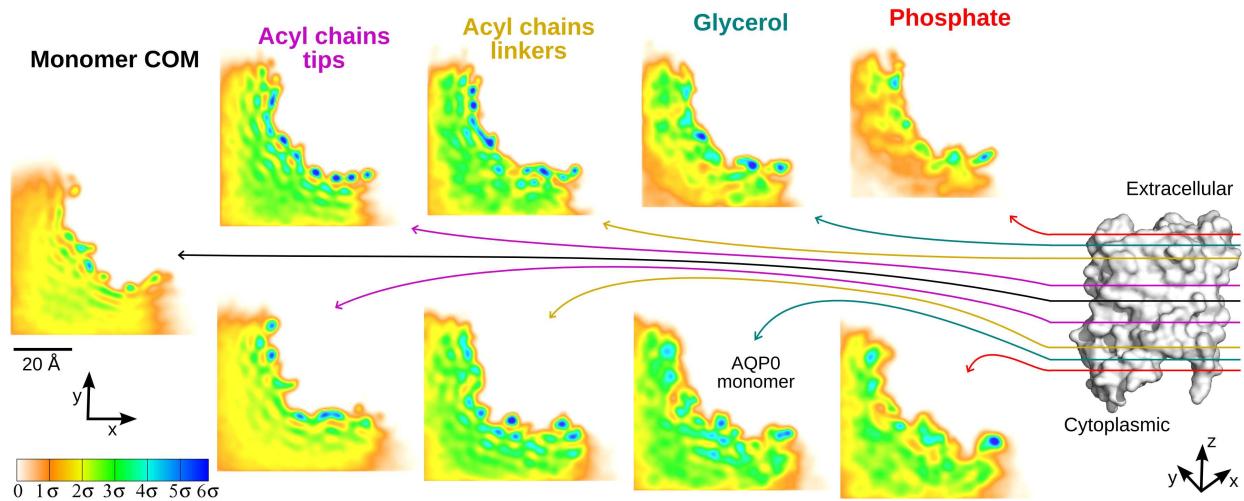
Figure S11. Comparison of the RMSF of AQP0 atoms derived from MD simulations with their B-factors from the electron crystallographic structure of AQP0 (1). The RMSF values and B-factors are color-coded according to the scales at the bottom.

RMSF derived from MD [Å]

Experimental B-factor [$Å^2$]

9

A. With crystallographic lipids

Monomer COM   **Acyl chains tips**   **Acyl chains linkers**   **Glycerol**   **Phosphate**

20 Å

y
x

0 1σ 2σ 3σ 4σ 5σ 6σ

Extracellular

AQP0 monomer

Cytoplasmic

z
y x

B. Without crystallographic lipids

Monomer COM   **Acyl chains tips**   **Acyl chains linkers**   **Glycerol**   **Phosphate**

20 Å

y
x

0 1σ 2σ 3σ 4σ 5σ 6σ

Extracellular

AQP0 monomer

Cytoplasmic

z
y x
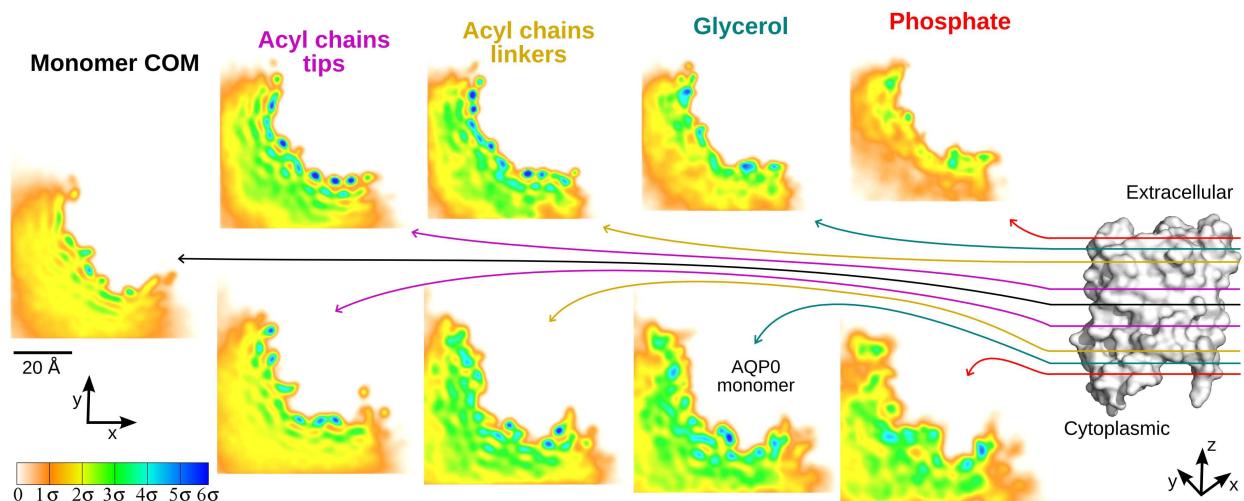
Figure S12. Lipid density beyond the annular shell of lipids, recovered from the MD simulations of a single AQP0 tetramer with (A) or without (B) the crystallographic lipids. The color maps at the left represent lateral projections (onto the xy membrane plane) of the lipid density, at the different z 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.
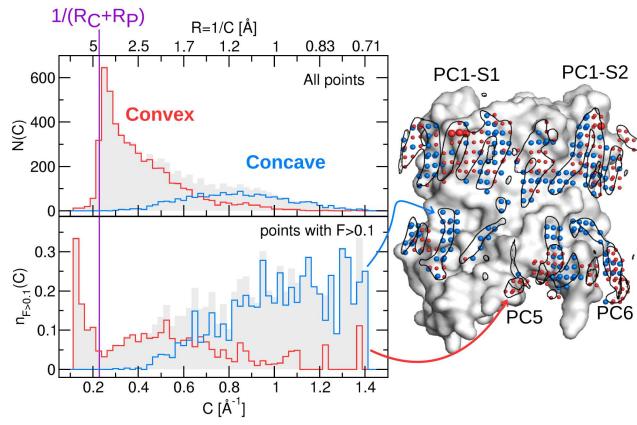
Figure S13. Protein surface curvature, C, and concavity computed from MD simulations. Left, top panel: Histogram of the curvature values $N(C)$ (gray), separated into contributions by convex (red) and concave (blue) surfaces. Left, bottom panel: density of curvature points with a high lipid density in 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 $R_C = 1.88\,\text{Å}$ and $R_p = 2.5\,\text{Å}$ 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.

11

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 $1.5\,\text{Å}$ and height $2\,\text{Å}$ 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).

14

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 $3.60\,\mathrm{nm}$ ($0.07\,\mathrm{nm}$ 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 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.

16

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:

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

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):

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

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 $\times$ 70 Å $\times$ 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:

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

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:

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

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.