

University of Groningen

Structural Determinants of the Supramolecular Organization of G Protein-Coupled Receptors in Bilayers

Periole, Xavier; Knepp, Adam M.; Sakmar, Thomas P.; Marrink, Siewert J.; Huber, Thomas

Published in:
Journal of the American Chemical Society

DOI:
[10.1021/ja303286e](https://doi.org/10.1021/ja303286e)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2012

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Periole, X., Knepp, A. M., Sakmar, T. P., Marrink, S. J., & Huber, T. (2012). Structural Determinants of the Supramolecular Organization of G Protein-Coupled Receptors in Bilayers. *Journal of the American Chemical Society*, 134(26), 10959-10965. <https://doi.org/10.1021/ja303286e>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Supporting Information

Structural determinants of the supramolecular organization of
G protein-coupled receptors in bilayers.

Xavier Periole,^{a,*} Adam M. Knepp,^b Thomas P. Sakmar,^b
Siewert J. Marrink,^a and Thomas Huber^{b,*}

^a Biomolecular Sciences and Biotechnology Institute and Zernike Institute for Advanced Materials, University of Groningen, Nijenborgh 7, 9747AG Groningen, The Netherlands;

^b Laboratory of Molecular Biology & Biochemistry, The Rockefeller University, 1230 York Avenue, New York, NY 10065, USA

*To whom correspondence should be addressed. E-mail: x.periole@rug.nl or hubert@rockefeller.edu.

CGMD simulations.

General setup. All simulations were performed with the Gromacs software package, version 4.0.X.¹ In the coarse-grained (CG) simulations the parameters of the MARTINI force field were used as described for the lipids² and for the proteins.^{3,4} The MARTINI force field defines a set of 20 interacting sites or building blocks that correspond to different chemical entities mapping on average four non-hydrogen atoms. The sites or beads interact through non-bonded potentials, including Coulomb and Lennard-Jones 12-6 terms. The interaction strengths are parameterized primarily to reproduce experimental partitioning coefficients. General setups associated with the MARTINI force field were used. These include shifted Coulomb and van der Waals potentials with a cut-off of 1.2 nm. A dielectric constant of 15 is used to screen effectively the electrostatic interactions. All the simulations were run with a 20-fs time step and the systems were weakly coupled⁵ to a Berendsen thermostat and barostat maintaining the temperature at 300 K and the pressure at 1 atm ($\tau_T=0.5$ ps and $\tau_p=1.2$ ps with a semi-isotropic scheme) unless otherwise indicated.

For the proteins, secondary structure was maintained using an elastic network.⁴ The CG model for inactive rhodopsin (including the reference structure used for the elastic network) used in the simulations (PMF and the model of rows-of-dimers) was based on the crystal structure 1U19.⁶

The retinal moiety molecule was not included in the models. The use of an elastic network in the CG model preserves the structure of the individual proteins, and therefore no specific deformation of the proteins was observed and/or associated with the absence of the ligands. Glycosylations were not included in the model.

Cys-palmitoylation of rhodopsin at positions 322 and 323. Note that the palmitoyl chains were not included in all simulations. Typically the model of rhodopsin used for the PMF determination did not include the chains except for the case in which H1/H8 interfaces are involved (See below). Palmitoyl chains were described by four C1 beads. All bond lengths between bead chains and the attachment to their respective

residue anchors were set at the conventional 0.47 nm and a 1250 kJ mol⁻¹ nm⁻² force constant was used. Cosine angle potentials ($\theta=180$ and $k_{\theta}=25$ kJ mol⁻¹) were used to describe the stiffness of the chains, except in the case of the farnesyl chain where $\theta=100$ and $k_{\theta}=10$ kJ mol⁻¹ were used.

Interpretation of time scale. Due to the smoothing of the energy landscape the dynamics observed in a CGMD simulation are generally faster. Accordingly, when reporting the simulation results with the MARTINI force field, a standard conversion factor of 4 is used, which is the effective speed-up factor in the diffusion dynamics of CG water compared with actual water.⁷ The same order of acceleration of the overall dynamics is also observed for a number of other processes, including the sampling of the local configurational space of a lipid,⁸ and the self-diffusion of lipids^{2,7} and transmembrane peptides.⁹ However, the speed-up factor might be quite different in other systems or for other processes. Particularly for protein systems, no extensive testing of the actual speed-up due to the CG dynamics has been performed, although protein translational and rotational diffusion was found to be in good agreement with experimental data in simulations of CG rhodopsin.¹⁰ Nevertheless, the time scale of the CGMD simulations has to be interpreted with care. The use of an effective time (factor four) is indicated by an * in the main manuscript and hereafter.

Simulations for PMF calculation. The two-rhodopsin system used in the PMF calculations was built up from one used in previous work.¹⁰ The initial system was constructed starting from an equilibrated one-rhodopsin system in which the protein was at the center and surrounded by lipids within an 11-nm square box. The system was then doubled in size by repeating the box in one dimension of the bilayer to form a rectangular boxed bilayer containing two rhodopsins. The final system (Fig. S1) thus contains two rhodopsins, 656 lipids (1,2-di(Δ 10-*cis*-eicosenoyl)-*sn*-glycero-3-phosphocholine; (C20:1)₂PC), and 10166 water beads (equivalent to 46664 actual water molecules), resulting in 20924 MARTINI beads and 16 additional beads for the palmitoyl chains.

The system was equilibrated and the proteins were brought to an inter-protein distance of 6.0 nm as shown in Fig. S1. The distance between rhodopsins was then

lowered by successive pulling and equilibration steps. Various restraints on the rhodopsin–rhodopsin orientation were then applied and umbrella sampling (US) simulations were started from these different conformations (see the SI section on Potential of Mean Force). When needed, conformations with a lipid-free dimer interface were generated manually by simply removing the lipids from the interface and reintroducing them into the bulk bilayer. The conformation at $d=6.0$ nm was used as starting conformation for the construction of the H4–H4, H4–H6, H4/H5–H4/H5 and H5–H5 dimer. The H1/H8–H1/H8 dimer deserves a more detailed description, since it has been observed in many early EM studies of 2D crystals¹¹ and more recently in high-resolution 3D cryogenic EM crystallographic densities.^{12,13} We built an atomistic model of the corresponding dimer, EM-dimer, by rigid-body fitting of rhodopsin into these high-resolution density maps using the Situs2.5 package.¹⁴ A similar interface was observed in recent X-ray crystallography studies that aimed at solving the structure of the photointermediate metarhodopsin I of rhodopsin.^{15,16} The interface observed in the 3D-crystals (X-ray-dimer) differs from the one obtained from EM studies by the value of the tilt angle between the two long axis of the receptors (direction of the membrane normal) (see Table S1). The difference is likely to result from the absence of a reasonable mimic of the lipid bilayer to enforce the vertical orientation of the receptors in the crystals. Simultaneously, CGMD spontaneous self-assembly simulations (see below) of photoreceptors were performed where we observed significant amounts of the H1/H8–H1/H8 dimer (CGMD-dimer). This CGMD-dimer compared extremely well with the EM-dimer, as shown in Table S1. The only significant difference between the EM-dimer and the CGMD-dimer is the distance between the two receptors (SI Table 1), which might result from the position of the C-terminus of rhodopsin in 1U19. The overall similarity of the interfaces found in these three studies is quite striking. Umbrella sampling simulations of the H1/H8 interface were started either from the spontaneously assembled CGMD dimer or from a configuration where the receptors were separated. Note also that for all the interfaces probed it was verified that the observed minimum was a stable conformation over a μs^* time scale even when the restraints were removed. No relevant rearrangement of the contact interfaces was observed.

The systems used for the PMF calculation run at a speed of 84 ns*/CPU/day on a typical personal computer. Accordingly an estimate of the CPU time used to produce the 6 PMFs discussed in the main manuscript, which accumulate 1,174 μs^* , is $\sim 14,000$ days or $\sim 335,000$ CPU hours. We can add to this the 300 μs^* of self-assembly simulation of the 16 and 64 rhodopsins systems (see the description below).

Systems for the self-assembly simulations. Systems with 16 and 64 embedded proteins were used to follow the self-assembly of receptors starting from an ideal dispersion in the membrane: the rhodopsins are placed on a 4-by-4 grid that maximizes the distance between them. The system containing 16 receptors is comparable to the one used in our previous study.¹⁰ The one containing 64 receptors was obtained by replicating the 16-receptors system on the directions of the membrane plane. In the case of the 16 receptors, the ten simulations were differentiated by the mean of the seed number of the random generator for the initial Maxwell–Boltzmann velocity distribution. A 1- μs^* simulation with position restraints applied on a central backbone bead (Gly121) of each receptor was used to randomize their relative orientation, while keeping their initial spatial distribution. The 16 proteins systems were ran for 20 μs^* each and the 64 receptors system was ran for 100 μs^* .

Simulations of the rows-of-dimers. The model of rows-of-dimers described in the main manuscript was built starting from an equilibrated H1/H8 rhodopsin dimer as used in the PMF calculation. To generate a system compatible with the data collected from the AFM images¹⁷ the following protocol was used. The simulation box was extended along its long side (x-axis) using a negative pressure (-5 atm). The pressure in the orthogonal direction (y-axis) in the plane of the bilayer (x,y) was coupled to a +5 atm pressure. A distance restraint between the receptors was added to maintain the integrity of the dimer while under the stress imposed by the deformation of the box. The box quickly elongated and slowly equilibrated (see Fig. S2) to a situation where the periodic images of the receptor dimer came close to the “real” one up to the point that they were separated by only a few lipids corresponding to a single layer (see Fig. S2). At this point, although the receptors are not in direct contact, it is clear that the receptor dimer is interacting with its

periodic image. Although this represents an obvious non-physical situation, it is reasonable to use this formulation for the construction of the system. From the end part of this simulation a conformation where the box had reached a length of 3.8 nm (in the direction of the rows-of-dimers, a) was selected and relaxed with a regular pressure coupling to 1 atm but keeping the $a=3.8$ nm. The additional restraints on the dimer were then removed and the system was further equilibrated. The system size in the “long” direction was then reduced by successively removing lipids going from the original 656 molecules to first 172 and then to 120. The lipids were removed from the edges of the box and the box dimension reduced accordingly before the system was equilibrated while maintaining the box length a at 3.8 nm and letting the other one, b , relax for times up to 4 μs^* .

This equilibrated system was then adjusted to the cell dimensions and rhodopsin orientation obtained from the AFM images with $a=3.8$ nm and $\gamma=85^\circ$. The other dimension, b , defines the repeat distance between rows-of-dimers (Fig. S2). In our model $b=10.89$ nm, which was chosen to be consistent with the average packing density of 48300 rhodopsins/ nm^2 as found in the same AFM images.^{17,18} According to this protein density and based on simple geometrical considerations one can show that a patch with two rhodopsins and lipids would have about 41.40 nm^2 of space and that the area for the lipids is 23.4 nm^2 . This corresponds to ~ 36 lipids/rhodopsin assuming 0.65 nm^2 area per lipid. In our equilibrated rhodopsin dimer system with $a=3.8$ nm and $b\approx 10.89$ nm, we obtained a unit cell containing 68 lipids, or 34 lipids per rhodopsin.

The rhodopsin-dimer surrounded by 68 lipids was further equilibrated for an additional 4 μs^* using a semi-isotropic pressure coupling scheme. The average box dimension were $a=10.93$ nm and $b=3.83$ nm. This system was used as a building block for the construction of our model of the rows-of-dimers. It contains two rows of 4 dimers each, which was obtained by repeating four times in the a -direction and two times in the b -direction (Fig. S3). The model of the disc membrane representing a double rows-of-

dimers contains eight rhodopsins, 544 (C20:1)₂PC molecules and 13288 water beads (corresponding to 53152 real water molecules) for a total bead count of 33624. In the following description we may refer to this system as 2×4-H1/H8–H1/H8-dimer (=2 rows with 4 dimers per row).

Potentials of Mean Force (PMFs).

Introduction. The PMFs of a pair of rhodopsins embedded in a (C20:1)₂PC lipid bilayer were determined as a function of the distance d between the two receptors. The relative orientation of the two receptors was controlled by the use of the virtual bond algorithm (described below). A total of ~ 1.2 ms* CGMD simulations were used to produce the final PMFs (described below). The formalism describing the approach used to obtain the PMFs and a few relevant technical details of the unbiasing procedure used are discussed below.

Description of the protocol to the PMFs. MD simulations using a perturbed potential energy function $U(\mathbf{R})$ can be used to generate configurations \mathbf{R} difficult to access from equilibrium MD simulations.

$$(1) \quad U(\mathbf{R}) = U_0(\mathbf{R}) + W_j(\mathbf{R})$$

Here we supplement the potential energy function $U_0(\mathbf{R})$ of a configuration \mathbf{R} (entirely defined by the CG model) with a perturbing potential $W_j(\mathbf{R})$ expressed as a sum of N_p potentials:

$$(2) \quad W_j(\mathbf{R}) = \sum_{p=1}^{N_p} W_j^p(\chi_p(\mathbf{R}), \chi_{p,j}^{(0)})$$

The potentials W_j^p are defined by the VBA restraints $\chi_p(\mathbf{R})$ for which $\chi_{p,j}^{(0)}$ is the reference value corresponding to the j^{th} umbrella window or trajectory (see Table S2, Fig. S4, and below).

Each umbrella trajectory, j , generates a biased probability density:

$$(3) \quad \rho_j^{(b)}(\mathbf{R}) \equiv \frac{1}{n_j} \sum_{l=1}^{n_j} \delta(\mathbf{R} - \mathbf{R}_{j,l})$$

with l running over the n_j conformations sampled from the j^{th} umbrella trajectory. $\mathbf{R}_{j,l}$ is the instantaneous configuration of the l^{th} snapshot from this trajectory, and δ is a Dirac delta function.

The unbiased probability densities are obtained from:

$$(4) \quad \rho_j^{(u)}(\mathbf{R}) = e^{\beta[W_j(\mathbf{R})-f_j]} \rho_j^{(b)}(\mathbf{R}) \quad \text{with } \beta = 1/RT$$

The optimal distribution of configurations is obtained from an optimal combination of the unbiased distributions:

$$(5) \quad \rho_0(\mathbf{R}) = C \sum_{j=1}^N p_j(\mathbf{R}) \rho_j^{(u)}(\mathbf{R})$$

For which WHAM (see below) gives the solutions for the weighting of each window (the constant C denotes a normalization constant):

$$(6) \quad p_j(\mathbf{R}) = \frac{n_j e^{-\beta[W_j(\mathbf{R})-f_j]}}{\sum_{k=1}^N n_k e^{-\beta[W_k(\mathbf{R})-f_k]}}$$

(4), (5) and (6) simplify to:

$$(7) \quad \rho_0(\mathbf{R}) = C \sum_{j=1}^N \frac{n_j}{\sum_{k=1}^N n_k e^{-\beta[W_k(\mathbf{R})-f_k]}} \rho_j^{(b)}(\mathbf{R})$$

(3) and (7) combine to:

$$(8) \quad \rho_0(\mathbf{R}) = C \sum_{j=1}^N \frac{1}{\sum_{k=1}^N n_k e^{-\beta[W_k(\mathbf{R})-f_k]}} \sum_{l=1}^{n_j} \delta(\mathbf{R} - \mathbf{R}_{j,l})$$

f_k values are obtained from a self consistent iterative procedure^{19,20} in which the initial values were set to zero.

$$(9) \quad e^{-\beta f_i} = \frac{\sum_{j=1}^N \sum_{l=1}^{n_j} e^{-\beta W_i(\mathbf{R}_{j,l})}}{\sum_{k=1}^N n_k e^{-\beta[W_k(\mathbf{R}_{j,l})-f_k]}}$$

(8) can be used to calculate an ensemble average of arbitrary quantities and be expressed as follows for a two dimensional case:

$$(10) \quad \rho(x_0, x_1) = \int d\mathbf{R} \rho_0(\mathbf{R}) C_{x_0}(x'_0(\mathbf{R})) C_{x_1}(x'_1(\mathbf{R}))$$

$$(11) \quad \rho(x_0, x_1) = \frac{1}{\Delta x_0 \Delta x_1 N_{tot}} C \sum_{j=1}^N \sum_{l=1}^{n_j} \frac{1}{\sum_{k=1}^N n_k e^{-\beta[W_k(\mathbf{R}_{j,l}) - f_k]}} C_{x_0}(x'_0(\mathbf{R}_{j,l})) C_{x_1}(x'_1(\mathbf{R}_{j,l}))$$

where Δx_m is the binning and the total number of sampled configurations is counted as:

$$(12) \quad N_{tot} = \sum_{k=1}^N n_k$$

Counter functions are defined as:

$$(13) \quad C_{x_m}(x'_m(\mathbf{R})) = \begin{cases} 1 & \text{if } x'_m(\mathbf{R}) \in [x_m - \frac{1}{2}\Delta x_m, x_m + \frac{1}{2}\Delta x_m) \\ 0 & \text{otherwise} \end{cases}$$

and allow when desired to define the section of the data that will be used. Here they permitted us to restrict rigorously the data to the conformations that were within a defined limit around the defined relative orientations of the receptors.

The final 2D-PMF can then be expressed as:

$$(14) \quad w(x_0, x_1) = -k_B T \ln[\rho(x_0, x_1)]$$

The 6-dimension (6-D) case was handled accordingly. Counter functions were used on angles to determine slices of the full 6-D PMF to look at specific relative orientation of the receptors. See Tables S2, S3–S7 for details on the restraints. In the present case x_0 can be assimilated to the distance between the receptors (no counter function is applied) and x_1 to one of the VBA angle restraints for which the exploration would be limited to the window defined by the counter function.

Note that the unspecified normalization constant C in equation (11) can be eliminated at this step by shifting the PMF to set an arbitrary point to zero. We chose to align the PMFs at large rhodopsin–rhodopsin distance.

Virtual Bond Algorithm (Biasing Potentials). The different interfaces between receptors probed in this work and presented in the main text were defined and controlled by the use of the virtual bond algorithm.²¹ By the mean of three anchors on each

molecule, the definition of one distance, d , two angles, θ_1 and θ_2 , and three dihedral angles, ϕ_1 , ϕ_2 , and ϕ_3 , allowed the definition of the relative orientation of the two receptors. These restraints were added to the topology of the system by the mean of harmonic potentials. ϕ_1 and ϕ_3 were replaced by regular angle, θ_1' and θ_2' , when more convenient. The complete definition of the restraints is shown in Fig. S4. The force constants and reference values used for the restraints are given in Table S2 together with other useful information. Harmonic restraints were used for distances and dihedral angles and a cosine angle potential for the regular angles.

Note that since the receptors are embedded into a membrane bilayer the vertical orientation of the receptor is naturally controlled. This reduces the actual number of restraints needed to maintain the relative orientation of the receptors (see Table S2). This also simplifies the task of sampling many degrees of freedom. We can assume that the relevant orientations of the receptor relative to the membrane normal are sampled. The reference values were derived by simple geometrical considerations when possible, or by average over a simulation of several hundreds of nanoseconds during which the receptor dimer was free of restraints.

Umbrella windows simulations performed. In Table S3–S7 the complete set of umbrella simulations used to determine the PMFs shown in the main text are listed. The reference values and the force constants used for the restraints are listed in Table S2 and S3–S7. It is important to note that the determination of the PMFs of the interfaces for H4/H6 and H4 models were more challenging than the other three, H1/H8, H4/H5 and H5. This is due to the larger buried surface area of the receptors in the H4–H6 and H4–H4 dimer models, which led to large energy barriers to solvate (lipidate) and desolvate (delipidate) these interfaces. It was therefore not possible to generate an equilibrated sampling at rhodopsin–rhodopsin distances where lipids had been trapped at the interface (when proteins are brought towards each other) or were challenged to squeeze in between the receptors (when proteins are pulled apart). For this reason simulations starting from different solvated interfaces (with or without lipids at the interface) were run for these interfaces. The selection of the windows to include in the calculation of the PMFs were carefully carried out based on multiple 10–20 μs^* long simulations at 4–8 different

distances (see Table S3–S7) around the interfacial region and with different starting conditions at the interface. This careful procedure ensured the exclusion of highly improbable situations that would otherwise be difficult to relax with short simulations and which would erroneously increase the depth of the wells relative to the energy barriers.

The comparison of the PMFs presented in the main text (Fig. 3) assumes that at long distance the relative orientation of the two receptors does not matter. In other words, that free energy of the system is insensitive to the relative orientation of the receptors at the larger distances considered, typically $d=6.0$ nm. We have checked this assumption by performing umbrella simulations covering the rotation (with an angle increment of 10°) of a receptor to mimic the transition from the orientation in as in the H4-H6 orientation to the one as in H4-H4 model. The distance d is then harmonically restrained at 6.0 nm. The free energy profiles (PMFs) in function of the relative angle of the two receptors while going from a H4–H6 orientation to a H4–H4 one, and *vice versa*, proved to be flat within ± 5 kJ/mol typical for the bootstrap error estimated of the reconstructed PMFs (*cf.* Fig. 3) and thereby validated our approach.

Unbiasing the umbrella simulations. The original weighted histogram analysis method (WHAM) introduced by Kumar *et al.*¹⁹ was used to unbias and combine the umbrella simulations. A few modifications and improvements were added to the implementation described by Souaille and Roux²⁰ and were included into a C++ code (thwham.cpp). Notably the code allows one to unbias simultaneously multiple reaction coordinates or dimensions. Moreover, a reorganized loop structure resulted in an order of magnitude accelerated algorithm. In the present case the six restraints of the virtual bond algorithm (VBA) (d , θ_1 , θ_2 , ϕ_1 , ϕ_2 and ϕ_3) were used as primary reaction coordinates together with two additional, θ_1' and θ_2' . The final PMFs were expressed in function of the d -dimension describing the distance between the two receptors. They are shown in the main manuscript as a function of the interfacial distance, d' , which is defined as the distance between the receptors to which the distance at the minimum of the PMF, d^{eq} , was subtracted (Fig. 3. and Fig. S4).

To take into consideration the restraint of the relative orientation of the receptors by the VBA we not only unbiased the US simulations according to the (biasing) potential added to the simulations (which obviously relates to the multi-dimensional aspect of the present PMFs mentioned above), but we explicitly restricted the analysis of an interface to the data (snapshot of the system) within a certain window around the reference orientation (see counter functions described above). This allowed us to focus the WHAM on the conformational space explored within a constant window size around a given orientation of the receptors and in the same time to avoid artifacts originating from the incomplete, and therefore inaccurate, sampling when going away from the reference orientation. Limits were applied on angles and dihedral angles used to restrain the orientation of the receptors. To summarize, a particular snapshot was considered for the determination of a PMF if the instantaneous value of a restrained angle was within a limit from its reference value as defined in VBA (given in Table S2). A 10°-limit was found to be a good compromise between the removal of border inaccuracy and withholding of sufficient sampling of the region of interest.

Noteworthy is the use of Dirac delta functions, equation (3), to describe the data instead of the more conventional use of histograms in such calculations. This approach has the major advantage to remove the well-known bin-size effect on the WHAM accuracy and to consider explicitly all the data points (snapshot of the system) in the determination of the PMF instead of one histogram per umbrella simulation, thereby making full use of the raw data at each iteration of the WHAM algorithm.

Limitations of the Model/Methodology. To put our results into perspective, it is important to point out some of the limitations underlying our model and the methodology used. First, the processes studied involve the slow diffusion of lipid and protein, which led to difficulty reaching complete convergence on some aspects of the data presented. Notably in the self-assembly simulations the lack of protein binding/unbinding events limited the spontaneous self-assembly simulations to only reflect the long- and medium-range interactions depicted by the PMFs. The interfaces having an energy barrier to binding are poorly sampled and the populations of the ones sampled do not reflect relative stabilities. In the PMFs the slow exchange of interfacial lipids with the bulk

lipids, alternatively, makes both a full lipidation and delipidation of some interfaces extremely challenging to sample at equilibrium even on time scales up to 20 μs^* per window. Equilibrium sampling is critical to obtain reliable PMFs. Note also that the restriction of the PMFs to slices of the hyper-surface (of the relative protein orientation) significantly reduced the need of conformational sampling. We have shown previously that for a single transmembrane helix (glycophorin A) embedded in a membrane bilayer it takes up to 8 μs^* to sufficiently sample the rotational degree of freedom to reach convergence for umbrella windows where the peptides are in contact. Therefore, the quantitative details of the PMFs presented in the manuscript have to be considered with care, but the qualitative features are consistent and significant. A second limitation may be the use of a CG model describing the protein-protein interactions. It has been observed that in an aqueous environment CG protein-protein interactions might be slightly over-stabilized²² but there is no similar evidence for membrane proteins. In fact we have recently reported studies of the association of glycophorin A (GpA)²³ and WALP peptides²⁴ in model membranes using the same MARTINI CG model and found that the free energy profile of the GpA peptide was essentially identical to one reported earlier using an atomistic force field²⁵ and that the estimated dimerization free energy of the WALP peptides agreed with the value obtained from fluorescence resonance energy transfer (FRET) experiments.²⁶ The rotational and translational diffusion of rhodopsin was also found to be in close agreement with experiments.¹⁰ It is also important to note that individual side chain-side chain association constants are overall in relatively good agreement with their atomistic homologues.²²

Analysis

Solvent accessible surface area. The protein burial, a_b , or buried solvent accessible surface area (ASA) of a protein dimer is defined for each interface at a time t by $a_b(t) = \text{ASA}_1(t) + \text{ASA}_2(t) - \text{asa}(t)$, where $\text{ASA}_i(t)$ is the ASA of the protein i ($i=1, 2$) isolated from the other and $\text{asa}(t)$ is the ASA of the proteins associated into a complex. In the main manuscript the average of $a_b(t)$ over a 800 ns* simulation is reported, \bar{a}_b .

ASA was computed using the double cube lattice algorithm²⁷ with a probe radius of 0.26 nm (vdW radius of the beads).

Cluster analysis. The cluster analysis of the receptor dimers reported in the main manuscript was performed using the so-called GROMOS approach as implemented in the GROMACS tools. The matrix of positional root-mean-square-difference (RMSD) of the backbone beads of the receptor dimers is first calculated (after fitting the dimer pairs using the same set of beads) and the number of neighboring dimers (RMSD < 0.4 nm) in the complete set is counted for each dimer conformation. The dimer with the higher number of neighbors together with its neighbors are removed from the pool of dimer conformations and define the first cluster. The process is repeated with the remaining pool of the dimer conformations to define the second cluster, and then for the third cluster and so on until the pool of conformations is empty. From the 601,200 possible pairs (120 possible pairs (16x15/2) for ten 20- μ s* simulations with configurations saved every 40 ns*) collected from the ten simulations of self-assembly with 16 receptors the cluster analysis was restricted to 10% of the conformations for which the center-of-mass (COM) distance between the receptors was less than 5.5 nm. The resulting 5,227 dimer conformations were symmetrized (the receptors 1 and 2 were swapped in the coordinate file so that receptor 2 becomes receptor 1 and *vice versa*) to correct for the bias due to the order of the receptors in the conformations used in the clustering analysis.

The schematics and some details of the ten most populated clusters are given in Fig S5.

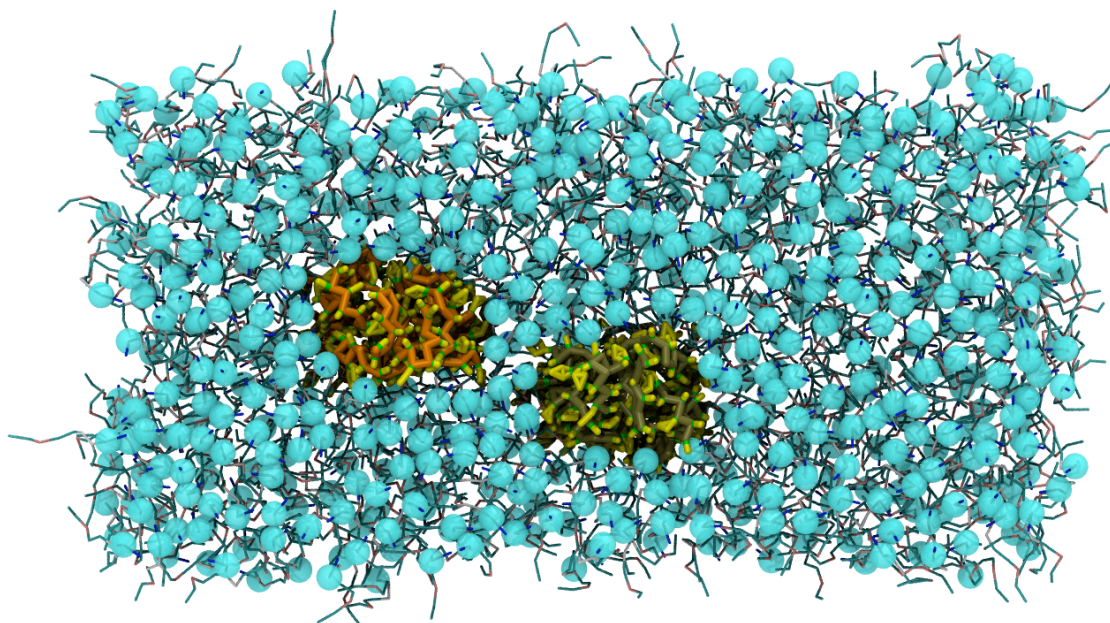


Fig. S1. The system used in the umbrella windows simulation to determine the potential of mean force (PMF) shown in the main text. The two proteins are here separated by a distance of 6.0 nm and are shown in a stick representation with orange and tan backbone particles and yellow side chains. The 656 lipids are shown in small stick representation and translucent cyan spheres highlight the phosphate beads. The 10,166 water beads are not represented.

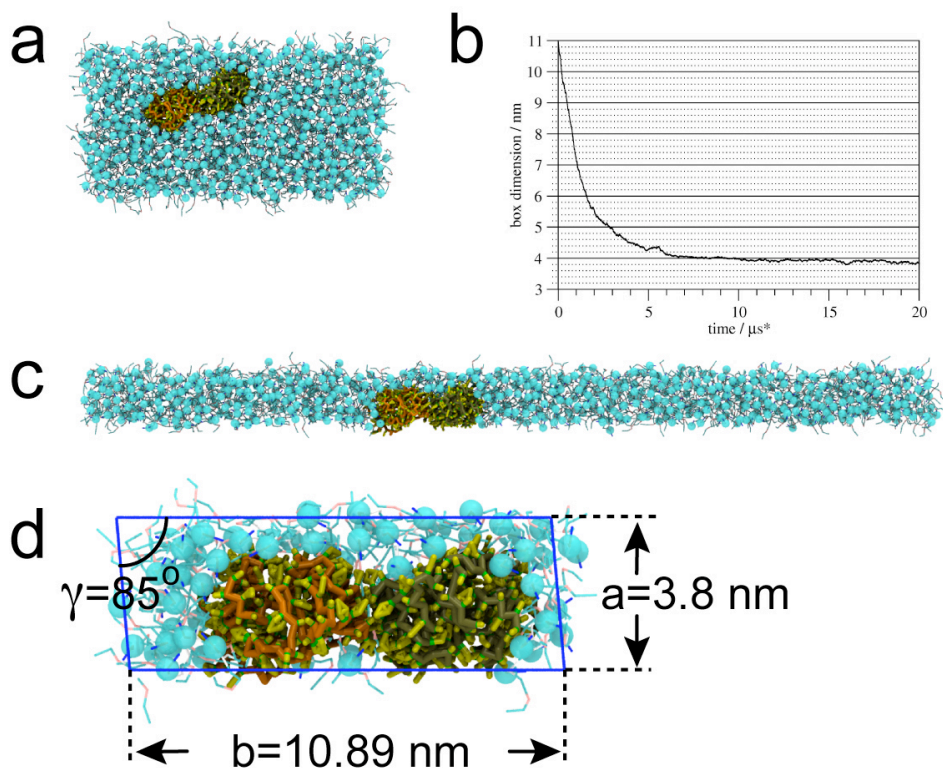


Fig. S2. Construction of the H1/H8 rhodopsin dimer unit cell to maintain agreement with the AFM experimental data.^{17,18} The views are from the extracellular side. (a) Initial conformation. The system is identical to the one shown in Fig. S1 with the difference that here the proteins interact through their H1/H8 interfaces. (b) Evolution of the a-dimension with time during the elongation procedure consisting of applying pressure in opposite directions along the dimensions, a and b. (c) Configuration of the system after 20 μs^* CGMD simulation. (d) Conformation after successively reducing the b-dimension (with $a=3.8$ nm and $\gamma=85^\circ$) to match eventually the average rhodopsin density observed in the AFM experiments. The resulting unit cell contains 68 lipid molecules (34 per rhodopsin) and respects the choices made of dimensions, $a=3.8$ nm, $b=10.89$ nm, and $\gamma=85^\circ$, as discussed in the text. The color code is identical to the one of Fig. S1.

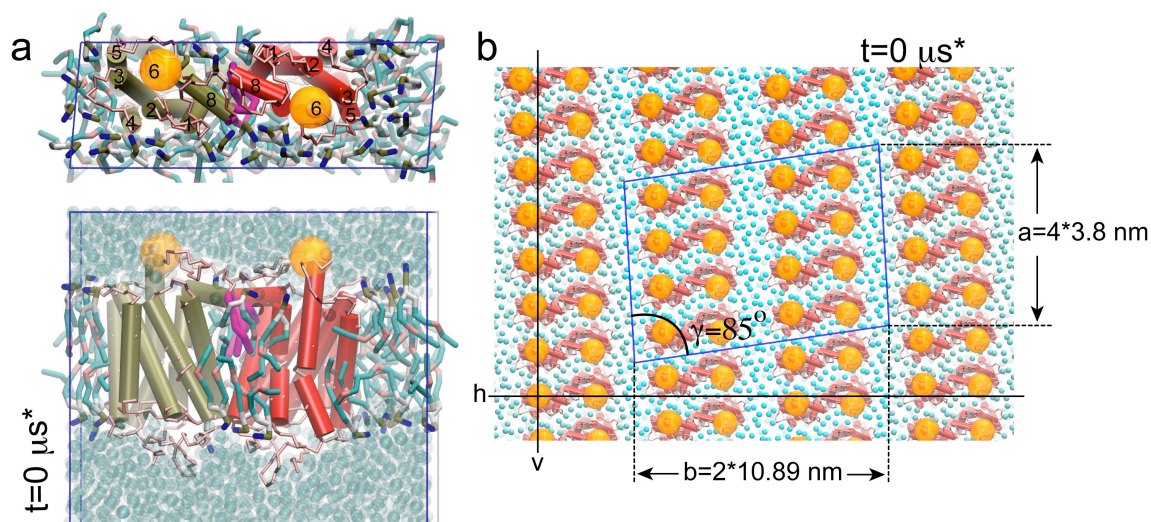


Fig. S3. Construction of the rows-of-dimers based on the H1/H8 rhodopsin dimer unit cell (see Fig. S2) and in agreement with the AFM experimental data.^{17,18} (a) The unit cell of the rhodopsin H1/H8 dimer conformation with the 68 lipids is shown this time from the intracellular space (top) and from the side. The protrusion from the H6 on the extracellular side is emphasized by an orange translucent sphere placed on Thr242. (b) The unit cell shown in panel (a) is replicated following $a=4*3.8$ nm, $b=2*10.89$ nm, and $\gamma=85^\circ$. The system was rotated into the membrane plane to arrive at the alignment of the Thr242 with the horizontal guide (line h). The line v helps the visualization of the derivation of the rows-of-dimers from the perfect alignment as observed in the AFM images. The receptors are shown in light red and the Thr242 in orange translucent spheres. The translucent cyan small spheres show the position of the phosphate beads.

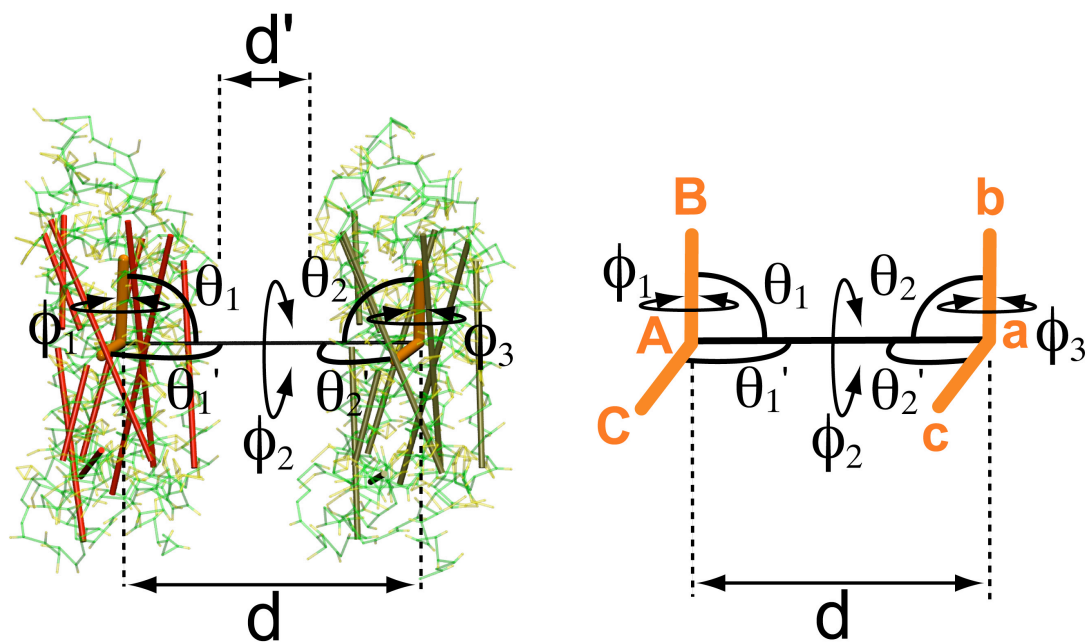


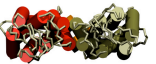

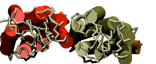







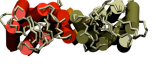




Fig. S4. Virtual Bond Algorithm (VBA) used to define and control the relative orientation of the two receptors. The left representation contains the two receptors and shows the 8 restraints defined. In the right panel the three anchor points of the receptors, (A, B, C) and (a, b, c), replace the receptor. A and a are the backbone bead of Cys187. B and b are the backbone bead of Gly121. C and c are the backbone bead of Gly51. d , θ_1 , θ_2 , ϕ_1 , ϕ_2 , ϕ_3 are the main restraints. They describe the distance between the receptors, d ; the tilt of long axis of each receptor relative to the receptor-receptor direction, θ_1 and θ_2 ; the rotation of the receptors around their long axis (parallel of the membrane normal), ϕ_1 and ϕ_3 , or θ_1' and θ_2' ; the relative orientation of the receptor's long axis, ϕ_2 . d' is the interfacial receptor distance and is defined as the distance, d , between the receptor to which the distance at the minimum of the PMFs was subtracted. The corresponding reference values and force constants for the different interfaces probed are listed in Table S2 and S3–S7.

cluster #	1	2	3	4	5
population (%)	9.1	7.6	6.9	6.8	5.9
top					
side					
bottom					
ϕ_1	-156	153	5	-179	11
ϕ_3	6	-154	13	-179	159

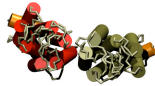

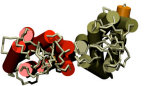
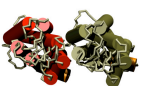

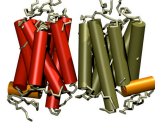



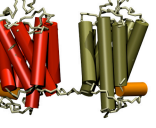





cluster #	6	7	8	9	10
population (%)	4.5	3.0	2.4	2.4	1.0
top					
side					
bottom					
ϕ_1	135	-146	-15	-54	171
ϕ_3	-179	120	-131	137	-163

Fig. S5. Ten most populated dimer interfaces of the receptor as found by a cluster analysis performed on conformations obtained in the 10 self-assembly simulations. For each cluster the population, a top and a side view of the representative conformation and the values of the (ϕ_1 , ϕ_3) of the central conformation (representative of their relative orientations) are given.

Model	EM-dimer	CGMD-dimer	X-ray-dimer
$d^{\text{eq}} / \text{nm}$	4.35	4.56 ± 0.07	4.24
θ_1	78	72 ± 2	73
θ_2	78	76 ± 2	73
ϕ_1	15	14 ± 4	-4
ϕ_2	21	9 ± 4	-30
ϕ_3	16	7 ± 4	-4

Table S1. Parameters of different H1/H8–H1/H8 rhodopsin dimers discussed in the main text. EM-dimer is the dimer generated by fitting rhodopsin structures into the recent high-resolution 3D-EM densities.¹³ CGMD-dimer was obtained from a typical dimer observed during self-assembly CGMD simulations, then equilibrated for 8 μs^* and the values reported were averages over the following 12 μs^* . X-ray-dimer is the model observed in X-ray crystallography experiments and reported in the PDB:2I35.¹⁵

VBA restraint in PMFs	anchors	H4–H6	H4–H4	H1/H8–H1/H8	H5–H5	H4/H5–H4/H5
d^{eq}	a–A	2.8	2.55	4.6	3.7	3.4
range of d sampled in the PMF		[2.4;6.0]	[2.1;6.0]	[4.3;7.0]	[3.4;6.5]	[3.1;6.0]
range of $d' = d - d^{eq}$ sampled in the PMF		[-0.4;3.2]	[-0.45;3.45]	[-0.3;2.4]	[-0.3;2.8]	[-0.3;2.6]
θ_1	b–a–A	-	-	73.3	-	-
θ_2	a–A–B	-	-	73.3	-	-
ϕ_1	a–A–B–C	-	-	10	180	-152.5
ϕ_2	b–a–A–B	-	-	-	-	-
ϕ_3	c–b–a–A	-	-	10	180	-152.5
θ_1'	a–A–C	90	90	-	-	-
θ_2'	c–a–A	90	-90 ^a	-	-	-
$k_d = 500/1000/5000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$		$k_\theta = 500 \text{ kJ mol}^{-1}$		$k_\phi = 300 \text{ kJ mol}^{-1} \text{ rad}^{-2}$		

Table S2. VBA parameters used to define and control the relative orientation of the receptors during the PMF calculations. The force constants and reference values of the restraints are given for each interface probed. “-“ indicates that the restraint was not used. The anchors (a, b, c, A, B, C) are the ones shown in Fig. S4. Distances are given in nm and (regular and dihedral) angles in degrees. Note that regular angles were described by cosine-based potentials instead of the harmonic potentials. ^a is a reminder that angle does not have a sign, it is used here to express the difference between the cases H4–H6 and H4–H4.

Distance / nm	2.4-2.8, dr=0.1	2.9	3	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	4.0 - 6.0, dr=0.1	times / μs^* simulated	used
RUN 1 $k_d=500/1000$	x	x	x	x	x	x	x	x	x	x	x	x	full	33.6	33.6
RUN 2 $k_d=500/1000/5000$	int	int	int	int	int	int	int	full	full	full	full	full	x	38.4	24
RUN 3 $k_d=500/1000/5000$	none	none	none	none	none	none	none	none	x	x	x	x	x	28.8	19.2
RUN 4 $k_d=500$		int	int	int	int									80	
RUN 5 $k_d=500$					none	none	none	none						80	
Total/$\mu\text{s}^*=$														260.8	76.8

Table S3. Set of simulations run for the determination of the PMF of H4–H6. The values used for k_d are indicated in the left column. When multiple values of k_d are given this indicates that a simulation was run at each value. Simulations are labeled “full”, “int”, or “none”, which indicates the type of receptor–receptor interface used as starting configuration of the system. “full” indicates that a fully solvated interface was present; “int” indicates that a few lipids were present at the interface; “none” indicates that the interface was free of interfacial lipids. The simulations or windows actually used for the PMF calculation are highlighted in yellow. The * indicates that times reported are effective times, which are scaled to correct for the increased dynamics observed in CG simulations (See Methods). All windows were simulated for $0.8 \mu\text{s}^*$ with the exception of RUN 5, which ran for $20 \mu\text{s}^*$.

Distance / nm	2.1-2.6, dr = 0.1	2.7	2.8	2.9	3	3.1	3.2	3.3	3.4	3.5	3.6 - 6.0, dr=0.1	times / μs^*	
												simulated	used
RUN 1 $k_d=500/1000$	x	x	x	x	x	x	x	x	x	x	full	38.4	38.4
RUN 2 $k_d=500/1000/5000$	x	x	x	x	x	x	full	full	full	full	x	9.6	9.6
RUN 3 $k_d=500/1000/5000$	none	none	none	none	none	none	x	x	x	x	x	26.4	21.6
RUN 4 $k_d=500/1000/5000$	x	int	int	int	int	int	int	int	int	int	x	21.6	19.2
RUN 5 $k_d=500$							none	none	none	none		80.0	
Total/$\mu\text{s}^* =$												176	88.8

Table S4. Set of simulations run for the determination of the PMF of H4–H4. All windows were simulated for 0.8 μs^* with the exception of RUN 5, which ran for 20 μs^* . See legend of Table S3 for details.

Distance / nm	4.3-4.8, dr=0.1	4.9	5	5.1	5.2	5.3	5.4	5.5	5.6	5.7	5.8	5.9	6.0 - 7.0, dr=0.1	times / μs^*	
														simulated	used
RUN 1 $k_d=500/1000$	ass	ass	ass	ass	ass	ass	ass	ass	ass	ass	ass	ass	x	27.2	27.2
RUN 2 $k_d=500/1000$	x	x	sep	sep	sep	sep	sep	sep	sep	sep	sep	sep	sep	33.6	33.6
RUN 3 $k_d=500/1000$	sep	sep	sep	sep	sep	sep	sep	sep	sep	sep	sep	sep	sep	76.0	76.0
RUN 4 $k_d=500/1000$	ass	ass	ass	ass	ass	ass	ass	ass	x	x	x	x	x	52.0	52.0
													Total / $\mu\text{s}^* =$	188.8	188.8
RUN 1-noPalm $k_d=500/1000$	ass	ass	ass	ass	ass	ass	ass	ass	ass	ass	ass	ass	x	27.2	27.2
RUN 2-noPalm $k_d=500/1000$	x	x	sep	sep	sep	sep	sep	sep	sep	sep	sep	sep	sep	33.6	33.6
RUN 3-noPalm $k_d=500/1000$	sep	sep	sep	sep	sep	sep	sep	sep	sep	sep	sep	sep	sep	76.0	76.0
RUN 4-noPalm $k_d=500/1000$	ass	ass	ass	ass	ass	ass	ass	ass	x	x	x	x	x	52.0	52.0
													Total-noPalm / $\mu\text{s}^* =$	188.8	188.8
													Total / $\mu\text{s}^* =$	377.6	377.6

0.8 μs^*
 2.0 μs^*

Table S5: Set of simulations run for the determination of the PMF of H1/H8 and H1/H8 with/without the Palm. The latter is a control simulation to which palmitoyl chains attached to the receptors at positions CYS322 and CYS323 were removed. All windows shaded were run for $0.8 \mu\text{s}^*$, except that the ones framed ran for $2 \mu\text{s}^*$. See legend of Table S3 for details.

Distance / nm	3.4-3.9, dr = 0.1	4	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9	5	5.1 - 6.5, dr=0.1	times / μs^*		
														simulated	used	
RUN 1																
$k_d=500/1000$	ass	ass	ass	ass	ass	ass	ass	ass	ass	ass	ass	ass	x	68.0	68.0	
RUN 2																
$k_d=500/1000$	x	x	x	x	x	x	sep	sep	sep	sep	sep	sep	sep	84.0	84.0	
													Total/ $\mu\text{s}^* =$	152.0	152.0	

SI Table 6: Set of simulations run for the determination of the PMF of H5–H5. All simulations were run for 2 μs^* . See legend of Table S3 for details.

References

- (1) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J Chem Theory Comput* **2008**, *4*, 435.
- (2) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. *J Phys Chem B* **2007**, *111*, 7812.
- (3) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S. J. *J Chem Theory Comput* **2008**, *4*, 819.
- (4) Periole, X.; Cavalli, M.; Marrink, S. J.; Ceruso, M. A. *J Chem Theory Comput* **2009**, *5*, 2531.
- (5) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. *Journal of Chemical Physics* **1984**, *81*, 3684.
- (6) Okada, T.; Sugihara, M.; Bondar, A. N.; Elstner, M.; Entel, P.; Buss, V. *Journal of Molecular Biology* **2004**, *342*, 571.
- (7) Marrink, S. J.; de Vries, A. H.; Mark, A. E. *J Phys Chem B* **2004**, *108*, 750.
- (8) Baron, R.; Trzesniak, D.; de Vries, A. H.; Elsener, A.; Marrink, S. J.; van Gunsteren, W. F. *Chem Phys Chem* **2007**, *8*, 452.
- (9) Ramadurai, S.; Holt, A.; Schafer, L. V.; Krasnikov, V. V.; Rijkers, D. T. S.; Marrink, S. J.; Killian, J. A.; Poolman, B. *Biophys J* **2010**, *99*, 1447.
- (10) Periole, X.; Huber, T.; Marrink, S. J.; Sakmar, T. P. *J Am Chem Soc* **2007**, *129*, 10126.
- (11) Schertler, G. F. X.; Hargrave, P. A. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 11578.
- (12) Krebs, A.; Edwards, P. C.; Villa, C.; Li, J. D.; Schertler, G. F. X. *Journal of Biological Chemistry* **2003**, *278*, 50217.
- (13) Ruprecht, J. J.; Mielke, T.; Vogel, R.; Villa, C.; Schertler, G. F. X. *EMBO J.* **2004**, *23*, 3609.
- (14) Wriggers, W.; Milligan, R. A.; McCammon, J. A. *J. Struct. Biol.* **1999**, *125*, 185.
- (15) Salom, D.; Lodowski, D. T.; Stenkamp, R. E.; Le Trong, I.; Golczak, M.; Jastrzebska, B.; Harris, T.; Ballesteros, J. A.; Palczewski, K. *Proceedings of the National Academy of Sciences of the United States of America* **2006**, *103*, 16123.
- (16) Lodowski, D. T.; Salom, D.; Le Trong, I.; Teller, D. C.; Ballesteros, J. A.; Palczewski, K.; Stenkamp, R. E. *J. Struct. Biol.* **2007**, *158*, 455.
- (17) Fotiadis, D.; Liang, Y.; Filipek, S.; Saperstein, D. A.; Engel, A.; Palczewski, K. *Nature* **2003**, *421*, 127.
- (18) Liang, Y.; Fotiadis, D.; Filipek, S.; Saperstein, D. A.; Palczewski, K.; Engel, A. *J. Biol. Chem.* **2003**, *278*, 21655.
- (19) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011.
- (20) Souaille, M.; Roux, B. *Comput Phys Commun* **2001**, *135*, 40.
- (21) Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. *J. Phys. Chem. B* **2003**, *107*, 9535.
- (22) de Jong, D. H.; Periole, X.; Marrink, S. J. *J Chem Theory Comput* **2012**.

- (23) Sengupta, D.; Marrink, S. J. *Physical Chemistry Chemical Physics* **2010**, *12*, 12987.
- (24) Schafer, L. V.; de Jong, D. H.; Holt, A.; Rzepiela, A. J.; de Vries, A. H.; Poolman, B.; Killian, J. A.; Marrink, S. J. *Proceedings of the National Academy of Sciences of the United States of America* **2011**, *108*, 1343.
- (25) Henin, J.; Pohorille, A.; Chipot, C. *J Am Chem Soc* **2005**, *127*, 8478.
- (26) Yano, Y.; Matsuzaki, K. *Biochemistry* **2006**, *45*, 3370.
- (27) Eisenhaber, F.; Lijnzaad, P.; Argos, P.; Sander, C.; Scharf, M. *J. Comput. Chem.* **1995**, *16*, 273.