Identification of Specific Lipid-binding Sites in Integral Membrane Proteins

Marc F. Lensink[1,2] , Cédric Govaerts[1], Jean-Marie Ruysschaert[1]

Structure and Function of Biological Membranes (SFMB)[1], and Genome and Network Bioinformatics (BiGRe)[2], Université Libre de Bruxelles (ULB), Boulevard du Triomphe – CP 263, B-1050 Brussels, Belgium

Running title: Identification of Specific Protein-Lipid Interactions

Address correspondence to: Marc Lensink, E-mail: marc.lensink@ulb.ac.be, Tel: +32 2 650 5411, Fax: +32 2 650 5425

Protein-lipid interactions are increasingly recognized as central for structure and function of membrane proteins. However, with the exception of simplified models, specific protein-lipid interactions are particularly difficult to highlight experimentally. We used here molecular dynamics simulations to identify a specific protein-lipid interaction in lactose permease, a prototypical model for transmembrane proteins. The interactions can be correlated with the functional dependence of the protein to specific lipid species. The technique is simple and widely applicable to other membrane proteins, and a variety of lipid matrices can be used.

INTRODUCTION

The molecular mechanisms underlying the influence of the lipid environment on the function of membrane proteins remain unclear. Lipid composition is known to have a modulatory effect on membrane protein activity, and for a number of membrane proteins a clear correlation was found between protein activity and bulk properties of the membrane bilayer such as fluidity [1]. Membrane proteins are anchored in their lipid environment through non-specific protein-lipid interactions [2,3], and different fluidic properties are therefore expected to play a modulatory role on membrane protein activity. However, there is also increasing evidence for specifically-bound lipids that are necessary to achieve biological function [4]. For example, a presence of phosphatidylethanolamine (PE) in the bilayer has been found to be essential for structure and function of a number of transmembrane transporter proteins [5-10]. Unfortunately crystallographic evidences of such dependency are exceptional. First, membrane protein structure remains rare, with, to this day, the structures of less than 200 unique proteins deposited in the Protein Data Bank (PDB). Second, membrane protein purification and crystallization require the use of detergent which tend to delipidate the target protein. Thus only a few examples of crystal structures show lipids, known to be important for function, specifically associated to the protein. The yeast cytochrome bc_1 complex contains a number of bound phospholipids, that suggest specific roles both for structural and functional integrity of the protein [11]. The crystal structure of the *Thermochromatium tepidum* photosynthetic reaction center shows one PE and seven detergent molecules on its molecular surface [12]. In the crystal structure of the *Rhodobacter sphaeroides* photoreaction center a specific interaction with cardiolipin has been found [13]. This binding site has subsequently been suggested to be a conserved feature of these reaction centers [14]. More recently, a tightly-bound cholesterol molecule was observed in the crystal structure of

a protein G-coupled receptor [15], triggering the question of a general presence of nonannular cholesterol binding sites for these proteins [16].

Molecular modeling and other computational studies are therefore playing an increasingly relevant role in the study of membrane proteins [17]; these include the use of molecular dynamics simulation to study dynamic events and conformational pathways [18]. The installment of membrane protein-targeted structural genomics and the rapid development of high-throughput structural biology techniques is expected to result in a significant increase in the number of available membrane protein structures in the near future [19]. But it is doubtful that these structures will show physiologically associated lipid molecules; the crystallization procedure typically delipidates the protein and only the strongest-bound lipid molecules may withstand the purification procedure and show sufficient electron density to be resolved at atomic level. Molecular modeling may play an important role here. With the increase in computing power, molecular dynamics simulations are now routinely applied and physics-based force fields are capable of treating a wide range of molecules and molecular interaction [20,21].

In this study, we use molecular modeling and dynamics to characterise a specific protein-lipid interaction between Lactose permease (LacY) and PE. LacY, which has been crystallized at a resolution of 2.95 Å [22], is a paradigm for the Major Facilitator Superfamily (MFS) [23], that represents as much as 25% of all membrane transport proteins, with over 15.000 sequence members identified to date [24]. Members of the MFS show a common architecture [25] and for a number of these transporters a presence of PE in the bilayer is required for proper structure and activity [26]. By carefully positioning the protein in a selection of lipid matrices and performing molecular dynamics simulations, we identify a limited number of strong interactions between specific amino acids and individual lipid molecules. The strength of the current technique lies in the fact that it leads to a molecular picture of the lipid-protein interactions involved. The only limitation is the number of membrane protein structures available with a sufficient resolution; a large number of lipid matrices in which proteins can be inserted are available and the technology required to create new ones mimicking all kinds of lipid membrane compositions is available and operational [20].

METHODS

Modeling Setup. We take the LacY crystal structure with the highest available resolution (2.95 Å), representing wild-type LacY in neutral pH, that is open on the cytoplasmic side and therefore susceptible to gradient sensing and sugar transport [22]. With lipid chain lengths near C_{18} often being most favorable for function [27,28], we employ POPC, POPE and POPG bilayers, that have been extensively used in MD simulations and for which parameter sets are available [29-31]. The bilayer thickness of these bilayers [32] shows good agreement with the calculated hydrophobic thickness of LacY [33], minimizing the effect of hydrophobic mismatch. Previous simulations of LacY, investigating the mechanism of sugar transport, have also been performed in a POPE bilayer [34].

Lactose permease (pdb-code 2cfq) was re-oriented with its principle axis aligned with the *z*-axis, rotated by 10 degrees over the *x*-axis and placed in the center of an equilibrated POPE bilayer [30] with all waters removed and of which the *x* and *y* coordinates had been expanded by a factor of 4. This orientation was found to yield best final alignment between phosphates and interfacial arginines. The entire system was subjected to 100 steps steepest descent energy minimization, applying 10^5 kJ/nm^2 position restraints on the protein non-hydrogen atoms. In subsequent iterations the system was restored to the reference area per lipid by shrinking the lipid *x* and *y* coordinates by 2% while deleting all lipids that had their phosphorus atom at a distance closer than 6 Å to any C_α atom of the protein [35]. Every iteration was accompanied by 100 steps steepest descent and after 8 iterations the deflation was increased to 5% per iteration.

The original box size of the resulting system of LacY and 298 POPE lipids was expanded by 1 nm perpendicular to the bilayer surface (now

*ca.*10×10×7 nm^3) and solvated with roughly 9000 water molecules. The atoms in the palmitoyl and oleoyl chains were given a Van der Waals radius of 6 Å to avoid solvation of the hydrophobic bilayer core. The system was neutralized by adding chloride ions [36], and subjected to 1000 steps steepest descent energy minimization and 100 ps MD using weak position restraining (10^3 kJ/nm^2 on the non-hydrogen protein atoms), followed by 10 ns free MD.

Previous simulations of PE systems have reduced the ethanolamine partial charges to correspond to the lysine parameterization [37]. In order to eliminate artificial association between the ethanolamine moiety and acidic side chains of the protein due to an overpolarization of the lipid partial charges, a new charge set was also developed by fitting atomic point charges to reproduce the electrostatic potential obtained from *ab initio* HF/6-31G* calculations [38], keeping symmetry considerations [39] and summing hydrogen charges onto the non-polar carbons [40]. For all purposes of this work, results obtained with both charge sets were found to be similar and all analyses presented henceforth were carried out using the latter charge set. Both charge sets are listed in Table 4.

Molecular dynamics simulations were performed with the Gromacs 3.3 package [41], using the Gromos96 43a2 force field [40], with Berger parameters for the lipid tails [42]. A time step of 2 fs was employed. All systems were coupled to a temperature bath of T = 310 K with a coupling constant τ_T = 0.1 ps. Protein, lipids and solvent (ions plus water) were coupled independently. SPC water was employed [43]. Pressure was maintained using semi-isotropic pressure coupling [44] at p = 1 bar with a coupling constant of τ_p = 1.0 ps. Van der Waals interactions were cut off at a distance of 1.4 nm, electrostatic interactions calculated with the particle mesh Ewald method [45], using fourth-order splining and a grid spacing of 0.12 nm. Equations of motion for the water molecules were solved analytically [46] and all covalent bonds in the system were constrained in the MD simulations [47].

The entire procedure of positioning, equilibration and simulation was repeated using POPE lipids with mono-methylated, di-methylated and tri-methylated ethanolamine (the tri-methylation resulting in POPC), and in a bilayer consisting of POPG lipids. The POPG bilayer was obtained by mutating every POPE lipid into POPG as before [48]. Methylation of the PE lipids was achieved by mutating a lipid amine hydrogen into methyl while at the same time modifying all related other parameters (bond lengths, angles, non-bonded parameters, *et cetera*). Additional *ab initio* calculations had shown it not to be necessary to modify the partial charges for the hydrogen to methyl mutation, which corresponds to the way the original POPE parameters were derived from POPC [30].

Validation of Equilibration Procedure. The orientational evolution of the α-helices in the simulation is used to validate the equilibration procedure and overall structural stability of the system. LacY contains twelve trans-membrane helices, of which two with a kink in the middle, and one additional short helix on the cytosolic side. Table 5 shows the residues defining the helices, and their angle with the bilayer normal, averaged over the four simulations with neutral lipids. Helical axes were calculated using a rotational least-squares fitting procedure [49]. It is known that TM helices may tilt or flex to match the hydrophobic thickness of a membrane [3,50]. We indeed observe this for the longer helices, *e.g.* for TM1 and TM4, and flex for TM3 and TM8. The shorter TM helices may span the bilayer without a tilt, *e.g.* TM5 and TM9. The observed behavior corresponds to what has been found by NMR experiments and MD simulations of the individual LacY helices TM1 and TM5 [50]. A secondary structure analysis shows all helices to retain their helical structure throughout the simulations (data not shown). Also no change in helix orientation (less than 10 degrees range of change in angle) was found, with the exception of TM6 and TM3.2, that varied within a 20 degree range. In addition, the relatively short helix 7.c showed some degree of motion, with a deviation up to 20 degrees around its average. TM6 is imperfect in the crystal structure and we find it to adopt fully α-helical structure in all simulations, most likely due to the stabilizing influence of the lipid environment. The same happens with

TM3.2, that is located next to TM6. The average helical angles and standard deviations were not found to differ significantly when calculated over the full trajectory or just the second half of it. Taken together with the correspondence to initial values we conclude that the positioning and equilibration procedure has had no destabilizing effect on the overall protein structure.

Detection of Lipid-mediated Salt Bridges. Salt bridge and hydrogen bond presence between LacY and lipids was determined using simple distance criteria. Residues involved in salt bridges and hydrogen bonds to individual lipids were determined using the criterium that at any point during the combined simulations the donor–acceptor distance fell below the threshold of 4 Å. The upper limit of 4 Å constitutes a weak interaction. Using the obtained residues, and the requirement that two of them must *simultaneously* be bound to the same lipid, lipid-mediated salt bridges were determined using a distance criterium of 4 Å of residue-phosphate (donor-phosphorus), 3 Å of residue-amine or residue-hydroxy (acceptor-donor) and 4 Å of residue-choline (acceptor-nitrogen) distance. Cumulative presence Δt_{cumul} of individual lipid-mediated salt bridges was calculated as combined fractional presence over the entire simulation time. The persistence factor F was calculated as

$$F_{atom} = \Delta t_{cumul} \times (MSF_{max} - MSF_{atom} + MSF_{min})$$

where *atom* is either the lipid nitrogen, hydroxy oxygen, or phosphorus atom, *MSF* the mean square fluctuation, and MSF_{min} and MSF_{max} equal 0.156 and 0.653 nm^2, resp. Lipid-mediated salt bridges were retained if they showed $\Delta t_{cumul} >$ 10.0%.

Residue Conservation. LacY sequences were obtained by querying the NCBI non-redundant sequence database of January 22nd, 2009, with BLAST [51], using the BLOSUM62 matrix and default parameters, and the *E. coli* LacY sequence as query. All sequences with an E-value better than 1.0e-10 were retained and aligned using MUSCLE [52], resulting in 92 unique sequences with the following (selected) keyword occurrences: *permease: 90; oligosaccharide/H+: 54; galactoside: 39; hypothetical protein: 38; lactose: 31; transport: 21; symporter: 20; lactose-proton: 20*. The resulting alignment can be found in Supplementary Figure S2.

RESULTS

Interactions between protein and lipids can manifest themselves in a number of ways: lipid acyl chains can settle on a hydrophobic surface patch created by one or two TM helices, the lipid carbonyl or phosphate groups can act as acceptor for hydrogen bonds emanating from the protein, and salt bridges may be formed between opposing charges, *e.g.* phosphate and arginine or lysine. Hydrophobic interactions are the weakest of these, and the strongest interactions are made by salt bridges, that are in fact a particularly strong form of a hydrogen bond.

In the following sections, we analyse the occurrence of individual and double hydrogen bonds and salt bridges. We then identify specific protein-lipid interactions from the presence of lipid-mediated salt bridges. The strongest of these is validated by additional simulations in lipid matrices with increasing degrees of methylation, and correlated with experimental data and sequence conservation in the family and superfamily.

Hydrophobic Interactions and Single Hydrogen Bonds. Annular lipids were found to show limited diffusion (as calculated from the mean square deviation of lipid tails), in accordance with spectroscopic data [53], but no clear transition in distribution between annular and bulk lipids was observed. We could therefore not identify any single lipid as being associated with the protein on the basis of spatial fluctuation and hydrophobic interaction.

LacY contains 17 negatively and 24 positively charged residues plus a number of otherwise hydrogen bond-capable residues; all these are potential candidates for protein-lipid specific interactions. We determined for all hydrogen bond-capable residues of the protein whether they were at one time or another favorably bonded to a hydrogen bond-complementary (but not necessarily opposingly-charged) lipid head group. For many of the candidate residues a large population of distances

indicative of binding is found, but it is impossible to distinguish between specific and non-specific protein-lipid interactions. As expected, the strongest interactions observed were those involving negatively (Asp, Glu) or positively charged residues (Arg, Lys), the latter of which are responsible for the anchoring of a membrane protein in its environment [3]. Interactions through the phosphate group are not lipid head group-specific, because they do not discriminate between different types of phospholipids. But also, single salt bridges and hydrogen bonds have a limited life time, with fluctuations on a 1 to 5 nanosecond time scale [2]. Lipids showing interactions through the amine, hydroxy or choline group would therefore show too fast an exchange rate with bulk lipids to call them specific.

Lipid-mediated Salt Bridges. Lipids that are simultaneously bound to two different residues, *e.g.*, through both their positively and negatively charged moieties, will show significantly longer residence times, and their diffusion away requires a two-step process. We therefore determined for every salt bridge and hydrogen bond, whether the same lipid was *simultaneously* bound to another salt bridge or hydrogen bond. Although cation-π interactions can be described using the current force field [31], and especially the PE amine group would be prone to such interactions, we have found no significant lipid-mediated salt bridges involving the π-clouds of aromatic residues. The list of lipid-mediated salt bridges is supplied in Table 1. As additional requirement we impose a presence of at least 10% of the simulation time (amounting to 1 ns), resulting in a total of nine lipid-mediated salt bridges, listed in Table 2.

Of the nine lipid-mediated salt bridges, five, including the three strongest, showing a $\Delta t_{cumul} >$ 15%, involve both an acidic and basic residue, and four of these occur in the PE bilayer (the fifth one with PC). Two bridges are found involving a PC lipid, and only one with PG. Most lipid-mediated salt bridges show similarly strong interactions, with persistence factor values lying between 5 and 10. An exception is the bridge involving Glu-215 and Lys-211, that shows a weaker interaction on the amine side, with $F_{donor} = 2.9$, related to an increased local fluctuation (Table 1). But unmistakably, the interactions with Asp-68 and Lys-69, two neighboring residues, stand out, with significantly larger persistence factors. In addition, this lipid-mediated salt bridge is weaker in the case of PC and may therefore be responsible for PE-preference. We validate the results by closer investigating of this bridge, both in terms of molecular details of the simulation and sequence conservation in the lactose permease family and major facilitator superfamily.

Binding of PE to Asp-68 and Lys-68 in Lipid Matrices with Increasing Degrees of Methylation. In independent studies of two MFS transporters, including LacY [54], transporter activity was investigated using PE lipids with various degrees of methylation. In both cases a triple methylation (*i.e.* PC-lipids) abolished activity completely, indicating a resident hydrogen on the ethanolamine moiety of PE to be essential [9,54]. We performed two additional simulations of LacY in lipid bilayer systems composed of mono- and di-methylated POPE lipids and analyze the four simulations together to provide a molecular basis for PE-specificity.

Fig. 1 (E and F) shows binding of Lys-69 to PE to occur already in the equilibration phase. Weak position restraints keep the protein heavy atoms in place, but the lipid is free to move and the phosphate group of PE is gradually moving closer to Lys-69, to form a salt bridge at the final stages of energy minimization (POPE), or the first step of position-restraint MD (Me-POPE, diME-POPE, POPC). The bridge is maintained throughout the remainder of the simulations, showing no significant difference in average distance.

The distance evolution between the lipid and Asp-68 is shown in Fig. 1 (A–D). Although a few lipids were found to bind to Asp-68 during the course of the simulations, in all four simulations a single lipid is continuously bound to Lys-69 and in every instance this is the same lipid that ultimately binds Asp-68. In the crystal and initial structure, Asp-68 is directed towards the inside of the protein, hydrogen-bonded to Lys-131. It is remarkable that POPC shows a movement of the choline group away from Asp-68 and all three PE

simulations an approach of the amine moiety to Asp-68 (Fig. 1B). After the side chain is released, the salt bridge is formed in the initial steps of the free MD simulation for POPE and mono-methylated POPE. Data for the full simulation (Fig. 1D) shows binding to occur progressively later upon increased methylation. Figure 1D shows initial binding of one of the two amine methyl groups at t = 1.5 ns, after which a rotation of the dimethylated amine group at t = 2 ns rotates the single hydrogen towards Asp-68. Binding of PC to Asp-68 (Fig. 1D) occurs only at t = 2.4 ns, and at an equilibrium distance more than 1 Å larger (0.37 nm *vs.* 0.25 nm), making the rather significant difference between strong and weak hydrogen bonding.

Our observation is that a phospholipid (PE or PC) is recruited by Lys-69, followed by a rotation of the Asp-68 side chain to bind its amine or choline group, thus forming a double salt bridge, or lipid-mediated salt bridge. Figure 2 shows an image of this lipid-mediated salt bridge.

Sequence Conservation in LacY and the MFS. The wide variety of substrates of MFS transporters is reflected in the large sequence variation going from one subfamily to the other, but all members of the MFS share a highly conserved cytoplasmic motif [55], that includes the acidic Asp-68, and the basic Lys-69, Arg-73 and Lys-74. Individual mutations have shown Asp-68, but also a resident positive charge, to be required for activity [55,56]. In addition, Asp-68 was found to be related to the proton gradient sensing mechanism but not to substrate transport [9,56,57]. Our strategy for detecting lipid-mediated salt bridges directly and unambiguously identifies Asp-68 as the most relevant PE-interacting residue. This is the first study where the PE-dependence of LacY [5] and the functional importance of Asp-68 [55] can be directly linked together.

For the residues involved in the three strongest lipid-mediated salt bridges of Table 2, residue conservation in lactose permease is listed in bold in Table 3. As opposed to the MFS, the multiple sequence alignment of (putative) lactose permease sequences is less restrictive, and still shows Asp-68 and Lys-69 to remain the only such bridge between two conserved residues. The other lipid-mediated salt bridges are likely to be only phosphate-specific, due to the limited conservation of the acceptor residues involved, *e.g.*, Asp-190 and Glu-215. Arg-73 and Lys-74 are located towards the periphery of the protein, at a small distance from Asp-68/Lys-69, and could therefore be involved in the initial recruitment of PE, shuttling a suitable phospholipid towards Asp-68 by the use of its flexible lysine side chain. In most LacY family members, a lysine is in fact found at position 73. Snorkeling actions of the lysine side chain are not an uncommon phenomenon [2,3], and could in this case be related to lipid recruitment.

CONCLUSION AND DISCUSSION

A Specific Interaction between D-68 and PE that contributes to the gradient sensing mechanism? Using molecular modeling and dynamics, we identify a specific protein-lipid interaction in LacY between Asp-68 and phosphatidylethanolamine (PE), by investigating the existence of lipid-mediated salt bridges of LacY embedded in five different lipid matrices. Our simulations show a consistent and strong hydrogen bond between (non-, mono-, and di-methylated) POPE amine groups and Asp-68, that is significantly weaker in the case of PC. In every instance the bond is formed to a free hydrogen of the amine group with the speed of formation being inversely related to the degree of methylation. Asp-68-bound PE is initially bound to Lys-69 through its phosphate entity before being recruited by Asp-68. The nearby residue Lys-74 may have a facilitating function in the recruitment of PE. This lipid-mediated salt bridge is the only such bridge existing between conserved residues in LacY.

Asp-68 and Lys-69 are also part of a highly conserved motif in the MFS; taken together with the PE-dependence of MFS transporters [26] and the involvement of Asp-68 in the energy-coupling mechanism [9,56,57], the interaction between Asp-68 and PE may constitute a specific protein-lipid interaction involved in ΔpH sensing. This suggests that not only secondary and tertiary structure elements [25], but also functional

elements such as the proton gradient sensing mechanism have been conserved in secondary transporters during evolution. The fact that Lys-69 has not been found to be crucial for transport is in accord with our findings; only a single resident positive charge is required in the conserved motif [55] and only a single positively-charged residue in principle suffices to recruit a phospholipid that can subsequently be shuttled towards Asp-68.

The role of alternative lipids. Although PE is reportedly required for proper LacY topogenesis and functioning [5,58], it has been proposed that it can be replaced by a non-endogenous lipid, monoglucosyl-diacylglycerol (GlcDAG), a lipid found in *Acholeplasma laidlawii* [59]. This may seem incompatible with the hypothesis proposed here considering the structural differences between PE and GlcDAG. We performed docking studies of the glucosyl moiety to identify binding modes with the LacY pocket. Surprisingly, the glucosyl can interact with both Asp-68 and Lys-69 in a PE-like fashion (Supplementary Figure S1). In the absence of the phosphate group, Lys-69 now interacts with the O_4 of the sugar ring while Asp-68 interacts with O_3, leading to a motif conformation very similar to the one observed in the POPE simulation. In addition, O_2 is found to be hydrogen-bonded to the backbone NH of Lys-69. Note that in diglucosyl-diacylglycerol (DGlcDAG) the O_2 atom is involved in the connection between the two sugar rings and would therefore not be available for binding, while binding of the second sugar ring to the motif would be prohibited by the bulkiness of DGlcDAG as a whole. While full-fledged MD simulations of LacY in GlcDAG and DGlcDAG bilayers should be done (a technically challenging process that goes beyond the scope of this paper), these results provide an explanation on why GlcDAG appears to allow LacY function and not DGlcDAG [26]. In contrast, and in accord with experimental data [54], the simulations of LacY in a POPG bilayer show no affinity of the glycerol side chain to the acidic D68 (Table 1), and it is found to bind the phosphate groups of itself and neighboring lipids.

Implications of the results on the functioning of membrane proteins. Our findings show that specific protein-lipid interactions may be identified through the identification of simultaneous occurrence of multiple (strong) interactions involving the same lipid. The occurrence of such interactions provides further evidence to the fact that specific protein-lipid interactions indeed play a role in the functioning of membrane proteins [4,10]. The significance of membrane proteins, that constitute approximately 25% of genomic sequences [19], is not reflected in the number of membrane structures that are currently available in the PDB. Yet membrane proteins represent 70% of current drug targets [60]. Specific lipid-binding residues may therefore also present a yet-unexplored area for drug targeting.

This is the first computational study ever to identify a specific protein-lipid interaction and provide the molecular basis for lipid species-specificity at the same time. The technique applied here can be used on any membrane protein structure, and embedded in a variety of lipid matrices, including mixtures with lipid species that do not form bilayer structures on themselves like cholesterol and cardiolipin. The approach is a thorough but simple application of molecular modeling and dynamics, a tried-and-tested technique.

LITERATURE REFERENCES

[1] Vigh, L., Escribá, P. V., Sonnleitner, A., Sonnleitner, M., Piotto, S., Maresca, B., Horváth, I. & Harwood, J. L. (2005), *Prog. Lipid Res.* **44**, 303–344.

[2] Deol, S. S., Bond, P. J., Domene, C. & Sansom, M. S. (2004), *Biophys. J.* **87**, 3737–3749.

[3] Nyholm, T. K., Ozdirekcan, S. & Killian, J. A. (2007), *Biochemistry* **46**, 1457–1465.

[4] Lee, A. G. (2005), *Mol. Biosyst.* **1**, 203–212.

[5] Bogdanov, M. & Dowhan, W. (1995), *J. Biol. Chem.* **270**, 732–739.

[6] Wang, X., Bogdanov, M. & Dowhan, W. (2002), *EMBO J.* **21**, 5673–5681.

[7] Zhang, W., Bogdanov, M., Pi, J., Pittard, A. J. & Dowhan, W. (2003), *J. Biol. Chem.* **278**, 50128–50135.

[8] Zhang, W., Campbell, H. A., King, S. C. & Dowhan, W. (2005), *J. Biol. Chem.* **280**, 26032–26038.
[9] Hakizimana, P., Masureel, M., Gbaguidi, B., Ruysschaert, J.-M. & Govaerts, C. (2008), *J. Biol. Chem.* **283**, 9369–9376.
[10] Powl, A. M., East, J. M. & Lee, A. G. (2008), *Biochemistry* **47**, 12175–12184.
[11] Lange, C., Nett, J. H., trumpower, B. L. & Hunte, C. (2001), *EMBO J.* **20**, 6591–6600.
[12] Nogi, T., Fathir, I., Kobayashi, M., Nozawa, T. & Miki, K. (2000), *Proc. Natl. Acad. Sci. USA* **97**, 13561–13566.
[13] McAuley, K., Fyfe, P. K., Ridge, J. P., Isaacs, N. W., Cogdell, R. J. & Jones, M. R. (1999), *Proc. Natl. Acad. Sci. USA* **96**, 14706–14711.
[14] Wakeham, M. C., Sessions, R. B., Jones, M. R. & Fyfe, P. K. (2001), *Biophys. J.* **80**, 1395–1405.
[15] Cherezov, V., Rosenbaum, D. M., Hanson, M. A., Rasmussen, S. G. F., Thian, F. S., Kobilka, T. S., Choi, H.-J., Kuhn, P., Weis, W. I., Kobilka, B. K. & Stevens, R. C. (2007), *Science* **318**, 1258–1265.
[16] Paila, Y. D., Tiwari, S. & Chattopadhyay, A. (2009), *Biochim. Biophys. Acta* **1788**, 295–302.
[17] Elofsson, A. & von Heijne, G. (2007), *Annu. Rev. Biochem.* **76**, 125–140.
[18] Aksimentiev, A., Brunner, R., Cohen, J., Comer, J., Cruz-Chu, E., Hardy, D., Rajan, A., Shih, A., Sigalov, G., Yin, Y. & Schulten, K. (2008), *Methods Mol. Biol.* **474**, 181–234.
[19] Carpenter, E. P., Beis, K., Cameron, A. D. & Iwata, S. (2008), *Curr. Opin. Struct. Biol.* **18**, 581–586.
[20] Guvench, O. & MacKerell, Jr., A. D. (2008), *Methods Mol. Biol.* **443**, 63–88.
[21] Klepeis, J. L., Lindorff-Larsen, K., Dror, R. O. & Shaw, D. E. (2009), *Curr. Opin. Struct. Biol.* **19**, 120–127.
[22] Guan, L., Mirza, O., Verner, G., Iwata, S. & Kaback, H. R. (2007), *Proc. Natl. Acad. Sci. USA* **104**, 15294–15298.
[23] Kaback, H. R. (2005), *C. R. Biol.* **328**, 557–567.
[24] Saier, Jr., M. H. (2000), *Microbiology* **146**, 1775–1795.
[25] Law, C. J., Maloney, P. C. & Wang, D. N. (2008), *Annu. Rev. Microbiol.* **62**, 289–305.
[26] Dowhan, W. & Bogdanov, M. (2009), *Annu. Rev. Biochem.* **78**, 515–540.
[27] Starling, A. P., East, J. M. & Lee, A. G. (1995), *Biochem. J.* **310**, 875–879.
[28] Yuan, C., O'Connell, R. J., Feinberg-Zadek, P. L., Johnston, L. J. & Treistman, S. N. (2004), *Biophys. J.* **86**, 3620–3633.
[29] Tieleman, D. P. & Berendsen, H. J. C. (1996), *J. Chem. Phys.* **105**, 4871–4880.
[30] Tieleman, D. P. & Berendsen, H. J. C. (1998), *Biophys. J.* **74**, 2786–2801.
[31] Lensink, M. F., Christiaens, B., Vandekerckhove, J., Prochiantz, A. & Rosseneu, M. (2005), *Biophys. J.* **88**, 939–952.
[32] Sprong, H., van der Sluijs, P. & van Meer, G. (2001), *Nat. Rev. Mol. Cell Biol.* **2**, 504–513.
[33] Lomize, M. A., Lomize, A. L., Pogozheva, I. D. & Mosberg, H. I. (2006), *Bioinformatics* **22**, 623–625.
[34] Yin, Y., Jensen, M. Ø., Tajkhorshid, E. & Schulten, K. (2006), *Biophys. J.* **91**, 3972–3985.
[35] Kandt, C., Ash, W. L. & Tieleman, D. P. (2007), *Methods* **41**, 475–488.
[36] Chandrasekhar, J., Spellmeyer, D. C. & Jorgensen, W. L. (1984), *J. Am. Chem. Soc.* **106**, 903–910.
[37] Marrink, S. J. & Mark, A. E. (2002), *Biochemistry* **41**, 5375–5382.
[38] Guest, M. F., Bush, I. J., van Dam, H. J. J., Sherwood, P., Thomas, J. M. H., van Lenthe, J. H., Havenith, R. W. A. & Kendrick, J. (2005), *Mol. Phys.* **103**, 719–747.
[39] Dupradeau, F.-Y., Cézard, C., Lelong, R., Stanislawiak, É., Pêcher, J., Delepine, J. C. & Cieplak, P. (2008), *Nucleic Acids Res.* **36**, D360–D367.
[40] Van Gunsteren, W. F., Billeter, S. R., Eising, A. A., Hünenberger, P. H., Krüger, P., Mark, A. E., Scott, W. R. P. & Tironi, I. G. *Biomolecular Simulation: The GROMOS96 Manual and User Guide* (Hochschulverlag AG an der ETH Zürich, Zürich, Switzerland, 1996).
[41] van der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E. & Berendsen, H. J. C. (2005), *J. Comput. Chem.* **26**, 1701–1719.
[42] Berger, O., Edholm, O. & Jähnig, F. (1997), *Biophys. J.* **72**, 2002–2013.
[43] Berendsen, H. J. C., Postma, J. P. M., Van Gunsteren, W. F. & Hermans, J. Interaction models for water in relation to protein hydration. In Pullman, B. (ed.) *Intermolecular Forces*, 331–342 (Reidel, Dordrecht, 1981).
[44] Berendsen, H. J. C., Postma, J. P. M., DiNola, A. & Haak, J. R. (1984), *J. Chem. Phys.* **81**, 3684–3690.

[45] Essman, U., Perela, L., Berkowitz, M. L., Darden, T., Lee, H. & Pedersen, L. G. (1995), *J. Chem. Phys.* **103**, 8577–8592.
[46] Miyamoto, S. & Kollman, P. A. (1992), *J. Comput. Chem.* **13**, 952–962.
[47] Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. (1997), *J. Comput. Chem.* **18**, 1463–1473.
[48] Lensink, M. F. (2008), *Methods Mol. Biol.* **443**, 161–179.
[49] Christopher, J. A., Swanson, R. & Baldwin, T. O. (1996), *Comput. Chem.* **20**, 339–345.
[50] Yeagle, P. L., Bennett, M., Lemaître, V. & Watts, A. (2007), *Biochim. Biophys. Acta* **1783**, 530–537.
[51] Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997), *Nucleic Acids Res.* **25**, 3389–3402.
[52] Edgar, R. C. (2004), *Nucleic Acids Res.* **32**, 1792–1797.
[53] Devaux, P. F. & Seigneuret, M. (1985), *Biochim. Biophys. Acta* **822**, 63–125.
[54] Chen, C.-C. & Wilson, T. H. (1984), *J. Biol. Chem.* **259**, 10150–10158.
[55] Jessen-Marshall, A. E., Paul, N. J. & Brooker, R. J. (1995), *J. Biol. Chem.* **270**, 16251–16257.
[56] Yamaguchi, A., Ono, N., Akasaka, T., Noumi, T. & Sawai, T. (1990), *J. Biol. Chem.* **265**, 15525–15530.
[57] Mazurkiewicz, P., Poelarends, G. J., Driessen, A. J. & Konings, W. N. (2004), *J. Biol. Chem.* **279**, 103–108.
[58] Bogdanov, M. & Dowhan, W. (1998), *EMBO J.* **17**, 5255–5264.
[59] Xie, J., Bogdanov, M., Heacock, P. & Dowhan, W. (2006), *J. Biol. Chem.* **281**, 19172–19178.
[60] Lundstrom, K. (2007), *J. Cell. Mol. Med.* **11**, 224–238.
[61] Chiu, S. W., Clark, M., Balaji, V., Subramaniam, S., Scott, H. L. & Jakobsson, E. (1995), *Biophys. J.* **69**, 1230–1245.
[62] Mulliken, R. S. (1955), *J. Chem. Phys.* **23**, 1833–1840.
[63] Bayly, C. I., Cieplak, P., Cornell, W. D. & Kollman, P. A. (1993), *J. Phys. Chem.* **97**, 10269–10280.

ACKNOWLEDGMENTS

MFL acknowledges support from the Région Wallonne of Belgium – DGTRE contract 515993. The BiGRe laboratory is a partner of the BioSapiens Network of Excellence funded under the sixth Framework programme of the European Commission (LSHG-CT-2003-503265). CG is a Chargé de Recherches of the Fonds National de la Recherche Scientifique.

FIGURE LEGENDS

Figure 1: Distances of Asp-68 and Lys-69 to the lipid. Distance of Asp-68 O_δ atoms to lipid N (A,B,C,D), and Lys-69 N_ζ to lipid P (E,F,G,H), for the last steps of the equilibration (A,B,E,F) and for the MD simulation (C,D,G,H). Weak position restraints on the protein heavy atoms are applied in the equilibration (A,B,D,E) and any change in distance there is solely due to lipid movement. In the free simulation (C,D,G,H) the entire system is free to move.

Figure 2: Final structure of the POPE simulation, showing the lipid-mediated salt bridge between Asp-68 and Lys-69. Distances indicated are between the closest hydrogen and oxygen atoms. Only amine hydrogens are shown.

Supplementary Data

Supplementary Figure 1. Critical distances in the interaction of GlcDAG with Asp-68 and Lys-69.

TABLES

Table 1: Lipid-mediated salt bridges

lipid-mediated salt bridge acceptor–donor[a]		Lipid[b]		RMSF (nm^2) Donor[c]	Acceptor[d]	Δt_{cumul}[e]	F_{donor}[f]	$F_{acceptor}$[g]
D44	N102	p	PC	0.288	0.214	10.14%	5.3	6.0
D44	N102	p	PG_1	0.536	0.375	5.04%	1.4	2.2
D44	N102	p	PG_2	0.534	0.375	6.64%	1.8	2.9
D68	R344	c	PE	0.505	0.506	9.41%	2.9	2.9
D68	S209	c	PE	0.505	0.506	3.42%	1.0	1.0
D68	K69	c	PE	0.213	0.213	36.84%	22.0	22.0
D68	K69	c	PC	0.318	0.204	28.57%	14.0	17.3
E139	N199	c	PE	0.286	0.282	4.97%	2.6	2.6
N166	Q167	p	PG_2	0.455	0.281	2.89%	1.0	1.5
D190	R73	c	PE	0.347	0.331	5.19%	2.4	2.5
D190	K74	c	PE	0.347	0.331	10.96%	5.1	5.2
D190	K188	c	PC	0.180	0.205	7.39%	4.6	4.5
D190	K74	c	PC	0.180	0.205	7.41%	4.7	4.5
N199	R73	c	PG_1	0.361	0.271	9.66%	4.3	5.2
N199	R73	c	PG_2	0.319	0.271	11.47%	5.6	6.2
H205	K69	c	PG_1	0.340	0.232	3.78%	1.8	2.2
E215	K211	c	PE	0.653	0.326	18.86%	2.9	9.1
E215	K211	c	PC	0.546	0.277	6.60%	1.7	3.5
E215	R218	c	PC	0.303	0.210	6.20%	3.1	3.7
Q219	K221	c	PC	0.443	0.175	3.71%	1.4	2.4
E255	Q256	p	PE	0.413	0.198	13.71%	5.4	8.4
E255	R259	p	PC	0.267	0.218	2.83%	1.5	1.7
E255	R259	p	PG_1	0.355	0.246	2.31%	1.0	1.3
Q340	Q412	c	PG_1	0.286	0.210	2.31%	1.2	1.4
E374	N371	p	PE	0.211	0.221	2.80%	1.7	1.6
E415	N284	c	PE	0.280	0.228	11.13%	5.9	6.5
E415	K335	c	PE	0.280	0.228	12.96%	6.9	7.5

(*a*) Hydrogen bond acceptor residue, binding amine, choline or hydroxy, and donor, binding phosphate

(*b*) Membrane side: (c)ytosol or (p)eriplasm, and lipid matrix: PE, PC or PG

(*c*) RMSF of lipid nitrogen or hydroxy oxygen

(*d*) RMSF of lipid phosphorus

(*e*) Cumulative presence of lipid-mediated salt bridge over simulation

(*f*) Lipid hydrogen bond donor persistence factor F_{donor}

(*g*) Lipid phosphorus persistence factor $F_{acceptor}$

Lipid-mediated salt bridges showing a cumulative presence of more than 2%. Bridges run from acceptor residue, over the lipid amine (PE), choline (PC), or primary (PG_1) or secondary (PG_2) hydroxy through phosphate, to donor residue. Lipid matrices used are POPE, POPC and POPG. Mean square fluctuation calculated over the lipid nitrogen, hydroxy oxygen and phosphorus atoms separately. Cumulative presence Δt_{cumul} and persistence factor F calculated as outlined in Methods.

Table 2: Lipid-bridged residue-residue contacts

acceptor	lipid	Δt_{cumul}	F_{donor}	$F_{acceptor}$	donor
Asp-68	PE	36.8%	22.0	22.0	Lys-69
Asp-68	PC	28.6%	14.0	17.3	Lys-69
Glu-215	PE	18.9%	2.9	9.1	Lys-211
Glu-255	PE	13.7%	5.4	8.4	Gln-256
Glu-415	PE	13.0%	6.9	7.5	Lys-335
Asn-199	PG	11.5%	5.6	6.2	Arg-73
Glu-415	PE	11.1%	5.9	6.5	Asn-284
Asp-190	PE	11.0%	5.1	5.2	Lys-74
Asp-44	PC	10.1%	5.3	6.0	Asn-102

Lipid-bridged residue-residue contacts, showing $\Delta t_{cumul} > 10.0\%$. The persistence factors F_{donor} and $F_{acceptor}$, calculated as outlined in the Methods, combine presence with spatial fluctuation. An increase in cumulative presence and a decrease in local fluctuation both lead to increased persistence factors.

Table 3: Residue conservation

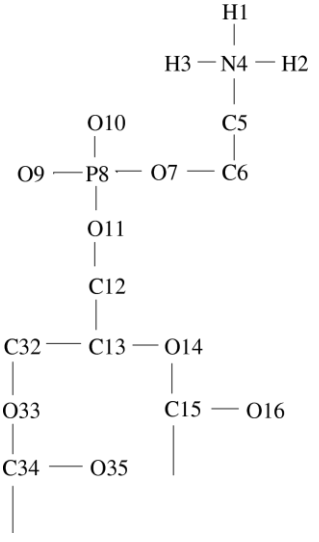
LacY[(a)]	similarity	Primary residue[(b)]	
D44	71.7%	E	40.2%
D68	94.6%	D	**94.6%**
K69	80.4%	K	**67.4%**
R73	89.1%	K	51.1%
K74	94.6%	K	79.4%
N102	76.1%	N	76.1%
E139	94.6%	E	94.6%
R142	62.0%	R	37.0%
N166	17.4%	P	78.3%
Q167	41.3%	N	28.3%
K188	69.6%	K	43.5%
D190	37.0%	E	27.2%
N199	20.7%	D	27.2%
H205	2.2%	P	18.5%
S209	58.7%	S	40.2%
K211	62.0%	K	**44.6%**
E215	23.9%	A	**25.0%**
R218	80.4%	K	47.8%
Q219	17.4%	M	33.9%
E255	46.7%	E	37.0%
Q256	23.9%	Q	23.9%
R259	34.8%	R	29.4%
N284	67.4%	N	67.4%
K335	89.1%	K	73.9%
Q340	29.4%	N	25.0%
R344	82.6%	R	63.0%
N371	10.9%	Y	15.2%
E374	82.6%	D	75.0%
Q379	33.7%	Q	32.6%
Q412	1.1%	R	5.4%
E415	6.5%	E	6.5%

(*a*) Residue in LacY sequence

(*b*) Primary residue in the alignment

Residue conservation in the lactose permease family, of the residues involved in lipid-mediated salt bridges that show a cumulative presence of more than 2%. Those residues that are involved in lipid-mediated salt bridges with a cumulative presence of more than 15% are listed in bold. Similarity score calculated by using the following similarity grouping: A,G,I,L,V; C,H,M; D,E; R,K; F,W,Y; N,Q; P; and S,T. In the alignment of LacY sequences, 152 out of 417 residues, or 37%, show a sequence similarity score above 75%. The primary residue is the most occurring residue at that position in the alignment, ignoring deletions and insertions.

Table 4: Lipid head group partial charges

atom	charge[(a)]	charge[(b)]	lipid head group
$H_{1,2,3}$[(c)]	0.4	0.252	
N_4	–0.5	–0.087	
C_5	0.3	0.107	
C_6	0.4	0.308	
O_7	–0.8	–0.468	
P_8	1.7	1.178	
$O_{9,10}$	–0.8	–0.787	
O_{11}	–0.7	–0.468	
C_{12}	0.4	0.16	
C_{13}	0.3	0.278	
O_{14}	–0.7	–0.48	
C_{15}	0.7	0.9	
O_{16}	–0.7	–0.62	
C_{32}	0.5	0.21	
O_{33}	–0.7	–0.48	
C_{34}	0.8	0.9	
O_{35}	–0.6	–0.62	

(*a*) Original partial charge

(*b*) Newly derived partial charge

(*c*) C in the case of methylated PE

Modified partial charges for the lipid head group. Although we qualitatively obtained the same results with previously published charges [61], these charges are based on a Mulliken population analysis of electron density [62] whereas a charge fitting to reproduce electrostatic potential is the currently more accepted standard [63]. This new charge set avoids preferential binding due to overpolarization.

Table 5: Transmembrane Helix Angles

helix		residues		angle[(a)]	angle[(b)]
TM	1	6	36	32.5	34.3 ± 3.9
TM	2	45	68	160.7	159.6 ± 2.8
2*TM	3.1	74	83	6.8	9.0 ± 4.2
	3.2	86	101	47.3	52.5 ± 7.4
TM	4	104	137	144.9	143.3 ± 2.4
TM	5	139	165	6.3	9.0 ± 2.4
TM	6	166	184	161.4	158.8 ± 5.5
	7.c	209	217	69.9	76.6 ± 7.0
TM	7	219	250	35.0	33.4 ± 2.5
2*TM	8.1	253	275	161.8	160.4 ± 3.8
	8.2	276	286	117.4	118.6 ± 3.9
TM	9	288	307	5.1	5.6 ± 2.5
TM	10	311	341	157.5	157.5 ± 2.5
TM	11	343	376	4.4	7.0 ± 2.8
TM	12	376	397	155.5	156.9 ± 3.1

([*a*]) initial angle, after positioning in the bilayer

([*b*]) average angle over the combined simulations

Transmembrane helix angles. Residues defining the TM helices and their angle with the bilayer normal (corresponding to the *z*-axis, pointing from cytosol to periplasm). The direction of the helix is taken into account when calculating its angle with the bilayer normal. For reference, the corresponding angles of the x-ray structure, after positioning in the bilayer, but before equilibration, are also listed. TM3 and TM8 show a kink in the middle. Helix 7.c is a short helix located in the cytosol, just before TM7 and under an angle of 97.4 ± 5.7 degrees with it.

Figure 1

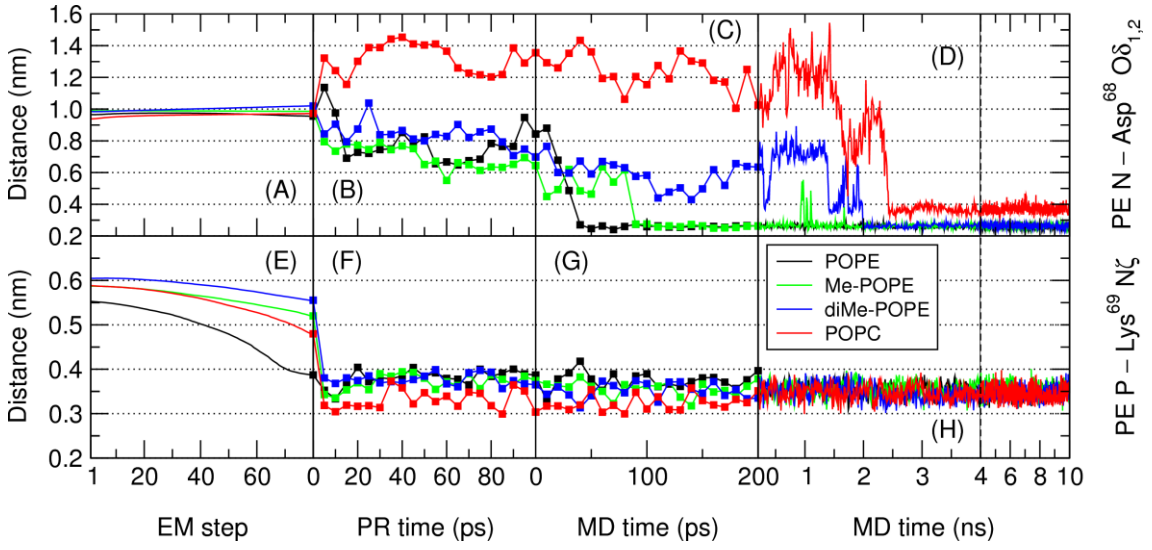

Distance (nm)
(A)
(B)
(C)
(D)
(E)
(F)
(G)
(H)
POPE
Me-POPE
diMe-POPE
POPC
PE N – Asp68 Oδ1,2
PE P – Lys69 Nζ
EM step
PR time (ps)
MD time (ps)
MD time (ns)

Figure 2