LAMBADA and InflateGRO2: Efficient Membrane Alignment and Insertion of Membrane Proteins for Molecular Dynamics Simulations

ARTICLE *in* JOURNAL OF CHEMICAL INFORMATION AND MODELING · SEPTEMBER 2012

Impact Factor: 3.74 · DOI: 10.1021/ci3000453 · Source: PubMed

2 AUTHORS:

Thomas H Schmidt
University of Bonn
10 PUBLICATIONS **47** CITATIONS
SEE PROFILE

Christian Kandt
Hochschule Bonn-Rhein-Sieg
44 PUBLICATIONS **824** CITATIONS
SEE PROFILE

JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

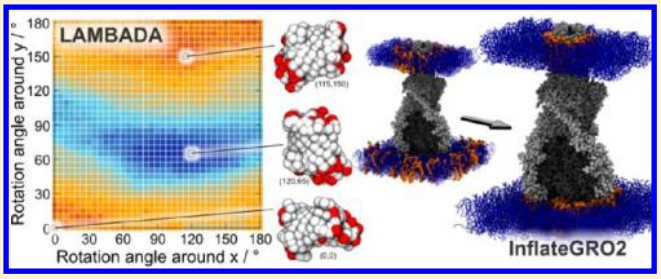
LAMBADA and InflateGRO2: Efficient Membrane Alignment and Insertion of Membrane Proteins for Molecular Dynamics Simulations

Thomas H. Schmidt and Christian Kandt*

Computational Structural Biology, Department of Life Science Informatics B-IT, Life & Medical Sciences (LIMES) Institute, University of Bonn, Dahlmannstr. 2, 53113 Bonn, Germany

Supporting Information

ABSTRACT: At the beginning of each molecular dynamics membrane simulation stands the generation of a suitable starting structure which includes the working steps of aligning membrane and protein and seamlessly accommodating the protein in the membrane. Here we introduce two efficient and complementary methods based on pre-equilibrated membrane patches, automating these steps. Using a voxel-based cast of the coarse-grained protein, LAMBADA computes a hydrophilicity profile-derived scoring function based on which the optimal rotation and translation operations are determined to align protein and membrane. Employing an entirely geometrical approach, LAMBADA is independent from any precalculated data and aligns even large membrane proteins within minutes on a regular workstation. LAMBADA is the first tool performing the entire alignment process automatically while providing the user with the explicit 3D coordinates of the aligned protein and membrane. The second tool is an extension of the InflateGRO method addressing the shortcomings of its predecessor in a fully automated workflow. Determining the exact number of overlapping lipids based on the area occupied by the protein and restricting expansion, compression and energy minimization steps to a subset of relevant lipids through automatically calculated and system-optimized operation parameters, InflateGRO2 yields optimal lipid packing and reduces lipid vacuum exposure to a minimum preserving as much of the equilibrated membrane structure as possible. Applicable to atomistic and coarse grain structures in MARTINI format, InflateGRO2 offers high accuracy, fast performance, and increased application flexibility permitting the easy preparation of systems exhibiting heterogeneous lipid composition as well as embedding proteins into multiple membranes. Both tools can be used separately, in combination with other methods, or in tandem permitting a fully automated workflow while retaining a maximum level of usage control and flexibility. To assess the performance of both methods, we carried out test runs using 22 membrane proteins of different size and transmembrane structure.



INTRODUCTION

Membrane proteins are key players in a number of fundamental biological processes, and their investigation through molecular dynamics (MD) computer simulations plays an increasingly important role in complementing experimental data and sparking new investigations.[1−3] At the beginning of each MD simulation stands the generation of a suitable starting structure. Whereas for water-soluble proteins this process is completed by solvating the simulation box with water and ions, membrane proteins require two additional working steps due the heterogeneous composition of the protein's microenvironment. These steps include (1) the alignment of membrane and protein followed by (2) the actual accommodation of the protein in the membrane.

To align protein and membrane, the methods currently available employ three different principal strategies, all based on the membrane protein-characteristic distribution of charged and hydrophobic surface residues known as the hydrophobic belt (HB),[4,5] indicating the protein surface sections directly exposed to the hydrophobic part of the membrane. (1) *Manual alignment*. Using molecular viewers or editors, the protein's hydrophobic belt is identified through visual inspection and the necessary rotation and translation steps required to align protein and membrane accordingly are carried out by hand. (2) *Semiautomatic alignment*. HB detection and prediction of the membrane position are performed computationally, whereas the necessary rotation and translation steps required for membrane/protein alignment are carried out manually. Computational tools applying this approach include the *orientation of proteins in membranes* (OPM) database[6] providing for some membrane proteins the predicted location of the lipid bilayer based on precomputed minimum transfer energy calculations where each protein was systematically transferred from a hydrophilic to a hydrophobic milieu;[7] IMPALA[8] using a Monte Carlo-based procedure to minimize a scoring function dependent on restraint energy, penetration depth, and angle against a reference vector; TMDET[9] employing a geometrical approach using PDB symmetry information and a scoring function based on protein hydrophobicity and conformational

Received: January 25, 2012

   dx.doi.org/10.1021/ci3000453 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

properties; or the algorithms implemented in GARLIC since version 1.6[10] resorting to hydrophobicity scoring functions. An explicit representation of the membrane is not accounted for in any of the aforementioned tools. (3) *Automatic alignment.* HB detection and prediction of the membrane position, as well as the according rotation and translation steps required for membrane/protein alignment are carried out computationally. At the time of writing the only tool available in this category is the web-based CHARMM-GUI membrane builder[11,12] resorting to the entries of the OPM database.

To accommodate the protein in the membrane two general strategies are currently in use,[13] both requiring knowledge of the protein's membrane-exposed regions or the protein in an orientation aligned with the lipid bilayer. Either the membrane is constructed around the protein or the protein is inserted into a pre-equilibrated membrane patch. Whereas the construction approach bears the advantage of obtaining membrane sizes tailored to the protein in question, long membrane equilibration times are required before production runs can be initiated. The insertion approach requires shorter equilibration times, which easily permits separate tests and validation of the membrane patch while restricting the user to the initial patch size or multiples thereof.

Methods employing the construction approach include the method of replacing lipid headgroup-sized pseudoatoms introduced by Woolf and Roux[13−15] and the construction option in the related CHARMM-GUI membrane builder[11,12] as well as constructing a bilayer with enlarged interlipid spacing and subsequent scaling of *XY* coordinates as proposed by Ash and co-workers.[13] Additionally, with the increase of computer power the strategy of lipid self-assembly around the protein of interest in sufficiently long MD simulations has gained importance particularly in coarse-grained MD for which a database of currently 382 DPPC-inserted membrane proteins is available.[16−18]

At the time of writing, methods available for inserting a protein into a pre-equilibrated membrane include the manual deletion of lipids overlapping with the protein; the usage of repulsive forces to drive out lipids from the protein volume as implemented in the related methods cylinder,[19] mdrun_hole,[20] and GRIFFIN,[21] each a further development of its predecessor; the insertion method of the CHARMM-GUI membrane builder;[11] the g_membed technique of compressing and expanding the protein in the membrane;[22] and the InflateGRO method[13] of employing lateral lipid translation within the membrane plane using scaling factors to first expand the membrane and, then, delete lipids within a user-defined distance cutoff around the protein, followed by a series of alternating steps of compression and energy minimization to bring the system back to natural dimensions. In this process the area per lipid serves as reference value to assess the required amount of shrinking steps. Using only energy minimization to remove steric clashes between lipids and protein in absence of any other solvent, the method provides a quick and easy way to embed even large membrane proteins. As long as the protein's membrane normal is parallel to the *Z*-axis, the method is also applicable to curved membrane systems.

InflateGRO has five distinctive drawbacks: (1) Membrane expansion and lipid exposure to vacuum can lead to a loss of equilibrium lipid conformations including a gel phase-like straightening of lipid tails during energy minimization or an untangling of entwined lipids through the initial expansion step.[22] (2) Unless the ideal combination of expansion factor and distance cutoff is determined manually for each system, the scaling factor and distance cutoff-dependent lipid deletion in the membrane-expanded system can lead to locally inhomogeneous lipid packing even though the overall area per lipid matches equilibrium values. (3) Due to an overestimation of the area occupied by protein—the underestimation of the protein area as stated in the original publication[13] is a typographical error—and the neglecting of protein-internal transmembrane cavities, there are inaccuracies in the area per lipid calculation which become most pronounced with membrane proteins of tilted or asymmetric transmembrane structure containing large cavities. (4) Using a residue name criterion to identify components affected (usually lipids) and unaffected by lateral translation (usually protein), InflateGRO's application flexibility is limited. For example, preparing lipid mixtures is laborious requiring workarounds such as renaming lipids between compression (all lipids share the same name) and energy minimization steps (each lipid species has its own name), whereas protein insertion into multiple membranes as necessary for, e.g., resistance nodulation division (RND) multidrug efflux pumps[23] is not possible. (5) InflateGRO is restricted to atomistic MD and not applicable to coarse-grained simulations. A more detailed review on the currently available methods to build starting structures for MD simulations of membrane proteins can be found in ref 3.

As already stated in ref 13, reproducibility and transferability of an experimental result independently from the person carrying out that particular experiment form one of the cornerstones of any scientific endeavor. Unfortunately, especially in generating starting structures for membrane protein simulations, this can be problematic as the influence the individual user has on the setup process can be large. It would therefore be desirable to aim for an increasingly automated, user-friendly approach to setup simulations minimizing the individual influence—offering maximum liberty in selecting protein and membrane components—while maximizing reproducibility and transferability of the results obtained. At the same time such an approach offers a vast application potential in automated large scale MD-based screening of membrane proteins—for both experimentally determined structures and homology models. At the time of writing, all available membrane/protein alignment methods require at least partial manual intervention or additional input precalculated by another application, whereas all accommodation methods require a prealigned protein orientation. Currently only the CHARMM-GUI membrane builder offers the option of a fully automated workflow of simulation setup. While representing a major step forward, practical usage is limited to proteins already contained in the OPM database; the membrane and lipid species provided by the server; as well as two membrane patch sizes and proteins of circular, transmembrane, cross-sectional area when pre-equilibrated membrane patches are to be used.[11]

Here we report two complementary, fast, efficient, and easy to use methods to generate starting structures for MD simulations of membrane proteins, automatically aligning protein and membrane as well as accommodating the protein in the membrane. Resorting to pre-equilibrated membrane patches, both tools can be used separately or in conjunction, permitting a fully automated workflow while retaining a maximum level of usage control and flexibility. Whereas many laboratories have sets of scripts to build membrane/protein systems with their desired simulation package, the tools we

dx.doi.org/10.1021/ci3000453 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

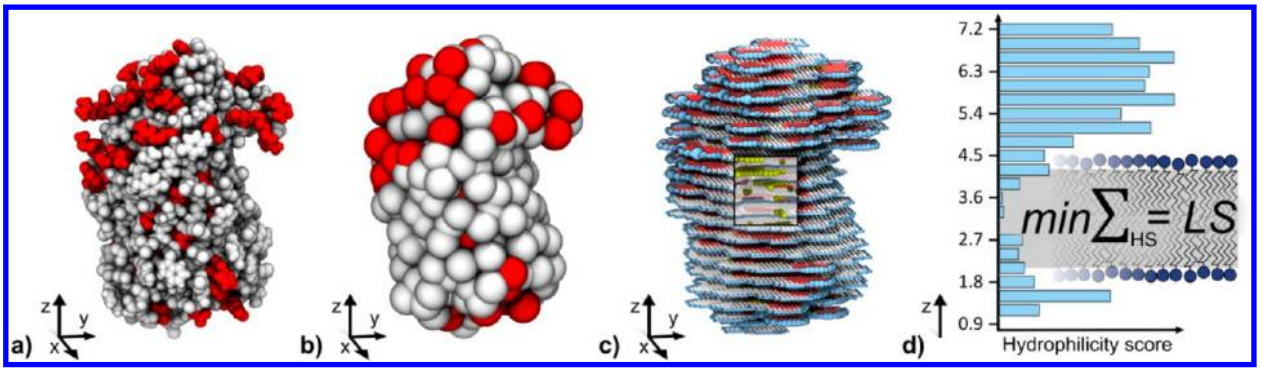


Figure 1. LAMBADA's protein/membrane alignment method based on calculating the LAMBADA score (LS). To compute LS, the protein (a) is first coarse-grained (b) and assigned hydrophilicity scores (HS) using a binary classification scheme of charged (red) and uncharged residues (white). Subsequently the protein is converted to a voxel representation (c, white and red spheres) based on which a 3D cast of the protein surface is computed (c, blue spheres and tubes) where cast voxels inherit the HS of adjacent protein voxels. Internal surfaces stemming from protein cavities (yellow spheres) are excluded in this process. Summing the cast voxel HS over each *XY* plane, a hydrophilicity profile along the *Z*-axis is computed (d). Calculating the HS sum of all *XY* voxel planes located within a membrane-thick sliding window, the least hydrophilic section of the protein surface is determined. Ths HS sum of this window position is defined as the LAMBADA score.

introduce here offer a clean, reproducible, and user-friendly approach to build starting structures for membrane protein MD simulations that anyone can use.

Using a grid-based cast of the coarse-grained protein surface, LAMBADA computes a hydrophilicity profile-based scoring function based on which the optimal rotation and translation operations are determined to align protein and membrane. Employing a purely geometrical approach, LAMBADA is independent from any precalculated data, and aligns even large membrane proteins within minutes on a regular workstation. Starting from aligned protein and membrane coordinates—as provided by LAMBADA—InflateGRO2 is an extension of the InflateGRO insertion method, addressing the shortcomings of its predecessor in a fully automated workflow adapting to each protein/membrane system to calculate optimal operation parameters. An improved calculation of the area occupied by protein within the membrane, that takes atomic van der Waals radii and protein-internal cavities into account, allows the automatic determination and deletion of an optimal number of lipids, preserving correct lipid densities in both leaflets. The subsequent phase of membrane expansion and alternating compression and energy minimization steps is restricted only to lipids still overlapping with the protein using automatically determined, system-optimized scaling factors to ensure a minimum disturbance of the membrane structure due to lipid exposure to vacuum. Applicable to atomistic and coarse grained structures in the MARTINI format, InflateGRO2 offers increased application flexibility as the new option of using GROMACS index files permits the easy preparation of systems exhibiting heterogeneous lipid composition as well as protein-embedding into multiple membranes. Even for large proteins the InflateGRO2 membrane insertion is completed within minutes on a regular workstation.

To assess validity and overall performance of both methods, we carried out test runs on a set of 22 membrane proteins of different TM structure. Licensed under the GNU GPL v2, LAMBADA and InflateGRO2 are available free of charge at: http://code.google.com/p/lambada-align/ and http://code.google.com/p/inflategro2/ or http://www.csb.bit.uni-bonn.de/downloads.html

MATERIALS AND METHODS

LAMBADA: Automatic Membrane/Protein Alignment.

Starting from two input files provided in the GROMACS GRO format, LAMBADA automatically aligns membrane and protein coordinates. The alignment procedure involves the two major working steps of (a) finding a protein orientation where the protein's HB runs parallel to the membrane plane, which is expected to coincide with the *XY* plane, as well as (b) translating the membrane patch in the *Z*-direction to achieve the best match between HB and the hydrophobic section of the membrane. The basis for both working steps is the detection of the least hydrophilic, membrane-thick section of the protein surface quantified in a numerical expression we call the LAMBADA score (LS).

Calculating the LAMBADA Score. First the protein (Figure 1a) is converted to a coarse-grained representation using a one bead per residue mapping where a 0.4 nm radius bead is placed at the geometrical center of each residue (Figure 1b). Asp, Glu, Arg, Lys, and His beads are considered "charged" (Figure 1b, red spheres), whereas the remaining beads are regarded as "non-charged" (Figure 1b, white spheres). Coarse-graining the protein bears the advantage of reducing model complexity, resulting in an overall faster performance and yielding a sharper HB detection by minimizing the influence of individual side chain orientations on the overall distribution of charged residues on the protein surface.

Subsequently, the CG protein is converted to a voxel representation using grid resolutions of 0.1 nm in the *X* and *Y* direction and 0.3 nm in *Z* (Figure 1c, white and red spheres), which we found giving the best results. Explicitly taking actual bead sizes into account, the conversion employs the same algorithm as described in detail in the InflateGRO2 part of the Materials and Methods section. Depending on its position, each voxel is assigned a hydrophilicity score (HS) of either +2000 times *Z*-resolution, when the voxel is located within the radius of a charged bead (Figure 1c, red spheres), or −1 times *Z*-resolution, when the voxel is located within the radius of a noncharged bead (Figure 1c, white spheres). Acting as a means of "contrast enhancement", a substantially higher score for charged residues was found to produce better results in HB detection and membrane/protein alignment. While increasing the HS for charged residues up to 2000 generated more

dx.doi.org/10.1021/ci3000453 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

accurate results, HSs exceeding 2000 yielded no further improvements. Taking the multiplier *Z-resolution* into account transforms the resulting HS into a relative numerical range, independent of the used *Z*-resolution which can be modified by the user.

To obtain the distribution of charged surface residues, a 3D cast of the voxelized protein is computed (Figure 1c, blue spheres and tubes). In this process, internal surfaces stemming from protein-internal cavities (Figure 1c, yellow spheres) are excluded employing the same *XY* plane-based cavity detection approach described in detail in the InflateGRO2 section. The surface cast is computed by checking each *protein voxel* within the same *XY* plane in all eight possible directions for empty neighbor voxels. If such a void voxel is found, it is classified as *cast voxel* and assigned the HS sum of all its neighboring protein voxels. That way enveloping voxel layers are generated (Figure 1c, blue spheres and tubes), based on which a hydrophilicity profile along the *Z*-axis is computed summing the HS of all cast voxels located in the same *XY* plane (Figure 1d). To obtain sharp results even for proteins with conically shaped transmembrane sections, where strong variations occur in protein circumference and HS sum per slice, LAMBADA normalizes for each *XY* plane the number of detected charged surface grid cells against the total number of surface voxels.

The obtained hydrophilicity profile (Figure 1d), is then processed further to determine the least hydrophilic section of the protein surface. To this end LAMBADA calculates the HS sum of all *XY* voxel planes located within a sliding window, whose size corresponds to the thickness of the hydrophobic part of the membrane. Either this thickness is determined automatically based on the provided membrane input structure—requiring the user to provide information on the lipid headgroup atoms encoded as GROMACS NDX index file—or, as done in the OPM database approach,[6] a default thickness of 2.4 nm is assumed which we found giving good results for membranes with similar leaflet thickness. Starting from the bottom *XY* plane of cast voxels, all possible window positions are tested, calculating for each the HS sum of the contained *XY* cast voxel planes. The window position exhibiting the lowest HS sum represents the least hydrophilic section of the protein surface and the HS sum of that window position is defined as *LAMBADA score* (Figure 1d, LS).

Orienting the Protein. As shown in Figure 2a, LS is a function of the protein orientation and becomes minimal when the protein's HB runs parallel to the *XY* plane. Therefore, orienting the protein with its HB parallel to the membrane is equivalent to finding the configuration of the protein where LS is minimal (LSmin). To this end, the protein is first brought in a standardized starting configuration oriented along its principal axes. Using the same algorithm as implemented in the GROMACS tool *editconf*,[24] this results in the first principal axis coinciding with the *X*-axis.

Subsequently LAMBADA determines the protein's LSmin configuration using a recursive optimization approach to test different protein orientations (Figure 3). Initially the algorithm uses an angular step size of 90° to search an *XY* angle range of 180° for the protein configuration where LS is minimal. Starting from the previous LSmin configuration, LAMBADA now uses an angular step size of 45° to search an *XY* angle range of 90° for the configuration where LS is minimal. Consistently using the previously determined LSmin configuration while continuously halving angular step size and *XY* angle range, the procedure is repeated until the angular step

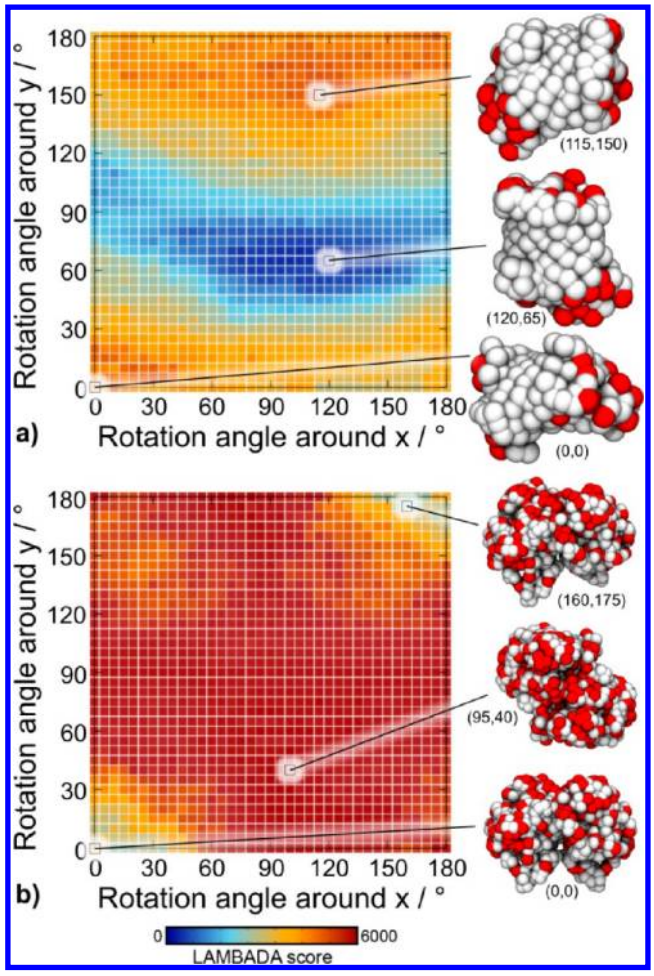


Figure 2. LAMBADA score (LS) is a function of the protein orientation and becomes minimal when the protein's hydrophobic belt runs parallel to the *XY* plane (a). When the LS distribution is computed for a water-soluble protein, all configurations yield LSs > 2000, indicating the absence of a hydrophobic belt (b). The outer membrane protein phospholipase A (PDB-ID: 1QD5) and the water-soluble NADPH binding site of the beef liver catalase (PDB-ID: 8CAT) were used as examples.

size falls below 5°. Decreasing the step size further was found not leading to further improvements in protein orientation. The final LSmin configuration has the protein oriented with its HB running parallel to the *XY* plane which is equivalent with the membrane plane in our approach. Figure 3 illustrates the procedure showing for the first four recursion steps the orientations of the first principal axis (Figure 3a) and the matrices of the resulting LAMBADA scores (Figure 3b), as well as the protein LSmin configurations compared to their first principal axis orientations (cylinders) and the HB plane (yellow spheres) (Figure 3c).

Translating the Membrane. Once the protein is oriented with its HB parallel to the *XY* plane, the membrane is translated in the *Z* direction so that the geometrical center of its *Z* coordinates matches the center of the sliding window position for which the LS was determined (Figure 1d). In the output structure, the *XY* dimensions of the membrane and the *Z* extension of the protein box are used.

Test Runs and Comparison to the OPM Database. To assess LAMBADA performance, we carried out test runs using 6.5 nm × 6.5 nm, 13 nm × 13 nm, or 19.5 nm × 19.5 nm pre-

dx.doi.org/10.1021/ci3000453 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

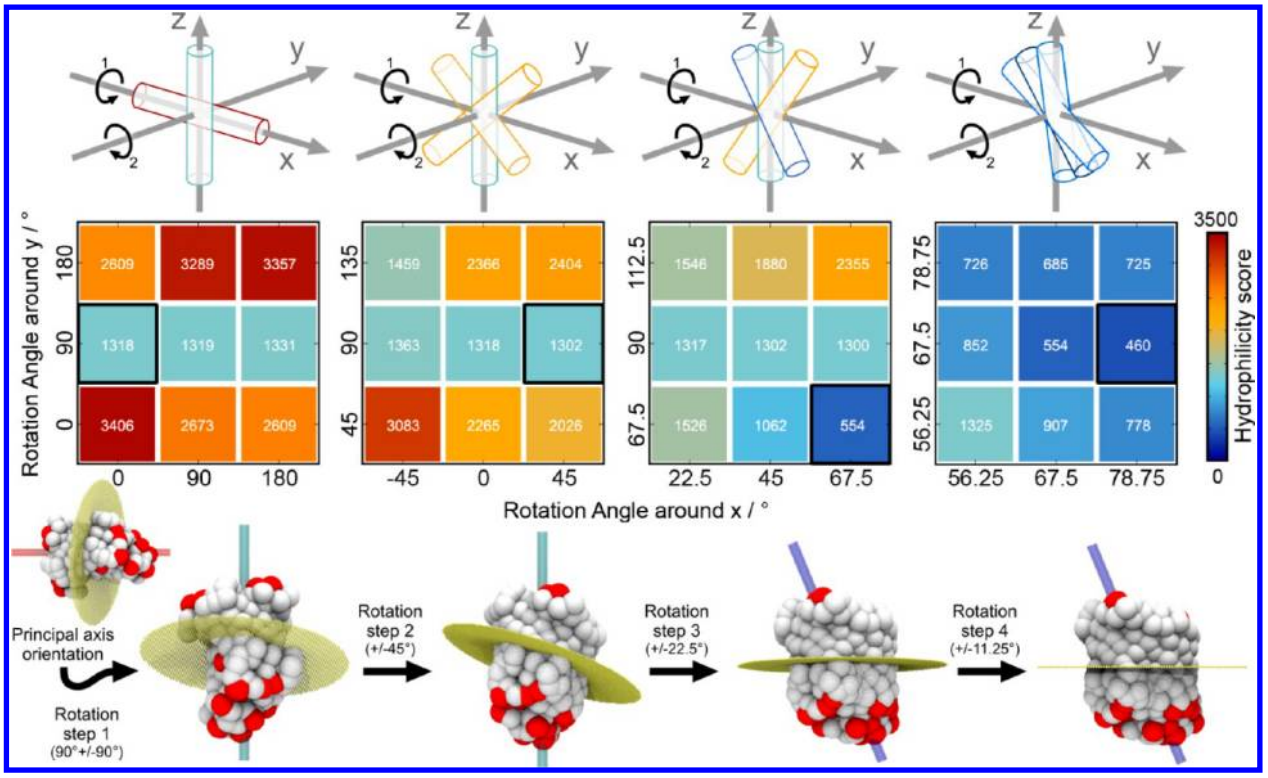


Figure 3. First four recursion steps of the LAMBADA protein orientation procedure. Instead of testing all possible configurations based on rotating the protein's first principal axis, LAMBADA uses recursive optimization procedure to detect a minimum in the LS landscape and find a protein orientation where the hydrophobic belt runs parallel to the *XY* plane. Starting from a standardized protein orientation where the first principal axis coincides with the *X*-axis and using increasingly smaller angular step sizes, an increasingly smaller configuration range is scanned around the protein configuration that was determined displaying the smallest LS in the previous iteration step. Here, the procedure is illustrated showing for the first four iteration steps the orientations of the first principal axis (a) and the matrices of the resulting LAMBADA scores (b) as well as the LSmin configurations together with their first principal axis orientations (cylinders) and the HB plane (yellow spheres) (c).

Table 1. LAMBADA Test Proteins

protein (PDB-ID)	residues	TM structure	internal TM cavity	alignment time[a]		orientation deviation (compared to OPM)
					+ orientation	
outer membrane protein OmpX (1QJ9)[59]	148	cylindrical (tilted)	no	2 s	36 s	9.991°
outer membrane phospholipase A (1QD5)[30]	257	cylindrical (tilted)	no	3 s	56 s	9.490°
outer membrane deacylase LpxR (3FID)[48]	296	cylindrical (tilted)	no	3 s	57 s	8.860°
Maltoporin LamB (1AF6)[33]	1263	irregular	yes	9 s	3 min 43 s	0.029°
outer membrane efflux duct TolC (1EK9)[29]	1284	conical	yes	9 s	4 min 16 s	0.302°
cell toxin α-hemolysin (7AHL)[35]	2051	irregular	yes	12 s	6 min 18 s	3.092°
potassium channel KcsA-Fab complex (1R3J)[31]	2136	conical	no	22 s	19 min 11 s	0.029°
potassium channel Kv1.2 (3LUT)[32]	2864	irregular	no	17 s	11 min 42 s	0.000°
multidrug efflux pump AcrB (2GIF)[38]	3108	irregular	yes	15 s	9 min 20 s	2.137°

[a]Performed using a single core on a 4 GB RAM DELL Precision 390 workstation (CPU: Intel Core2 6700 2 × 2.66 GHz).

equilibrated bilayer patches based on refs 25 and 26 and 19 membrane proteins of different size and transmembrane structure (Table 1; Supporting Information Table 1). To compare our orientations with the ones in the OPM database, we calculated the angle between the OPM and LAMBADA membrane planes after superimposing the corresponding protein coordinates.

LAMBADA is licensed under the GNU GPL v2 and available free of charge at http://code.google.com/p/lambada-align/ or http://www.csb.bit.uni-bonn.de/lambada.html

InflateGRO2: Automatic Protein Insertion. Starting from a GRO file containing protein and membrane in an aligned orientation, InflateGRO2 first determines the number of lipids to be removed by calculating the area occupied by the protein in each leaflet (Figure 4). This is done projecting all protein atoms located within a *Z*-range of ±0.5 nm around the average *Z*-position of the lipid phosphorus atoms in each leaflet's headgroup region (Figure 4a) on a 2D grid using a grid cell resolution of 0.01 nm (Figure 4b). Looping over all protein atoms, the algorithm checks for each atom the surrounding grid cells contained in a square-shaped section of the 2D grid as spanned by the atom's van der Waals radius. Grid cells located within the range of the atom's van der Waals radius are assigned a value of 1.

dx.doi.org/10.1021/ci3000453 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

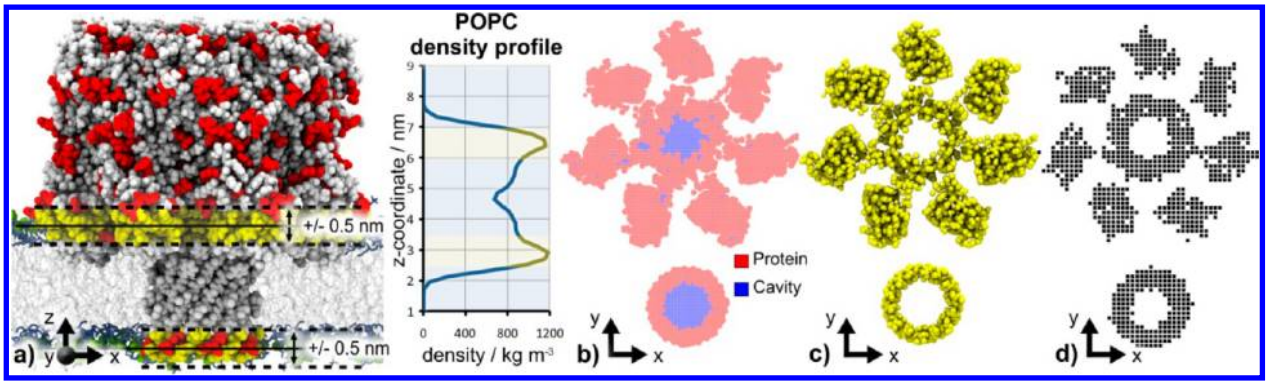


Figure 4. Calculation of the area occupied by the protein in each leaflet. To calculate the area occupied by the protein, based on which the amount of lipids to delete and the area per lipid is determined, InflateGRO2 projects protein atoms located in the lipid headgroup regions (a) on 0.01 nm 2D grids taking into account both the protein atoms' van der Waals radii and cavities (b). For comparison, the respective protein sections are additionally shown as a space-filling CPK visualization (c) as well as the protein representation used in the initial version of InflateGRO (d). InflateGRO2 calculation of the area occupied by the protein is illustrated using the cell toxin α-hemolysin as an example.

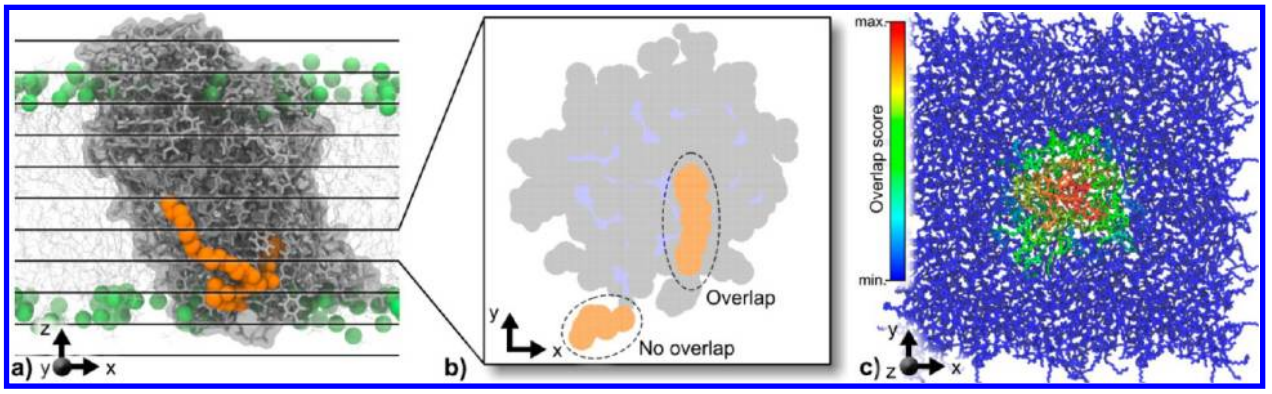


Figure 5. Ranking lipids by protein overlap. Dividing the protein transmembrane range into 0.5 nm Z slices, InflateGRO2 determines for each lipid its overlap with the protein quantified as individual overlap score (a). Projecting protein and lipid atoms on a 2D grid for each slice, the amount of overlapping 0.01 nm grid cells is determined (b) which, summed over all slices, constitutes a lipid's overlap score based on which all lipids in the system are ranked (c). Lipid overlap score calculation is illustrated using the lipid A deacylase LpxR as an example.

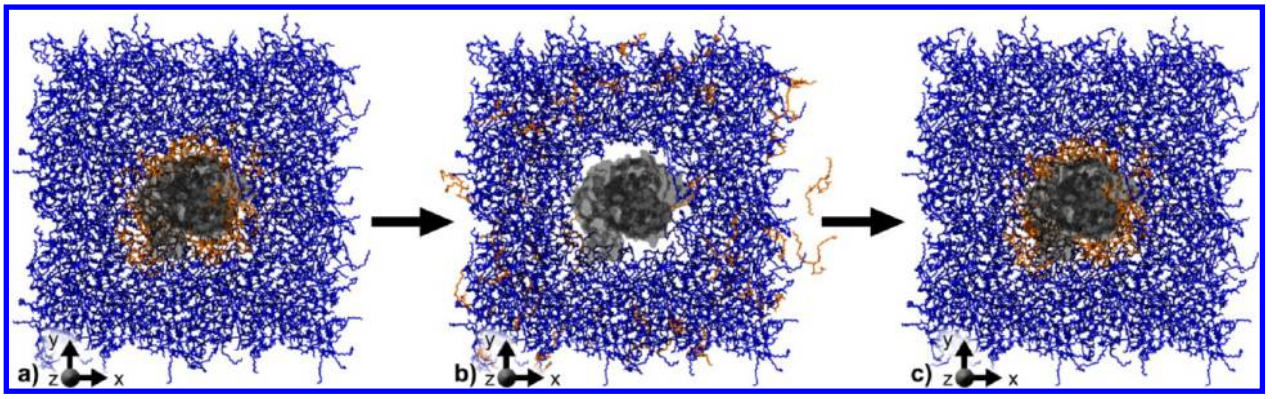


Figure 6. Embedding the protein (gray) into the membrane. InflateGRO2 subjects lipids exhibiting an overlap score > 0 (a, orange) to an initial expansion step around the protein geometrical XY center (b) using an automatically determined scaling factor that is as small as possible. Subsequently these lipids are brought back to natural dimensions in a series of alternation compression and energy minimization steps (c). Lipids exhibiting an overlap score of 0 (blue) are excluded from any interaction with any other system components until the procedure is concluded with a final energy minimization. The InflateGRO2 embedding procedure is illustrated using the lipid A deacylase LpxR as an example.

Following the projection of protein atoms, protein-internal cavities are identified which are subsequently included in the area occupied by the protein. To this end, every cell in each of the two 2D grids that is not occupied by protein is initially defined as cavity, resulting in *true* and *false* cavity cells. To isolate true cavity cells, the algorithm uses an iterative approach of "washing out" false cavity cells employing the following procedure: Starting in the lower left corner (X_{min}, Y_{min}), grid cells are scanned rowwise checking each grid cell's (X_i, Y_j) neighbors in the directions (X_{i-1}, Y_j) and (X_i, Y_{j-1}). If at least one of these neighbors is either *empty* or undefined (lying outside the 2D grid), grid cell (X_i, Y_j) is considered empty (false cavity cell). Starting from each corner of the 2D grid ((X_{max}, Y_{max}), (X_{max}, Y_{min}), (X_{min}, Y_{max})), the process is carried

dx.doi.org/10.1021/ci3000453 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Table 2. InflateGRO2 Test Proteins and Bilayers

protein (PDB-ID)	residues	TM structure	TM cavity	lipid composition and membrane type	calculated scaling factors (expansion, compression)	insertion time[a]
outer membrane protein OmpX (1QJ9)[59]	148	cylindrical (tilted)	no	POPC (n = 128), planar	2.0 0.966	1 min 03 s
outer membrane phospholipase A (1QD5)[30]	257	cylindrical (tilted)	no	POPC (n = 128), planar	2.0 0.966	1 min 25 s
outer membrane deacylase LpxR (3FID)[48]	296	cylindrical (tilted)	no	POPC (n = 128), planar	1.7 0.974	1 min 24 s
potassium channel KcsA (1R3J)[31]	412	conical	no	POPE/POPC (n = 512 /512), planar	2.1 0.964	7 min 00 s
Maltoporin LamB (1AF6)[33]	1263	irregular	yes	POPC (n = 512), planar	1.4 0.983	5 min 20 s
outer membrane efflux duct TolC (1EK9)[29]	1284	conical	yes	POPC (n = 512), planar	1.4 0.983	4 min 32 s
cell toxin α-hemolysin (7AHL)[35]	2051	irregular	yes	POPC (n = 512), planar	2.0 0.966	6 min 08 s
potassium channel KcsA-Fab complex (1R3J)[31]	2136	conical	no	POPC (n = 1152), planar	2.0 0.966	9 min 43 s
potassium channel Kv1.2 (3LUT)[32]	2864	irregular	no	POPC (n = 512), planar	2.3 0.959	8 min 39 s
multidrug efflux pump AcrB (2GIF)[38]	3108	irregular	yes	POPC (n = 512), planar	1.3 0.987	7 min 19 s
complex of AcrAB−TolC[49]	5412	irregular/ conical	yes	POPC/DLPC (n = 512), planar	2.0 (both membranes) 0.966 (both membranes)	58 min
coarse-grained mechanosensitive channel MscL (2OAR)[50]	625	cylindrical	yes	DPPC (n = 877), vesicular	9.5 0.894	13 min 17 s

[a]Performed using a single core on a 4 GB RAM DELL Precision 390 workstation (CPU: Intel Core2 6700 2 × 2.66 GHz).

out four times, eliminating in most cases all false cavity cells (Figure 4b). The neighbor search procedure is repeated until the total cavity area stops changing. If false cavity cells are still present at this point an optional checking of grid cells in (X_{i-1}, Y_{j-1}) direction can be activated.

As evident from space-filling CPK visualizations of the respective protein sections (Figure 4c) and the protein representation used in the initial version of InflateGRO (Figure 4d), the InflateGRO2 approach yields a highly detailed representation of the area occupied by the protein, permitting the exact calculation of both the optimal number of lipids to be deleted and the area per lipid in each leaflet.

To determine which lipids will be removed, they are ranked based on their protein overlap computing for each lipid its individual overlap score (OLS) (Figure 5). To this end, the Z range, defined by the protein transmembrane region, is divided into 0.5 nm thick slices (Figure 5a). For each slice the contained protein and lipid atoms are projected on a 2D grid as described above in order to determine the number of clashing grid cells occupied by lipid and protein or lipid and cavity (Figure 5b). A lipid's OLS is the amount of co-occupied grid cells summed over all Z slices and permits ranking all lipids in the system based on their protein overlap (Figure 5c). The number of top-ranking lipids corresponding to the previously calculated number of lipids to be deleted is then removed from the system.

To embed the protein into the membrane, lipids exhibiting an OLS > 0 (Figure 6a, orange) are subjected to an initial expansion step around the geometrical XY center of the protein using an automatically determined scaling factor that is as small as possible (Figure 6b). The scaling factor is determined in an iterative process of applying an initial scaling factor of 1.1 and checking for any protein−lipid overlapping grid cells. If no co-occupied grid cells are found, the scaling factor is used else the process is repeated, incrementally increasing the scaling factor by 0.1 until no clashes are found. After the expansion step, the affected lipids are brought back to natural dimensions in a series of 20 alternating steps of compressive XY lipid translation and energy minimization using a corresponding and automatically determined scaling factor (Figure 6c). During the entire process, lipids with an OLS = 0 (Figure 6, blue) are excluded from any interactions with any other system components until the procedure is concluded with a final energy minimization.

The InflateGRO method differentiates two types of components in a simulation system: mobile parts affected by lateral translation (usually lipids) and immobile parts unaffected by lateral translation (usually protein). In InflateGRO2 component identification is based on GROMACS NDX index files[24,27] which represent an easy and interactive means to quickly generate user-defined combinations of residues or atoms.

To assess InflateGRO2 performance and demonstrate the improvements compared to the preceding version, we have performed test runs inserting 12 membrane proteins of different TM structure into single or multiple phospholipid membrane patches of uniform or heterogeneous lipid composition (Table 2). During the compression phase, steepest descent energy minimization was performed using GROMACS 4.0[24,27,28] employing a maximum of 80 integration steps, a force tolerance of 1000 kJ mol^{-1} nm^{-1}, and an initial step size of 0.01 nm. During minimization protein atoms and nonoverlapping lipids are defined as GROMACS freeze-groups and do not change their conformations while the calculation of molecular interactions is restricted to overlapping lipids among themselves and toward the protein. Electrostatic interactions between the protein and lipids subjected to lateral translation

dx.doi.org/10.1021/ci3000453 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

were calculated using a simple distance cutoff of 1 nm, whereas van der Waals interactions were computed using twin range cut-offs of 1 and 1.4 nm. Except for the final energy minimization concluding the insertion process and taking into account all system components, interactions between translated lipids and unmodified lipids were excluded. InflateGRO2 is licensed under the GNU GPL v2 and along with application examples and tutorials available free of charge at http://code.google.com/p/inflategro2/ or http://www.csb.bit.uni-bonn.de/inflategro.html

RESULTS AND DISCUSSION

Compared to water-soluble proteins, molecular dynamics simulations of membrane proteins require two additional working steps to build a suitable starting structure: the alignment of protein and membrane followed by the seamless accommodation of the protein into the membrane. To automate these steps, we have developed two fast and complementary tools based on pre-equilibrated membrane patches whose membrane normal coincides with the *Z*-axis. Whereas LAMBADA automatically aligns protein and membrane coordinates independently from any precalculated data, InflateGRO2 automatically performs the membrane embedding process with minimal disturbance of the membrane structure. Both tools can be used separately, in combination with other methods, or in tandem permitting a fully automated workflow while retaining a maximum level of usage control and flexibility.

LAMBADA: Automatic Membrane/Protein Alignment. Starting from protein and planar membrane 3D coordinates in the GROMACS GRO file format, LAMBADA aligns membrane and protein by automatically finding a protein orientation where the hydrophobic belt is parallel to the *XY* plane, which is assumed to coincide with the membrane plane of the input bilayer, followed by translating the membrane patch in *Z*-direction to achieve the best match between HB and the hydrophobic section of the bilayer. To this end, the algorithm employs an entirely geometrical approach where, using a voxel-based cast of the coarse-grained protein, hydrophilicity profiles along the *Z*-axis are computed to determine the least hydrophilic, membrane-thick section of the protein surface (Figure 1). Quantified in a single numerical parameter, the resulting LAMBADA score (LS) is a function of the protein orientation and becomes minimal when the HB runs parallel to the *XY* plane (Figure 2). As evident from Figure 2a, a membrane protein exhibits several, equivalent LS minima due to protein configurations differing in their rotation angles around the membrane normal which can (as in, e.g., the outer membrane channel TolC;[29] Figure 7) but does not have to coincide with the protein's first principal axis (as in phospholipase A[30] shown in Figures 2 and 3). Minimizing LS in a recursive optimization procedure then yields the correct protein orientation (Figure 3). The alignment process is completed by determining and applying the optimal *Z* translation vector for the membrane matching the detected HB position (Figure 1d).

Performance. To assess LAMBADA performance, we carried out example applications using a set of 19 membrane proteins representing different types of cylindrical, conical, or irregularly shaped transmembrane sections in titled or straight orientation, with and without transmembrane cavities (Table 1 and Supporting Information Table 1, columns 1−4). As summarized by postalignment snapshots (Figure 7, upper row, Supporting Information Figure 1) and corresponding hydrophilicity profiles (Figure 7, lower row), LAMBADA determines for each protein the optimal orientation and correctly aligns hydrophobic belt (Figure 7 and Supporting Information Figure 1, yellow protein atoms) with the hydrophobic section of the membrane (Figure 7 and Supporting Information Figure 1, gray sticks). To assess the quality of the predicted LAMBADA orientations, we compared our findings to the OPM database and computed the difference in membrane orientation after superimposing protein coordinates (Table 1 and Supporting Information Table 1, column 7). Whereas for ten proteins the orientation differences range from 0° (KcsA-Fab,[31] Kv1.2,[32] maltoporin LamB,[33] and bacteriorhodopsin[34]) up to 3.1° (α-hemolysin[35]), in nine cases orientation differences were observed ranging from 6° (P-glycoprotein[36]) up to 21° (M2 proton channel[37]). However, as evident from visual inspection and postalignment snapshots comparing the LAMBADA and OPM output structures and highlighting the HB residue distribution, the LAMBADA predictions yield a better (Supporting Information Figure 1 2RLF; Supporting Information Figure 2a and b) or equally valid match of HB and the membrane (Supporting Information Figure 1: 2GZM, 3V3Y, 3G5U, 2QI9, 2R6G; Supporting Information Figure 2c).

Using a single core of an Intel Core2 CPU 6700 (2 × 2.66 GHz) on a 4 GB RAM DELL Precision 390 workstation, processing times for the example proteins range from 36 s for the 148 residue outer membrane protein OmpX up to 23 min 56 s for the 6729 residue photosystem I (Table 1 and Supporting Information Table 1, column 6). The 10 min longer processing time for the 2136 residue ion channel KcsA-Fab comprising 1000 residues less than the multidrug efflux pump AcrB[38] (9 min 20 s) has two reasons. For one, due to its 3D structure KcsA-Fab exhibits a large surface area which is a main factor determining processing time. Second, for this protein, LAMBADA found two minima in the scoring function, displaying identical LS values. Considering both configurations in the optimization procedure leads to a further increase of the processing time. When LAMBADA's orientation search is deactivated and the proteins are preoriented, processing times range from 2 s for OmpX up to 34 s for photosystem I (Table 1 and Supporting Information Table 2, column 5).

Comparison and Applications. Like all previously published methods for HB detection or the prediction of membrane protein orientation,[6−10] LAMBADA is also based on a membrane protein's characteristic hydrophobic belt distribution of charged and uncharged surface residues[4,5] and employs a scoring function to align membrane and protein. However, the composition of the scoring function is different in LAMBADA in several regards. Instead of using multiclass hydrophobicity scales differentiating between several categories of amino acids,[39−43] LAMBADA follows a strategy similar to TMDET[9] employing a simple binary charged or neutral hydrophilicity scale considering only Asp, Glu, Lys, Arg, and His residues as charged but nevertheless yielding sharp results for all proteins tested (Figure 1 and 7). Instead of resorting to molecular surfaces,[9] projecting residues on a cylinder[10] or using force field-based criteria,[6−8] surface residues are identified employing a voxel-based cast of the coarse-grained protein (Figure 1a−c). Next to a high level of accuracy, this approach also bears the advantage of explicitly taking protein-internal cavities into account, while coarse-graining the protein offers increased program performance and poses an alternative to the restriction to ß carbons as implemented in e.g. GARLIC[10] to minimize an

dx.doi.org/10.1021/ci3000453 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

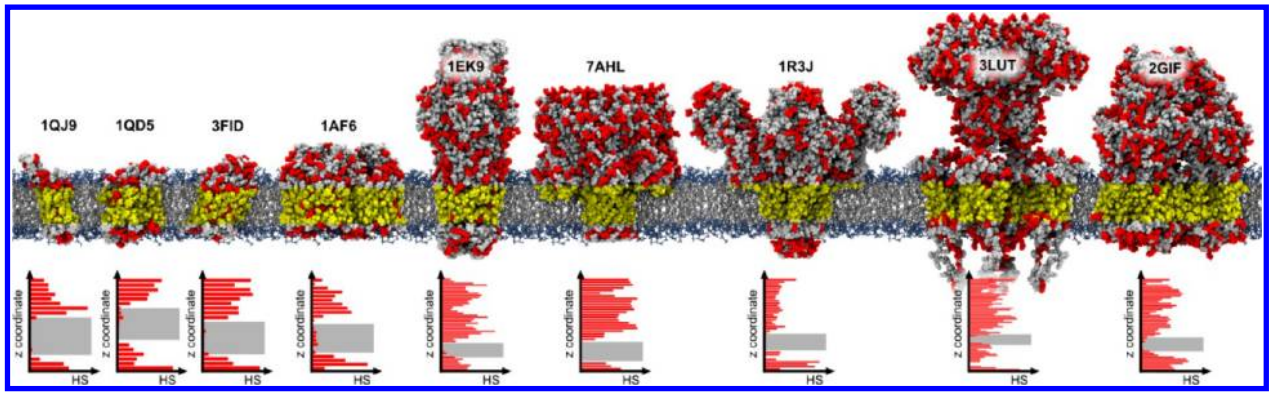


Figure 7. LAMBADA output of nine test proteins representing different TM shapes and structures in an aligned POPC bilayer. Charged residues are highlighted as red spheres; lipid head groups are colored in blue. The second row shows the corresponding hydrophilicity score profiles along the *z*-axis of each simulation box. The hydrophobic belt detected for each protein is highlighted in yellow (upper row) and gray (lower row).

individual side chain's influence on the overall distribution. LAMBADA's sliding *Z* window approach to determine the least hydrophilic, membrane-thick section of the protein surface (Figure 1d) permits the option of either using a standard value for membrane thickness as in refs 6−9 or determining the membrane thickness automatically from the bilayer input structure. Moreover, LAMBADA's recursive procedure of minimizing the scoring function based on rotating the protein's first principal axis (Figure 3) requires significantly less computation steps than systematically testing out all possible rotation angle orientations as in TMDET[9] or generating an ensemble of Monte Carlo conformations as in IMPALA.[8]

As shown in Figure 7, Table 1, Supporting Information Figure 1, and Table 1, LAMBADA correctly aligned each of the 19 representative test proteins with the POPC bilayer patch used in our tests. One might argue whether 19 examples is a sufficient number of test cases. However, while it is always desirable to have as many test cases as possible, the number of possible cases is naturally limited by the method of validation applied. As an orientation comparison between LAMBADA and OPM predictions alone is insufficient to judge the quality of the membrane/protein alignment (Table 1, Supporting Information Table 1) an additional visual inspection/molecular graphics rendition of the postalignment structures is required, highlighting the position of the membrane and the protein's hydrophobic belt (Figure 7, Supporting Information Figure 1 and 2). Assuming an average of ten proteins per figure, testing all 281 currently known unique membrane protein structures[44,45] would require a total amount of 28 figures which we consider lying outside the scope of a paper. Furthermore, given that our test proteins cover a wide range of membrane protein sizes and trans-membrane shapes, the quality of the LAMBADA predictions is encouraging, particularly as our orientations either coincide with the predictions of the OPM database or provide a better orientation in three cases including the outer membrane protein OmpX, whose hydrophobic belt indicates a tilted in-membrane orientation (Supporting Information Figure 2a), and the M2 proton channel whose hydrophobic belt and 4-fold symmetry indicate an orientation with the protein's first principal axis running parallel to the membrane normal (Supporting Information Figure 1 2RLF). Future applications to other proteins will show how representative these findings are.

All methods currently available for membrane/protein alignment methods require at least partial manual intervention[6−10] or additional input precalculated by another application.[11,12] LAMBADA is the first tool to perform the entire alignment process automatically and providing as output the explicit 3D coordinates of the aligned protein and membrane instead of only indicating the predicted HB position.[6−10] As long as planar membrane patches are used, the algorithm's independence from any precalculated data permits the unrestricted processing of any membrane and protein structure the user provides—including self-built bilayer patches, protein homology models, or newly released crystallographic structures not yet incorporated in for example the OPM database. LAMBADA is also currently the only membrane/protein alignment tool permitting a seamless integration into the simulation setup workflow of the GROMACS MD simulation package.[24,27,28] As a command line tool executed by a single command requiring either two or three input files LAMBADA can be easily scripted allowing on one hand the automated generation of starting structures for MD simulations of membrane proteins when coupled to insertion tools such as g_membed,[22] GRIFFIN,[21] InflateGRO,[13] or InflateGRO2. On the other hand LAMBADA can also be employed as structural analysis tool quantifying the distribution of charged surface residues while the characteristic differences in LAMBADA score distribution between membrane (Figure 2a) and water-soluble proteins (Figure 2b) bear application potential in scanning structural databases and automatically determining the protein type.

Although in its current version LAMBADA is restricted to the format of GROMACS structure and index files, output files can be easily converted to other file formats making our method applicable to other MD packages as well. Like any other modeling technique, the results of LAMBADA membrane/protein alignments should be checked carefully before proceeding in the simulation setup. The LAMBADA package, inclusive example applications and tutorials, is available free of charge at http://www.csb.bit.uni-bonn.de/lambada.html or http://code.google.com/p/lambada-align/

InflateGRO2: Automatic Protein Insertion. InflateGRO2 automatically embeds membrane proteins into lipid bilayer patches by first calculating for each leaflet the area occupied by the protein (Figure 4). Ranking all lipids in the system based on their protein overlap (Figure 5), the number of top-ranking lipids matching the previously calculated area per protein is determined and removed from the system. To minimize disturbance of the membrane structure and retain a maximum

dx.doi.org/10.1021/ci3000453 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

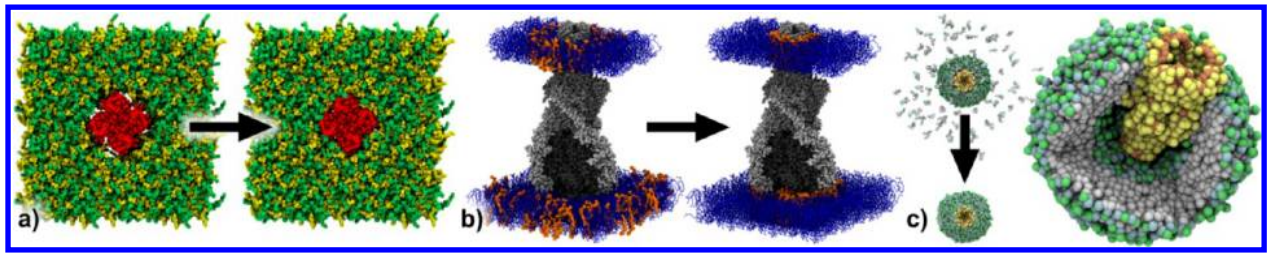


Figure 8. InflateGRO2 application flexibility. Preparation of heterogeneous lipid systems. Using GROMACS index files, InflateGRO2 permits the easy and straightforward preparation of systems of heterogeneous lipid composition (a); the embedding of proteins spanning more than one membrane (b); and the preparation of starting structures for coarse-grained simulations in the MARTINI format (c). InflateGRO2 application flexibility is illustrated using the potassium channel KcsA (a), the multidrug efflux pump AcrAB-TolC (b), and the mechanosensitive channel MscL (c) as examples.

of equilibrium lipid conformations, only lipids still overlapping with the protein are subsequently subjected to lateral translation in an initial expansion step followed by a series of alternating compression and energy minimization step until the embedding process is complete (Figure 6). During that process the scaling factors for extension and compression are determined automatically to ensure lipid exposure to vacuum is reduced to a minimum. Embedded in an automated workflow, InflateGRO2 offers improved accuracy and enhanced application flexibility addressing the shortcomings of its predecessor[13] as discussed below.

Changes and Improvements. Unless the ideal combination of expansion factor and distance cutoff is determined manually for each system, the initial version's approach of lipid deletion using a protein α-carbon−lipid phosphorus distance cutoff criterion after membrane expansion can lead to local inhomogeneities in lipid packing such as high lipid densities near the protein but lower densities at the periphery of the membrane, even though the overall area per lipid matches equilibrium values. Furthermore, as the distance cutoff is proportional to the circularity of the hole created by lipid deletion[3] the lack of standardized distance cutoffs can lead to an inaccurate representation of the protein's TM-shape which in turn also affects lipid packing.

Replacing the initial version's distance cutoff-based and scaling factor-dependent lipid deletion with determining and removing the exact number of overlapping lipids matching the area occupied by the protein (Figures 4 and 5), InflateGRO2 circumvents the problem of inhomogeneous lipid packing and imprecise protein shape representation in a procedure that is both more accurate and standardized requiring no manual testing of parameter combinations.

To maximize accuracy in detecting the protein's TM-shape, the underlying algorithm computing the area occupied by the protein has been rewritten, introducing three major modifications: (1) Reflecting the physicochemical properties of lipid bilayers for which a heterogeneous distribution of lipid density and rigidity along the membrane normal is characteristic (Figure 4a, right), area per protein calculation is now restricted to protein atoms located in the lipid headgroup regions (Figure 4a, left) instead of using both halves of the entire transmembrane *Z*-range[13] including protein atoms interacting with the flexible lipid tails. Nevertheless the protein's transmembrane section is still considered in its entirety when calculating each lipid's protein overlap (Figure 5). (2) The arbitrary, user-defined resolution of the 2D grid on which transmembrane protein atoms are projected is replaced with a standard resolution of 0.01 nm grid size, which in combination with (3) the explicit consideration of the previously neglected atomic van der Waals radii and protein-internal cavities (Figure 4b) leads to an improved representation of the protein shape and the area occupied by the protein as evident from comparing InflateGRO2 results for α-hemolysin[35] with space filling CPK visualizations (Figure 4c) and corresponding representations in InflateGRO1 (Figure 4d). Furthermore, beyond yielding a more accurate area per lipid calculation, the improved transmembrane shape representation and slice-based calculation of the area occupied by protein could also be employed for computing protein cross-sectional areas, complementing the three currently available methods based on single conformations[11] or trajectories.[46,47]

The InflateGRO method of subjecting a membrane to expansion and alternating steps of compression and energy minimization can lead to a loss of equilibrium lipid conformations as entwined lipids become untangled through the initial expansion, and energy minimizing vacuum exposed lipids eventually results in a gel phase-like straightening of the lipid tails.[22] Restricting expansion and compression to a subset of relevant lipids using automatically computed system-optimized scaling factors (Figures 5 and 6), InflateGRO2 reduces lipid vacuum exposure to a minimum preserving as much as possible of the equilibrated membrane structure. As illustrated by the lipid A deacylase LpxR,[48] of the lipids still present after the initial lipid removal only those exhibiting an overlap score > 0 are expanded around the protein's geometrical TM center (Figure 6a) while nonoverlapping lipids remain unchanged and their interactions with other system components deactivated during the compression phase (Figure 6b) until a final systemwide energy minimization process completes the embedding process (Figure 6c). To avoid clashes between mobile and stationary lipids after completing the compression phase, we found that automatically computing a system-optimized compression factor ensures every single lipid exactly reaches its initial position in a number of predefined steps (default is 20), giving good results and working without problems for each of the systems tested (Table 2).

Using a residue name criterion to identify mobile and immobile components affected (usually lipids) and unaffected by lateral translation (usually protein), application flexibility is limited in the initial version of InflateGRO. For example, preparing lipid mixtures is laborious, requiring workarounds such as renaming lipids between compression (all lipids share the same name) and energy minimization steps (each lipid species has its own name), whereas protein insertion into multiple membranes as required for e.g. multidrug efflux pumps

dx.doi.org/10.1021/ci3000453 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

of the resistance nodulation division family[23] is not possible. Addressing these restrictions, InflateGRO2 employs GROMACS NDX index files[24] to define mobile and immobile system components. Representing an easy and interactive means to generate any user-defined combinations of atoms, NDX files permit the easy preparation of lipid mixtures as exemplified with the KcsA potassium ion channel[31] inserted in a 1:1 POPC/POPE bilayer (Figure 8a) or embedding proteins spanning multiple membranes as illustrated by the AcrAB-TolC multidrug efflux pump[49] inserted into DLPC and POPC bilayers (Figure 8b). Furthermore the initial version's restriction to atomistic MD is now expanded to coarse-grained structures in the MARTINI format as demonstrated by the mechanosensitive channel MscL[50] embedded in a DPPC vesicle[51] (Figure 8c). While protein insertion works well for curved bilayers as long as the protein's transmembrane vector coincides with the *Z* axis, the user should be aware that the area per lipid values reported will be wrong due to InflateGRO2's approach of detecting upper and lower leaflet by comparing each lipid's *Z* coordinate of a representative headgroup atom (phosphorus by default) to the average *Z* position of these atoms.

Performance and Comparison. To assess InflateGRO2 performance, we carried out test runs on a set of twelve membrane proteins of different sizes and TM shape which we inserted into single and multiple planar or vesicular bilayers of homogeneous or heterogeneous lipid composition using a single core of an Intel Core2 CPU 6700 (2 × 2.66 GHz) on a 4 GB RAM DELL Precision 390 workstation. Except for the coarse-grained ion channel MscL inserted in a vesicle and the multi drug efflux pump complex AcrAB-TolC inserted in a double membrane, all preceding protein/membrane alignments were carried out using LAMBADA. With processing times ranging from 1 min 3 s for the 148 residue outer membrane protein OmpX up to 58 min for the 5412 AcrAB-TolC complex (Table 2), InflateGRO2 performance is faster than its predecessor[3] yielding excellent embedding results for each protein well-illustrated by four proteins of cylindrical (Figures 6 and 8c), conical (Figure 8a and b), and irregular transmembrane structure (Figure 8c) displaying straight (Figure 8) or tilted membrane orientation (Figure 6), embedded into homogeneous (Figures 6 and 8b) and heterogeneous bilayers (Figure 8a) and a double membrane (Figure 8b) as well as a coarse-grained vesicle (Figure 8c).

As recently reviewed in ref 3, of the currently available and pre-equilibrated membrane patch-based insertion methods, GRIFFIN[21] offers the highest accuracy while exhibiting the steepest learning curve using repulsive forces to drive out lipids from the protein volume, g_membed's[22] protein-compressing and expanding approach is the fastest method currently available, providing the highest levels of application flexibility and accessibility, whereas the CHARMM-GUI membrane builder[11] offers a maximum degree of automation but is restricted to two membrane sizes and proteins with circular transmembrane cross sections when pre-equilibrated membrane patches are to be used. Building on and extending its predecessor, InflateGRO2 bundles the strengths of currently available protein insertion methods, complementing existing tools with a combination of properties not available before. Through the implemented procedures of protein shape representation and cavity detection, overlapping lipid detection and removal as well as minimum membrane structure disturbance, InflateGRO2 offers a GRIFFIN-like level of accuracy. Focusing on relevant lipids only and computing the area per lipid/protein only once, InflateGRO2 reaches a performance level comparable to that of g_membed with whom it also shares a similar level of application flexibility. Automatically calculating its operation parameters, InflateGRO2 approaches a level of automation similar to the CHARMM GUI membrane builder without posing any restrictions on the protein and membrane patch input structures. However, the user should be aware that unlike GRIFFIN the current version of InflateGRO2 cannot automatically handle proteins with lipid-filled TM cavities. Where this task arises, lipids have to be placed manually[52] and defined alongside the protein as one index group. Furthermore the concern raised in the g_membed publication that the InflateGRO method always includes a loss of membrane hydration is also valid for the new version. However, given that the diffusion rate of water[53] is approximately ten times higher than that of lipids[54] and hydrophobic regions of the bilayer can easily be excluded from water placement during resolvation,[13] we consider the consequences of lipid hydration loss during protein insertion negligible for lipid structure especially since InflateGRO2 now minimizes membrane structure disturbance as discussed above. Beyond that, in compressing the protein structure g_membed too creates local vacuum conditions which can lead to lipid dehydration or water intrusion between the compressed protein and membrane after initial lipid removal.

As a single, platform-independent PERL executable InflateGRO2's fully automatized workflow ensures high level reproducibility offering a high degree of accessibility and usability. In combination with LAMBADA, InflateGRO2 permits the full automation of the entire simulation set up procedure of membrane protein insertion starting from separate protein and bilayer patch structure files yielding a single structure file holding the protein in aligned HB orientation seamlessly accommodated in the membrane and ready for salvation and simulation initiation. Although like g_membed, InflateGRO2 is primarily written with the GROMACS package in mind; due to the improved modularized structure of its source code, which we release under the GNU GPL v2, InflateGRO2 can be easily adapted to other MD engines such as NAMD[55,56] or CHARMM.[57,58] Like any other modeling technique, the results of InflateGRO2 membrane protein insertion should be checked carefully before proceeding in the simulation setup. InflateGRO2 along with example applications and usage tutorials is available free of charge at http://www.csb.bit.uni-bonn.de/lambada.html or http://code.google.com/p/inflategro2/.

■ CONCLUSIONS

We have developed two, efficient, complementary, and easy to use tools for building starting structures for molecular dynamics simulations of membrane proteins, automating the working steps of membrane/protein alignment and seamless protein accommodation in the membrane. Independent from any precalculated data, LAMBADA automatically determines the correct protein orientation and membrane position, outputting in a single file the explicit 3D coordinates of the aligned protein and membrane patch. LAMBADA's underlying scoring function bears also further application potential in structural analysis and automatic distinction between membrane and water-soluble proteins. Addressing the shortcomings of its predecessor, InflateGRO2 performs the protein embedding process in an entirely automated workflow offering optimal lipid packing,

dx.doi.org/10.1021/ci3000453 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

maximum preservation of the membrane's equilibrium lipid conformations as well as a high level of application flexibility that permits the easy preparation of, e.g. lipid mixtures, coarse-grained structures, or protein spanning multiple membranes. Both tools can be used separately, in combination with other methods, or in tandem permitting a fully automated workflow while retaining a maximum level of usage control and flexibility.

■ ASSOCIATED CONTENT

Supporting Information

One table and two figures describing the results of ten additional LAMBADA test proteins and comparing the LAMBADA predictions with the corresponding entries in the OMP database. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: kandt@bit.uni-bonn.de. Phone: 49 228 2699 324. Fax: 49 228 2699 341.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Martin Raunest and Nadine Fischer as well as Christian Czekalski and Dr. Frank Wennmohs for proof reading and helpful comments. This work was financially supported by the Ministerium für Innovation, Wissenschaft und Forschung des Landes Nordrhein-Westfalen. C.K. is a junior research group leader funded by the NRW Rückkehrerprogramm.

■ ABBREVIATIONS

CG, coarse-grain; HB, hydrophobic belt; HS, hydrophilicity score; GLD, distance between the geometrical centers of the bilayer leaflets; LS, LAMBADA score; OLS, overlap score; TM, transmembrane

■ REFERENCES

(1) Kandt, C.; Matyus, E.; Tieleman, D. P. Protein Lipid Interactions from a Molecular Dynamics Simulation Point of View. In *Structure & Dynamics of Membranous Interfaces*; Nag, K., Ed.; Wiley: Hoboken, NJ, 2008; pp 267−282.

(2) Kandt, C.; Monticelli, L. Membrane protein dynamics from femtoseconds to seconds. *Methods Mol. Biol.* **2010**, *654*, 423−440.

(3) Schmidt, T. H.; O'Mara, M. L.; Kandt, C. Molecular Dynamics Simulations of Membrane Proteins: Building Starting Structures and Example Applications. *Curr. Phys. Chem.* **2012**, *2* (4), 363−378.

(4) Rees, D. C.; DeAntonio, L.; Eisenberg, D. Hydrophobic organization of membrane proteins. *Science* **1989**, *245* (4917), 510−513.

(5) Wallin, E.; Tsukihara, T.; Yoshikawa, S.; von Heijne, G.; Elofsson, A. Architecture of helix bundle membrane proteins: an analysis of cytochrome c oxidase from bovine mitochondria. *Protein Sci.* **1997**, *6* (4), 808−815.

(6) Lomize, M. A.; Lomize, A. L.; Pogozheva, I. D.; Mosberg, H. I. OPM: Orientations of proteins in membranes database. *Bioinformatics* **2006**, *22* (5), 623−625.

(7) Lomize, A. L.; Pogozheva, I. D.; Lomize, M. A.; Mosberg, H. I. Positioning of proteins in membranes: A computational approach. *Protein Sci.* **2006**, *15* (6), 1318−1333.

(8) Basyn, F.; Charloteaux, B.; Thomas, A.; Brasseur, R. Prediction of membrane protein orientation in lipid bilayers: a theoretical approach. *J. Mol. Graphics Modell.* **2001**, *20* (3), 235−244.

(9) Tusnady, G. E.; Dosztanyi, Z.; Simon, I. Transmembrane proteins in the Protein Data Bank: identification and classification. *Bioinformatics* **2004**, *20* (17), 2964−2972.

(10) Zucic, D.; Juretic, D. Precise annotation of transmembrane segments with garlic - a free molecular visualization program. *Croat. Chem. Acta* **2004**, *77* (1−2), 397−401.

(11) Jo, S.; Kim, T.; Im, W. Automated Builder and Database of Protein/Membrane Complexes for Molecular Dynamics Simulations. *PLoS One* **2007**, *2* (9), e880.

(12) Jo, S.; Lim, J. B.; Klauda, J. B.; Im, W. CHARMM-GUI Membrane Builder for Mixed Bilayers and Its Application to Yeast Membranes. *Biophys. J.* **2009**, *97* (1), 50−58.

(13) Kandt, C.; Ash, W.; Tieleman, P. Setting up and running molecular dynamics simulations of membrane proteins. *Methods* **2007**, *41* (4), 475−488.

(14) Woolf, T. B.; Roux, B. Molecular-Dynamics Simulation of the Gramicidin Channel in a Phospholipid-Bilayer. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91* (24), 11631−11635.

(15) Woolf, T. B.; Roux, B. Structure, energetics, and dynamics of lipid-protein interactions: A molecular dynamics study of the gramicidin A channel in a DMPC bilayer. *Proteins: Struct., Funct., Genet.* **1996**, *24* (1), 92−114.

(16) Scott, K.; Bond, P.; Ivetac, A.; Chetwynd, A.; Khalid, S.; Sansom, M. Coarse-Grained MD Simulations of Membrane Protein-Bilayer Self-Assembly. *Structure* **2008**, *16* (4), 621−630.

(17) Sansom, M. S. P.; Scott, K. A.; Bond, P. J. Coarse-grained simulation: a high-throughput computational approach to membrane proteins. *Biochem. Soc. Trans.* **2008**, *36*, 27−32.

(18) Bond, P. J.; Holyoake, J.; Ivetac, A.; Khalid, S.; Sansom, M. S. P. Coarse-grained molecular dynamics simulations of membrane proteins and peptides. *J. Struct. Biol.* **2007**, *157* (3), 593−605.

(19) Shen, L. Y.; Bassolino, D.; Stouch, T. Transmembrane helix structure, dynamics, and interactions: Multi-nanosecond molecular dynamics simulations. *Biophys. J.* **1997**, *73* (1), 3−20.

(20) Faraldo-Gomez, J. D.; Smith, G. R.; Sansom, M. S. P. Setting up and optimization of membrane protein simulations. *Eur. Biophys. J.* **2002**, *31* (3), 217−227.

(21) Forrest, L. R.; Staritzbichler, R. S., R.; Anselmi, C.; Faraldo-Gomez, J. D. GRIFFIN: A Versatile Methodology for Optimization of Protein-Lipid Interfaces for Membrane Protein Simulations. *J. Chem. Theory Comput.* **2011**, *7* (4), 1167−1176.

(22) Wolf, M. G.; Hoefling, M.; Aponte-Santamaria, C.; Grubmuller, H.; Groenhof, G. g_membed: Efficient Insertion of a Membrane Protein into an Equilibrated Lipid Bilayer with Minimal Perturbation. *J. Comput. Chem.* **2010**, *31* (11), 2169−2174.

(23) Saier, M. H.; Paulsen, I. T. Phylogeny of multidrug transporters. *Semin. Cell Dev. Biol.* **2001**, *12* (3), 205−213.

(24) Van der Spoel, D.; Lindahl, E.; Hess, B.; Van Buuren, A. R.; Apol, E.; Meulenhoff, P. J.; Tieleman, D. P.; Sijbers, A. L. T. M.; Feenstra, K. A.; Van Drunen, R.; Berendsen, H. J. C. *Gromacs User Manual*, version 4.5; www.gromacs.org (accessed 2010).

(25) Poger, D.; Van Gunsteren, W. F.; Mark, A. E. A New Force Field for Simulating Phosphatidylcholine Bilayers. *J. Comput. Chem.* **2010**, *31* (6), 1117−1125.

(26) Poger, D.; Mark, A. E. On the Validation of Molecular Dynamics Simulations of Saturated and cis-Monounsaturated Phosphatidylcholine Lipid Bilayers: A Comparison with Experiment. *J. Chem. Theory Comput.* **2010**, *6* (1), 325−336.

(27) Van der Spoel, D.; Lindahl, E.; Hess, B.; Van Buuren, A. R.; Apol, E.; Meulenhoff, P. J.; Tieleman, D. P.; Sijbers, A. L. T. M.; Feenstra, K. A.; Van Drunen, R.; Berendsen, H. J. C. *Gromacs User Manual*, version 4.0; www.gromacs.org (accessed 2005).

(28) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008**, *4* (3), 435−447.

(29) Koronakis, V.; Sharff, A.; Koronakis, E.; Luisi, B.; Hughes, C. Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. *Nature* **2000**, *405* (6789), 914−919.

(30) Snijder, H. J.; Ubarretxena-Belandia, I.; Blaauw, M.; Kalk, K. H.; Verheij, H. M.; Egmond, M. R.; Dekker, N.; Dijkstra, B. W. Structural evidence for dimerization-regulated activation of an integral membrane phospholipase. *Nature* **1999**, *401* (6754), 717−721.
(31) Zhou, Y.; MacKinnon, R. The occupancy of ions in the K+ selectivity filter: charge balance and coupling of ion binding to a protein conformational change underlie high conduction rates. *J. Mol. Biol.* **2003**, *333* (5), 965−975.
(32) Ma, J. P.; Chen, X. R.; Wang, Q. H.; Ni, F. Y. Structure of the full-length Shaker potassium channel Kv1.2 by normal-mode-based X-ray crystallographic refinement. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107* (25), 11352−11357.
(33) Wang, Y. F.; Dutzler, R.; Rizkallah, P. J.; Rosenbusch, J. P.; Schirmer, T. Channel specificity: structural basis for sugar discrimination and differential flux rates in maltoporin. *J. Mol. Biol.* **1997**, *272* (1), 56−63.
(34) Belrhali, H.; Nollert, P.; Royant, A.; Menzel, C.; Rosenbusch, J. P.; Landau, E. M.; Pebay-Peyroula, E. Protein, lipid and water organization in bacteriorhodopsin crystals: a molecular view of the purple membrana at 1.9 angstrom resolution. *Structure* **1999**, *7* (8), 909−917.
(35) Song, L. Z.; Hobaugh, M. R.; Shustak, C.; Cheley, S.; Bayley, H.; Gouaux, J. E. Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore. *Science* **1996**, *274* (5294), 1859−1866.
(36) Jordan, P.; Fromme, P.; Witt, H. T.; Klukas, O.; Saenger, W.; Krauss, N. Three-dimensional structure of cyanobacterial photosystem I at 2.5 angstrom resolution. *Nature* **2001**, *411* (6840), 909−917.
(37) Schnell, J. R.; Chou, J. J. Structure and mechanism of the M2 proton channel of influenza A virus. *Nature* **2008**, *451* (7178), 591−595.
(38) Seeger, M. A.; Schiefner, A.; Eicher, T.; Verrey, F.; Diederichs, K.; Pos, K. M. Structural asymmetry of AcrB trimer suggests a peristaltic pump mechanism. *Science* **2006**, *313* (5791), 1295−1298.
(39) Eisenberg, D.; Schwarz, E.; Komaromy, M.; Wall, R. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J. Mol. Biol.* **1984**, *179* (1), 125−142.
(40) Hessa, T.; Meindl-Beinker, N. M.; Bernsel, A.; Kim, H.; Sato, Y.; Lerch-Bader, M.; Nilsson, I.; White, S. H.; von Heijne, G. Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature* **2007**, *450* (7172), 1026−1030.
(41) Jayasinghe, S.; Hristova, K.; White, S. H. Energetics, stability, and prediction of transmembrane helices. *J. Mol. Biol.* **2001**, *312* (5), 927−934.
(42) Kyte, J.; Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **1982**, *157* (1), 105−132.
(43) Wimley, W. C.; Creamer, T. P.; White, S. H. Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides. *Biochemistry* **1996**, *35* (16), 5109−5124.
(44) White, S. H. The progress of membrane protein structure determination. *Protein Sci.* **2004**, *13* (7), 1948−1949.
(45) White, S. H. Biophysical dissection of membrane proteins. *Nature* **2009**, *459* (7245), 344−346.
(46) Raunest, M.; Kandt, C. dxTuber: Detecting Protein Cavities, Tunnels & Clefts Based on Protein and Solvent Dynamics. *J. Mol. Graphics Modell.* **2011**, *29* (7), 895−905.
(47) Samuli Ollila, O. H.; Louhivuori, M.; Marrink, S. J.; Vattulainen, I. Protein shape change has a major effect on the gating energy of a mechanosensitive channel. *Biophys. J.* **2011**, *100* (7), 1651−1659.
(48) Rutten, L.; Mannie, J. P. B. A.; Stead, C. M.; Raetz, C. R. H.; Reynolds, C. M.; Bonvin, A. M. J. J.; Tommassen, J. P.; Egmond, M. R.; Trent, M. S.; Gros, P. Active-site architecture and catalytic mechanism of the lipid A deacylase LpxR of Salmonella typhimurium. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106* (6), 1960−1964.
(49) Symmons, M. F.; Bokma, E.; Koronakis, E.; Hughes, C.; Koronakis, V. The assembled structure of a complete tripartite bacterial multidrug efflux pump. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106* (17), 7173−7178.
(50) Yefimov, S.; van der Giessen, E.; Onck, P. R.; Marrink, S. J. Mechanosensitive Membrane Channels in Action. *Biophys. J.* **2008**, *94* (8), 2994−3002.
(51) Marrink, S. J.; Mark, A. E. Molecular dynamics simulation of the formation, structure, and dynamics of small phospholipid vesicles. *J. Am. Chem. Soc.* **2003**, *125* (49), 15233−15242.
(52) Fischer, N.; Kandt, C. Three ways in, one way out: Water dynamics in the trans-membrane domains of the inner membrane translocase AcrB. *Proteins: Struct., Funct., Bioinf.* **2011**, *79* (10), 2871−2885.
(53) Mark, P.; Nilsson, L. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *J. Phys. Chem. A* **2001**, *105* (43), 9954−9960.
(54) Anezo, C.; de Vries, A.; Holtje, H.-D.; Tieleman, P.; Marrink, S.-J. Methodological Issues in Lipid Bilayer Simulations. *J. Phys. Chem. B* **2003**, *107* (35), 9424−9433.
(55) Nelson, M.; Humphrey, W.; Kufrin, R.; Gursoy, A.; Dalke, A.; Kale, L.; Skeel, R.; Schulten, K. Mdscope - a Visual Computing Environment for Structural Biology. *Comput. Phys. Commun.* **1995**, *91* (1−3), 111−133.
(56) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26* (16), 1781−1802.
(57) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4* (2), 187−217.
(58) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30* (10), 1545−1614.
(59) Schulz, G. E.; Vogt, J. The structure of the outer membrane protein OmpX from Escherichia coli reveals possible mechanisms of virulence. *Structure* **1999**, *7* (10), 1301−1309.