ProtSqueeze: Simple and Effective Automated Tool for Setting up Membrane Protein Simulations

Semen O. Yesylevskyy*

Department of Physics of Biological Systems, Institute of Physics of National Academy of Science of Ukraine, Prospect Nauky 46, Kiev-28, 03680, Ukraine

Received December 13, 2006

The major challenge in setting up membrane protein simulations is embedding the protein into the pre-equilibrated lipid bilayer. Several techniques were proposed to achieve optimal packing of the lipid molecules around the protein. However, all of them possess serious disadvantages, which limit their applicability and discourage the users of simulation packages from using them. In the present work, we analyzed existing approaches and proposed a new procedure of protein insertion into the lipid bilayer, which is implemented in the ProtSqueeze software. The advantages of ProtSqueeze are as follows: (1) the insertion algorithm is simple, understandable, and controllable; (2) the software can work with virtually any simulation package on virtually any platform; (3) no modification of the source code of the simulation package is needed; (4) the procedure of insertion is as automated as possible; (5) ProtSqueeze is distributed for free under a general public license. In this work, we present the architecture and the algorithm of ProtSqueeze and demonstrate its usage in case studies.

INTRODUCTION

Atomistic molecular dynamics (MD) simulations of membrane proteins have become increasingly popular in recent years.[1−3] In such works, the protein is embedded into the lipid membrane and solvated in water containing the necessary number of the counterions. Several popular MD packages, such as GROMACS,[4] CHARMM,[5] NAMD,[6] and AMBER,[7] are routinely used for these simulations. The main challenge in setting up membrane protein simulations is embedding the protein into the lipid bilayer. Ideally, one should achieve an optimal packing of the lipid molecules around the protein. This presumes the matching of the hydrophobic surfaces of the protein and the surrounding lipids and the favorable interactions of the lipid heads with the hydrophilic parts of the protein.

Below, we briefly review methodological and technical aspects of existing protein insertion techniques. It is necessary to note that some characteristics of the discussed techniques, such as complexity or user-friendliness, are rather subjective and reflect the personal experience and attitude of the author. Some information is obtained from personal communications due to the lack of technical details in publications.

Several approaches were proposed for building the protein−lipid bilayer systems.[8−14] The first approach is the "brute force" insertion of the protein.[15,16] This approach is described in various online resources;[17−19] however, to our knowledge, it was never published as a separate method. There is very little specific information about the performance and the quality of this technique in the literature. In this approach, the protein is "solvated" by the pre-equilibrated bilayer by removing all lipids, which overlap with the protein. Subsequent long equilibration (up to hundreds of nanoseconds according to personal communications) is supposed to repack remaining lipids around the protein in an optimal way. Such long equilibration is not always feasible, especially if the system is large. The "brute force" method can, in principle, produce unexpected artifacts if the protein is asymmetric and the volumes, substituted by the protein in both monalayers, are very different. In this case, the number of removed lipids depends strongly on the particular structure of the bilayer and orientation of the protein. As a result, one monolayer is very likely to contain fewer lipids than necessary to fill the equilibrated periodic box, while the other monolayer will contain an excessive number of lipids. Thus, both monolayers will be under uncontrollable surface tension, which can distort the results of the simulation. To our knowledge, no special attention was given to this possibility in existing simulations of asymmetric proteins.

In the second approach, developed in the Roux lab,[20] the bilayer is built around the protein by adding the lipids one by one with subsequent removal of unfavorable contacts. The lipids are taken from the library compiled from the pre-equilibrated pure bilayers.[9,10,13,14] This method allows reducing the subsequent equilibration time significantly but has specific disadvantages. The main disadvantage is a complicated and laborious multistage procedure of building the bilayer,[20] which is only possible in the CHARMM[5] software. To our knowledge, no efforts were made to transfer this method to other simulation packages. This limits the applicability of the technique significantly. Other disadvantages are an inability to handle mixed bilayers and multiple proteins without additional programming (scripting) efforts and a dependence on the lipid libraries, which should be prepared separately.[9,10,13,14,20]

The third approach exploits the idea of growing the hole in the pre-equilibrated bilayer and inserting the protein when the hole reaches the necessary size. In the pioneering method,

* Author e-mail: yesint3@yahoo.com.

10.1021/ci600553y CCC: $37.00
Published on Web 07/25/2007

a weak repulsive radial force was applied to the lipids to form a cylindrical pore.[11,12] This technique was greatly improved in the subsequent work of Faraldo−Gomez et al.[8] They developed the rigorous technique, which allows growing complex holes for highly asymmetric membrane proteins. However, both the theoretical basis and implementation of this technique are quite complex.[8] It presumes the usage of at least three different programs (GRASP,[21] HOLE,[22,23] and GROMACS[4]), which requires the researcher to be familiar with all of them. The core program of the GROMACS package is modified to implement the hole-growing algorithm, which causes the whole spectrum of portability and compatibility problems since official GROMACS distribution does not contain these modifications. Finally, the users of other MD programs cannot use this method without prohibitively hard programming efforts.

The analysis of existing methodologies of setting up membrane protein simulations convinced us that the "ideal" software for this purpose should possess the following characteristics:

1. The protein insertion algorithm should be simple, understandable, and easily controllable.

2. The software should be compatible with (or be easily portable to) all major simulation packages. Compatibility with new versions of the packages should be easy to implement.

3. No modifications of the source code of the simulation package should be made. The software should only use a conventional interface of the simulation package available for the user.

4. The procedure of protein insertion should be as automated as possible.

We used these guidelines to develop the ProtSqueeze—easy to use software for setting up membrane protein simulations, which can work with virtually any simulation package.

In this work, we present the architecture and the algorithm of ProtSqueeze and demonstrate its usage in the case studies.

METHODS

The Architecture and the Algorithm of ProtSqueeze. ProtSqueeze utilizes the power of the widely used model building and visualization software VMD,[24] which is available free of charge, portable, and compatible with virtually any simulation package. The core of ProtSqueeze is a TCL script, which is executed by the built-in TCL interpreter of VMD. The main script calls the simulation package and system commands. ProtSqueeze is distributed for free under a general public license (GPL). It is available for download at http://www.geocities.com/yesint3/protsqueeze.html.

The algorithm used by ProtSqueeze consists of three stages—preparation, squeezing, and expansion. The flow chart of the algorithm is shown in Figure 1. All user input is supplied before the core script runs; however, in Figure 1 and in the description given below, different types of user-defined parameters and files are described in the corresponding step of the algorithm to assist understanding. Once the script starts, its execution is completely automatic.

Preparation Stage. In the preparation stage, the protein is aligned with the bilayer and the lipids, which overlap with the protein too much, are removed:

1. The user prepares the files in the Protein Data Bank (PDB) format containing the coordinates of the protein and the pre-equilibrated lipid bilayer. The script takes care of incorrect atom and residue numbering when the structures are combined. The *z* axis of the bilayer structure is presumed to be aligned with the bilayer normal. User-supplied structures should be compatible with the chosen simulation package.

2. The user determines the method of aligning the protein and the bilayer. Currently two methods are available:

In the first method, the reference structure from the database of Orientations of Proteins in Membranes (OPM)[25] is used. This database contains modified PDB structures of virtually all known membrane proteins, which are aligned properly with the bilayer normal (*z* axis) and shifted along the *z* axis in such a way that $z = 0$ corresponds to the center of the bilayer. The user-supplied bilayer is shifted to ensure that its center corresponds to $z = 0$. The protein is positioned in the *xy* plane in the geometrical center of the bilayer. Its position at the *z* axis is determined by aligning the protein with the reference structure from the OMP database.

The second method is based on the fact that many membrane proteins possess well-defined residues, which mark two membrane−water interfaces (for example, in the KcsA channel, such residues are arranged in "collars"[26]). The user supplies the list of residues from each "collar". The positioning of the protein and the bilayer in this method is the same as in previous ones except for the last step. The protein is positioned at the *z* axis in such a way that the distances in the *z* direction between the "collar" residues and the average positions of the lipid head groups of the corresponding leaflet are minimal. This aligns the hydrophobic part of the protein with the hydrophobic core of the bilayer.

3. For each lipid in the bilayer, the number of atoms which form the steric clashes with the protein is determined. If more than N_{crit} percent of the total number of atoms in the lipid clashes with the protein, the lipid is deleted. The parameter N_{crit} is user controllable and defaults to 50%.

Remaining lipids may still have some atoms clashing with the protein; however, the number of such atoms is smaller than N_{crit} percent of the total number of atoms in the lipid molecule. In practice, the remaining steric overlaps are usually observed in the lipid tails or heads, which are bent significantly toward the protein interior. These steric overlaps are removed in the subsequent stages.

Squeezing Stage. During the squeezing stage, the protein is squeezed in the *xy* plane to get rid of any remaining steric overlaps with the lipids:

4. The squeezing factor f is set to 1.

5. New *x* and *y* coordinates of all protein atoms are computed as

$$X_{new} = X_{init} f$$

$$Y_{new} = Y_{init} f$$

where X_{init} and Y_{init} are initial coordinates of the protein measured relative to the *z* axis ($x = 0$, $y = 0$). In the case of a significantly tilted protein, the squeezing is performed not relative to the *z* axis but relative to the precomputed pseudo-

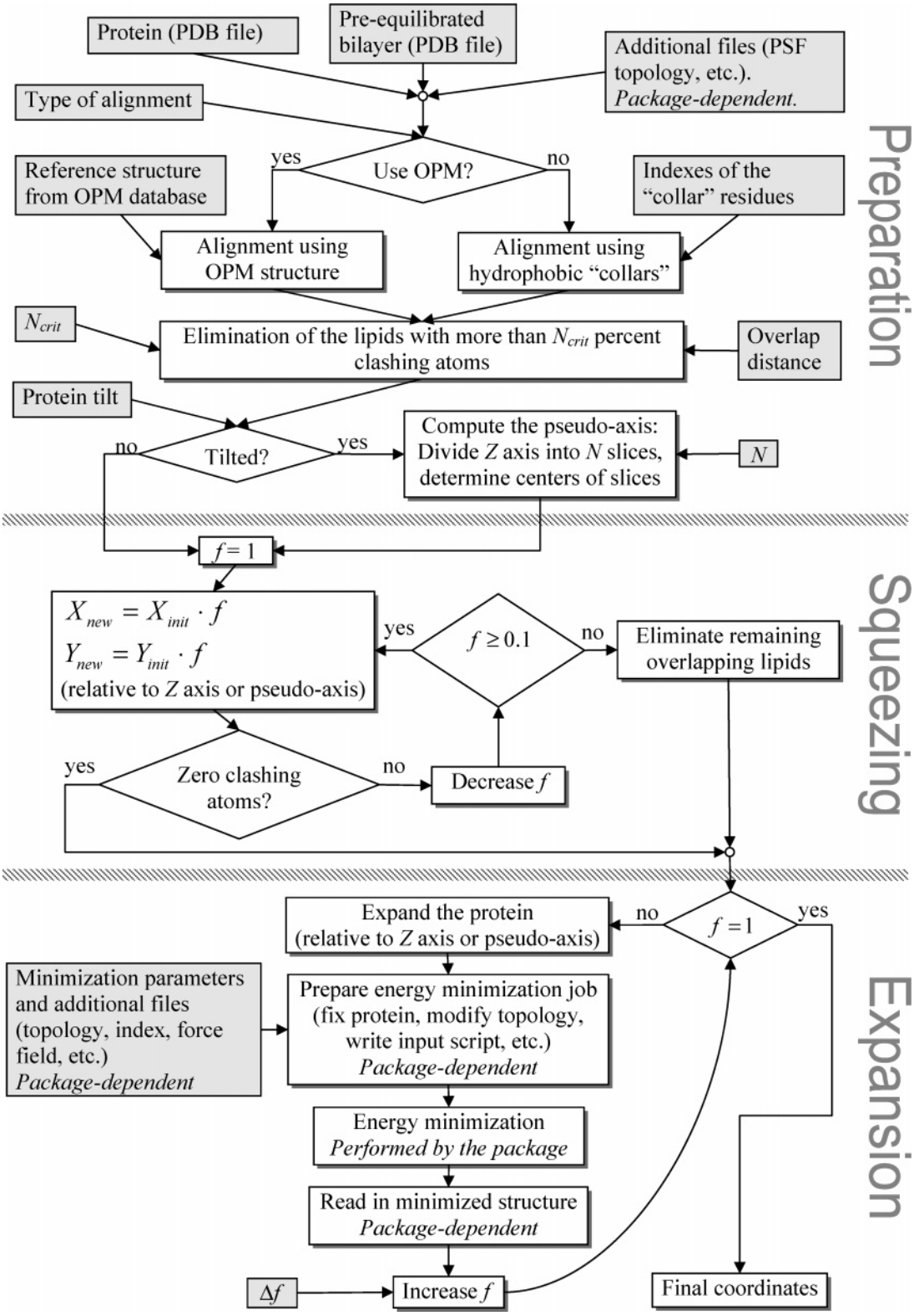

Figure 1. Flow chart of the ProtSqueeze algorithm. Shaded boxes show the user input. Package-dependent steps are indicated.

axis of the protein (see below). The protein is considered to be tilted if the user-defined flag is supplied.

6. The total number of atoms in all lipids, which still clash with the protein, N_{cl} is computed.

7. If $N_{cl} = 0$, then go to the next step; otherwise, reduce f by 0.1, and go to step 6.

8. If f reaches the value of 0.1, all lipids, which still possess steric clashes with the protein, are removed from the system.

Expansion Stage. After squeezing, the protein atoms are congested in a nonphysical way, but they still interact with the lipids according to the nonbond potentials set up in the MD force field. The idea of the expansion stage is to expand the protein back to its native size in small steps and to run the energy minimization for the lipids at each step. This expels incorrectly positioned lipids (or their parts) from the volume, occupied by the protein.

9. The input for the energy minimization job is written using the current coordinates of the system. All protein atoms are marked as frozen (they serve as a repulsive core only). The user should make sure that all files, which are required by the simulation package to run energy minimization, are present. This may include topologies, force field parameters, run input scripts, index files, and so forth. Details of this step depend on the simulation package used.

10. An appropriate program from the simulation package is called to perform energy minimization. The criterion of convergence of the minimization procedure (the value of residual force, the number of minimization steps, etc.) depends on the simulation package used.

11. The energy-minimized structure is read in. This step depends on the simulation package. If necessary, the system is translated to move the origin of the coordinates back to zero.

12. The squeezing factor f is increased, and the protein is expanded according to this new value. This procedure is essentially the inversion of the squeezing step 5. The increase of the factor f is user-defined. The default value is 0.02, which ensures a smooth energy minimization in the next step. If f increases too fast, the energy minimization is likely to fail.

13. Steps 9−12 are continued until the protein expands to its native size ($f = 1$).

14. Final coordinates are saved.

Treatment of Tilted Proteins. There are a number of membrane proteins, which are tilted significantly to the normal of the bilayer. The squeezing and expansion of such proteins should be performed not around the global z axis but around the tilted pseudoaxis of the protein (the prefix pseudo here indicates that this axis is not identical to true symmetry axis or axis of inertia). In order to allow the insertion of tilted proteins, the following steps are added at the end of the preparation stage (after step 3):

1. The protein is divided into N narrow slices perpendicular to the z axis. The parameter N is user-defined and defaults to 20.

2. The geometric center of all backbone atoms, which fall into the given slice, is determined. This point is used as a center of squeezing and expansion for the corresponding slice in subsequent steps.

Portability Issues. The algorithm of ProtSqueeze is completely generic and can be used, in principle, with any simulation package. However, it is not practical to create a single generic script which is compatible with several simulation packages. The reason for this is the completely different architectures and input schemes of different packages. For example, the topology file for GROMACS can be prepared before ProtSqueeze runs. ProtSqueeze only adds one line to the topology, which determines the number of lipids retained in the system. In contrast, the topology file for NAMD cannot be prepared in advance. ProtSqueeze should modify it during operation by means of the psfgen plug-in of VMD.

Thus, we decided to provide distinct ProtSqueeze implementations for each simulation package. Currently, implementations for GROMACS and NAMD are tested and supported. Compatibility with any other simulation package can be achieved by introducing only minor changes to the main TCL script. The users of other simulation packages can consider ProtSqueeze as a kind of "do-it-yourself" kit, distributed under a GPL license.

Relation to Other Techniques. ProtSqueeze utilizes the same idea of the growing hole as that of the algorithms described in refs 8 and 12 in the expansion stage. However, the force, which drives the lipids out of the volume, occupied by the protein is different. In our method, the lipids are pushed by the same forces, which facilitate the lipid−protein interaction during subsequent MD simulations. No other artificial forces are introduced. This ensures that the resulting structure is free from any unexpected artifacts caused by a possible inadequate choice of the pushing force. Our technique is also much easier to use and to comprehend because ProtSqueeze does not compute the molecular surface of the protein and does not perform complex projection of the expansion forces as it is done in previous growing-hole algorithms.[8,12]

Test Cases. The ProtSqueeze technique was tested by setting up two simulations of very different membrane proteins embedded into different bilayers and simulated by different MD packages.

Test Case 1: Insertion of KcsA Potassium Channel into Palmitoyloleoylphosphatidylcholin (POPC) Bilayer Using GROMACS. The KcsA channel is a good example of an α-helical protein, which forms the voids in upper and lower monolayers of very different sizes. The GROMACS 3.3.1 simulation package[4] was used with the ffgmx force field, modified to incorporate the parameters for lipids, of Berger et al.[27] ProtSqueeze was executed within VMD 1.8.3.[24]

The POPC lipid bilayer consists of 200 lipids (100 per monolayer). The lipids in each monolayer were initially arranged into a regular grid (initial box width was 9.14 nm). The bilayer was solvated by 7432 water molecules (∼37 per lipid). The bilayer was equilibrated for 30 ns under constant pressure, temperature, and chemical composition (NPT) conditions at 300 K. A Berendsen thermostat with a relaxation time of 0.1 ps and a Berendsen barostat with a relaxation time of 0.5 ps were used. Semi-isotropic pressure coupling was used. The PME method was used for computing long-range electrostatics. The last 5 ns of the run were used to compute the equilibrium area per lipid.

The PDB structure 1BL8[26] was used to build the model of the KcsA channel. Missing side chains were added by the Swiss PDB Viewer 3.7.[28] Two ions were inserted into the selectivity filter of the channel and another ion into the channel cavity, as described elsewhere.[29,30] The protein was aligned with the bilayer using the method of interface "collar" residues.

In ProtSqueeze, the lipid atoms were considered to overlap with the protein if they were within 2.2 Å of the protein atoms. The lipids were deleted if more than 50% of their atoms overlapped with the protein.

Post-insertion equilibration was performed as following. In order to equilibrate the lipids without distortion of the protein, we imposed position restraints on all heavy atoms of the protein and the potassium ions in the pore of the channel. The system was solvated with SPC water molecules,[31] and 11 Cl counterions were added at random positions to neutralize the charge of the protein and potassium ions. Lipids and water molecules were kept free during the simulation. The parameters given above were used, except for the relaxation time of the Berendsen barostat, which was set to 10 ps. This suppressed wild fluctuations of the pressure, which were caused by the congested lipid tails and the

existence of the voids in the bilayer. The system was equilibrated for 10 ns.

Test Case 2: Insertion of the Outer Membrane Enzyme PagL into Palmitoyloleoylphosphatidyletanolamin (POPE) Bilayer Using NAMD. The outer membrane enzyme PagL from *Pseudomonas aeruginosa*[32] is a prominent example of a β-barrel protein, which is tilted significantly in the bilayer. The protein was embedded as a monomer. The NAMD 2.5[6] simulation package was used with CHARMM 27 force field. ProtSqueeze was executed within VMD 1.8.5.[24]

The POPE bilayer contains 96 lipids (48 per monolayer). It was constructed using the Membrane plug-in of VMD. The Membrane plug-in produces the bilayer solvated by a minimal amount of water, which is just enough to cover the lipid heads. Such a solvation state is not typical for the free lipid bilayer in the fluid or liquid crystal state. Thus, we added additional slabs of water molecules. The total number of water molecules is 3548 ($\sim$37 per lipid).

The bilayer was equilibrated for 3 ns under NPT conditions at a temperature of 300 K. The Langevin dynamics with a damping coefficient of 10 ps^{-1} were used for temperature coupling. A Nose−Hoover Langevin piston barostat with a period of 200 ps and a decay parameter of 100 ps was used. The PME method was used for computing long-range electrostatics.

The PDB structure 2ERV was used to build the model of enzyme PagL. The protein was aligned with the bilayer using the reference structure from the OPM database.[25] The protein is neutral, so no counter-ions were added. In ProtSqueeze, the lipid atoms were considered to overlap with the protein if they were within 2.5 Å of the protein atoms. The lipids were deleted if more than 50% of their atoms overlapped with the protein.

In order to perform the post-insertion equilibration, the system was resolvated. The solvated system was energy-minimized and subject to 3.25 ns of MD simulation with the parameters described above. The positions of all heavy atoms of the protein were restrained to avoid distortion of the protein.

Monitoring the Quality of Insertion. The quality of the systems produced by protein insertion techniques is rather hard to evaluate. The ultimate quality criterion is the comparison of the properties of the protein−membrane system with experimental data. However, this criterion cannot be used in the majority of studied systems. In the present study, we only evaluate the quality of the insertion algorithm itself. We used the volume of the voids, left by the removed lipids, as a quality criterion. This shows how much the pre-equilibrated bilayer is perturbed by the inserted protein. During post-insertion equilibration, the lipids are filling the voids by moving in the *xy* plane and the total volume of the system decreases. Thus, it is enough to monitor the size of the simulation box in the *xy* plane to estimate the empty volume, which is filled by the lipids during equilibration. Since the simulation box is kept square in the *xy* plane, the width of the box is monitored during the simulations.

RESULTS

Test Case 1. The equilibrium area per lipid of the pre-equilibrated pure POPC bilayer was 60.01 $Å^2$. In the preparation stage of the ProtSqueeze run, 161 lipids were retained in the system. Some lipid tails still overlapped with the protein. The protein was squeezed to 0.4 of its initial size until these overlaps vanished. Subsequent expansion pushed all overlapping tails out of the volume occupied by the protein. The process of expansion is shown in Figure 2a−c. It is clearly seen that the lipid tails, which protrude into the cavity, occupied by the protein, are progressively pushed out by the expanding protein. The reader is also advised to see the movie of expansion provided in the Supporting Information. The lipid tails, which were pushed by the protein during expansion, are perturbed a lot (the data of visual inspection). The post-insertion equilibration removes these perturbations and fills the voids, which are left by the removed lipids.

Figure 3 shows the evolution of the width of the simulation box during the whole simulation. Initial equilibration of the pure bilayer is marked by large fluctuations of the box size, which is usual for bilayer simulations. The box size stabilizes after $\sim$25 ns at a mean value of 7.74 nm (area per lipid 60.01 $Å^2$). After insertion of the protein, remaining lipids begin to fill the voids left by the removed lipids, and the size of the box decreases. This decrease is smooth and takes about 5−7 ns. The new equilibrium width of the box (measured as the average over last 3 ns of the run) is 7.64 nm. The change in the box dimensions after insertion of the protein is surprisingly small (only 1 Å), which means that ProtSqueeze produces the system with quite small voids in the bilayer. The time needed to equilibrate the system after insertion of the protein is about 5−10 ns, which is quite tolerable.

Test Case 2. It is believed that the PagL protein forms a functional dimer in the bilayer.[32] We embedded the single monomer of the enzyme into the membrane to show the ability of ProtSqueeze to handle unusual significantly tilted proteins. The functional and biological aspects of insertion of this protein in the monomeric form are not addressed in the present study.

Since the POPE bilayer produced by the membrane plug-in of VMD is assembled from the pre-equilibrated pieces, the box size stabilizes after $\sim$2 ns of simulation after fast initial fluctuations. The last 0.75 ns of the run were used to compute the equilibrium area per lipid, which appeared to be 78.7 $Å^2$.

A total of 81 lipids were retained in the system after the preparation stage of the ProtSqueeze run. The protein was squeezed to 0.1 of its initial size until all remaining overlaps vanished. Such a small value is expectable because the protein is tilted substantially, while the lipid tails are approximately parallel to the bilayer normal. The process of expansion is shown in Figure 2d−f. Despite the complex tilted shape of the protein, the lipid tails are successfully pushed out of the resulting cavity.

Figure 3b shows the evolution of the width of the simulation box during both pre- and post-insertion equilibration. The final box size average over the last 0.5 ns of the post-insertion run is 6.08 nm, which is only $\sim$0.5 Å smaller than the value obtained for the pure bilayer. This result is consistent with the change in box width obtained in the previous test. Our results show that ProtSqueeze can handle significantly tilted proteins and produces systems which can be smoothly equilibrated in a few nanoseconds.

DISCUSSION

The Advantages of ProtSqueeze. MD is now a well-tested and mature technique, which is used in very diverse

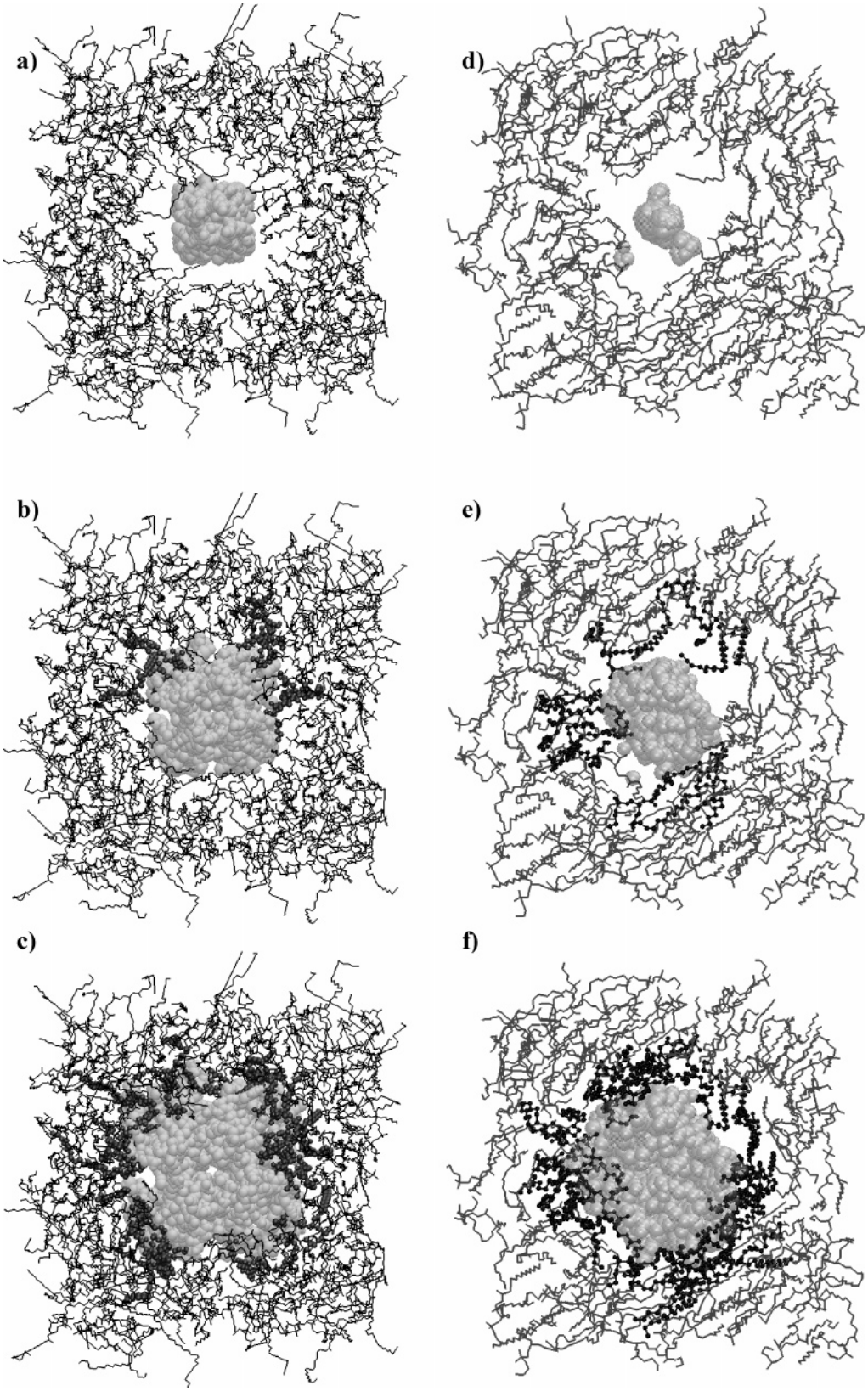

Figure 2. Expansion phases of the ProtSqueeze runs of the KcsA channel in the POPC bilayer (a, b, and c) and monomeric enzyme PagL in the POPE bilayer (d, e, and f). The lipids are shown as lines; the protein is in the space-fill representation. The hydrogen atoms are not shown. The positions of the lipids shown as balls and sticks are adjusted to accommodate the protein. The squeeze factors are: (a) 0.4, (b) 0.75, (c) 1.0, (d) 0.1, (e) 0.6, and (f) 1.0. The long (pseudo) axis of the protein is aligned with the line of sight; thus, the plane of the membrane does not lie in the plane of the picture in the case of the tilted PagL protein in d, e, and f. The scale of parts a, b, and c and d, e, and f is different due to different sizes of the systems.

research areas ranging from material science to rational drug design. The complexity of the systems studied by MD is increasing rapidly. This leads to an increasing complexity of setting up MD simulations. The systems, which contain membrane proteins, are especially hard to set up due to the inability of the standard simulation packages to embed the

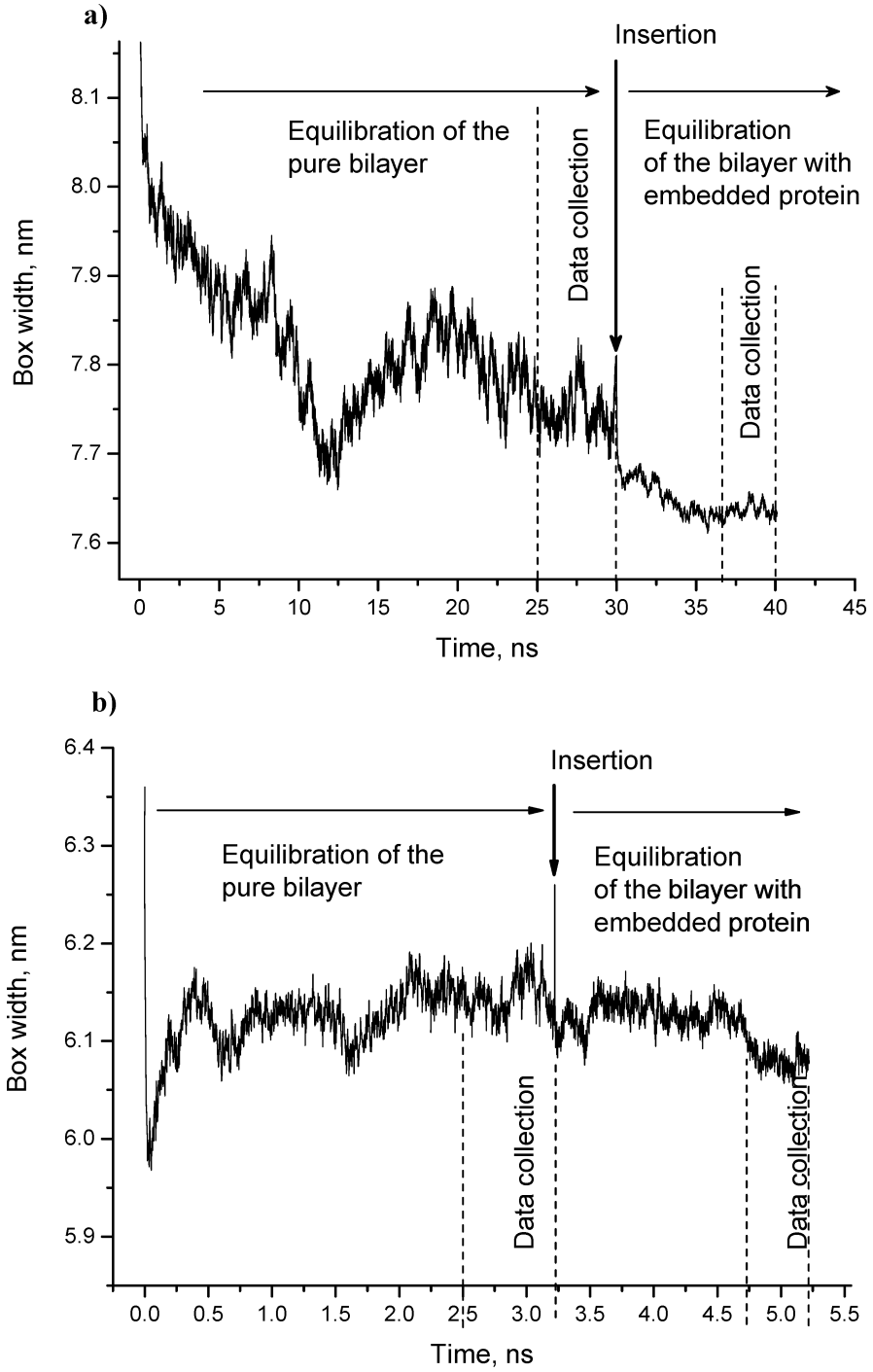

Figure 3. Evolution of the box width during the simulations. (a) KcsA channel in the POPC bilayer; (b) enzyme PagL in the POPE bilayer.

protein into the lipid bilayer. As it was stated in the Introduction, existing techniques of embedding the proteins into the bilayer have many drawbacks. The most severe are the complexity of the algorithms, laborious multistage operation, and limited portability. ProtSqueeze is developed in order to overcome the disadvantages of existing techniques. The idea of the algorithm is very simple—the expanding protein pushes the lipids out of the excluded volume by means of the same nonbond interactions, which facilitate the protein−lipid interactions in ordinary MD simulations. There are no artificial forces, which are used in several other techniques to grow the hole for the protein.[8] This eliminates additional empirical parameters and ensures the absence of possible artifacts caused by an inadequate choice of these parameters. The simplicity of the algorithm makes it easy to understand and control.

The operation of ProtSqueeze is as automated as possible. Setting up the ProtSqueeze run is actually no harder than setting up the energy minimization of the protein−lipid system in the chosen simulation package. However, the intended user can interfere into any stage of insertion easily by editing the ProtSqueeze script. This makes the method very flexible and suitable for further improvements and modifications by the community.

Due to the usage of VMD as a host application, ProtSqueeze is portable to all major computer platforms. Compatibility

with virtually any simulation package (other than GROMACS and NAMD, which are supported now) can be achieved with only minor programming efforts. The source code of the simulation package should not be modified to ensure compatibility with ProtSqueeze, which makes the migration to newer versions of the simulation packages effortless.

The test studies described in the present work demonstrate that ProtSqueeze can facilitate the insertion of the proteins, which are usually viewed as "hard targets" for membrane protein simulations. The method can handle substantially asymmetric proteins (like KcsA channel), which require different numbers of lipids to be removed from the upper and lower monolayers of the membrane. Tilted proteins (like enzyme PagL) can be inserted with no additional efforts, since the pseudo-axis of the tilted protein is determined automatically. Our method is shown to work smoothly with two very popular simulation packages, GROMACS and NAMD, which ensures its immediate availability for a large part of the modeling community.

Limitations and Special Cases. Although the ProtSqueeze is quite generic, the insertion of particular "exotic" types of membrane proteins and assemblies requires additional efforts from the user. Some proteins form two or more distinct disconnected voids in the bilayer, while ProtSqueeze can only form a single void. A typical example of such a protein is rat mono-amine oxidase A.[33] The transmembrane segments of this particular protein are simple and symmetric, so the brute-force technique can be suitable for it. In the case of more complex proteins or macromolecular assemblies that form several disconnected voids, the following method can be used.

1. The protein or assembly is manually cut into the pieces, which form only one void each.

2. ProtSqueeze is run successively with each of them, but with the same bilayer (automatic alignment of the protein in the *xy* plane should be disabled).

3. The pieces are glued together again.

Another complicated case is illustrated by the proteins which form the transmembrane rings with a large central hole. Examples of such structures are multimeric pores formed by bacterial toxins (anthrax toxin, aerolysin, and pneumolysin).[34] The central hole of such assemblies can be so large that it accommodates several lipids from the original bilayer. Some of these lipids would not be deleted, since they do not overlap with the protein atoms in the final steps of squeezing. In order to use ProtSqueeze with these assemblies, the user should manually fill the central hole with some dummy atoms, run ProtSqueeze, and then remove the dummy particles.

These special cases show that the procedure of insertion of an arbitrary protein or assembly into the bilayer cannot be considered absolutely trivial. However, ProtSqueeze can assist the user even in very complicated cases.

A Note on Obtained Areas per Lipid. The value of the area per lipid of 60.01 Å^2 for the pure POPC bilayer is reasonably close to the experimental value of 64.0 Å^2 [35] and the results of other simulations.[36] It should be noted, however, that other values of the area per lipid parameter for POPC can be found in the literature as well.[37]

The pre-equilibrated pure POPE bilayer with an area per lipid of 78.7 Å^2 is likely to be the meta-stable intermediate rather than the true equilibrium state because of a short equilibration time. That is why the resulting bilayer with embedded protein is also likely to be the meta-stable intermediate, and the box dimensions can change further if the postinsertion run is prolonged.

The goal of the present study is not to study the equilibrium properties of the bilayers, but to test the protein insertion technique. Thus, the bilayers produced in both simulations are acceptable for us. In the present study, we discuss neither the reason of discrepancies in the area per lipid nor the quality of the used force field and simulation parameters.

CONCLUSIONS

We developed the ProtSqueeze algorithm, which is dedicated to setting up membrane protein simulations. ProtSqueeze allows insertion of the protein into the pre-equilibrated lipid bilayer by growing the hole, which accommodates the protein. No artificial forces are used to grow the hole. The protein is first squeezed in the *xy* plane until any overlaps between the lipids and the protein vanish. The positions of the lipids, which overlap with the protein, are then adjusted by expanding the protein back to its original size. The method can work with proteins of any shape and size. The method produces the structure with very few voids in the bilayer, so equilibration over 5−10 ns is enough to re-equilibrate the bilayer after insertion of the protein. The major advantages of the implementation of ProtSqueeze over other similar tools are platform independence, compatibility with virtually any simulation package, simplicity and effectiveness of the algorithm, and a completely automated operation. The software is free (available for download at http://www.geocities.com/yesint3/protsqueeze.html) and easy to use and modify. ProtSqueeze can become a useful tool for molecular simulations of the membrane proteins.

ACKNOWLEDGMENT

The author is grateful to the "Entry" company (Kiev) and to the administration of computational cluster of Kiev Taras Shevchenko National University for providing computer time for this work. Prof. Alexander P. Demchenko is deeply acknowledged for critical reading of the manuscript and valuable comments.

Supporting Information Available: An mpeg video illustrating expansion. This amterial is available free of charge via the internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Saiz, L.; Klein, M. L. Computer simulation studies of model biological membranes. *Acc. Chem. Res.* **2002**, *35*, 482−489.

(2) Tielman, D. P.; Marrink, J. S.; Berendsen, H. J. C. A computer perspective of membranes: molecular dynamics studies of lipid bilayer systems. *Biochem. Biophys. Acta* **1997**, *1331*, 235−270.

(3) Allen, M. P. Introduction to Molecular Dynamics Simulation. In *Computational Soft Matter: From Synthetic Polymers to Proteins*; Attig, N., Binder, R., Grubmuller, H., Kremer, K., Eds.; John von Neumann Institute for Computing: Julich, Germany, 2004; Vol. 23, pp 1−28.

(4) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible and Free. *J. Comput. Chem.* **2005**, *26*, 1701−1718.

(5) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: a program for macromo-

lecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.
(6) Kale, L.; Skeel, R.; Brunner, R.; Bhandarkar, M.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulten, K. NAMD2: Greater Scalability for Parallel Molecular Dynamics. *J. Comput. Phys.* **1999**, *1*, 283−312.
(7) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E., III; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsui, V.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. *AMBER6*; University of California, San Francisco: San Francisco, CA, 1999.
(8) Faraldo-Gomez, J. D.; Smith, G. R.; Sansom, M. S. P. Setting up and optimization of membrane protein simulations. *Eur. Biophys. J.* **2002**, *31*, 217−227.
(9) Petrache, H. I.; Grossfield, A.; MacKenzie, K. R.; Engelman, D. M.; Woolf, T. B. Modulation of glycophorin A transmembrane helix interactions by lipid bilayers: molecular dynamics calculations. *J. Mol. Biol.* **2000**, *302*, 727−746.
(10) Roux, B.; Woolf, T. B. Molecular dynamics of Pf1 coat protein in a phospholipid bilayer. In *Biological membranes: a molecular perspective from computation and experiment*; Merz, K. M., Ed.; Birkhauser: Boston, MA, 1996; pp 555−587.
(11) Shen, L.; Bassolino, D.; Stouch, T. Transmembrane helix structure, dynamics, and interactions: multi-nanosecond molecular dynamics simulations. *Biophys. J.* **1997**, *73*, 3−20.
(12) Tieleman, D. P.; Berendsen, H. J. C. A molecular dynamics study of the pores formed by Escherichia coli OmpF porin in a fully hydrated palmitoyloleoylphosphatidylcholine bilayer. *Biophys. J.* **1998**, *74*, 2786−2801.
(13) 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*, 11631−11635.
(14) 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*, 92−114.
(15) Treptow, W.; Maigret, B.; Chipot, C.; Tarek, M. Coupled Motions between Pore and Voltage-Sensor Domains: A Model for Shaker B, a Voltage-Gated Potassium Channel. *Biophys. J.* **2004**, *87*, 2365−2379.
(16) Aksimentiev, A.; Schulten, K. Imaging α-hemolysin with molecular dynamics: ionic conductance, osmotic permeability, and the electrostatic potential map. *Biophys. J.* **2005**, *88*, 3745−3761.
(17) Gromacs Workshop at CSC. Membrane simulations. https://extras.csc.fi/chem/courses/gmx2007/Erik_Talks/membrane_simulations.pdf (accessed Mar 27, 2007).
(18) Explicit Membrane Protein Simulations in NAMD/VMD. http://mccammon.ucsd.edu/∼rlaw/ctbp_workshop_rlaw.htm (accessed Mar 27, 2007).
(19) NAMD tutorials. Building Gramicidin A. http://www.ks.uiuc.edu/Research/namd/tutorial/NCSA2002/hands-on/ (accessed Mar 27, 2007).
(20) Roux Lab. Membrane builder. http://thallium.bsd.uchicago.edu/RouxLab/membrane.html (accessed Mar 27, 2007).
(21) Nicholls, A.; Bharadwaj, R.; Honig, B. GRASP: graphical representation and analysis of surface properties. *Biophys. J.* **1993**, *64*, 166−170.
(22) Smart, O. S.; Goodfellow, J. M.; Wallace, B. A. The pore dimensions of gramicidin A. *Biophys. J.* **1993**, *65*, 2455−2460.
(23) Smart, O. S.; Neduvelil, J. G.; Wang, X.; Wallace, B. A.; Sansom, M. S. P. HOLE: a program for the analysis of the pore dimensions of ion channel structural models. *J. Mol. Graphics* **1996**, *14*, 354−360.
(24) Humphrey, W.; Dalke, A.; Schulten, K. VMD - Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.
(25) Lomize, M. A.; Lomize, A. L.; Pogozheva, I. D.; Mosberg, H. I. Orientations of Proteins in Membranes database. *Bioinformatics* **2006**, *22*, 623−625.
(26) Doyle, D. A.; Cabral, J. M.; Pfuzner, R. A.; Kuo, A.; Gulbis, J. M.; Cohen, S. L.; Chait, B. T.; MacKinnon, R. The structure of the potassium channel: molecular basis of K+ conduction and selectivity. *Science* **1998**, *280*, 69−77.
(27) Berger, O.; Edholm, O.; Jahnig, F. Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure and constant temperature. *Biophys. J.* **1997**, *72*, 2002−2013.
(28) Guex, N.; Peitsch, M. C. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* **1997**, *18*, 2714−2723.
(29) Compoint, M.; Carloni, P.; Ramseyer, C.; Girardet, C. Molecular dynamics study of the KcsA channel at 2.0 A resolution: stability and concerted motions within the pore. *Biophys. Biophys. Acta* **2004**, *1661* 26−39.
(30) Compoint, M.; Picaud, F.; Ramseyer, C.; Girardet, C. Targeted molecular dynamics of an open-state KcsA channel. *J. Chem. Phys.* **2005**, *122*, 134707.
(31) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. Interaction models for water in relation to protein hydration. In *Intermolecular Forces*; Pullman, B., Ed.; Reidel: Dordrecht, The Netherlands, 1981.
(32) Rutten, L.; Geurtsen, J.; Lambert, W.; Smolenaers, J. J. M.; Bonvin, A. M.; van der Ley, P.; Egmond, M. R.; Gros, P.; Tommassen, J. Crystal structure of the outer membrane enzyme PagL. **2006**, in press .
(33) Ma, J.; Yoshimura, M.; Yamashita, E.; Nakagawa, A.; Ito, A.; Tsukihara, T. Structure of rat monoamine oxidase a and its specific recognitions for substrates and inhibitors. *J. Mol. Biol.* **2004**, *338*, 103−114.
(34) S. J. Tilley; Saibil, H. R. The mechanism of pore formation by bacterial toxins. *Curr. Opin. Struct. Biol.* **2006**, *16*, 230−236.
(35) König, B.; Dietrich, U.; Klose, G. Hydration and structural properties of mixed lipid/surfactant model membranes. *Langmuir* **1997**, *13*, 525−532.
(36) Mukhopadhyay, P.; Vogel, H. J.; Tieleman, D. P. Distribution of Pentachlorophenol in Phospholipid Bilayers: A Molecular Dynamics Study *Biophys. J.* **2004**, *86*, 337−345.
(37) Kučerka, N.; Tristram-Nagle, S.; Nagle, J. F. Structure of Fully Hydrated Fluid Phase Lipid Bilayers with Monounsaturated Chains. *J. Membr. Biol.* **2005**, *208*, 193−202.

CI600553Y