Modeling α-helical pore-forming-protein nanopores in Martini

Lorenzo Gaifas

Supervisors: Siewert-Jan Marrink, Giovanni Maglia

Daily supervisor: Bart Bruininks

University of Groningen

December 3, 2019

Abstract

In nanopore analytics, individual molecules block the entrance of a single nanopore, thus generating a detectable blockade of ionic current which can be studied to identify properties of the analyte. Because of its high resolution and relative low cost and simplicity, this method is gaining more and more popularity, employing both man-made inorganic pores and biological, protein-based ones. Pore-forming-proteins are often used as biological nanopores, having several advantages on their artificial counterparts, such as being easier to produce as well as being potentially more customizable in shape and charge. However, our knowledge on transmembrane protein structure and folding is not yet refined enough to allow us to design biological nanopores with fine-tuned properties. Here, we study the pore-forming-protein FraC with the use of coarse-grained molecular dynamics, in order to better understand the mechanisms of its pore and to find a common ground with experimental data on ionic current and single-molecule analytics. Knowledge on this protein has already been gathered in the field of nanopore analytics, which makes it a good candidate for this study. FraC models were created with two different versions of Martini force field – Martini 2 and Martini 3 – to compare the applicability of each version to the study of pore-forming-proteins and transmembrane proteins as a whole. Results indicate that Martini 3 is better than its older version at simulating the behaviour of transmembrane pores, though it still needs some parameters to be tuned properly. With either force field, modeling FraC proved to be not trivial, as there are several factors whose contributions are still poorly understood, such as the role of sphingomyelins.

Contents

1 Introduction **4**
- 1.1 Nanopore analytics . . . 4
- 1.2 FraC . . . 5
- 1.3 The Martini force field . . . 6
  - 1.3.1 Martini 2 . . . 7
  - 1.3.2 Martini 3 . . . 8
  - 1.3.3 Proteins in Martini . . . 8

2 Methods **10**
- 2.1 Generating a CG protein . . . 10
  - 2.1.1 Starting structure . . . 10
  - 2.1.2 Coarse-graining with Martinize2 . . . 10
- 2.2 System setup . . . 10
- 2.3 Gromacs simulation setup . . . 11
- 2.4 Mutant design . . . 12
- 2.5 Ionic flux simulations . . . 13
- 2.6 Analysis . . . 14
  - 2.6.1 Data collection and preparation . . . 14
  - 2.6.2 Pore size . . . 14
  - 2.6.3 Flux . . . 14
  - 2.6.4 Graphs and representation . . . 15
  - 2.6.5 Scripting . . . 15

3 Results **17**
- 3.1 Reference simulations . . . 17
  - 3.1.1 Martini 3 . . . 18
  - 3.1.2 Martini 2 . . . 18
- 3.2 Effect of sphingomyelins . . . 20
  - 3.2.1 Martini 3 . . . 20
  - 3.2.2 Martini 2 . . . 20
- 3.3 The membrane height problem . . . 22
- 3.4 N-term secondary structure . . . 25
  - 3.4.1 Martini 3 . . . 25
  - 3.4.2 Martini 2 . . . 26
- 3.5 The role of the Elastic Network . . . 27
  - 3.5.1 Martini 3 . . . 27
  - 3.5.2 Martini 2 . . . 29
- 3.6 Force Field comparison . . . 29
- 3.7 Scripts and computational tools . . . 30

3.7.1 Martinize2 30
3.7.2 Insane.py 30
3.7.3 HOLEtraj 30
3.7.4 Fluxer.py 31

4 Discussion and Conclusions **32**
4.1 Importance of sphingomyelins 32
4.2 Membrane height 33
4.3 Helix vs Coil N-term 33
4.4 Role of the Elastic Network 33

5 Future Perspectives **34**
5.1 Sphingomyelins 34
5.2 Computational Electrophysiology 34
5.3 Mutants 35

1 Introduction

In the recent years, single molecule (SM) biochemistry has become a widely used approach to detect molecules and study molecular mechanisms with unprecedented detail[1]. Among the fields in SM studies, nanopore analytics is one of the youngest, growing exponentially in popularity and range of applications. Nanopores are nano-scale biological or artifical structures which form a hole through a surface – a biological membrane or a layer of synthetic material – thus allowing the selective passage or insertion of molecules depending on their properties.

1.1 Nanopore analytics

In nanopore analytics, the presence of molecules in the lumen of the pore is detected by applying a potential between the two sides of the membrane and then measuring the ionic current though the pore; whenever an analyte enters the pore, it partially blocks the lumen, thus reducing the flow of ions and resulting in a detectable reduction of ionic current[2–4]. The amount and duration of this blockade depends on the size, shape and charge of the analyte and its interaction with the lumen of the pore. This method was succesfully employed to perform high resolution tasks such as DNA sequencing[5] and identify folded proteins with minimal structural differences[3].

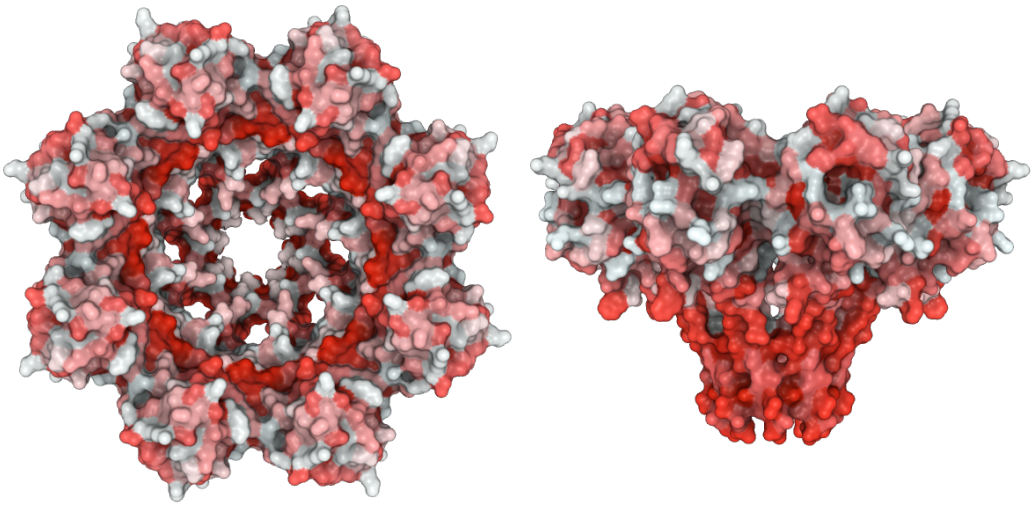
Future perspective include small peptide recognition and protein sequencing. However, differently from DNA, polypeptides present a non-uniform charge profile and aminoacids have a wider range of sizes compared to nucleobases, making it difficult to find or design a nanopore able to translocate and succesfully detect a wide variety of amino acid sequences.

To this end, intelligent design of biological nanopores appears to be a promising direction. Protein-based nanopores are easier to produce than their artificial counterparts, which makes them more attractive for industrial and commercial use. Compared to solid-state nanopores, they also provide the advantage of a more customizable and better controlled geometry and internal surface chemistry, down to atomic resolution. Specifically, the use of protein-based nanopores may allow for a modular approach in which a library of domain and subdomain sequences could be combined to design ad-hoc structure depending on the specific analyte and experimental conditions.

In this project, we analyze FraC, a toxin PFP produced by the sea anemone *Actinia fragacea*, with the use of molecular dynamics and the coarse-grained (CG) force field (FF) Martini in order to better understand the mechanisms of its pore and to find a common ground with experimental data on ionic current.

1.2 FraC

Pore-forming-toxins (PFTs) can be divided in two subcategories: those that rely on α-helical transmembrane domains (α-PFTs) and those that form a cylindrical pore with transmembrane β-sheets (β-PFTs). FraC belongs to the α-PFT family, employing helices to form a cone-shaped pore.



(a) *Top view of the FraC nanopore.* ***(b)*** *Side view of the FraC nanopore.*

Figure 1: *Top and side view of the surface of the FraC nanopore. Aminoacids are colored based on their hydrophobicity (white for hydrophobic, red for hydrophilic). The cone-shaped hydrophilic α-helics constitute the transmembrane domanin, while the rest is mostly hydrophobic and sits on top of the membrane.*

The pore-formation mechanism is still unknown, although several hypotheses have been advanced based on intermediate X-ray structures and chemical considerations[6,7]. The protein seems to form a homo-multimeric complex, likely an octamer (Figure 1), with the water-soluble domains forming a toroidal shape in the *trans* side of the membrane, while *cis* side the pore is outlined by the transmembrane α-helices.

The FraC nanopore is particularly stable, which is one of the reasons of its efficacy as a toxin in lysing cells. Due to this property, FraC is an attractive candidate for nanopore analytics, where stability of the pore is key in order to reduce background current fluctuations and improve the resolution of the experiment. Several groups have already utilized FraC in different applications, as in identification of folded protein[3,8].

As previously mentioned, other than its stability, FraC is interesting because of its α-helical nature and thus conical shaped pore. This property

may allow for wider applicability in detection of differently sized analytes, which would sit at different depths in the pore lumen, while still being able to provide a good resolution for all of them.

Another poorly understood aspect of FraC worthy of mention is the role of Sphingomyelins in the target membrane: there are currently conflicting opinions on their role in the FraC pore formation and/or stabilization, but most seems to agree that they play an important role in the efficacy of the toxin.

1.3 The Martini force field

> You cannot coarse-grain a protein.
>
> Alex H. de Vries, 2007 ca.

In this project, FraC nanopores were investigates through molecular dynamics simulations using the Martini CG force field. The force field was initially developed for lipids and membranes[9], but was subsequently extended to macromolecules such as proteins and nucleic acids[10]. Using a CG force field allows sampling of much longer time scales – or bigger systems – which would be heavily restrictive for such a big system if a full-atom force field was to be used.

Martini uses a 4 to 1 mapping approach, were on average 4 heavy atoms (non-hydrogen) are represented by a single CG particle. This drastically reduces the amount of particles in the system, significantly improving computation time and efficiency.

This improved computational efficiency comes at the cost of resolution and precision. A coarse grained system will not be able to accurately represent interactions that require atomistic resolution; depending on the application, this trade-off may be irrelevant orof vital importance. Several years of Martini papers have shown that proteins and lipid membranes can be accurately represented without resulting in unrealistic behaviour. However, a protein with a peculiar transmembrane domain such as FraC may prove difficult to correctly characterize using coarse-grained molecular dynamics.

1.3.1 Martini 2

There are four main types of interaction sites: polar (P), nonpolar (N), apolar (C) and charged (Q) which are further subdivided in several subtypes.

Interaction between particles are defined by bonded and non bonded parameters, which together comprise the Hamiltonian equation for the Martini force field.

Nonbonded interactions are defined with a Lennard-Jones (LJ) potential energy function:

$$U_{\mathrm{LJ}}(r) = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r}\right)^{12} - \left(\frac{\sigma_{ij}}{r}\right)^{6}\right] \tag{1}$$

with σ representing the closest distance of approach between two particles and ϵ the strength of their interaction. In Martini 2 there are three sizes of beads – normal (N), small (S) and tiny (T). Small beads are typically used in ring-like structures to maintain their geometry without compromising their sterinc hindrance. Tiny beads are a special case, used mainly only in DNA.

All interactions have $\sigma = 0.47$ nm, except for S-S interactions, which have $\sigma = 0.43$ nm and interactions between charged particles (Q) and the most apolar types (C1-2) which have $\sigma = 0.62$ nm[9]. The value of ϵ is defined according to the Interaction Table of the Martini FF[9], which was parametrized according to partitioning coefficients in water/octanol of numerous compounds.

Interactions between charged particles are defined by a Coulombic potential energy function:

$$U_{\mathrm{el}}(r) = \frac{q_i q_j}{4\pi\epsilon_0\epsilon_{\mathrm{r}} r} \tag{2}$$

with q representing the charges of the two particles and explicit screening implemented with the relative dielectric constant $\epsilon_{\mathrm{r}} = 15$. ϵ_{r} represents the dielectric constant of vacuum.

Bonded interactions, angles and dihedrals are described by several equations, depending on the use case and necessities. The most common way of describing a bond is through a weak harmonic potential:

$$V_{\mathrm{bond}}(R) = \frac{1}{2}K_{\mathrm{bond}}(R - R_{\mathrm{bond}})^2 \tag{3}$$

with an equilibrium distance R_{bond} and a force constant of K_{bond}. The LJ interactions are normally excluded between bonded particles, since their distance is on average closer than nonbonded particles.

A weak harmonic potential of the cosine type is also often used to represent angles:

$$V_{\mathrm{angle}}(\theta) = \frac{1}{2}K_{\mathrm{angle}}[\cos(\theta) - \cos(\theta_0)]^2 \tag{4}$$

Depending on the type of chain, K_{angle} has a different value and a different equilibrium bond angle θ_0.

An improper dihedral angle potential is often used to prevent out of plane distortions:

$$V_{\text{id}}(\theta) = K_{\text{id}}(\theta - \theta_{\text{id}})^2 \tag{5}$$

1.3.2 Martini 3

Compared to its predecessor, Martini 3 has a few important changes. It is, however, still under development; therefore, results obtained with this force field are to be examined carefully.

This version of the force field utilizes the third type of interaction site – tiny bead (T) – in many more molecules other than DNA. It also introduces varying values of σ for all the possible combinations of beads, fixing several issues of its predecessor. In Martini 2, the N-N interactions and N-S interactions use the same σ, which creates unintended, invisible barriers between S and N beads that can lead to several unexpected behaviours[11].

Martini 3 also extends the interaction table with additional subtypes, smoothing the transition between bead types and improving the adaptability of the FF to new molecule types. Furthermore, all interaction levels underwent revision and reparametrization.

Martini 2 has several others issues[11], and Martini 3 attempts to solve most of them. A complete reference to the new version is unfortunately not yet available.

1.3.3 Proteins in Martini

Although proteins were not the initial scope of the Martini force field[9], they were subsequently added and refined along the years[10].

A few important adjustments are needed for correctly describing proteins, such as the use of structure-dependant bead types and the implementation of an elastic network. While the former is quite established and less relevant for the purpose of this report, the latter is often problematic, particularly when dealing with protein whose conformation can undergo drastic changes.

An elastic network consists of additional medium-long range interactions, usually described by harmonic potentials, which consolidate the secondary and tertiary structure of the protein, preventing unfolding. Such network is necessary with Martini proteins because – especially in Martini 2 – most of the interactions involved in proteins folding are too weak to keep the protein from unfolding.

Another approach which showed promising results in maintaining proteing fodling in Martini systems is the Gō-like model[12,13], which uses Lennard-Jones-like potentials to keep the proteing folding while still allowing for unfolding if forces permit it. This model was not tested in this project, but it is one of the natural follow-ups.

2 Methods

2.1 Generating a CG protein

2.1.1 Starting structure

Atomistic coordinates of the FraC protein were taken from the Protein Data Bank, PDBID: 4TSY[6]. Extra atoms such as lipids, water and other crystallization additives were removed manually using Pymol[14].

Atoms that were missing from the crystal structure (N-term) were also added using the Pymol build function.

Some residues presented double conformations, which are sometimes not handled properly by Martinize2. To avoid this problem, one of the two conformations was removed manually from the PDB file.

2.1.2 Coarse-graining with Martinize2

CG structures were generated using Martinize2[15]. Minor changes to the code were needed to properly generate the elastic network between merged chains.

Additionally, position restraints were added for the backbone beads with the `-p backbone` option to maintain their position fixed during energy minimization and the first equilibration steps. In some cases, position restraints were added for some lipid head groups by manually editing the lipid's topology.

N-term secondary structure was defined upon input in Martinize2 using the `-ss` option, using as a template the string automatically generated by `dssp` and adapting it to the needed secondary structure.

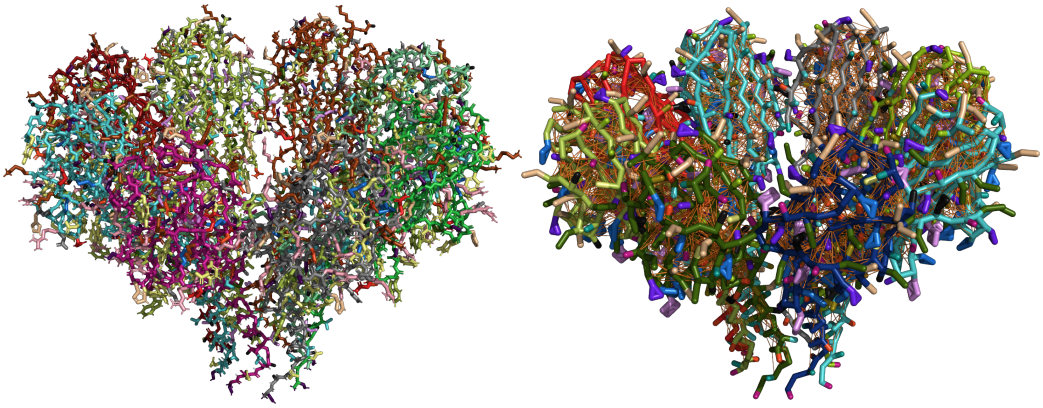
The elastic network was generated using the `-elastic` option of Martinize2. Superfluous bonds were then removed using the domELNEDIN script (official link currently unavailable). Adjustments to the script were needed in order to work on Martinize2 topologies and to be called from command line (see Supplemental Information).

Additional intermolecular elastic bonds were added in the main topology file as intermolecular interactions. Exclusions for non bonded interactions were also implemented for each of these bonds.

2.2 System setup

Each system was generated using insane[16]. The core options used were:

```
-pbc cubic -d 6.0 -dm -8 -z 14.5 -salt 0.15 -sol W -orient
```



(a) *Full-atom model of FraC.* ***(b)*** *CG model of FraC.*

Figure 2: *Comparison of atomistic and martinized models of FraC. Colors serve no purpose other than highlighting different backbone chains and different aminoacid sidechains. In the CG model, the elastic network is represented by thin orange lines.*

which determine box size, salt concentration and positioning of the protein in the membrane. The `-orient` option was sometimes removed and protein orientation was corrected manually because of inconsistencies in the automatic orientation performed by insane.

Lipid bilayers were generated with two additional options. To create a uniform bilayer with DOPC in both upper (`-u`) and lower (`-l`) leaflet:

```
-u DOPC -l DOPC
```

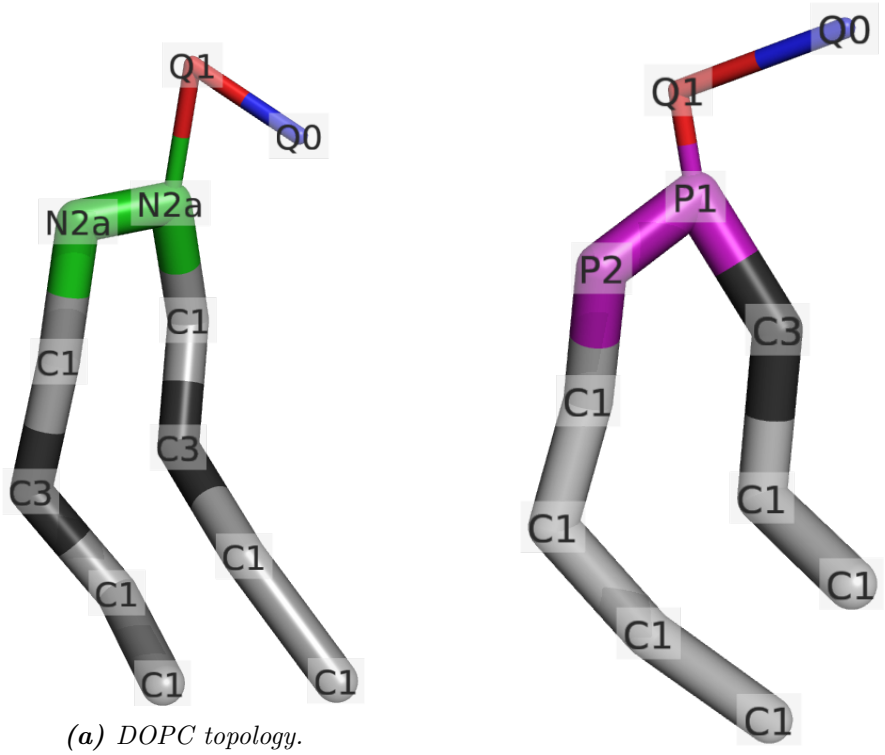
To create a bilayer with equally mixed DOPC and DPSM in both leaflets:

```
-u DOPC:1 -u DPSM:1 -l DOPC:1 -l DPSM:1
```

The topology files for the lipids used in these bilayers come from the Martini website (cgmartini.nl): DOPC as base lipids used in each membrane and DPSM as sphingomyelins.

2.3 Gromacs simulation setup

Gromacs simulations were run using mostly the same setup for each simulation. First, energy minimization was performed with all the position restraints on the protein backbone – and lipid heads, if applicable – activated. Afterwards, several equilibration steps were needed for the system to reach a sufficiently stable state for production. Each of these steps has



(a) DOPC topology.

(b) DPSM topology.

Figure 3: *Structure of DOPC and DPSM, the two lipids used in simulation. Beads are labelled with their Martini mapping.*

slighly different settings, and all of them can be found in the Supplemental Infromation.

The main production parameters are as following:

Before the production run, several equilibration steps were required to prevent the system from exploding, starting with an NVT system at low temperature and slowly progressing to an NpT with higher temperature and bigger time-step.

Each simulation was checked for convergence via ‘gmx energy‘, making sure bilayer properites such as bilayer thickness and area per lipid reached a state of equilibrium.

2.4 Mutant design

Mutants were designed based on inputs from previous work by prof. G. Maglia’s lab. Fusion proteins were created using PyMOL as a molecular building software. Scripts to automate the mutant creation and simulation

Parameter	Value
time step	20 fs
cutoff scheme	Verlet
coulomb type	reaction field
T-coupling	velocity-rescale
coupling groups	protein, lipids and solvent
reference T	310 K
P-coupling	parrinello-rahman (semiisotropic)
reference P	1.0 bar

Table 1: *Gromacs parameters for the production run of all the simulations.*

setup were designed with the help of Dr. T.H. Wassenaar and can be found in the Supplemental Information.

The hydrophilic core used was based on the homomeric yeast protein Lsm3 (PDBID: 3BW1)[17]. The transmembrane helix sequence was instead taken from FraC and attached to the N-term of the core.

Several linker sequences were tried, together with different conbinations of secondary structures and small changes to the helix sequence. Most of these sequences were either randomly picked to test different combinations, or based on insights from the lab of prof. G. Maglia.

2.5 Ionic flux simulations

To study the behaviour of ionic current through the pore, simulations were set up with a flat electric field perpendicular to the membrane. Such field is easily implemented in the gromacs simulation parameter file with the option:

```
E-z = 0 x 0
```

with x equal to the electric field strength in V/nm. Additionally, electrostatics were changed to PME to correctly capture the interactions between ions.

2.6 Analysis

Data analysis was performed mostly using built-in gromacs tools and python3. Simulations were analyzed via individual scripts designed to take care of each task, applied sequentially by a main script to each trajectory. All the scripts can be found in the supplemental information.

2.6.1 Data collection and preparation

Simulation data was saved as `.xtc` files. Before analysis, each trajectory was centered around the protein and fitted to the bilayer, in order to provide a consistent framework for data analysis.

Afterwards, most of the analysis was performed using python and the module MDAnalysis[18,19].

2.6.2 Pore size

The pore size was obtained via the python module HOLE[20], which calculates the pore radius at each Z value throughout the whole trajectory.

From this data, values of the lowest radius in each frame were used to describe the width of the pore.

Unfortunately, the resulting data alwayts contained a (low) percentage of corrupted frames, which resulted in incredibly high and obviously wrong values for the pore radius. These frames were not accounted for, and ignored in the calculation of the running average.

2.6.3 Flux

Flux of particles through the pore was obtained using fluxer[21], which calculates the number of crossing events through an xy-plane membrane. Though unnecessary for other purposes, this analysis required simulations to have an increased save frequency to have a good accuracy.

Trajectories were prepared for this step by removing the periodicity with `gmx trjconv`; the fluxer script requires this step in order to be able to correctly calculate the amount of times a particle traversed through neighbouring periodic images.

Flux of water was calculated in both directions to increase sampling and improve statistical accuracy. Flux of ions was intead calculated separately in both directions and for both positive and negative ions.

Though curated data for ion fluxes will not be presented in this report – not enough time was spent on this topic – some generic conclusions will be drawn from the preliminary results.

2.6.4 Graphs and representation

Data obtained with programs mentioned above were then manipulated using numpy and represented with matplotlib. Running averages were used to improve readability. Figures and movies were rendered using VMD[22] or PYMOL.

2.6.5 Scripting

Analysis was performed in parallel using several subscript and one main script to call all the others. The programming language used were python and bash (and in one occasion Tcl). The analysis script can be found in the Supplemental Information; a schematic representation of the structure of the script is displayed in the diagram in Figure 4.

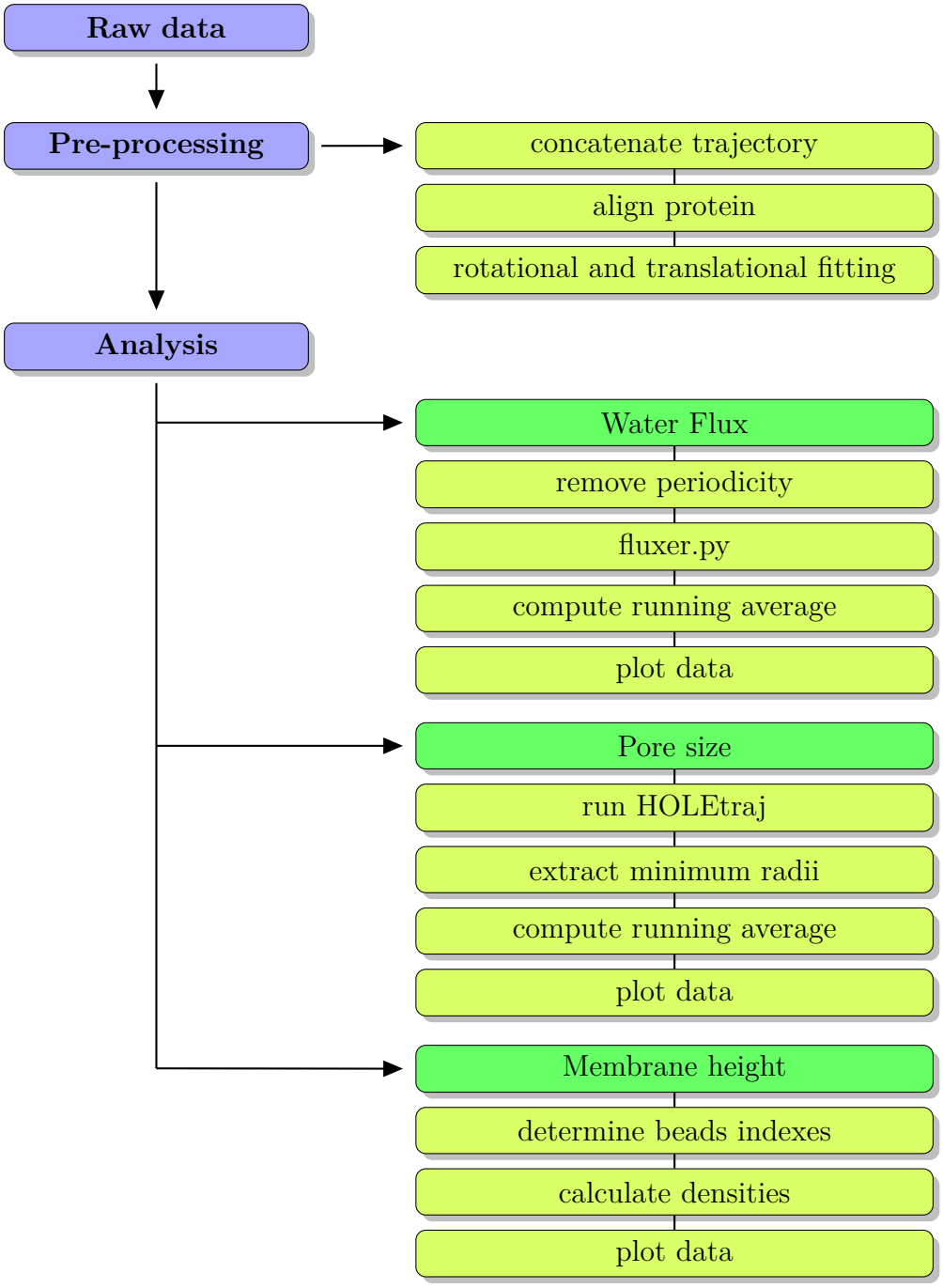


Figure 4: *Schematic representation of the scripted workflow for production and analysis.*

3 Results

Result are here presented divided in a few sections, to highlight the influence – or lack thereof – of individual changes to the model. Pore size and water diffusion flux are used as comparison properties to asses the goodness of the model. Parameters that change between simulations are: Martini version, membrane composition, elastic network, secondary structure of the N-term.

All simulations were produced in triplicates with randomly generated initial velocities. For each set, the reported figure representats only a single simulation out of the triplicates, unless simulations with the same setup showed relevant differences; in that case, multiple figures are shown and discussed case by case.

3.1 Reference simulations

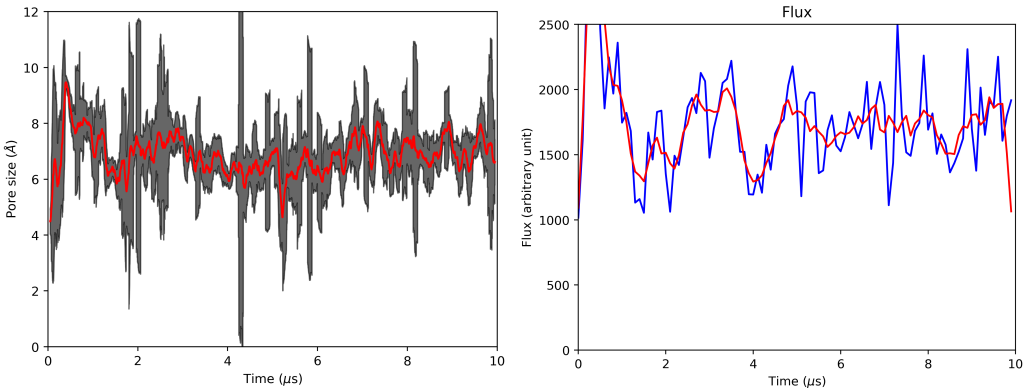
The following plots are results from the reference simulations, and as such they are used as comparison for all the other results. They show pore size and flux through the FraC pore in a symmetrical and uniform DOPC membrane.

The simulation in Martini 3 required the introduction of an elastic network, as opposed to Martini 2. Preliminary simulations showed that the latest version of the force field does not provide strong enough intermolecular interactions to keep monomers of the protein together throughout the simulation. Thus, the reference simulation for Martini 3 needed a partial elastic network to prevent the oligomer from falling apart. This elastic network was added between monomers in the water-soluble domain of the protein. The transmembrane domains can still move freely.

Disclaimer: It's important to point out that the HOLE program did not perform as expected. The resulting array contained often corrupted frames with inconsistent box sizes and pore radii; other frames were missing altogether. These problems were partially corrected for by adding back in the missing frames and by smoothing the data by showing a running average. The script in the Supplemental Information shows exactly how these problems were accounted for.

3.1.1 Martini 3

Results for the Martini 3 reference simulation are very consistent among the triplicates. The pore radius is not particularly stable, but the average size is in agreement with the literature value of about 6 Å (Figure 5a). This result is, again, confirmed by the flux analysis, which shows a high and consistent diffusion throughout the whole simulation (Figure 5b).



(a) *Pore size.* ***(b)*** *Water flux.*

Figure 5: *Reference simulation for Martini3. Pore radius is fluctuating a lot, but overall stays around the literature value of* 6 *Å. Flux is consistently high, indicating that the pore is open and allows constant passage of water molecules.*

3.1.2 Martini 2

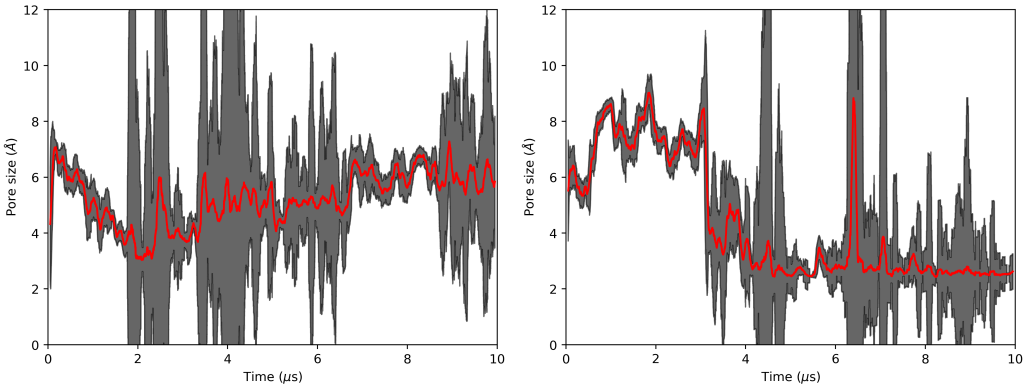
In this case, results are different between triplicates of this system. A pattern of behaviour can however be identified among the simulations – also taking into account previously run trajectories which used different equilibration protocols.

The pore radius fluctuates significantly during the simulation, often far from the radius calculated from the crystal structure of about 6 Å. Apart from this, these simulations can run for several microseconds without encountering severe issues (Figure 6a).

However, if at any time the pore closes completely, the trasmembrane helices get stuck in that conformation, never able to open the pore again for the rest of the simulation. A good example of this can be seen in Figure 6b.

It is important to note that the line at about 2.5 Å corresponds to a perfectly closed pore (which was determined through visual inspection and other, fully-closed simulations). This shifted ”zero value” is likely to be a

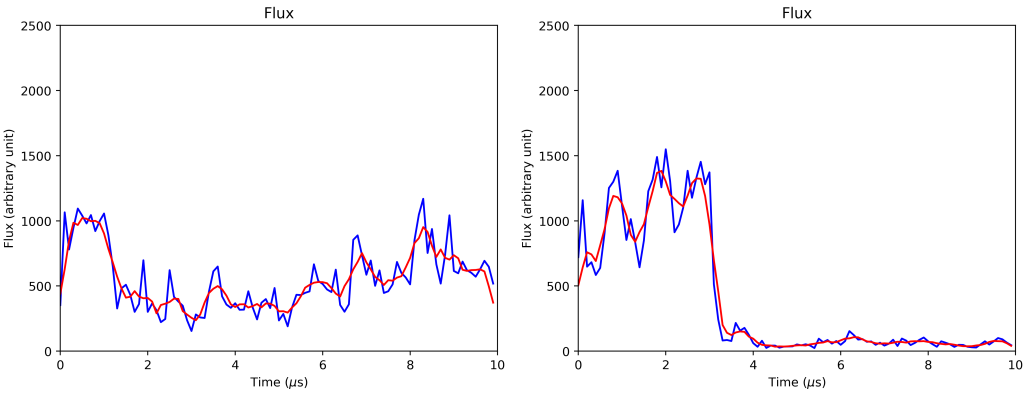
result of some level of incompatibility of the HOLE program with coarse-grained force fields – even though the higher VDW radius of Martini beads was taken into account.



(a) *Pore size, pore remains open.* ***(b)*** *Pore size, pore closes.*

Figure 6: *Pore size in reference simulation for Martini2. Radius is overall very unstable. If it ever closes (Figure 6b), it gets stuck and never opens again.*

Water diffusion through the pore agrees with these graphs, giving a clear distinction between open and close pore (Figure 7).



(a) *Water flux, pore remains open.* ***(b)*** *Water flux, pore closes.*

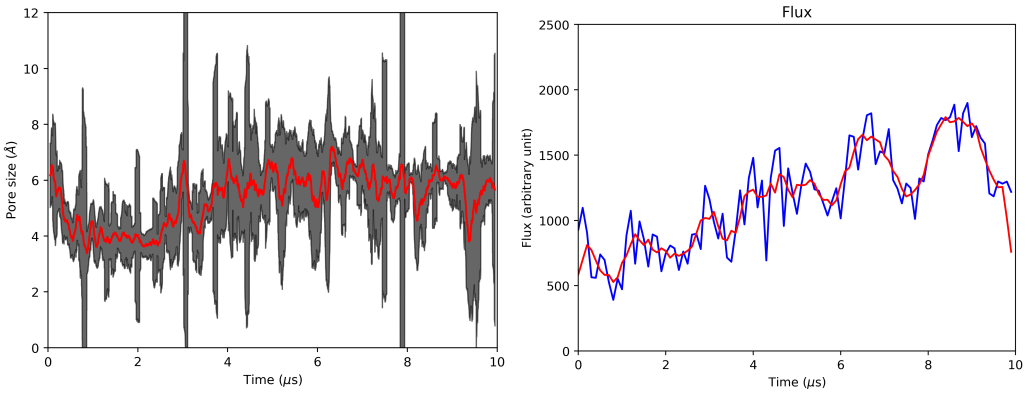
Figure 7: *Water flux in reference simulation for Martini2. This data confirms what was shown by the pore size: the pore is unstable and eventually closes, completely blocking the water flux.*

3.2 Effect of sphingomyelins

To study the effect of sphingomyelins, a system similar to the reference was used, the only difference being the membrane composition; instead of pure DOPC, the bilayer is made of 50% DOPC and 50% sphingomyelins.

3.2.1 Martini 3

Results seem to overall suggest that the stable form of the pore has – on average – a smaller radius compared to the pure DOPC membrane (Figure 8). The pore, however, is not stable enough over the course of the simulation to draw sure conclusions about this. A longer simulation time would guarantee that the pore is indeed at aequilibrium. Overall, the pore size – and more clearly the water flux – appears to tend to the 6 Å mark.

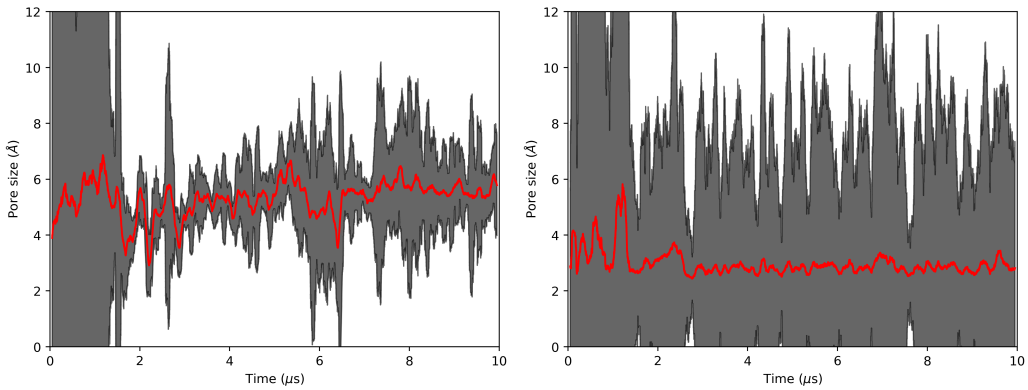


(a) *Pore size.* ***(b)*** *Water Flux.*

Figure 8: *Pore size and water flux in 1:1 DOPC:DPSM membrane with Martini 3. The pore appears smaller compared to the reference simulation, though its size increases towards the second half of the simulation.*

3.2.2 Martini 2

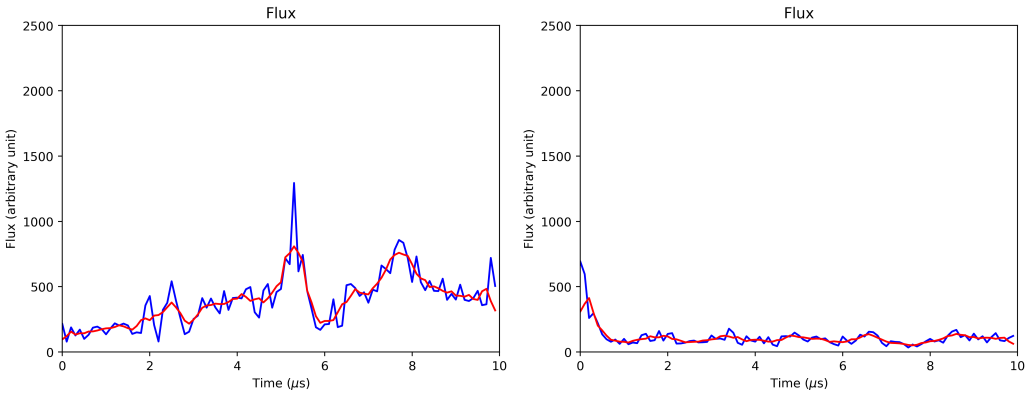
The pore does not show significantly different behaviour when embedded in a 50% sphingomyelins membrane; In fact, it shows the same trends as the reference simulation, with some pores staying open for the whole 10 μs of the simulation and others getting stuck in the close conformation (Figure 9).



(a) *Pore size, pore remains open.* ***(b)*** *Pore size, pore closes.*

Figure 9: *Pore size in 1:1 DOPC:DPSM membrane with Martini 2. Similarly, to the reference simulation, if at any time the pore closes, it remains stuck in the closed conformation (Figure 9b)*

Again, these findings are confirmed by the water flux through the pore (Figure 10).



(a) *Water flux, pore remains open.* ***(b)*** *Water flux, pore closes.*

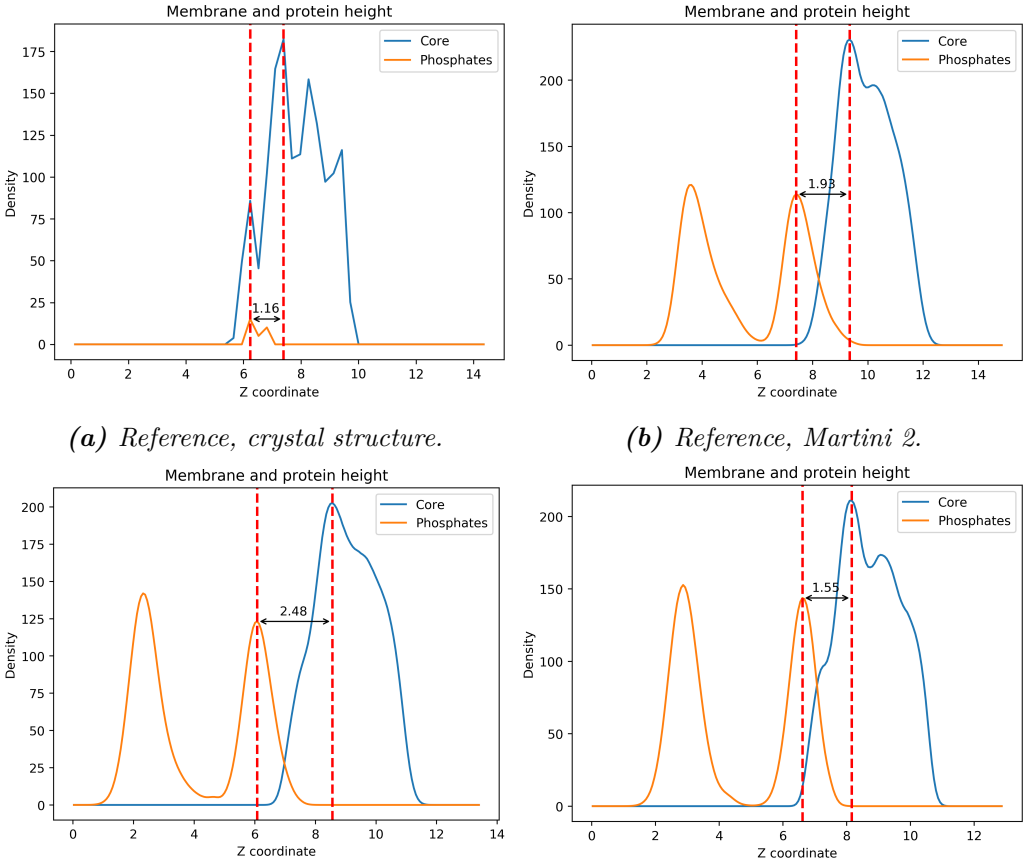
Figure 10: *Water flux in 1:1 DOPC:DPSM membrane with Martini 2. Results are in agreement with pore size (Figure 9)*

3.3 The membrane height problem

Previous simulations in Martini 3 showed promising results. However, visual inspection of the trajectories revealed an interesting difference compared to both the Martini 2 simulations and the crystal structure: the head groups of lipids adjacent to the protein – and subsequently of the surrounding membrane – were sitting at a different height compared to the protein.

To illustrate this, density maps of backbone beads and phosphate groups were calculated along the Z axis and the peaks were used to calculate the average height difference of membrane and protein. Figures 11b and 11c show respectively the membrane height in the reference simulations of Martini 2 and Martini 3. Figure 11a shows the same graph relative to the crystal structure, which was resolved with several lipids inserted in binding pockets with possibly relevant roles[6].

To examine possible problems this height difference might cause, a workaround was designed to counteract this issue. Intermolecular elastic bonds were added between the 24 lipids present in the crystal structure and backbone beads of their binding pockets. Such bonds were tuned to keep the lipids inside the binding pockets and hopefully drag the protein deeper in the membrane by creatingan hydrophobic layer between the transmembrane helices and the lipid bilayer.



(a) *Reference, crystal structure.*

(b) *Reference, Martini 2.*

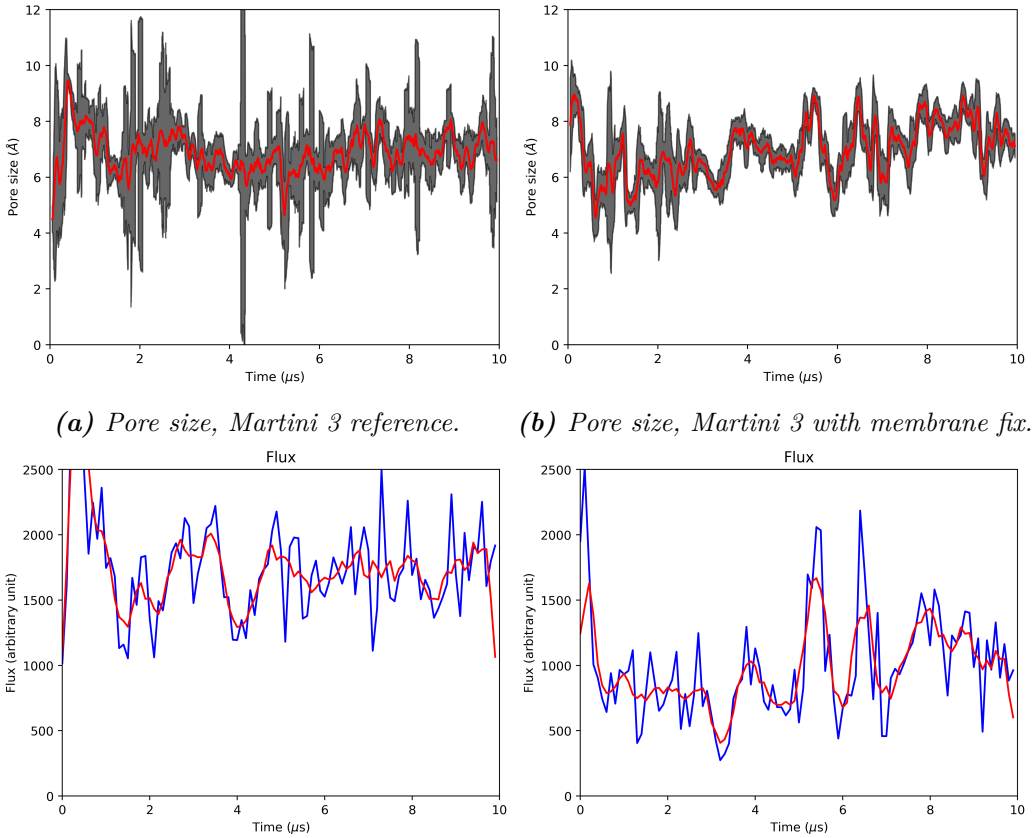
(c) *Reference, Martini 3.*

(d) *Martini 3, with lipid height fix applied.*

Figure 11: *The membrane height problem. The graphs show the z positioning of the protein core – in blue – relative to the lipid head groups. Compared to the Martini 2 reference simulation (11b) and the positioning of the lipid heads in the crystal structure (11a), the Martini 3 protein is considerably shifted along the z axis (11c). Once the membrane height fix is in place, FraC positioning in Martini 3 resembles much more that of Martini2 and the crystal structure (11d).*

Pore size and water flux were then measured for such altered systems and compared to reference simulations. Results do not show a significant difference in pore size (Figures 12a and 12b); however, they show a difference in pore stability, as highlighted by the significantly lower running standard deviation.

Moreover, water diffusion through the pore decreased to almost half, compared to the reference simulation (Figures 12c and 12d). Possibly, the stronger vibrations in the pore in the reference simulation may have caused higher agitation of the water molecules, promoting an overall stronger flux. Further investigation is needed to make sure the probram ran as expected.



(a) *Pore size, Martini 3 reference.* ***(b)*** *Pore size, Martini 3 with membrane fix.*

(c) *Water flux, Martini 3 reference.* ***(d)*** *Water flux, Martini 3 with membrane fix.*

Figure 12: *Behaviour of the pore in Martini 3 simulations, before (left) and after applying the membrane fix. While the pore size appears to become more stable, the water flux is strangely hindered by change in membrane height.*

For completenes's sake, the same modification was applied to Martini 2, resulting in very subtle change in membrane height and no significant differences in pore size and flux.

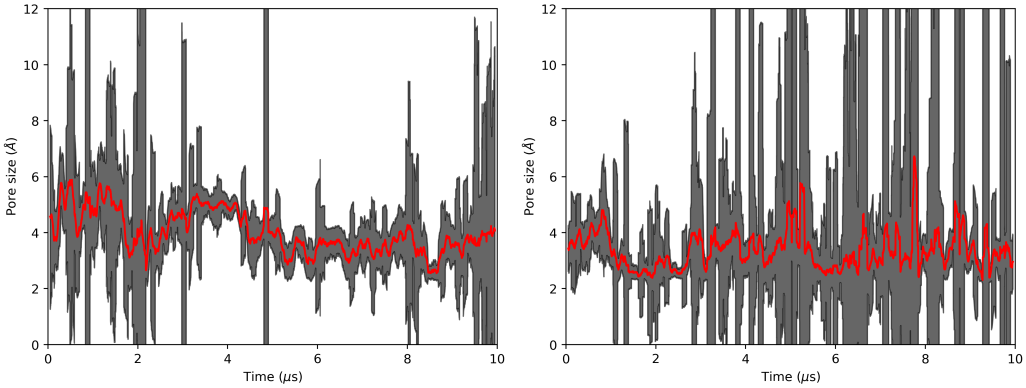
3.4 N-term secondary structure

The coordinates reported in the crystal structure were incomplete: the first few aminoacids of the N-term were missing from the PDB, and thus their conformation unknown. According to the literature, these residues are part of the transmembrane α-helix; however, their positioning at the N-term and the difficulty in resolving their coordinates with X-ray crystallography suggests they may have a less structured conformation.

To explore the implications of this possibility, simulations were run with the same setup as aforementioned, but changing the secondary structure of the N-term from helical to random coil.

3.4.1 Martini 3

A clear effect of changing the N-term secondary structure in Martini 3 the drastically reduced pore size (Figure 13a). Moreover, a similar behaviour to that of the reference Martini 2 simulation arises in one of the triplicates, where the pore closes and never gets unstuck again, although with very high, rapid fluctuations (Figure 13b). Visual inspection of the trajectory suggests that this is caused by the now overly flexible N-term repeteadly blocking and freeing the entrance of the pore.

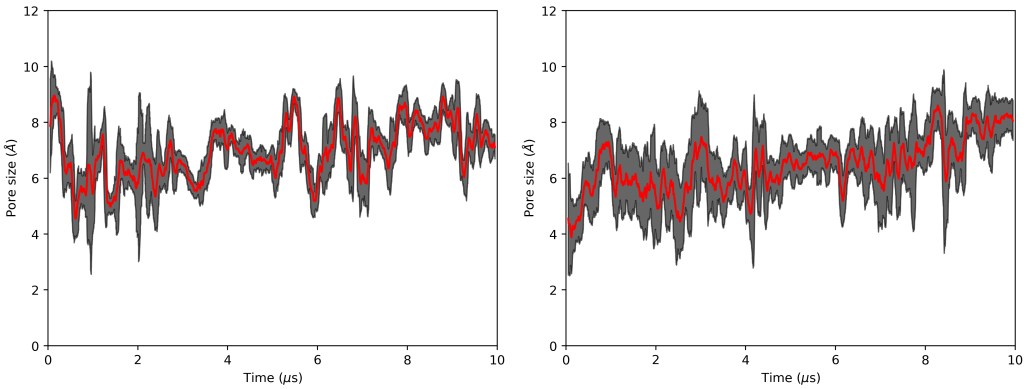


(a) *Pore size, coil N-term, simulation 1.* ***(b)*** *Pore size, coil N-term, simulation 2.*

Figure 13: *Pore size in Martini 3 with the secondary structure of the N-term set as random coil. Pore size is drastically reduced, and sometimes the lumen completely closes (13b), similarly to simulations in Martini 2.*

Interestingly, this behaviour changes when the random coil N-term is applied to simulations from the previous section – those with an elastic network between lipid and protein to fix the membrane height.

With these parameters, compared to the simulation with helix N-term (Figure 14a), the random coil N-term leads to slightly higher running standard deviation(Figure 14b); however, the pore remains wide open, without ever getting stuck into a closed conformation.

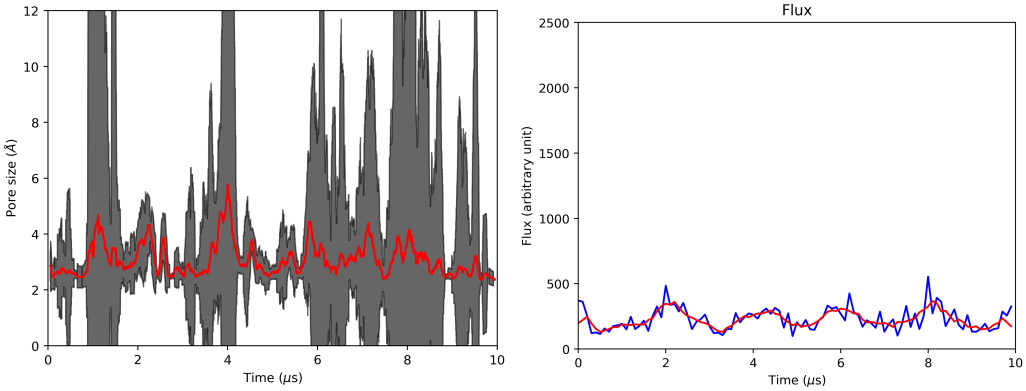


(a) *Pore size, coil N-term, membrane fix.* ***(b)*** *Pore size, helix N-term, membrane fix.*

Figure 14: *Pore size in Martini 3 with both the membrane fix and the random coil N-term applied. With the membrane fix applied, the secondary structure of the N-term seems to have little to no effect on the pore.*

3.4.2 Martini 2

Using a random coil N-term in Martini 2 yields a similar effect to that of Martini 3: the pore still closes completely, but is now able to partially reopen, albeit only shortly (Figure 15), probably due to the now much more flexible opening. This change is, however, not enough to compensate the overall stickiness of Martini 2.



(a) Pore size, coil N-term, Martini 2. *(b) Water flux, coil N-term, Martini 2*

Figure 15: *Pore size and water flux in Martini 2 with random coil N-term. Behaviour is not significantly different from the reference simulation, as the pore stays mostly closed, with negligible water flux.*

3.5 The role of the Elastic Network

The role of the elastic network in the Martini force field is to maintain the secondary, tertiary and sometimes quaternary structure of the protein so the pore does not lose its shape during the simulation. A well-balanced use of this system is tricky to achieve: an overly restrictive network would strongly bias the model towards a specific conformation, not allowing for structural rearrangement that may be needed for proper behaviour of the pore.

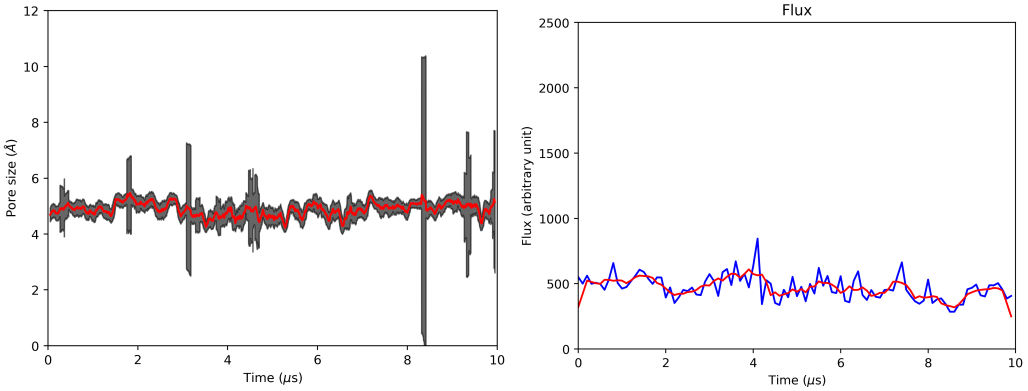
This can have the advantage of more reliably capture the behaviour of a protein with known active form structure; it will, however, inhibit the ability to describe proteins with unknown structure, such as mutants, greatly reducing the prediction capability of the model.

On the other hand, an elastic network that is too loose – or absent – would lead to a very unstable protein with unrealistic conformations, rendering the whole model useless.

In this section, systems with stronger elastic network were set up to analyze the effect of such a restriction on the pore and consider its viability for future modeling of pores such as FraC. Such system is also useful to see how a very stable pore range looks like compared to the rest of the simulations.

3.5.1 Martini 3

As expected, a full elastic network extending to the end of the α-helices, results in a stiff, semi-closed pore which remains in the same conformations as the crystal structure (Figure 16).



(a) *Pore size, full elastic network.* ***(b)*** *Water flux, full elastic network.*

Figure 16: *Pore size and water flux in Martini 3 with the secondary and tertiary structure of Frac completely controlled by the elastic network. Fluctuations in pore size are minimized, and radius is locked at a value similar to the crystal structure. Water flux is significantly reduced compared to the reference simulation.*

Results for pore size in the absence of an elastic network are not shown as a graph, as they are completely useless: the pore simply does not exist, with each monomer slowly drifting away from the rest and breaking the necessary multimeric conformation (Figure 17).

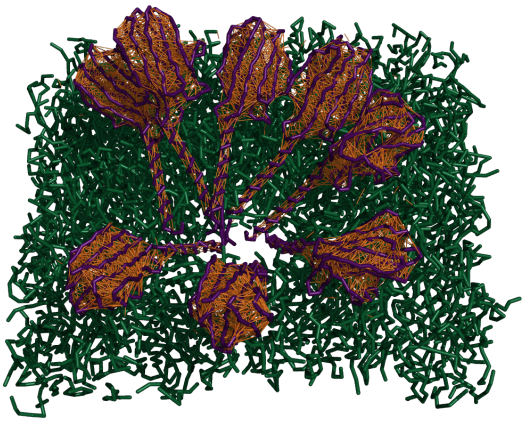
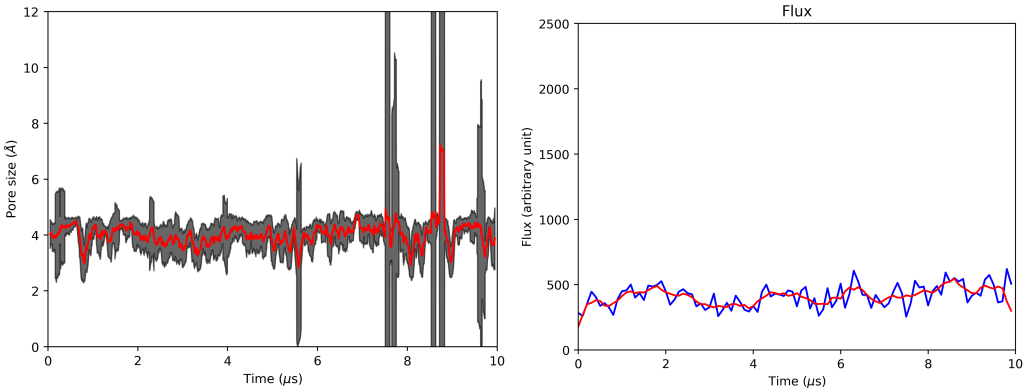


Figure 17: *Snapshot of the Martini 3 simulation without elastic network between monomers after a few microseconds. The monomers drift apart and FraC is no longer able to maintain a pore in the membrane.*

3.5.2 Martini 2

Martini 2 behaves very similarly to Martini 3 in the presence of a stricter elastic network, showing significant stiffness, a semi-closed pore and a stable, though small, flux.



(a) *Pore size, full elastic network.*

(b) *Water flux, full elastic network.*

Figure 18: *Pore size and water flux in Martini 2 with a strict elastic network. Behaviour is similar to Martini 3: pore is semi-closed, more stable and flux is heavily reduced.*

3.6 Force Field comparison

Comparison of Martini 2 and Martini 3 during these experiments highlighted some important differences between the two versions of the force field. Visual inspection of Martini 2 simulations revealed that the so called "stickiness" problem has important effects on hydrophobic environments such as the transmembrane domain of a pore. Helices tend to stick together once they touch each other, not allowing for recovery of the open-pore conformation once the lumen gets closed. This is also confirmed by several results in the previous sections. Conversely, such stickiness has a positive impact on intended protein-protein interaction: in Martini 3, when the intermolecular elastic network is absent, the monomers interact to weakly and soon enough the complex falls apart.

Martini 3 is still showing some uninteded behaviour. The "membrane height problem" may be caused by a wrongly tuned hydrophobicity of some residues at the water-membrane interface – particularly Tryptophan, which in its current state might be more hydrophilic than intended. Further investigation and testing is required; this result will likely be useful for further development of the Martini 3 interaction table.

Both force fields still present a common limitation: an elastic network is needed to retain some level of folding – especially in Martini 3, where backbone-backbone and hydrophobic interactions seem to be particularly small. Unfortunately, the presence of an elastic network prevents major conformational changes; for this reason, the Gō-like model may be a better approach[12,13].

3.7 Scripts and computational tools

3.7.1 Martinize2

Martinize2, although still in development, was able to generate both Martini 2 and Martini 3 topologies without the need of different commands. Moreover, when used with the Martini 3 force field, it did not require any additional adjustments on improper dihedrals and other parameters – steps that are instead currently required if using the original Martinize.

At the time of this project, Martinize2 was unable to merge chains and then add an elastic network between monomers. However, it was as simple as swapping two sections of code for it to work as intended.

A feature that was sorely missing was the capability to easily override the automatically generated secondary structure (which was needed for manually setting the N-term); at the time of writing, this feature is finally being added to the program.

3.7.2 Insane.py

Insane is a quick and easy tool to build membranes with embedded proteins. Using it for pore proteins can lead to problems, as it is unable to put solvent inside the pore itself. The problem can be solved by either solvating the system after building it or by having longer equilibration steps with position restraints for any non-solvent particle.

3.7.3 HOLEtraj

Although powerful, this tools turned out to be very slow and unreliable. Several frames gave corrupted results – although the frames are not corrupted themselves – and some others were completely missing from the resulting numpy array. This led to significant disturbance in data and graphs. These problems may be caused by slight incompatibilities with coarse-grained force field, since the program was developed with only atomistic force field in mind.

A more likely reason could be the unusual and unstable nature of the pore: instead of the more common pores, mostly cilindrical and clearly delineated,

the program was used on a cone-shaped pore with a dynamically changing surface.

Calculation of the pore size requires Van der Waals radii specific to the force field used. As such files were not available for Martini 2 and Martini 3 values, they needed to be crated from scratch. Resulting files can be found in the Supplemental Information.

3.7.4 Fluxer.py

The Fluxer script worked as intended. The need of a high time resolutions to correctly assess jumps through the membrane required more frequent saving of the trajectory, but overall it did not slow down the production run significantly. Analysis via this script was rather quick and gave additional data (such as claffication of jumps by type) that was not used in this project, but could be useful for future applications.

Overall, I would prefer fluxer.py over HOLEtraj for similar experiments in the future: it appears more realiable, consistent and exhaustive. Of course, the two tools measue different properties, but for the purposes of computational electrophysiology their scope overlaps significantly.

4 Discussion and Conclusions

This study showed some strengths and weaknesses of the Martini force field for the study of pore-forming-proteins and possibly other trasmembrane proteins.

Overall, the model was able to capture the dynamic nature of the FraC protein nanopore. The Martini 3 force field showed good improvements on the stickiness problem compared to the old versions, although some issues may still linger, particularly to interactions between hydrophobic regions.

Results also highlight the importance of secondary structure bias – especially at the termini – on the overall behaviour of dynamic regions of the protein; closely related to this problem is how big an effect can have a different elastic network. These aspects make modeling dynamic α-helical nanopores extremely tricky.

As a generalization, it proves necessary to carefully control and inspect dynamically structured termini in coarse-grained simulations, since small changes in the model may have a big impact on the results.

More detailed insights on the several aspects of the present model are outlined below.

4.1 Importance of sphingomyelins

Contrary to what can be drawn from literature research, the presence or absence of sphingomyelins in the membrane did not show a strong effect on the pore formation and/or its efficacy. This agrees with the one of the current hypothesis, stating that sphingomyelins might be involved in the formation of the polymer and in membrane recognition, rather than in pore stability and function.

However, preliminary analysis on density maps of sphingomyelin distrubution revealed that there is no preferential binding of sphingomyelins to the transmembrane domain of FraC, which may suggest that either Martini or the current models of FraC and DPSM do not provide enough detail to capture such interactions.

Simulations in Martini 3 suggest that a small reduction of the pore radius could be an effect of the presence of sphingomyelins, but there is not enough data to confirm this; either way, it seems too small of a change to warrant the significant importance sphingomyelins in PFPs efficacy reported in literature[23].

Overall, Martini 2 and Martini 3 did not show significant differences on the effect of sphingomyelin on the nanopore.

4.2 Membrane height

It's hard to draw conclusions about the role of the membrane height. One effect of forcing the membrane to the supposedly correct height appears to be the higher stability of the pore. However, this may be simply the result of the rigidity introduced by the lipids that were anchored to the sides of the transmembrane helices.

Overall, results suggest that both models should be able to capture the variability of the pore size, and both may be applicable to electrophysiology experiments as long as this issue is kept in mind and sufficiently considered.

Such considerations apply only to Martini 3 models, since Martini 2 did not give rise to the membrane height to begin with.

4.3 Helix vs Coil N-term

N-term secondary structure has a significant impact on the behaviour of the pore, both relating to pore stability and radius.

A random coil N-term gives the Martini 2 systems the ability of breaking free from the closed pore state, although not enough to let the pore open upompletely again. Martini 3 shows an opposite trend, with the coil N-term seemingly causing the closure of the pore in a fashion similar to that of Martini 2.

Interestingly, when the coil N-term is introduced in the Martini 3 simulation with the membrane height fix in place, these two changes together show the best results: the pore is open at around 6 Å, with the coil N-term increasing the medium-term stability of the pore (although increasing slightly the short-term deviation).

4.4 Role of the Elastic Network

When the protein is constructed with a much more restrictive elastic network, the resulting pore is extremely stable but is forced to mintain its initial conformation. This means that the protein is strongly biased towards the crystal structure conformation, which might be different from the *in vivo* conformation.

Such a model could be useful for computational electrophysiology of protein with known *in vivo* conformation; while still unable to provide an absolute quantitative value for predicted current, it could still assess relative current values as well as giving qualitative insights about the internal properties of the pore lumen.

On the other hand, it has very limited application to protein with unknown or partially known structure – such as mutants – whose conformation may have a strong impact on other qualities of the pore.

5 Future Perspectives

Some further investigation on FraC is needed in order to be able to accurately characterize and describe the protein in a molecular dynamics simulation. Particularly, the role of sphingomyelins is not yet clear. *In vitro* experiments agree on the critical role of sphingomyelins, but no clear explanation of the mechanism is available.

5.1 Sphingomyelins

To this end, it can be useful to design some additional experiments to investigate how this interaction happens. Some preliminary tests were tried as part of this project: the initial hypothesis was that sphingomyelins are embedded in the regularly spaced cavities on the outside of FraC, at the interface between water and membrane. Though with the caveat that the force field might be misrepresenting them, sphingomyelins appeared to have no preference for the protein over the rest of the membrane.

A better understanding can possibly be reached through *in vitro* experiments. For example, to understand if sphingomyelins have an important role on membrane attack rather than pore efficacy, it may be useful to embed the proteins inana SM-free membrane and then slowly add sphingomyelins while detecting changes in the current.

5.2 Computational Electrophysiology

A natural follow up to the present work is to test the passage of current through the pore via computational electrophysiology. Preliminary tests showed that this is possible by using integrated functions of gromacs. Results, though incomplete, showed that ions – particularly the positive ones – had the tendency of getting stuck in the pore, possibly attracted by the negatively charged and polar side chains of the lumen. This resulted in a higher flux of negative ions through the pore (almost double) compared to the positive ones.

5.3 Mutants

Initially the goal of this project, the generation and systematic analysis of mutants showed very soon several problems. The current theoretical model of FraC is limited and still presents a few challenges, such as the sphingomyelins problem. Similarly, the third version of the Martini force field needs further development, since as of now it still presents some issue. Additionally, a better system may be necessary to replace the elastic network approach, such as the Gō-like model.

A relevant issue highlighted by the N-term problem is the importance of the bias in the secondary and tertiary structure. A possible way to circumvent it is to testseveralt conformations for each mutant, but there's a limit to how much we can predict.

Once these problems are solved and the Wild Type FraC is properly characterized, new steps can be taken to analyze mutants, comparing them against the Wild-Type. Computational electrophysiology using fluxer.py showed promising results, and may be a good way to analyze new mutants.

References

[1] Jordanka Zlatanova and Kensal van Holde. Single-molecule biology: What is it and how does it work? *Mol. Cell*, 24(3):317–329, November 2006. (4)

[2] Stefan Howorka and Zuzanna Siwy. Nanopore analytics: Sensing of single molecules. *Chem Soc Rev*, 38(8):2360–2384, August 2009. (4)

[3] Gang Huang, Kherim Willems, Misha Soskine, Carsten Wloka, and Giovanni Maglia. Electro-osmotic capture and ionic discrimination of peptide and protein biomarkers with FraC nanopores. *Nature Communications*, 8(1):935, October 2017. (4, 5)

[4] Kherim Willems, Veerle Van Meervelt, Carsten Wloka, and Giovanni Maglia. Single-molecule nanopore enzymology. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, 372(1726), August 2017. (4)

[5] James Clarke, Hai-Chen Wu, Lakmal Jayasinghe, Alpesh Patel, Stuart Reid, and Hagan Bayley. Continuous base identification for single-molecule nanopore DNA sequencing. *Nat Nanotechnol*, 4(4):265–270, April 2009. (4)

[6] Koji Tanaka, Jose M. M. Caaveiro, Koldo Morante, Juan Manuel González-Mañas, and Kouhei Tsumoto. Structural basis for self-assembly of a cytolytic pore lined by protein and lipid. *Nat Commun*, 6:6337, February 2015. (5, 10, 22)

[7] Nejc Rojko, Mauro Dalla Serra, Peter Maček, and Gregor Anderluh. Pore formation by actinoporins, cytolysins from sea anemones. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1858(3):446–456, March 2016. (5)

[8] Shidi Zhao, Laura Restrepo-Pérez, Misha Soskine, Giovanni Maglia, Chirlmin Joo, Cees Dekker, and Aleksei Aksimentiev. Electro-Mechanical Conductance Modulation of a Nanopore Using a Removable Gate. *ACS Nano*, 13(2):2398–2409, February 2019. (5)

[9] Siewert J. Marrink, H. Jelger Risselada, Serge Yefimov, D. Peter Tieleman, and Alex H. de Vries. The MARTINI force field: Coarse grained model for biomolecular simulations. *J Phys Chem B*, 111(27):7812–7824, July 2007. (6, 7, 8)

[10] Luca Monticelli, Senthil K. Kandasamy, Xavier Periole, Ronald G. Larson, D. Peter Tieleman, and Siewert-Jan Marrink. The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J Chem Theory Comput*, 4(5):819–834, May 2008. (6, 8)

[11] Riccardo Alessandri, Paulo C. T. Souza, Sebastian Thallmair, Manuel N. Melo, Alex H. de Vries, and Siewert J. Marrink. Pitfalls of the Martini Model. *J. Chem. Theory Comput.*, September 2019. (8)

[12] Adolfo B. Poma, Marek Cieplak, and Panagiotis E. Theodorakis. Combining the MARTINI and Structure-Based Coarse-Grained Approaches for the Molecular Dynamics Studies of Conformational Transitions in Proteins. *J. Chem. Theory Comput.*, 13(3):1366–1374, March 2017. (9, 30)

[13] Sebastian Thallmair, Petteri A. Vainikka, and Siewert J. Marrink. Lipid Fingerprints and Cofactor Dynamics of Light-Harvesting Complex II in Different Membranes. *Biophysical Journal*, 116(8):1446–1455, April 2019. (9, 30)

[14] Schrödinger, LLC. The PyMOL molecular graphics system, version 1.8. November 2015. (10)

[15] P.C.Kroon, J.Barnoud, T.A.Wassenaar and S.J.Marrink. Martinize2 and vermouth: The ultimate resolution transformation tools. Manuscript in preparation. Source at: https://github.com/marrink-lab/vermouth-martinize, 2019. (10)

[16] Tsjerk A. Wassenaar, Helgi I. Ingólfsson, Rainer A. Böckmann, D. Peter Tieleman, and Siewert J. Marrink. Computational Lipidomics with insane: A Versatile Tool for Generating Custom Membranes for Molecular Simulations. *J. Chem. Theory Comput.*, 11(5):2144–2155, May 2015. (10)

[17] N. Naidoo, S. J. Harrop, M. Sobti, P. A. Haynes, B. R. Szymczyna, J. R. Williamson, P. M. G. Curmi, and B. C. Mabbutt. Crystal structure of Lsm3 octamer from Saccharomyces cerevisiae: Implications for Lsm ring organisation and recruitment. *J.Mol.Biol.*, 377:1357–1371, 2008. (13)

[18] Richard J. Gowers, Max Linke, Jonathan Barnoud, Tyler J. E. Reddy, Manuel N. Melo, Sean L. Seyler, Jan Domański, David L. Dotson, Sébastien Buchoux, Ian M. Kenney, and Oliver Beckstein. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics

Simulations. *Proceedings of the 15th Python in Science Conference*, pages 98–105, 2016. (14)

[19] Naveen Michaud-Agrawal, Elizabeth J. Denning, Thomas B. Woolf, and Oliver Beckstein. MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *J Comput Chem*, 32(10):2319–2327, July 2011. (14)

[20] Oliver S. Smart, Joseph G. Neduvelil, Xiaonan Wang, B. A. Wallace, and Mark S. P. Sansom. HOLE: A program for the analysis of the pore dimensions of ion channel structural models. *Journal of Molecular Graphics*, 14(6):354–360, December 1996. (14)

[21] Manuel N. Melo, Clément Arnarez, Hendrik Sikkema, Neeraj Kumar, Martin Walko, Herman J. C. Berendsen, Armagan Kocer, Siewert J. Marrink, and Helgi I. Ingólfsson. High-Throughput Simulations Reveal Membrane-Mediated Effects of Alcohols on MscL Gating. *J. Am. Chem. Soc.*, 139(7):2664–2671, February 2017. (14)

[22] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, 14(1):33–38, February 1996. (15)

[23] Peter Schön, Ana J. García-Sáez, Petra Malovrh, Kirsten Bacia, Gregor Anderluh, and Petra Schwille. Equinatoxin II Permeabilizing Activity Depends on the Presence of Sphingomyelin and Lipid Phase Coexistence. *Biophysical Journal*, 95(2):691–698, July 2008. (32)