

MAPPING THREE DIMENSIONAL INTERACTIONS BETWEEN
BIOMOLECULES AND ELECTRIC FIELDS

By

Joseph Patrick Brian, P.E.

B.S. Chemical Engineering, 1980

P.E. Chemical Engineering, 1985

P.E. Electrical Engineering, 1988

A Dissertation

Submitted to the Faculty of the

J.B. Speed School of Engineering

of the University of Louisville

in Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy
in Chemical Engineering

With Specialization in
Computational Chemistry / Molecular Dynamics

Department of Chemical Engineering
University of Louisville
Louisville, Kentucky

April 2020

Copyright 2020 by Joseph Patrick Brain

All rights reserved

MAPPING THREE DIMENSIONAL INTERACTIONS BETWEEN
BIOMOLECULES AND ELECTRIC FIELDS

By

Joseph Patrick Brian, P.E.

B.S. Chemical Engineering, 1980

P.E. Chemical Engineering, 1985

P.E. Electrical Engineering, 1988

Dissertation approved on

April 14, 2020

by the following dissertation Committee:

Dissertation Director Name

Second Committee Member Name

Third Committee Member Name

Fourth Committee Member Name

DEDICATION

First and foremost, I would like to publicly thank my
Lord and Savior Jesus Christ
for saving me,

and orchestrating the events of my life to make the earning of this degree possible.

I couldn't have done it without him.

I thank him because
I know there will be no Ph.D.'s in heaven,
only children of God,
and that he gave his life on the cross
that I could be one of them,

and for allowing me to kneel in his presence
on the evening of 30 November 2009.

ACKNOWLEDGMENTS

I would like to thank my advisor, mentor and friend, Dr. Vance Jaeger for teaching me GROMACS, Unix, Metadynamics, PLUMED, CHARMM-GUI, AMBER and all the tricks he learned at the Max Plank Institute for Biophysical Chemistry in Gottingen, Germany.

My best friend Joe Mahoney for being there through thick and thin. There is a friend who sticks closer than a brother. Prov 18:24b

My parents Joe and Marlyn for bringing me into this world and caring for me.

My undergraduate professors at the University of Kentucky, especially Dr. Grieves.

My former employer, BFGoodrich / Westlake Chemical Corporation, who gave me a wonderful career, and indulged my quest to never stop learning.

My boss for most of my career, Jim Best, for being a fantastic boss, always having my back, never limiting my opportunities, and took the time to write one of my letters of recommendation.

Dr. Timothy Dowling, for his time and encouragement during my first serious time in a college classroom in 34 years and writing one of my letters of recommendation.

Dr. Mahendra Sunkara, for believing I still had plenty of gas in the tank and making a way for me to enter the program and earn my way through on such short notice.

Dr. Joshua Spurgeon, my first advisor, for always being accommodative of my needs to simultaneously deal with certain personal issues at that time related to the worst personal situation of my life.

The support of the faculty and staff of the Department of Chemical Engineering and the Conn Center for Renewable Energy Research even through the rough patches, especially Dr. Fried.

My first compatriot post-docs, Dr. Bijandra Kumar and Dr. Sudesh Kumari.

My classmates and friends at school, especially Farnaz Minooei, Tim Dubbs, Marzieh Moradi and Mike Martin.

Harrison Simrall, Assistant Director of IT Research Computing for all his help when I needed software packages updated on the Cardinal Research Cluster.

My committee members for their time and support.

ABSTRACT

MAPPING THREE DIMENSIONAL INTERACTIONS BETWEEN BIOMOLECULES AND ELECTRIC FIELDS

Joseph Patrick Brian

April 14, 2020

The fields of computational chemistry (CC) and molecular dynamics (MD) simulations on supercomputers have benefitted greatly in recent years largely due to the fortuitous development of graphical processing units (GPU's) developed to make the visual depictions in video game consoles nearly lifelike with the incorporation of light raytracing algorithms. The mathematics of calculating ray traces are identical in nature to those needed for calculating the particle trajectories of atoms in MD simulations. GPU's are vastly superior to familiar computer CPU's for this work, and the concomitant increases in calculation speeds are making all-atom simulations, recently impractical for all but the smallest systems viable. In MD, biostructures consisting of hundreds and thousands of biomolecules needed to simulate the life functions involved with cell respiration, having a million or more individual atoms are now becoming feasible. With all-atom simulations of biostructures now come new opportunities to study with some measure of fidelity heretofore unknowable engineering details of the role non-uniform electric fields play in the conformations and functions of these structures that enable and control cell respiration and maintenance. This dissertation details new simulation and post processing capabilities developed by others and me that are collectively enabling exciting new understanding of these processes that hold the promise of improving clinical treatment for patients today.

TABLE OF CONTENTS

Dedication	iii
Acknowledgments	iv
Abstract	v
Table of Contents	vi
List of Tables	vii
List of Figures	viii
Introduction	1
Section 1	11
Insights into the Molecular Mechanisms of Electroporation from Computational Electrophysiology	11
Supporting Material.....	46
Section 2	78
Post-Processing and Visualization (Electric Fields and Computational X-Rays) .	78
Section 3	?
Effects of Salt, Water, and Protein Force Fields on Protein Folding Thermodynamics in Molecular Simulations	?
Supporting Material.....	?
Section 4	?
Nitrogenase	?
Supporting Material.....	?
Appendix A: Listings of programs developed to make this work possible	?
Vita	last page

LIST OF FIGURES

FIGURE	PAGE
1. Initial attempt to verify integrity of data after voxel integration	5
2. Final attempt to plot integrated charges using transparent pixels.....	6
3. First two partial attempts at plotting integrated data with non-transparent pixels ..	6
4. First computational x-ray	7
5. Example of three-dimensional electric field vector map for water moiety	7
6. Double membrane geometry.....	16
7. Simulation workflow to generate CompEL production runs	24
8. Distribution of core water molecules in induced pores	27
9. Membrane charge distribution, electric field and potential	29
10. Ion permeation rates vs. pore size and charge imbalance – no SCMTR	30
11. Pore disintegration dynamics after removal of charge imbalance	35
12. Simulation x-y plane box area vs. charge imbalance	36
13. SCMTR Molecule	48
14. History of pore core water molecules in upper and lower membranes	49
15. Details of electrical parameters for pre-pore stage	50
16. Results of GROMACS potential analysis for a 40 ns production run	51
17. Vertical alignment of choline moieties extending into center of pore	52
18. Ion permeation rates vs. pore size and charge imbalance – with SCMTR	53
19. Large system after 125 ns	54

LIST OF FIGURES

FIGURE	PAGE
20. Plots for production run history with +/- 4 e- charge imbalance	59
21. Plots for production run history with +/- 5 e- charge imbalance	60
22. Plots for production run history with +/- 6 e- charge imbalance	61
23. Plots for production run history with +/- 7 e- charge imbalance	62
24. Plots for production run history with +/- 8 e- charge imbalance	63
25. Plots for production run history with +/- 9 e- charge imbalance	64
26. Signature pattern of water molecule counts.	68
27.	
28.	
29.	
30.	
31.	
32.	
33.	
34.	
35.	
36.	
37.	
38.	
39.	
40.	

INTRODUCTION

Molecular dynamics had its birth as computational structural biology in 1968 when Michael Levitt programmed the coordinates of one of the two known protein structures at the time into a computer. The structure of the enzyme lysozyme had only been determined three years earlier from x-ray crystallography, after the structure of myoglobin had first been determined in 1959. Pictures of both had been published in the journal Scientific American. He'd read both articles, the first as a teenager and the structures of the molecules had always fascinated him. Working in Israel in an unfilled post-doc position for a year as a condition to be accepted to the Ph.D. program at Kings College in London, he had just finished building the first physical model of a protein, Lysozyme. Because he had the listing of all the atom coordinates needed to build the physical model, upon its completion he immediately set about building a computer model of it.

Also, in 1968 Shneior Lifson and Arieh Warshel published the first consistent force-field (mathematical) model of the bond strain energies and Lennard-Jones parameters for several n-alkanes and cycloalkanes. Levitt and Lifson applied this energy modeling technique to bigger molecules of more general structure to include the elements of hydrogen, carbon, nitrogen, oxygen and sulfur so it could be applied to their computerized models of lysozyme and myoglobin. Their 1969 paper of this work on macromolecular energy minimization established the foundation of what is now practiced in the much

broader realm of MD, which contributes greatly to knowledge of properties of non-biological systems as well.

At the time, the significance was that the technique provided first revelation of the functional structure of these proteins. X-ray crystallography revealed the protein crystal structure, but to begin to attempt to understand how the proteins worked, they first needed to have a knowledge of their structure when in their biologically active form. Unlike most other areas of chemistry related research, biologically active systems are dependent upon being in their biologically active form, which only exist between very narrow constraints of the environment. Hydration, pH, salinity, concentration, temperature and electric fields cannot be maintained while once active structures are prepared for electron microscopy to discover in detail how they work. This is an extreme problem, and extreme methods have been developed to try to overcome them.

Transmission electron cryomicroscopy (commonly known as cryo-EM) is probably the most capable of these methods, requiring hyperfast cryofixation of samples in a bath of liquid ethane at cryogenic temperatures to solidify the structures before they have time (of which they need fantastically little) to complete the free energy (and physical conformation) rearrangement driven by the change in temperature. Unfortunately, the technique is only applicable to structures that are relatively static in shape, another way of saying constant over time. But many biostructures are very dynamic, changing shape to open or close membrane channels, to propagate membrane potentials along axions, dock and undock cofactors to enzymes, etc., etc., all generally taking place in the sub-microsecond time frame. Critical functions that will likely never be imaged in their native environment at the molecular scale.

These are the inaccessible to direct observation frontiers that supercomputers and advances in force-field development and simulation algorithms are beginning to conquer at the all-atom level of detail. This dissertation contains some advances in these frontiers made possible by huge speed advances, new simulation software capabilities introduced only last year, and lots and lots of my own hard work. Even then, the most significant outcome of this work, computational x-rays, was not planned or imagined in advance.

I had been working on lipid bilayer membrane (BLM) systems for about a year and a half, specifically modelling the behavior of BLMs to develop pores in the presence of sufficient transmembrane electric fields. This procedure known as electroporation is used clinically for patient treatments for various purposes. Section 1 of this dissertation contains my published research in the area. The research and published material I read during that work on its practice suggested to me three things. One, that there was no good theoretical explanation of the behavior that was observed, which led to the fact that; Two, it was practiced mostly as a black art, somewhat blindly attempting to increase treatment efficacy without increasing the rate of cell lysis upon being subjected to the treatment. Thirdly, I had recognized that the simulation programs presently had no ability to investigate the role non-uniform electric fields were playing in this behavior. I had a conviction they were playing a more significant and interesting role than anyone seemed to realize, one I was very interested in determining for myself.

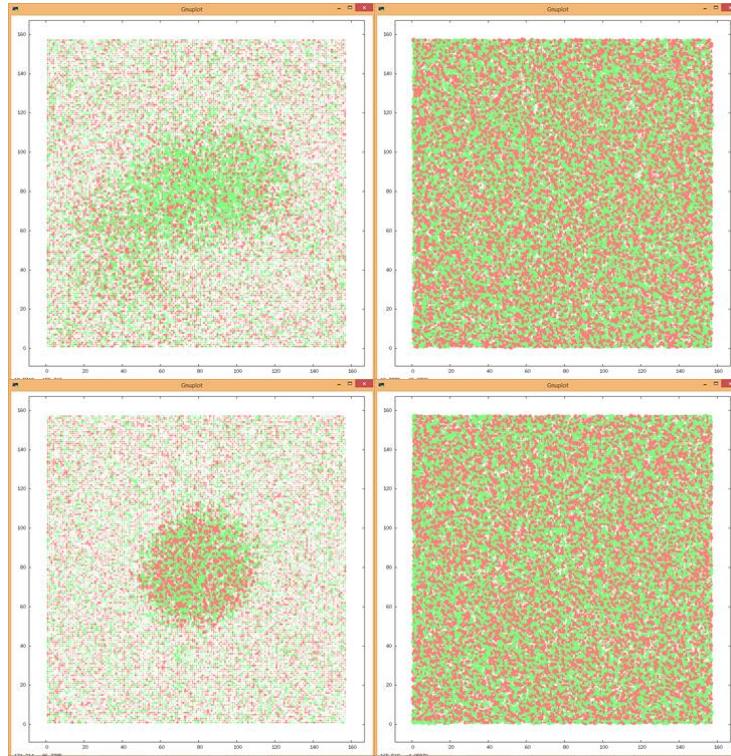
This was the genesis of my dissertation proposal that had the same title as this work. I knew I had two significant hurdles to overcome to make my analysis possible. One, to develop an automated technique to remove the extraneous movement from the simulation trajectories caused by the algorithm artifacts like center of mass motion removal and

fluctuations in simulation box dimensions created by the semi-isotropic pressure barostat. Fortunately, the work I had done to analyze the trajectories in my first first-author paper had produced some new techniques that I knew could achieve this requirement. Secondly, since the calculation of the electric field map scales as N^6 , I knew the brute force method of creating this map at the resolution I had originally imagined was out of the question. I had an idea that since I could divide the volume of the simulation into rectilinear voxels, and once I had removed the extraneous motion, I could integrate the average charge in each voxel over the span of the simulation. This would leave me with voxel center coordinates (where I would reckon the average charge) that were orthogonal to one another along each axis. I was hoping I could exploit this geometrical symmetry created by my approach to the problem to reduce the scaling cost by about 2. I had intended to test the validity of this approach before making my proposal. I was encouraged to leave that till later and first finish writing the proposal. This detail was left until it was time to do these calculations.

Briefly, my original attempt at this approach failed. It also involved the most complex indexing scheme I have ever had to devise to work through the volume space and properly reckon the relative position of every voxel in the space index wise, without mishandling the ever-changing relative position of the periodic boundaries to my voxel of interest at that iteration of the calculations. In problems of this complexity, without validated tools, you cannot assume anything. Was the problem in the data or was it in my algorithm? Had I integrated everything correctly? I had no idea. I had to revert to a graphing program to check that the data going in looked OK. No problem, the data set only contained over 77,000,000 voxels and was only 100MB in size. My graphing program couldn't plot it because it couldn't even read it. There is a reason scientific computing is done on

supercomputers in a Unix operating system. It's called size, also known as "big data" and it comes with its own special set of problems. Record length limits, matrix formats, utility program limitations, disk I/O throttling on supercomputer clusters servicing the whole organization, and coding errors are some of the challenges that must be overcome. In simulation work, any of these real challenges that compromise the data integrity in any way give rise to the well-known problem of: Garbage In / Garbage Out. To avoid wasting time working with Garbage data that has yet to have been recognized as such requires constant vigilance, testing and verification of even the most mundane aspects of the data processing and analysis tasks.

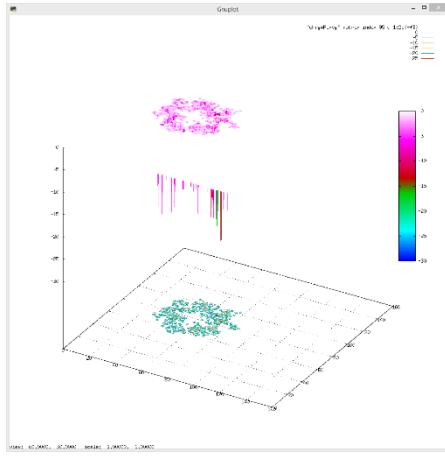
It was while I was developing a method to visually check the integrity of the data that I inadvertently created the first computational x-ray. Because the original file was too



large, I plotted the charge polarities in one slice of the simulation box in the region where I expected to see some of the first pore and then a slice where I hoped to see I was in the ion bath shown in **Fig. 1**. I was OK with the second image, but I had forgotten which pore I had centered on in this simulation. I already had

Fig. 1: Initial attempt to verify integrity of data after trajectory integration.

in my head that the top pore was the one I had centered so I was slightly unexcited about the first, but after I plotted the next two, I was relieved with what I saw and then realized the first image was to be expected when it wasn't the focus of the centering process. These were rather crude images, so I spent some time trying to come up with a way of getting a better overall view of the data. To do this throughout the volume, I needed to be able to plot transparent pixels whose non-transparent fraction would cumulatively add as other



transparent pixels were plotted on top of them. I struggled a lot to get try to get GNUPLOT to do so, since its native transparency does not act this way. You can see indications of a throated pore in **Fig. 2**, but this was the point where I had to give up on the notion of using transparent pixels.

Fig. 2: Final attempt to plot integrated charges using transparent pixels.

I decided to use non-transparent pixels and try setting the viewing angle in such a way to minimize the number of pixels that would overlap. The result had some promise except I knew the charge colors couldn't be changing sign like this. I decided to set the color by the moiety of the molecules instead, initially using the cholines, phosphates and esters and the second image of **Fig. 3** was more promising. I had come up with a way

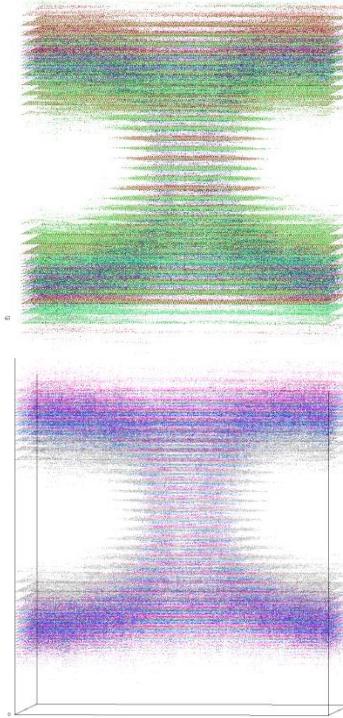
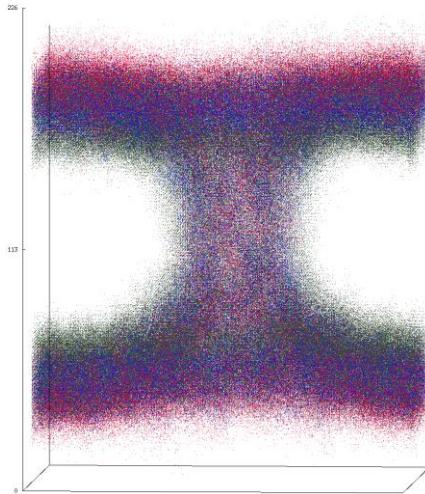


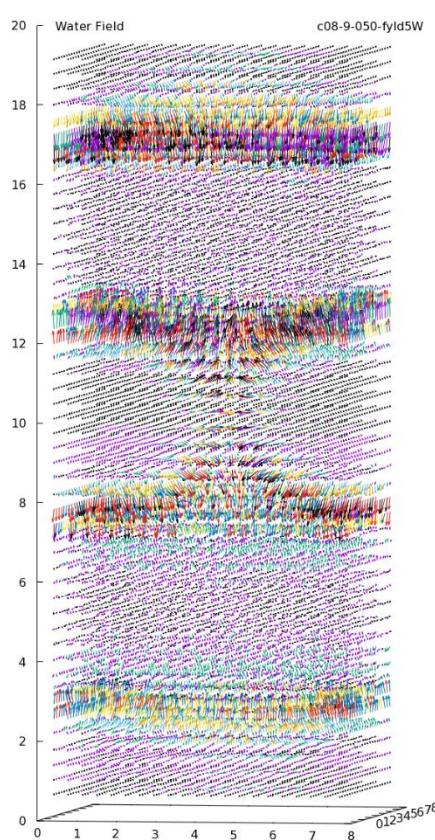
Fig. 3: First two partial attempts at plotting integrated data with non-transparent pixels.

to read only specific layers of the data file to get around the size problem, so there was now no problem with printing every layer. It would take a bit of time, but I decided to try it. The



raster lines would appear one at a time. You could count them as they did. When this image (**Fig. 4**) was about 40% complete, I started almost shaking inside. Not because of the image so much, but because I was realizing the potential of what was appearing before my eyes. The rest of this story and the results are contained in Section 2 of this dissertation.

Fig. 4: First computational x-ray; red-cholines, blue-phosphates, black-esters.



Now that I knew my integrated data was good. I could press on with the electric field work. To have another rendering of the electric field map with which I could check the validity of my new calculation approach, I wrote a routine to do the calculations by the brute force method at a much lower resolution. The results immediately came out good. While I worked on a way to visually render the first results (**Fig. 5**), I started recalculating them all again at twice the resolutions since I could calculate it would take only 48 hours per moiety to do the calculations. This was not my eventual

Fig. 5: Example of three-dimensional electric field vector map for water moiety.

desired resolution, but with 14 nodes to use, the timeframe was only a weekend.

I validated my brute force algorithm against GROMACS, and plotted the output data consisting of an electric field map and a potential map (**Figure 5**). The electric field map renders a vector space so GNUPLOT could do a fair job of that. The potential map renders a scalar space, which is much better visualized with equipotential surfaces and isolines. GNUPLOT doesn't have that native capability, and although when I do plot them that way they will be interesting and informative, remember that what exerts force in nature is the electric field, not the integral of that field over distance. The rest of this story and the results are also contained in Section 2 of this dissertation.

Especially in biological problems, even with all the recent advances in simulation speeds, many problems still require simulation times vastly larger than can be performed. One class of these problems, calculating free energy surfaces of protein folding in response to changing environmental conditions, like salt species and concentration variations are an important example. To overcome these limitations and explore such surfaces in reasonable timeframes, a type of MD has been developed called Metadynamics. To learn the specifics of the most advanced form of this approach, parallel bias metadynamics, simulations were performed to determine the effects of salt, water, and protein force fields on protein folding thermodynamics in molecular simulations. Section 3 of this dissertation contains my published research from this work.

An important aspect of chemical engineering seeks to create catalysts to increase the rates of reactions to make commercially valuable compounds. Maximization of product selectivity and minimization of energy expenditure to do so are major economic drivers of

this work. In biology, these catalysts are known as enzymes, and nature often uses yet unknown processes in fantastically selective and efficient enzymes to do so. Nitrogenase is an enzyme produced by certain bacteria types that performs the reduction of atmospheric nitrogen to form ammonia, a critical component for synthesis of all amino acids and nucleotides. The Haber-Bosch (HB) industrial process mimics this reaction but requires pressures of 2200 to 3600 psi and temperatures between 752 and 932 °F, consuming 1-2% of the world's entire energy production and 3-5% of the world's entire natural gas production. Without the fertilizers produced by the HB process, insufficient food could be grown to keep the world's population fed. Scientists around the world have aggressively been trying to discover the secrets of nitrogenase activity for the last 50 years. Again, this inaccessible to direct observation frontier is one where MD and CC have much to offer to help unravel these secrets.

From what is known, the entire nitrogenase puzzle is too complex to describe for the scope of this work. Briefly, the nitrogenase enzyme works in tandem with a co-enzyme, nitrogenase reductase that binds adenosine-triphosphate (ATP) molecules, before docking with nitrogenase to thus transfer freed electrons to the main enzyme before undocking in a repeating cycle. Collectively these enzymes use three types of metal cofactors to transport and retain these free electrons. The docking/undocking cycle operates through physical changes in docking region shape and charge. Nitrogenase must accumulate 8 free electrons to reduce one molecule of nitrogenase, electrons that change the local electric field and hence also influence the shape of the enzyme.

We have started the important work investigating these conformation changes. Simply building the model of this large system is no small challenge. For a first iteration,

the CC program Gaussian was used to calculate the charge distribution in the metal cofactors in their base (uncharged) state. In the enzyme, most of these cofactors are held in their active position by ligand bonds. To achieve additional electronic isolation of the metal cofactor where the donated electrons reduce the bound nitrogen molecule, a chelated molecule, homocitrate is also entirely held in place by ligand bonds. After months of work overcoming all these challenges, I rebuilt the entire model from scratch, attempting to script as much of the work as possible to remove as many opportunities for human error as possible. The smallest version of this system we have built consists of over 300,000 atoms. One mistake on only one of them is all it takes to invalidate the results of just the basic MD calculations, which take about a month to complete 1 μ s of simulation. The details of the challenges and the to be published results of our research are contained in Section 4 of this dissertation.

There are many more detailed results from these four sections than can reasonably be individually discussed. However, for the sake of completeness, any summarized discussion of result aspects can be verified in the A appendix where all the detailed results are included for all 4 sections. Section 1 alone has 144 pages of plots detailing the simulation histories of the production runs reported in that paper.

Hopefully, this introduction has provided an effective outline and perspective for appreciating the skills and techniques I have learned as well as those I have developed to accomplish this body of work as I begin my trek in the field.

SECTION 1

Insights into the Molecular Mechanisms of Electroporation from Computational Electrophysiology

AUTHOR NAMES. J. Patrick Brian[†], Joel R. Fried[†], Vance W. Jaeger^{†*}

AUTHOR ADDRESS. Department of Chemical Engineering, University of Louisville,
Louisville, KY 40292

KEYWORDS. Electroporation, Computational Electrophysiology, Lipid Membranes,
Molecular Dynamics, Ion Channels

ABSTRACT

The transport of ions and small molecules across lipid membranes can be facilitated or enhanced by exposure to a transmembrane electrical potential, creating pores in a

process known as electroporation. To study this phenomenon and how it affects membrane transport characteristics, a recently developed method known as computational electrophysiology (CompEL) has been employed. The CompEL algorithm allows the modeling of quasi-steady state membrane behavior using molecular dynamics (MD) by maintaining constant charge imbalance and concentration gradients of chemical species across the membrane. Thus, CompEL prevents pore closure during simulations, enables simulations with finite quantities of electric flux and provides access to larger simulation timescales. The effects of transmembrane charge imbalance on pore formation kinetics, pore size, pore stability, and ion permeation rates have been studied. Additionally, the enhancement of pore formation and stability through the inclusion of synthetic chloride ion membrane transporter (SCMTR) molecules has been studied. Our results show total pore area is remarkably correlated with the electric flux through the membrane irrespective of pore size distribution and that SCMTR molecules demonstrate an enhancement of pore persistence and stability.

STATEMENT OF SIGNIFICANCE

Though electroporation is widely practiced in research and clinical environments, the fundamental theoretical molecular level mechanisms of pore formation, stability and longevity are poorly understood. Advances in multiple facets of molecular dynamics have enabled simulations of previously unobtainable fidelity. All atom simulations incorporating a recently developed method known as computational electrophysiology demonstrate convincingly that pore size is established mostly by the magnitude of the applied electric flux. This insight suggests certain logical modifications to electroporation

protocols and instrument design to increase treatment efficacy while simultaneously decreasing cell mortality.

INTRODUCTION

Electroporation is often used to induce pore formation(1-3) in lipid membranes to facilitate the transfection of cells with DNA and to deliver macromolecules such as proteins(4, 5). Although the efficacy of electroporation has been repeatedly demonstrated(6-11), molecular level structures during and after pore development, pore characteristics, transport properties, stability, lifetime, decay and disintegration are not well understood(12). Since electroporation involves the interaction of many thousands of molecules dynamically rearranging at once in an electric field, it is difficult to study the molecular biophysics experimentally at high resolution(13). MD simulations, on the other hand, are well suited to provide molecular-level structures of dynamically rearranging biomolecules.

It is well known that molecular simulations of lipid membranes are sensitive to several factors(14-18). Consequently, modeling complex behavior in lipid bilayer systems requires techniques that accurately reflect real life conditions. Previously published MD simulations of electroporation have induced pore formation by the addition of an electric field term in the Hamiltonian(19, 20). While this method is easily implemented in many simulation packages, it is an imperfect solution to the problem for two main reasons. First, the applied electric field causes ions to irreversibly permeate through the membrane, which changes the effective transmembrane potential during the simulation. Second, to simulate a concentration gradient, a vacuum layer must be introduced to separate the ion baths. This

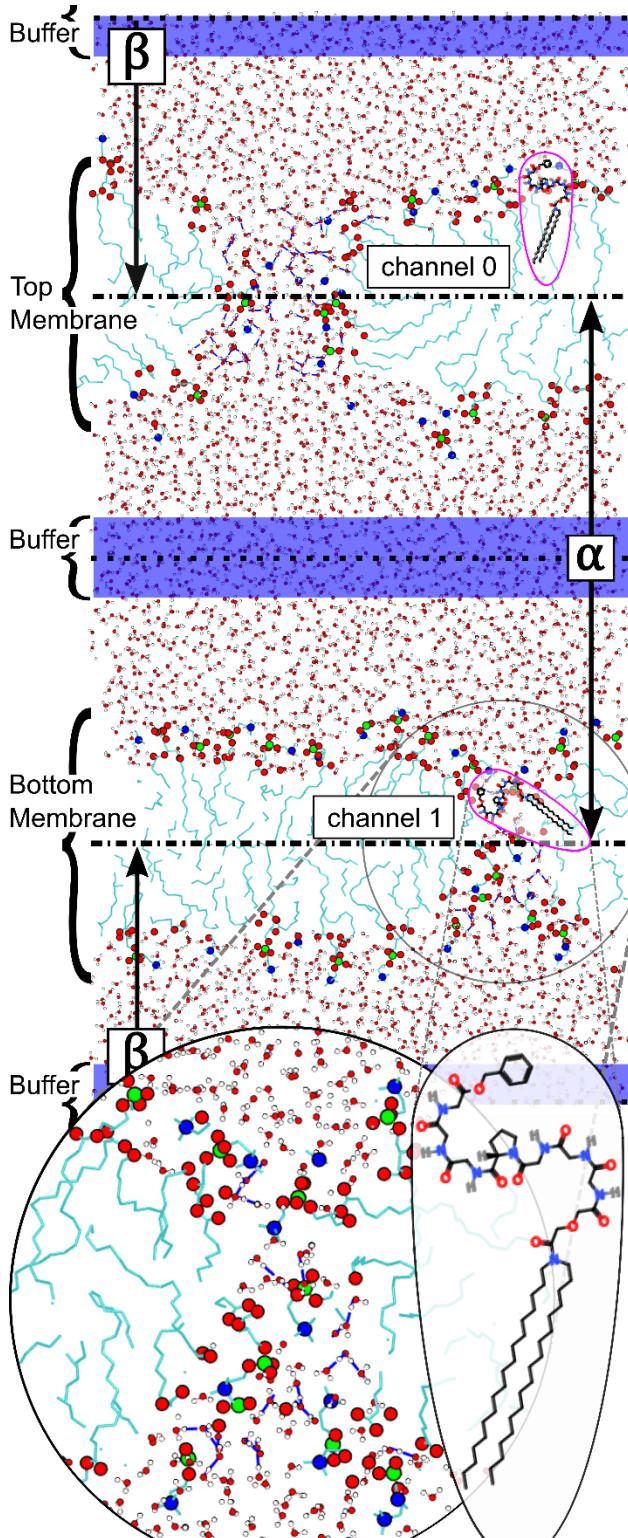


Fig. 6: Double membrane geometry. The α region is a volume in the center of the simulation box between the membranes. Phosphorus is green, nitrogen is blue, and oxygen is red. Faint blue lines in the pore are hydrogen bonds. **Oval inset:** SCMTM molecule. Magenta ovals show typical positions in the membrane and in pores.

vacuum layer causes changes in the structuring of water near the interface, and the simulation of a vacuum layer requires the Canonical Ensemble (NVT) rather than the preferred Isobaric-Isothermal Ensemble (NPT).

To overcome these disadvantages, we have used a recently developed method called computational electrophysiology(21) that modifies the behavior of an MD simulation. CompEL uses a resourceful simulation geometry in which two membranes are present, denoted as the “double membrane” geometry as illustrated in **Fig. 6**. This double-membrane construction creates two independent ion baths (regions α and β) with no vacuum layer between. Thus, the simulation boundaries are free of discontinuities, and no vacuum layer

must be introduced. Therefore, a barostat can be used to control the lateral pressures on the membrane with a semi-isotropic algorithm unlike many previous simulations(22-24). Because of these advantages, CompEL has previously been used to study ion permeation through protein ion channels(25-27). We hypothesize that the NPT ensemble coupled with the CompEL algorithm will model more realistic behavior during simulations of electroporation, because it allows for the expansion or contraction of the membrane surface area, which is important in the evolution, maintenance, and fidelity of pore size and formation/disintegration kinetics.

The ability to modulate membrane behavior in the presence of electric fields and ion gradients could be useful in enhancing electroporation efficacy. To study how membrane anisotropy affects electroporation, a synthetic ion transporter known as SCMTR has been included. SCMTRs are a class of molecules designed to mimic the structure and function of natural membrane chloride ion channels. Several SCMTR structures have been synthesized and tested with different efficacy in encouraging anion transport(28). Typically, their structures consist of a heptapeptide capped on one end with two lipid-like tails and on the other end with a moiety selected to impart hydrophobicity. The class of SCMTRs includes variations on this theme with varying tail lengths, tail saturation, tail cis-trans geometry, amino acid sequences and headgroup types. When added to liposomes with a high internal chloride ion concentration, SCMTRs rapidly and selectively induce the release of chloride ions from the liposome into the surrounding bath (10:1 Cl-/K⁺ ion permeations(29)). Recent studies on SCMTRs in planar lipid membranes indicate that channel formation at low transmembrane voltages is a relatively rare event, sometimes requiring several minutes for channels to form(30).

SCMTRs have only been studied computationally once before(31) and kinetic mechanisms for their observed behavior have yet to be postulated. Prior experimental experience with SCMTR molecules(30) led us to hypothesize that their ion transport enhancing characteristics might extend to electroporation as well. The oval inset of **Fig. 1** and **Fig. S1** in the Supporting Material show the SCMTR type used in this work. The chemical formula for this SCMTR is ((C₁₈H₃₇)₂NCOCH₂OCH₂CO-GGGPGGG-OCH₂Ph).

Reversible electro-permeabilization of lipid membranes has revealed a 10:1 specificity for cations(32), but has not described the selectivity mechanism. Several classes of molecules such as SCMTRs(29), hydraphiles(33) and lariats(34) promote membrane pore formation based upon ionic concentration and charge gradients.

The CompEL double membrane geometry, while eliminating vacuum boundary discontinuities, doubles the number of atoms in a simulation, causing the simulation to run at about half the speed, since the simulations presented here were run near the linear scaling regime. However, in the optimal geometry with two pores there are twice as many locations for membrane-crossing events to occur, meaning CompEL does not diminish overall computational efficiency. CompEL tracks the concentration of ions in each of the baths separated by the lipid membranes. When an ion crosses a membrane, CompEL detects a change in the concentration gradient, and reestablishes the set-point concentrations in each bath by removing an ion of the same species from the buffer region and placing it back in the buffer region of the chamber with the depleted ion concentration. Thus, the concentration gradient over the membrane is controlled, and a quasi-steady state is maintained. With CompEL's ability to address shortcomings of previous simulation

methods, we aim to study the formation, structure, and dynamics of pores formed during electroporation and the transport of ions through these pores.

CompEL is therefore is an invaluable method that we believe will enable the first development of a fully three-dimensional theoretical understanding of the role the non-uniform electric field plays in stabilizing the equilibrium membrane structure that contains pores. Establishing the initial (uniform) electric field (and finite flux) with a finite transmembrane charge imbalance ensures that the resulting non-uniform electric field must consist of that same initial finite electric flux. For a brief review of the relevant electrostatics discussed in this paper, see the **electrostatics** section in the Supporting Material where it is explained that electric flux is proportional to transmembrane charge imbalance. With a better understanding of the role electric flux has in determining pore size and stabilizing membrane structure, new electroporation methods could potentially exploit this knowledge to grow pores to an equilibrium size for the applied flux.

Finally, electroporation as practiced, involves timescales beyond current computational capabilities, especially in cases of macromolecule transport. However, by enhancing pore formation with additional electrical potential, MD simulations can elucidate transitions between membrane and pore states and the equilibrium characteristics of these states(35, 36). With a first principles theoretical approach, a clearer understanding of the molecular biophysics underlying pore formation, membrane permeation, pore dynamics, and the specific interactions that lead to experimental observables can be gained.

Molecular mechanisms of pore formation are conceptualized here as the competition between forces that stabilize and destabilize membrane structures. The lateral

compressive and tensile forces between atoms in lipid bilayer membranes vary more than two orders of magnitude through the thickness of the membrane(37-42). The compressive and tensile forces in equilibrium with one another, establish the planar geometry of the membrane(43). Membrane structures experience a time-averaged uniform electric field (normal to the membrane surface) due to equal distributions of oppositely charged ions on either side of the membrane. In the presence of sufficient transmembrane electrical potential, (~2.0 V) we hypothesize the following driving forces for electroporation:

- 1) The external electric field torques the lipid headgroup dipoles in the anti-aligned leaflet away from anti-alignment, thus strengthening the effective magnitude of the external electric field locally.
- 2) This local disruption of the normal headgroup dipole alignments reduces the energy barrier for the intrusion and growth of continuous water columns into the membrane hydrophobic region. In these columns, the water molecule dipoles are aligned with the external field, thus conducting, intensifying and extending it non-uniformly in the direction of the opposite leaflet(44-46).
- 3) The intensified electric field induces further self-reinforcing interactions that evolve to develop a continuous hydrophilic region between the outer surfaces of the two leaflets, commonly referred to as a pre-pore (i.e. water penetrates the membrane).
- 4) The primary driving force is the concentration of the electric flux in the center of the pore by the progressive alignment of water molecule dipoles.

5) The intensification of electric flux in the center of the pore is balanced by a commensurate decrease in the electric flux in the hydrophobic region beyond the annulus of the pore.

By this process, the membrane structure rearranges to form a new, stable pore in equilibrium with the non-uniform electric field. Around the pore's center, the lipid leaflets are drawn into a toroidal geometry as a portion of the relaxed electrical stress transforms into increased leaflet mechanical stress. The pore geometry that minimizes the Gibbs free energy of the conformation while balancing the electric and mechanical stresses is toroidal(12, 47, 48), and as the diameter grows, an open pore forms in the membrane. Ultimately, the equilibrium size of the pore is established by the electric flux through the membrane.

METHODS

Build CHARMM lipids membrane and SCMTR

To begin our simulation campaign, a simulation box with a single membrane geometry was created using CHARMM(49) lipids. The CHARMM-GUI web server(50) was used to create a single lipid bilayer membrane of 174 DOPC (Dioleoylphosphatidylcholine) molecules and 11506 water molecules with the CHARMM36(51) force field for lipids and mTIP3P(52) for water. Steepest descent minimization was performed for up to 5000 steps, and the single membrane was equilibrated in the NPT ensemble with semi-isotropic pressure coupling for 10 ns in GROMACS2018(53-62). CHARMM-GUI was also used to create a topology file for the

SCMTR molecule using the CHARMM general force field(63, 64) (CGenFF) (see **Fig. S1**).

MD parameters

GROMACS2018 was used for all MD simulations. The standard GROMACS simulation parameters for CHARMM36 lipids(65) were used as indicated by the CHARMM-GUI webserver. Hydrogen-heavy atom bond lengths were constrained with the LINCS algorithm. A Verlet integrator was used, and Lennard-Jones interactions were cutoff with a force-switch modifier between 1.0 and 1.2 nm. Electrostatics were calculated using the particle mesh Ewald (PME) method with real space a cutoff of 1.2 nm. No dispersion corrections were used. Production runs were performed using a 2 fs integration timestep. Temperature was controlled at 303.15 K with the Nose-Hoover thermostat(66) and pressure was controlled at 1 bar with Parrinello-Rahman barostat(67) unless otherwise noted. Representative MD parameter file for GROMACS have been included in the Supplemental Information.

Place SCMTR in membrane

Several modifications to the geometry produced by CHARMM-GUI were needed. First, the GROMACS utility *genion* was used to replace 22 water molecules with 11 Na ions and 11 Cl ions using a standard CHARMM ion force-field(68). The salt molarity of the resulting bath solutions was about 50 mM (ionic molarity of about 100 mM) since cations and anions were always swapped 1:1 between baths to produce the span of charge gradients. These values were selected to match a previous study by Burkhardt et al.(31) The *-membed* option of GROMACS *mdrun* program was used to insert a dimer of SCMTR molecules into the membrane in the active conformation suggested by previous

experimental results(69, 70). This step resulted in the removal of 14 lipids from the bilayer to make room for the SCMTR dimer, 6 from the upper leaflet and 8 from the lower leaflet. This system was equilibrated for 20 ns with semi-isotropic pressure coupling using the Berendsen barostat(71) to converge the density of the simulation box. After this step, vacuum layers of 3.0 nm were added in the z -dimension and ions were rearranged to create a $\pm 6 e^-$ charge imbalance across the membrane (an excess of 6 Na ions on one side of the membrane and an excess of 6 Cl ions on the other). Then steepest descent energy minimization was performed for up to 50000 steps in GROMACS.

Build double membrane system

To produce the double membrane geometry, the energy minimized system described above was copied, rotated 180 degrees about the x axis and translated along the z-axis while removing any vacuum layer to create the continuous double membrane system. Ions were swapped between the α and β regions (see **Fig. 1**) to set an initial charge imbalance of $\pm 6 e^-$. Up to 50000 steps of energy minimization and 100 ps of NPT equilibration were used to collapse any remaining gap between the two halves of the box. Finally, 40 ns of NVT equilibration was conducted to ensure a good starting configuration for the CompEL algorithm.

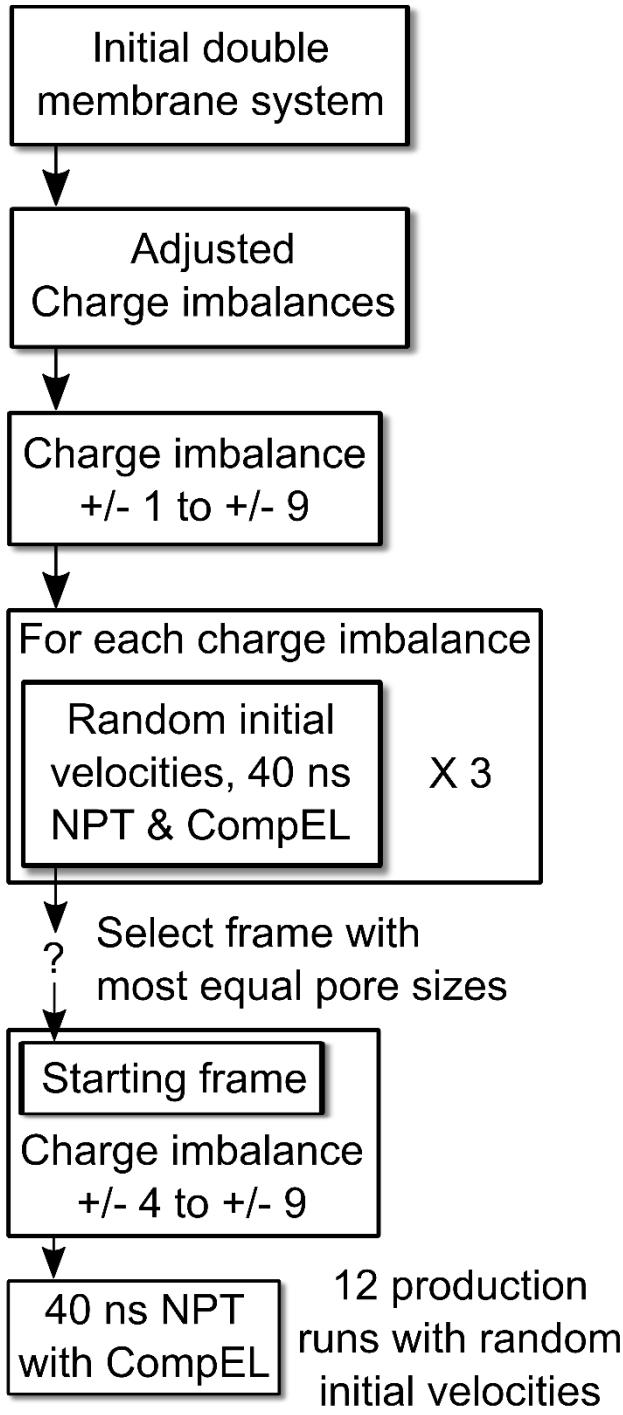
Build system with pore in each membrane

Five copies of the double membrane configuration were assigned independent initial velocity distributions. In this geometry, pore formation can be induced by introducing a relatively high transmembrane voltage (~ 2.0 V) caused by an ion charge gradient of $\pm 6 e^-$. This voltage was selected, because the timescale of pore formation at lower voltages is too long to consistently observe pore formation. For this work we coined

and use the term headgroup channel to define any continuous column of lipid hydrophilic moieties forming between the membrane leaflets, which we then differentiate as pores or pre-pores as described later. Simulations were run for up to 40 ns using the NVT ensemble. Once a stabilized pore formed in the simulation, a frame was selected that had no ions in the center of the pore. The half of the double membrane system without a pore after this step was then removed. The remaining half of the system with the pore in the membrane was copied, rotated 180 degrees about the x axis and translated along the z-axis to create the continuous double membrane system with a pore in each membrane. Ions were swapped between the α and β regions (see **Fig. 1**) to keep the charge imbalance at $+/- 6 e^-$. Energy minimization and 100 ps of NPT equilibration were applied to collapse any remaining gap between the two halves of the box.

Production simulations workflow

This double membrane open-pore system became the initial frame used to create simulations spanning a larger range of transmembrane ion charge gradients of $+/- 1$ to $+/- 9 e^-$. After appropriately swapping ions to create the nine initial configurations, three instances of each were simulated in the NPT ensemble for 40 ns to allow the pores to reach their equilibrium size. For each charge imbalance, a frame from near the end of one of these simulations was selected in which the pores were nearly equal in size and no ions were present in the center of the pores. Visual analysis of simulations of this system with smaller charge imbalances, ($+/- 1$ to $+/- 3 e^-$), prompted us to classify them as pre-pore. For the systems with headgroup channels that we considered to be at the pore stage of development ($+/- 4$ to $+/- 9 e^-$), 12, independent 40 ns simulations were performed in the NPT ensemble using random initial velocities. These 72 simulations were duplicated



identically in all respects except that the SCMTR molecules, while still present, were not associated with the pore. This second set of simulations allowed us to identify changes to pore behavior induced solely by the presence of SCMTR molecules. **Fig. 2** illustrates this workflow once the starting system had been assembled.

Count water molecules

A method was developed to accurately count the water molecules in the core of a pore on a frame by frame (one every 20 ps) basis from the trajectories. Because of fluctuations in the coordinates of the center of a pore caused by the semi-isotropic pressure control and center of mass motion removal, some challenges arose which were overcome using the GROMACS

Fig. 2: Simulation workflow to generate CompEL production runs from assembled double membrane system.

select routine. The *select* routine statement that enabled the extraction of this data along with a full explanation of its operation is presented in the Supporting Information. Our

definition of the pore's core is bounded vertically (+/- 0.75 nm) from the center of geometry of the headgroup nitrogen, phosphorous and oxygen atoms within a horizontal radius of the pore center. To determine a proper value for this radius, the average radius of the core water molecules, which also define the radial extent of lipid headgroups in the core, is increased by 50%. Without this limit, water molecules associated with membrane curvature away from the pore would be counted as well, increasing water molecule counts an average of ~6.5%. For systems with no developed pore, the +/- 0.75 nm setting created a so-called “live zero” pore water count in the typical range of 1-3 molecules. Live zero measurements are implemented by setting selection criteria parameters to maximize the explicit discernment between noise and the anticipated signal when it is at a known value of zero. Using a live zero setting permits any development in the membrane interior, like water wires or initial movement of headgroups toward the interior to immediately be detected and located for accurate quantification to produce the graph of pore core water molecules.

SCMTR-pore and pore-flux interactions

To quantify SCMTR's effects on pore formation kinetics, 240 simulations of 40 ns on double membrane systems starting with no pores at +/- 6 e⁻ charge imbalance were conducted. To quantify SCMTR effects on pore stability, starting frames (equal pore sizes in each membrane) were selected from the production runs at +/- 4 e⁻ charge imbalance (see **Fig. 2**) both with and without SCMTR's associated with the pore. Ions were swapped to eradicate the charge imbalance. Then ten production runs of 40 ns, each with random initial velocities, were conducted for both starting frames. The trajectories were analyzed in VMD(72) to determine pore closure times. To quantify the effects of charge imbalance

on simulation box x - y dimensions (area per lipid), the average area of each of the 144 production runs was calculated and then the standard deviation was calculated for each of the 12 types of production runs (six different charge imbalances with and without SCMTM's).

RESULTS AND DISCUSSION

Transmembrane charge imbalance

Three initial 40 ns simulations were conducted for each of the nine charge imbalances. Many simulation animations across the span of charge imbalances were visualized in VMD. Visual inspection prompted us to classify the channels in simulations with charge imbalances of $\pm 3 e^-$ or less as pre-pores. The pre-pore structure comes into being once the water and headgroup intrusions from each leaflet meet in the hydrophobic region, forming a headgroup channel. We distinguish the pre-pore and pore regimes as well as pore size by counting the number of water molecules in the central (1.5 nm) transmembrane region. For now, we define pores as headgroup channels where half or more of the water molecule hydrogen bonds are with other water molecules, thus establishing an objective criterion for classifying water channels as pores or pre-pores in the computation realm. Subsequently, a set of 144 simulations of 40 ns each was conducted with transmembrane charge imbalances spanning from ± 4 to $\pm 9 e^-$. Pore size, pore size distribution, ion flux and ion charge selectivity were quantified as a function of transmembrane ion charge imbalance. Then the variance attributable to the presence of the SCMTM molecules in the pores was determined for each of these observables. **Fig. 3** illustrates the pore size distribution over the span of simulated transmembrane charge imbalances.

The average number of water molecules present in the pores is a linear function of the charge imbalance. No matter how water molecules are distributed between the two membrane pores, the total number of water molecules over the two pores (i.e. the sum of the number of waters in both membranes) remains the same. However, **Fig. 3** obfuscates some detail from our simulations by averaging over the full 40 ns trajectory. In some simulations, one of the two pores close, yet the closure is always balanced by equivalent growth in the remaining pore. **Fig. S2** illustrates a case where a pore closes, yet somewhat

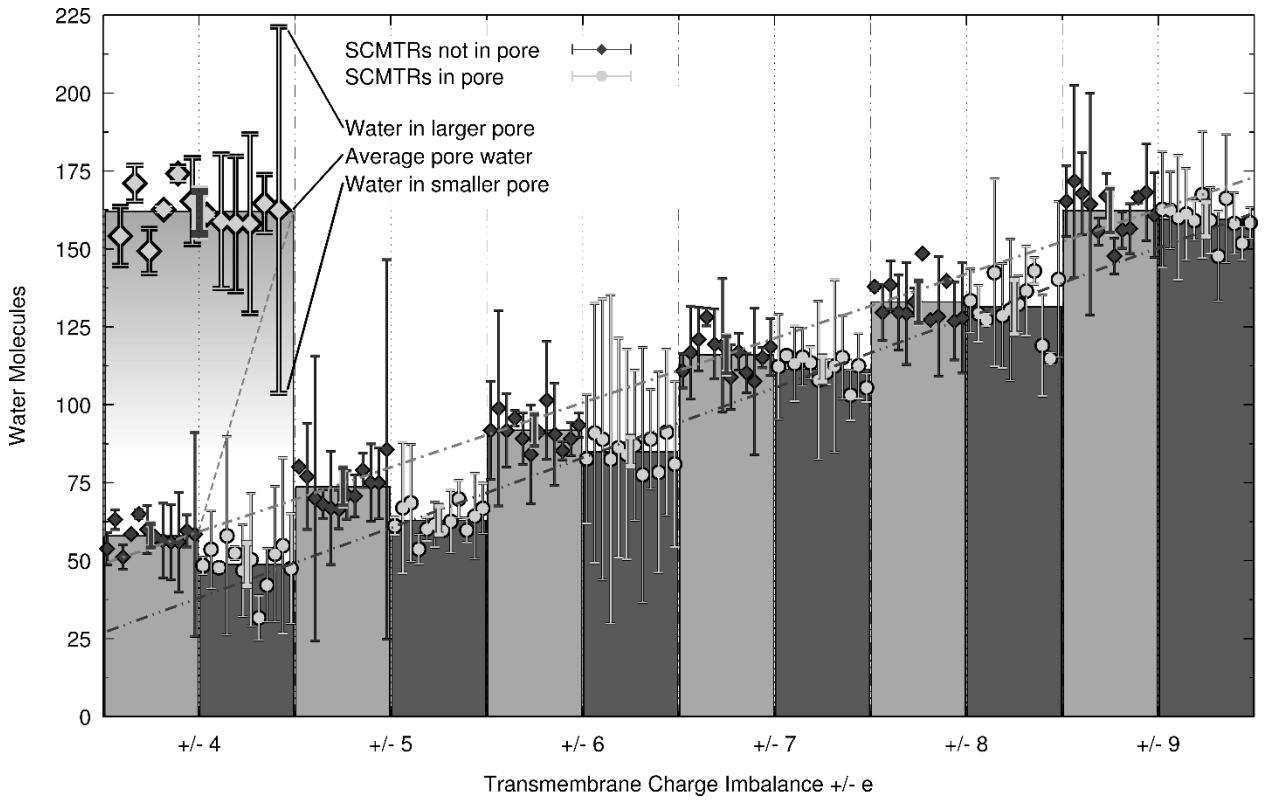


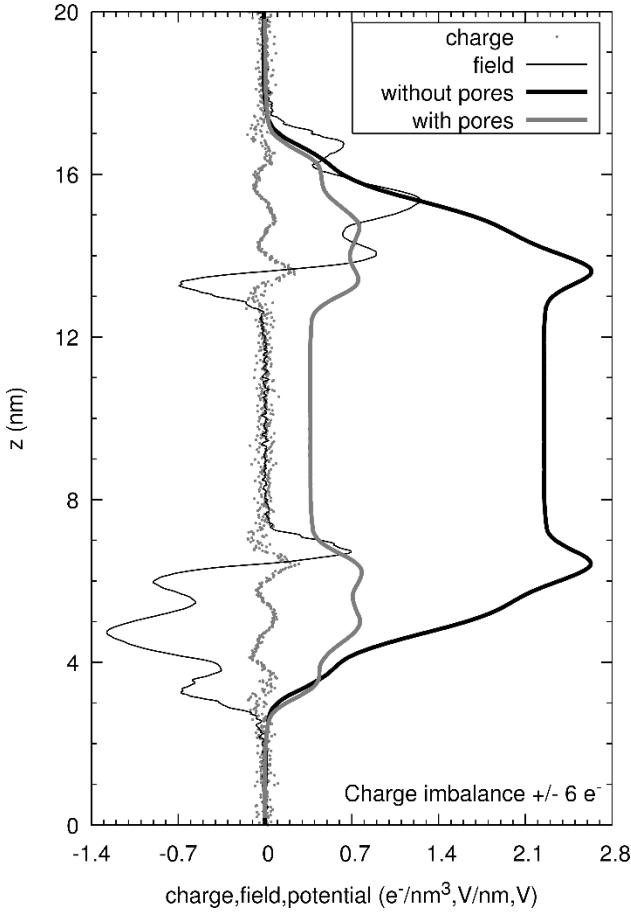
Fig. 3: Distribution of core water molecules in induced pores. The magnified inset explains how the data are presented. Bars represent the average number of water molecules in both pores for a given charge imbalance. Light bars indicate a SCMTR-free pore, and dark bars indicate SCMTRs associated with the pore. The central whisker is the standard deviation of this average over 12 observations. The dashed lines show the linear regression for each pore type with regression coefficients of $R^2=0.9930$ (SCMTRs not in pore) and $R^2=0.9932$ (SCMTRs in pore).

counterintuitively, the average water molecules for each pore in that simulation are nearly identical.

Another curiosity in **Fig. 3** is the clear and consistent difference in the pore size over the span of charge imbalances attributable to the presence of SCMTR molecules in the pore. During visual inspections of animations, we developed a hypothesis to explain this difference. While the simulations started with the SCMTR molecules dimerized, the structure of the dimer evolves when it is coupled with a pore. Often the heptapeptide chains are wrapped like two belts around sections of the circumference of the pore, hydrogen bonding to headgroups and occupying headgroup sites that would normally be hydrogen bonded to water molecules. This substitution explains the very consistent difference in pore water molecule counts. In later discussion, other results indicate that this heptapeptide conformation with the pore headgroups also enhances pore stability and longevity.

The initial double membrane geometry (before pore formation) exposes each membrane to a roughly uniform electric flux, because ions freely diffuse through the solvent bath. The average transmembrane potential resulting from this uniform flux was calculated using the GROMACS *potential* utility as shown in **Fig. 4** for a system with a transmembrane charge imbalance of +/- 6 e⁻. Pore formation kinetics are enhanced by using a high transmembrane voltage of about 2.0 V. With this voltage pores tend to form in under 40 ns. With smaller voltages, pores take significantly longer to form. Since no external electric field is applied in our simulations, pore formation and growth can only develop to a point of equilibrium with the finite electric flux created by the ion charge imbalance between the baths. For a more detailed view of simulation, pre- and post-pore electrical properties see **Fig. S3** and **Fig. S4**. That the total pore size was so highly

correlated with the total electric flux irrespective of the pore size distribution was not



anticipated. Visual inspection of simulation trajectories also revealed a pronounced alignment of the choline moieties (lipid headgroups) in the pore in the direction of the electric field (see Fig. S5) even at transmembrane potentials ostensibly as small as 0.4 volts. The explanation for this misleading and possibly unappreciated disparity lies in how membrane potentials are currently calculated in MD.

Fig. 4: Membrane charge distribution, electric field and potential (before pore development) and potential (after development). The flat character of the potential curve in the water bath regions reveals there is no external electric field present except for that produced by the ion charge imbalance that exists between the two baths.

Analytical solutions for determining the electric field of uniform charge distributions only exist for conformations that are highly symmetrical. Averaging the analysis of electric properties over the x-y plane (as implemented in the GROMACS *potential* utility, being predicated on the assumption of planar symmetry) cannot elucidate important details about the three-dimensional evolution of pore structures as their coulombic interactions are transitioning from a uniform to a non-uniform electric field. We are not aware of a simulation analysis tool that permits a full three-dimensional determination of the non-uniform electric field and its interaction with the membrane to

form pores by analysis of the simulation trajectory data. We address the need for such a utility in our conclusions.

Ion fluxes

During each simulation, ion permeation events were recorded for each pore. These data were matched with the actual pore size in each membrane and plotted in three dimensions to visualize interactions and dependencies/correlations. The results for pores without SCMTRs are displayed below in **Fig. 5**, illustrating relationships between transmembrane charge imbalance, pore size, and ion permeation events. The corresponding figure for pores with SCMTRs is shown in **Fig. S6**.

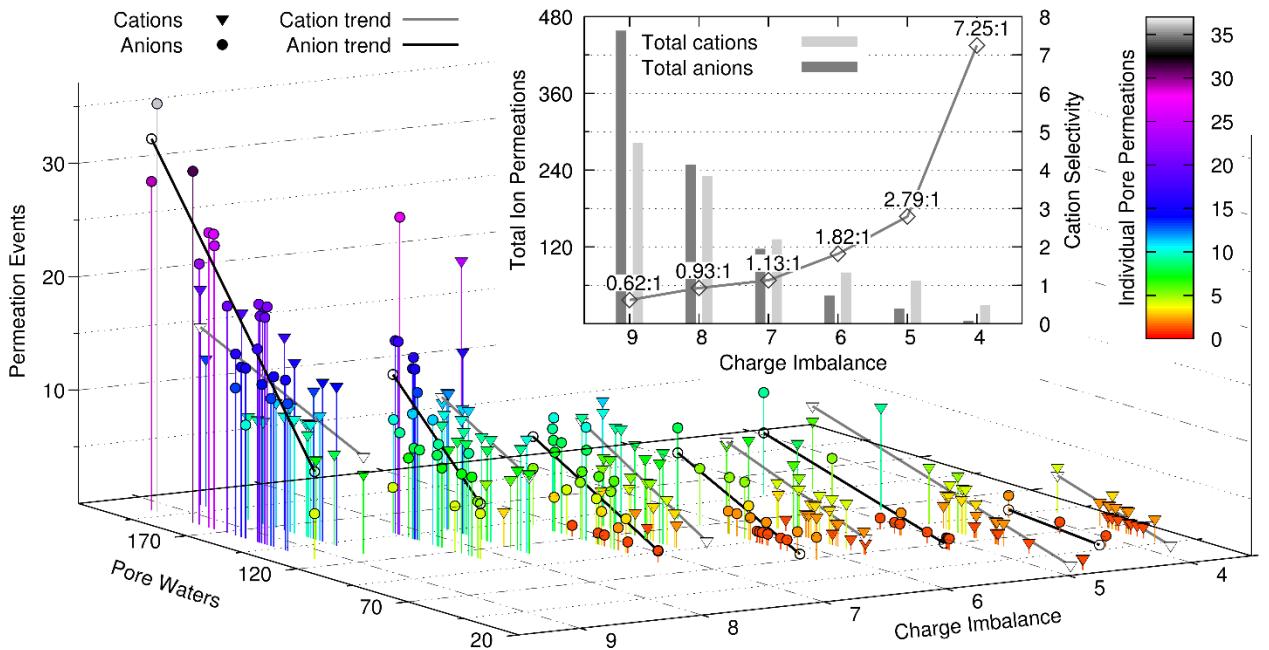


Fig. 5: Ion permeation rates vs. pore size and charge imbalance for plain pores (no SCMTR present). Data for each charge imbalance (integer values) are split left and right to separate permeation of cations and anions. Note relatively constant trend of cation diffusion vs. pore size compared to steadily increasing trend of anion diffusion. Inset plot displays the total ion permeations by charge for each charge imbalance and associated variance in cation selectivity.

Along the y-axis (charge imbalance), the integer values are split left and right to separate the permeations by ionic charge, right for cations and left for anions. At charge imbalances of $+/- 4 e^-$ and $+/- 5 e^-$, some small pores with cation permeations had no corresponding anion permeations. This is the reason for the apparent lack of anion points for some pore sizes where a plotted cation data point is visible. By summing all the permeation events by ion type for each of the charge imbalances, the total ion flux versus charge imbalance can be compared, which the regression coefficients of **Fig. 3** demonstrated, strongly correlates with total pore size. These total permeations are plotted as bars on the inset plot. The ratio of the total of the cation to anion permeations is the pore cation selectivity, plotted with the diamond points.

The difference in the total permeations of opposite charged ions is the net flow of electrical current through the pore. By averaging the total permeations of each ion type for each charge imbalance these currents can be calculated and another nuance of the pore cation selectivity becomes more apparent. Over the span of these charge imbalances, the direction of the net current flow reverses. Since each simulation was 40 ns in length, each net pore permeation accounts for a current of about four picoamperes. The distribution of the charge imbalance in the simulations was such that the electric field always pointed from the α -region (positive charge) to the β -region (negative charge) (see **Fig. 1**). Positive electric current flow is defined as the flow of positive charges in the direction of the electric field. Knowing the direction of the electric field, the net currents were calculated at charge imbalances from $+/- 4 e^-$ to $+/- 9 e^-$ and averaged +4.2, +7.2, +6, +2.5, -5.2 and -29.2 pA respectively. The last value was the difference between an anionic current of -76.3 pA and the cationic current of 47.1 pA. These values correlate well with experimentally measured

values(73). If this simulation scenario were carried out in a planar bilayer experiment, the patch clamp amplifier would show a current reversal between the $+/- 7 e^-$ and $+/- 8 e^-$ charge imbalances. This indicates that the anion flux has exceeded the cation flux. Since the CompEL algorithm automatically swaps crossed ions back, thereby maintaining the established charge imbalance, these simulations provide insight into the mechanisms establishing the steady state fluxes. In experimental conditions, no net current flow simply means equal and opposite counter ion diffusion, which also means decreasing concentration and charge gradients.

Nevertheless, clearly pore size and charge imbalance effect a significant change in pore selectivity, yet a larger challenge is to identify and understand the all the mechanisms responsible for this change. We believe the feature that stands out in **Fig. 5**, the constantly increasing slopes of the anion trend lines, are an important clue. Compared to the nearly invariant slopes of the cation trend lines over the span of increasing charge imbalance, determining the causes for these differences is essential.

All other things being equal, considering ions of equal concentration interacting only with water in the absence of electric fields, differences in fluxes through any surface should vary only with respect to their individual diffusion coefficients. Animations of pre-pore and pores mechanics reveal cations entering the pore freely, whereas there is clearly an energy barrier for anion entry into smaller pores. At larger pore sizes, diffusion appears to be unhindered for both ion species, since larger pore dimensions make it possible for ions to pass through the pore interacting only with water molecules and the electric field. In the absence of electric fields and molecular interactions, different ions would exhibit selectivity for the species with the higher diffusion rate. To understand by what degree this

effect is present in our simulation data, a system with 11,480 water molecules, 11 chlorine ions and 11 sodium ions was simulated for 100 ns to determine the diffusion rates of the ions in this system. The GROMACS *msd* program calculated diffusion coefficients of $2.33 \times 10^{-5} \text{ cm}^2 \text{ sec}^{-1}$ for sodium and $4.0 \times 10^{-5} \text{ cm}^2 \text{ sec}^{-1}$ for chlorine at 303.15K. This equates to chloride ions having a ~73% higher diffusion rate in the simulations with our set of force fields. Thus, for 283 cations observed to cross the membrane, about 489 anions would be expected to diffuse through the pore if it has reached the purely diffusive regime. The actual value of 458 is about 94% of 489, so the steadily increasing trend of the anion permeations is not inconsistent with the difference in the ion diffusion rates. Applying the same analysis to the permeation data for pores with SCMTRs, (**Fig S6**) 597 anion and 226 cation permeations are observed at the same charge imbalance. For 226 cation permeations, 394 anion crossings would be predicted when the system is in the purely diffusive regime. This is well short of the 597 observed crossings. We speculate the higher electric field in the pore resulting from the pore centric concentration of the electric flux will be identified as the cause of these transport rates in excess of a purely diffusion-based regime.

Potential simulation size artifacts

To assess how the finite size of the simulation box affected our results, a set of simulations were conducted with a system four times the size of the original (doubled in the *x*- and *y*-dimensions and identical in the *z*-dimension). Since pore formation in the original system required a charge imbalance of +/- 6 e⁻, the larger system was created with a charge imbalance of +/- 24 e⁻, anticipating an equivalent electric flux would lead to similar pore formation dynamics. While the large and small systems had the same transmembrane voltage, pores were surprisingly formed much less often in the larger

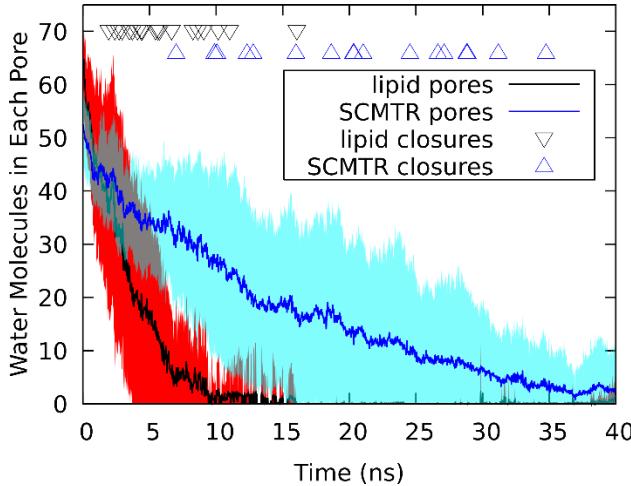
systems. As expected, the +/- 24 e⁻ charge imbalance gave rise to a pore with equivalent increases in area. The pore area is proportional to the charge imbalance and thus the electric flux. The higher relative flux of chloride ions through the pore compared to sodium ions moving in the opposite direction caused a significant net flux of water molecules from one bath to the other, because solvated ions moving through the pore are accompanied by their water solvation shell. **Fig. S7** depicts the extreme outcome of this scenario in which one bath shrinks while the other grows. Relative chloride flux between the larger and smaller systems is significantly greater than a factor of four. We hypothesize that the chloride ion flux increases due to changes in the ion concentration gradient caused by the net flux of water. Therefore, CompEL may not be a viable method to observe quasi-steady state ion flux through sufficiently larger pores. However, modifications to the CompEL protocol could be made in the future to control not only the number of ions in each bath but also the number of water molecules.

SCMTR effects on pore formation

While investigating pore formation kinetics and the effects of SCMTR on pore formation, we unexpectedly observed that NPT simulations exhibit a clear pattern of much lower pore formation rates compared to the NVT ensemble, even though the NPT ensemble is needed for simulations to reach their equilibrium box dimensions. The campaign of 240 simulations that were used to probe pore formation kinetics were each 40 ns in length. The initial 40 simulations, conducted in the NPT ensemble, developed only one pore over the total 1600 ns. Next, the same 40 simulations were repeated in the NVT ensemble using the equilibrium box size for membranes with pores (so that pore development would not be constrained by available area per lipid), and seven pores were formed. Of these seven pores,

four formed with SCMTM molecules incorporated and three formed without. Another set of 40 NVT simulations were run using the equilibrium box size for membranes without pores (at the risk of making lipid molecules and pores compete for area), and counterintuitively produced 16 pores. The next 40 NVT simulations were run at the largest box dimensions seen in a frame in an NPT trajectory and again counterintuitively produced five pores. Finally, 80 NVT simulations were performed at the original equilibrium box size (for membranes with pores), but the charge imbalance was reduced from $+/- 6 e^-$ to $+/- 5 e^-$, producing three pores. In total, the data indicate that when simulating pore formation, the choice of ensemble and the surface area per lipid are both critical variables. The NVT ensemble, more crowded lipid membranes, and higher transmembrane potentials all encourage pores to form more quickly.

SCMTM effects on pore stability

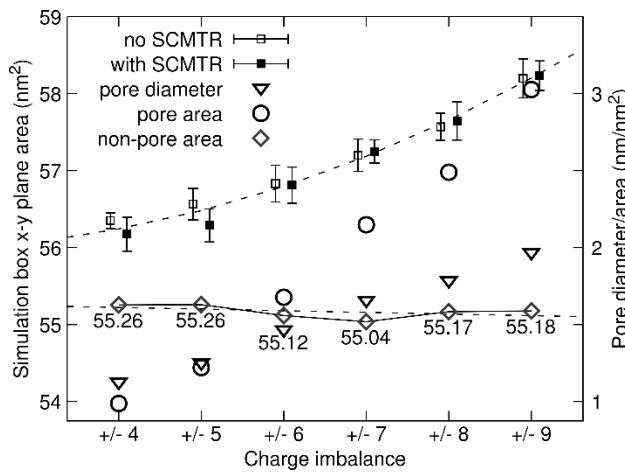


To quantify stability of pores formed by electroporation, pore disintegration kinetics were explored. Ten simulations were conducted for systems containing two equally sized pores at equilibrium with a $+/- 4 e^-$ charge imbalance for pores with and

Fig. 6: Pore disintegration dynamics after removal of charge imbalance. We hypothesize hydrogen bonding between SCMTM molecules and lipid headgroups in the pore play a key role in slowing the disintegration of membrane pores in the absence of any electric field. Shaded regions indicate the standard deviation over 10 simulations and the trendlines are averages.

without SCMTRs. At time zero the charge imbalance was eliminated and all but one of the twenty pores without SCMTRs disintegrated within 12 ns. The average lifetime for the non- SCMTR pores was 6.2 ns. The decay of the SCMTR-containing pores was slower by a factor greater than three, with an average pore lifetime not less than 22 ns, and two pores were still open after 40 ns. For these simulations, the pore water molecules were counted as an indicator of pore size, and pore closure events were tabulated. These data are presented in **Fig. 6** which clearly indicates enhanced pore stability and longevity attributed to the SCMTR molecules' inclusion in the pore structure. As previously discussed, we also attribute this increased stability to the SCMTRs hydrogen bonding to multiple lipid headgroups in the pore. We hypothesize that SCMTR's enhancement of ion permeations in previous experiments(29) can be explained by the stabilizing effects observed in simulations. However, we see insufficient evidence to confirm or dispute that SCMTR enhances pore formation.

Area per lipid of membranes with pores



We had hypothesized that the NPT ensemble would more realistically simulate all phases in the lifetime of membrane pores formed in response to transmembrane electric fields. Semi-isotropic pressure control permits simulated membranes to expand or

Fig. 7: Simulation x-y plane box area vs. charge imbalance. Error bars are standard deviations of the average x-y plane area for each of 12 simulations. Note how the total pore area as measured by water molecules fully accounts for changes in simulation cross-sectional area.

contract as necessary to reach an equilibrium surface area with their surroundings. These fluctuating areas were extracted from the full trajectories of all 144 production runs and plotted in **Fig. 7**. As hypothesized, our simulations show that different simulation box sizes are necessary to allow pores to reach their equilibrium sizes without being constrained as the charge imbalance changes.

To determine if counting core water molecules is an accurate measure of the pore size, we checked for the self-consistency of multiple measurements. First, we used the counts of core water molecules to calculate pore area. We define the pore area as the number of water molecules in the core divided by the bulk density of water divided by the height of the core, which was defined as 1.5 nm in our counting algorithm. In **Fig. 3**, we demonstrate that the number of core water molecules is extremely highly linearly correlated with the electric flux. Second, we calculated the difference in box x-y area that is caused by the formation of a pore, which we use as a second measure of the pore area as in **Fig. 7**. The consistency of these two measures indicates that counting core water molecules is in fact an accurate measure of pore size. Furthermore, this supports our earlier claim that the equilibrium pore size is a function of the electric flux which in our simulations is proportional to charge imbalance.

For simulations with charge imbalances of ± 4 and $\pm 9 e^-$ (averages of 50 and 152 water molecules), the pore areas are calculated to be 0.99 and 3.03 nm^2 . A quadratic regression of the data in Figure 7 predicts box x-y areas of 56.2 and 58.2 nm^2 at charge imbalances ± 4 and $\pm 9 e^-$. If the area of the pore completely accounts for differences in box x-y area, then the remaining area containing lipids is found by subtracting the pore area from the x-y box dimensions, and this value should be nearly constant. We thus

calculate the lipid area to be 55.3 and 55.2 nm² (at +/- 4 and +/- 9 e⁻), which are statistically the same when considering the error bars. This result is consistent for all intermediate charge imbalances.

CONCLUSIONS

MD simulations of lipid membrane pore formation in response to electric fields induced by transmembrane charge gradients have been conducted using the CompEL method. These simulations have provided molecular-level insight into restructuring of lipid membranes in response to electric fields. The CompEL method is a step forward in more realistically modeling membrane structures, dynamics and steady state properties. Aside from the ability to model steady state transport through membranes it enables studying membrane interactions with controlled, steady, finite amounts of electric flux. A key finding of our work is that the electric flux determines the equilibrium pore size.

We have developed and implemented a new method of determining pore size by counting pore core water molecules. However, our method requires significant proficiency with the GROMACS *select* commands, and these commands are much more complicated than any examples we have seen published. This method has enabled us to verify a compelling correlation in MD between equilibrium pore size and the electric flux. Because pore growth to equilibrium size is not instantaneous, we speculate initial formation times of individual pores significantly effect pore size distributions observed in practice. If so, the connection of pore size to electric flux could be useful for improving the experimental and clinical practice of electroporation, because to the best of our knowledge previous work has not considered this variable to be a contributing factor to pore size distributions.

At smaller pore sizes chloride ions seem to encounter an energy barrier at the pore entrance. As pore sizes increase, lower probabilities of ion-pore wall interactions allow observed chloride ion fluxes to approach purely diffusive transport. We hypothesize the unquantified non-uniform electric field gradients are key variables affecting pore permeation dynamics when fluxes exceed rates consistent with unperturbed diffusion. We believe accurate electric field mapping in and around pores will be required before a better theoretical explanation of these observed ion fluxes can be postulated.

In future work, we plan to develop a tool to accurately reconstruct the exact nature of interactions among electric fields and biomolecules from the existing trajectory data. An accurate, three-dimensional description of these interactions as well as the three-dimensional electric field map would be important theoretical developments, leading to useful insights that could be exploited in practical applications. Furthermore, since coulombic interactions are known to be used by nature for control and modulation of many essential membrane specific processes (endo- and exocytosis as well as transport of species via membrane proteins), a more complete understanding of how the membrane itself interacts with the local electric field is beneficial.

Although the NPT ensemble was crucial to establishing the link between electric flux and pore size, we noted it also has a profound effect on pore formation rates. Trying to account for the extreme variation in observed formation rates accounted for the bulk of the 240 simulations performed hoping to ascertain pore formation kinetics. Our results indicate that close to the equilibrium area per lipid, pore formation rates observed using the NVT ensemble counterintuitively increase as that area becomes more constrained. We speculate being more constrained could be creating simulation artifacts manifesting as if

the membrane is experiencing more extreme pressure fluctuations in the X-Y plane, contributing to larger fluctuations of the initial pore formation energy barrier.

Author Contributions

JPB designed the study, carried out the simulations, developed the post-processing techniques and algorithms, analyzed the data, interpreted the data and wrote the paper with advice and feedback primarily from VWJ and additional review by JRF. JPB, JRF and VWJ approved the final version of the manuscript.

Acknowledgements

JR Fried, Department of Chemical Engineering, VW Jaeger, Assistant Professor and Advisor, H Simrall, Assistant Director of IT Research Computing

REFERENCES

1. Shi, J. F., Y. F. Ma, J. Zhu, Y. X. Chen, Y. T. Sun, Y. C. Yao, Z. G. Yang, and J. Xie. 2018. A Review on Electroporation-Based Intracellular Delivery. In *Molecules*.
2. Nakamura, H., and S. Watano. 2018. Direct Permeation of Nanoparticles Across Cell Membrane: A Review. *Kona Powder Part J*(35):49-65.
3. Kar, S., M. Loganathan, K. Dey, P. Shinde, H. Y. Chang, M. Nagai, and T. S. Santra. 2018. Single-cell electroporation: current trends, applications and future prospects. *J Micromech Microeng* 28(12).
4. Boukany, P. E., A. Morss, W. C. Liao, B. Henslee, H. C. Jung, X. L. Zhang, B. Yu, X. M. Wang, Y. Wu, L. Li, K. L. Gao, X. Hu, X. Zhao, O. Hemminger, W. Lu, G. P. Lafyatis, and L. J. Lee. 2011. Nanochannel electroporation delivers precise amounts of biomolecules into living cells. *Nat Nanotechnol* 6(11):747-754.
5. Potter, H. 2003. Transfection by electroporation. *Curr Protoc Cell Biol Chapter 20:Unit 20 25.*
6. Saulis, G. 1997. Pore disappearance in a cell after electroporation: Theoretical simulation and comparison with experiments. *Biophys J* 73(3):1299-1309.
7. Li, S., J. Cutrera, R. Heller, and J. Teissie. 2014. Electroporation protocols : preclinical and clinical gene medicine. *Humana Press* ;, New York.
8. Kim, K., and W. G. Lee. 2017. Electroporation for nanomedicine: a review. *J Mater Chem B* 5(15):2726-2738.
9. Betters, E., R. M. Charney, and M. I. Garcia-Castro. 2018. Electroporation and in vitro culture of early rabbit embryos. *Data Brief* 21:316-320.
10. Cruz-Ramon, J., F. J. Fernandez, A. Mejia, and F. Fierro. 2019. Electroporation of germinated conidia and young mycelium as an efficient transformation system for *Acremonium chrysogenum*. *Folia Microbiol* 64(1):33-39.
11. Wang, L., L. J. Yang, X. Wen, Z. Y. Chen, Q. Y. Liang, J. L. Li, and W. Wang. 2019. Rapid and high efficiency transformation of *Chlamydomonas reinhardtii* by square-wave electroporation. *Bioscience Rep* 39.
12. Mukherjee, P., S. S. P. Nathamgari, J. A. Kessler, and H. D. Espinosa. 2018. Combined Numerical and Experimental Investigation of Localized Electroporation-Based Cell Transfection and Sampling. *Acs Nano* 12(12):12118-12128.
13. Germann, M., T. Latychevskaia, C. Escher, and H. W. Fink. 2010. Nondestructive Imaging of Individual Biomolecules. *Phys Rev Lett* 104(9).
14. Lyubartsev, A. P., and A. L. Rabinovich. 2016. Force Field Development for Lipid Membrane Simulations. *Bba-Biomembranes* 1858(10):2483-2497.
15. Sandoval-Perez, A., K. Pluhackova, and R. A. Bockmann. 2017. Critical Comparison of Biomembrane Force Fields: Protein-Lipid Interactions at the Membrane Interface. *Journal of Chemical Theory and Computation* 13(5):2310-2321.
16. Sajadi, F., and C. N. Rowley. 2018. Simulations of lipid bilayers using the CHARMM36 force field with the TIP3P-FB and TIP4P-FB water models. *Peerj* 6.

17. Leonard, A. N., R. W. Pastor, and J. B. Klauda. 2018. Parameterization of the CHARMM All-Atom Force Field for Ether Lipids and Model Linear Ethers. *J. Phys. Chem. B* 122(26):6744-6754.
18. Park, S., and W. Im. 2019. Analysis of Lipid Order States and Domains in Lipid Bilayer Simulations. *Journal of Chemical Theory and Computation* 15(1):688-697.
19. Tarek, M. 2005. Membrane electroporation: A molecular dynamics simulation. *Biophys J* 88(6):4045-4053.
20. Kong, Z., H. B. Wang, L. J. Liang, Z. S. Zhang, S. B. Ying, Q. Hu, and J. W. Shen. 2017. Investigation of the morphological transition of a phospholipid bilayer membrane in an external electric field via molecular dynamics simulation. *J Mol Model* 23(4).
21. Kutzner, C., H. Grubmuller, B. L. de Groot, and U. Zachariae. 2011. Computational Electrophysiology: The Molecular Dynamics of Ion Channel Permeation and Selectivity in Atomistic Detail. *Biophys J* 101(4):809-817.
22. Hyvonen, M. T., T. T. Rantala, and M. AlaKorpela. 1997. Structure and dynamic properties of diunsaturated 1-palmitoyl-2-linoleoyl-sn-glycero-3-phosphatidylcholine lipid bilayer from molecular dynamics simulation. *Biophys J* 73(6):2907-2923.
23. Im, W., and B. Roux. 2002. Ions and counterions in a biological channel: A molecular dynamics simulation of OmpF porin from Escherichia coli in an explicit membrane with 1 M KCl aqueous salt solution. *J Mol Biol* 319(5):1177-1197.
24. Skelton, A. A., V. M. Khedkar, and J. R. Fried. 2016. All-atom molecular dynamics simulations of an artificial sodium channel in a lipid bilayer: the effect of water solvation/desolvation of the sodium ion. *J Biomol Struct Dyn* 34(3):529-539.
25. Zachariae, U., C. Kutzner, H. Grubmuller, and B. de Groot. 2012. Computational Electrophysiology Reveals Ion Channel Permeation in Atomistic Detail. *Biophys J* 102(3):22a-22a.
26. Zachariae, U. 2016. Computational Electrophysiology: Close-UPS of Ion Permeation and Migration in Membrane Proteins. *Biophys J* 110(3):376a-376a.
27. Kutzner, C., R. T. Ullmann, B. L. de Groot, U. Zachariae, and H. Grubmueller. 2017. Ions in Action - Studying Ion Channels by Computational Electrophysiology in GROMACS. *Biophys J* 112(3):139a-139a.
28. Djedovic, N., R. Ferdani, E. Harder, J. Pajewska, R. Pajewski, M. E. Weber, P. H. Schlesinger, and G. W. Gokel. 2005. The C- and N-terminal residues of synthetic heptapeptide ion channels influence transport efficacy through phospholipid bilayers. *New J Chem* 29(2):291-305.
29. Schlesinger, P. H., R. Ferdani, J. Liu, J. Pajewska, R. Pajewski, M. Saito, H. Shabany, and G. W. Gokel. 2002. SCMTTR: A chloride-selective, membrane-anchored peptide channel that exhibits voltage gating. *J Am Chem Soc* 124(9):1848-1849.
30. Minooei, F., M. D. Martin, J. R. Fried, and J. P. Brian. 2018. Electrophysiological measurements reveal that a succinyl linker enhances performance of the synthetic chloride channel SCMTTR. *Chem Commun* 54(37):4689-4691.

31. Burkhardt, J. B., A. A. Skelton, and J. R. Fried. 2013. The water-channel forming ability of heptapeptide-based anion channels: insights from molecular dynamics simulations. *Soft Matter* 9(17):4444-4454.
32. Chernomordik, L. V., and I. G. Abidor. 1980. Cation Selectivity of Blm in a Stress State. *Biofizika+* 25(5):918-919.
33. Gokel, G. W. 2000. Hydraphiles: design, synthesis and analysis of a family of synthetic, cation-conducting channels. *Chem Commun*(1):1-9.
34. Gokel, G. W. 1992. Lariat Ethers - from Simple Sidearms to Supramolecular Systems. *Chem Soc Rev* 21(1):39-47.
35. Bennett, W. F. D., C. K. Hong, Y. Wang, and D. P. Tieleman. 2016. Antimicrobial Peptide Simulations and the Influence of Force Field on the Free Energy for Pore Formation in Lipid Bilayers. *Journal of Chemical Theory and Computation* 12(9):4524-4533.
36. Tang, J. C., H. R. Yin, J. L. Ma, W. F. Bo, Y. Yang, J. Xu, Y. Y. Liu, and Y. B. Gong. 2018. Terahertz Electric Field-Induced Membrane Electroporation by Molecular Dynamics Simulations. *J Membrane Biol* 251(5-6):681-693.
37. Cevc, G. M., D. 1987. Phospholipid Bilayers - Physical Principles and Models. John Wiley & Sons: New York.
38. Ollila, O. H. S. V. 2010. Molecular Simulations and Biomembranes; From Biophysics to Function. M. S. P. Sansom, Biggin, P. C., editor. The Royal Society of Chemistry, London, pp. 26-55.
39. Orsi, M., M. G. Noro, and J. W. Essex. 2011. Dual-resolution molecular dynamics simulation of antimicrobials in biomembranes. *J R Soc Interface* 8(59):826-841.
40. Cantor, R. S. 1997. Lateral pressures in cell membranes: A mechanism for modulation of protein function. *J. Phys. Chem. B* 101(10):1723-1725.
41. Patra, M. 2005. Lateral pressure profiles in cholesterol-DPPC bilayers. *Eur Biophys J Biophys* 35(1):79-88.
42. Ollila, S., M. T. Hyvonen, and I. Vattulainen. 2007. Polyunsaturation in lipid membranes: Dynamic properties and lateral pressure profiles. *J. Phys. Chem. B* 111(12):3139-3150.
43. Pera, H., J. M. Kleijn, and F. A. M. Leermakers. 2014. Linking lipid architecture to bilayer structure and mechanics using self-consistent field modelling. *J Chem Phys* 140(6).
44. Galanis, J., and Y. Tsori. 2014. Interface initiation and propagation in liquid demixing with electric fields. *J Chem Phys* 141(21).
45. Galanis, J., and Y. Tsori. 2014. Phase separation dynamics of simple liquids in non-uniform electric fields. *J Chem Phys* 140(12).
46. Galanis, J., and Y. Tsori. 2019. Surface tension and domain growth in nonuniform electric fields. *Eur Phys J-Spec Top* 227(18):2675-2687.
47. Woo, S. Y., and H. Lee. 2017. Effect of lipid shape on toroidal pore formation and peptide orientation in lipid bilayers. *Phys Chem Chem Phys* 19(32):21340-21349.
48. Tieleman, D. P., H. Leontiadou, A. E. Mark, and S. J. Marrink. 2003. Simulation of pore formation in lipid bilayers by mechanical stress and electric fields. *J Am Chem Soc* 125(21):6382-6383.

49. Pastor, R. W., and A. D. MacKerell. 2011. Development of the CHARMM Force Field for Lipids. *J Phys Chem Lett* 2(13):1526-1532.
50. 2019. CHARMM-GUI. charmm-gui.org. Effective simulation input generator and more.
51. Klauda, J. B., R. M. Venable, J. A. Freites, J. W. O'Connor, D. J. Tobias, C. Mondragon-Ramirez, I. Vorobyov, A. D. MacKerell, and R. W. Pastor. 2010. Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *J. Phys. Chem. B* 114(23):7830-7843.
52. Boonstra, S., P. R. Onck, and E. van der Giessen. 2016. CHARMM TIP3P Water Model Suppresses Peptide Folding by Solvating the Unfolded State. *J. Phys. Chem. B* 120(15):3692-3698.
53. GROMACS User Manual version 2018 (www.gromacs.org).
54. Miyamoto, S., and P. A. Kollman. 1992. Settle - an Analytical Version of the Shake and Rattle Algorithm for Rigid Water Models. *J Comput Chem* 13(8):952-962.
55. Berendsen, H. J. C., D. Vandervelde, and R. Vandrunen. 1995. Gromacs - a Message-Passing Parallel Molecular-Dynamics Implementation. *Comput Phys Commun* 91(1-3):43-56.
56. Essmann, U., L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. 1995. A Smooth Particle Mesh Ewald Method. *J Chem Phys* 103(19):8577-8593.
57. Hess, B., H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije. 1997. LINCS: A linear constraint solver for molecular simulations. *J Comput Chem* 18(12):1463-1472.
58. Lindahl, E., B. Hess, and D. van der Spoel. 2001. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J Mol Model* 7(8):306-317.
59. Van der Spoel, D., E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. 2005. GROMACS: Fast, flexible, and free. *J Comput Chem* 26(16):1701-1718.
60. Hess, B., C. Kutzner, D. van der Spoel, and E. Lindahl. 2008. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation* 4(3):435-447.
61. Pronk, S., S. Pall, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess, and E. Lindahl. 2013. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29(7):845-854.
62. Pall, S., M. J. Abraham, C. Kutzner, B. Hess, and E. Lindahl. 2015. Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS. *Lect Notes Comput Sc* 8759:3-27.
63. Vanommeslaeghe, K., E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, and A. D. MacKerell. 2010. CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J Comput Chem* 31(4):671-690.
64. Mukherjee, G., K. Vanommeslaeghe, and A. Mackerell. 2016. Ongoing developments in the CHARMM general force field. *Abstr Pap Am Chem S* 252.
65. 2019. Force fields in GROMACS.

66. Hoover, W. G. 1989. Generalization of Noses Isothermal Molecular-Dynamics - Non-Hamiltonian Dynamics for the Canonical Ensemble. *Phys Rev A* 40(5):2814-2815.
67. Parrinello, M., and A. Rahman. 1981. Polymorphic Transitions in Single-Crystals - a New Molecular-Dynamics Method. *J Appl Phys* 52(12):7182-7190.
68. Yoo, J., and A. Aksimentiev. 2018. New tricks for old dogs: improving the accuracy of biomolecular force fields by pair-specific corrections to non-bonded interactions. *Phys Chem Chem Phys* 20(13):8432-8449.
69. Schlesinger, P. H., N. K. Djedovic, R. Ferdani, J. Pajewska, R. Pajewski, and G. W. Gokel. 2003. Anchor chain length alters the apparent mechanism of chloride channel function in SCMTR derivatives. *Chem Commun*(3):308-309.
70. Pajewski, R., R. Ferdani, J. Pajewska, N. Djedovic, P. H. Schlesinger, and G. W. Gokel. 2005. Evidence for dimer formation by an amphiphilic heptapeptide that mediates chloride and carboxyfluorescein release from liposomes. *Org Biomol Chem* 3(4):619-625.
71. Berendsen, H. J. C., J. P. M. Postma, W. F. Vangunsteren, A. Dinola, and J. R. Haak. 1984. Molecular-Dynamics with Coupling to an External Bath. *J Chem Phys* 81(8):3684-3690.
72. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: Visual molecular dynamics. *J Mol Graph Model* 14(1):33-38.
73. Palankar, R., B. E. Pinchasik, B. N. Khlebtsov, T. A. Kolesnikova, H. Mohwald, M. Winterhalter, and A. G. Skirtach. 2014. Nanoplasmonically-Induced Defects in Lipid Membrane Monitored by Ion Current: Transient Nanopores versus Membrane Rupture. *Nano Lett* 14(8):4273-4279.

Supporting Information

Electrostatics

Electric charges exert forces at a distance on surrounding charges via a first order tensor called the electric field, often represented by electric field lines whose density through a plane normal to their direction is equivalent to the strength of the electric field. Electric flux is the average density of these field lines through a defined surface, finite as well as infinitesimal, without which the strength and direction of non-uniform electric fields cannot be described.

Since the external electric field the membranes are exposed to in these simulations is generated solely by the transmembrane charge imbalance, each membrane is also exposed to a finite quantity of electric flux. When the electric flux is distributed uniformly over the surface of the membrane, the electric field is uniform, and we can simply and accurately say the membrane is exposed to a specific transmembrane potential, being the product of the uniform electric field to which the membrane is exposed and the thickness of the membrane.

Once the membrane structure interacts with this external field, having overcome the formation energy barrier, and develops persistent non-uniform structure, non-uniform electric fields develop, meaning they now vary in intensity and/or direction as a function of position. However, since the only source of the original external electric field was charge imbalance, and the CompEL method restores that charge imbalance every time an ion crosses through a pore, the membrane remains exposed to a finite, unchanging amount of

electric flux. These conditions thereby enable us to investigate and study the equilibrium conformation of membrane pores with electric flux.

On all graphs, Charge Imbalance on the x-axis is equivalent to and scales identically with Electric Flux.

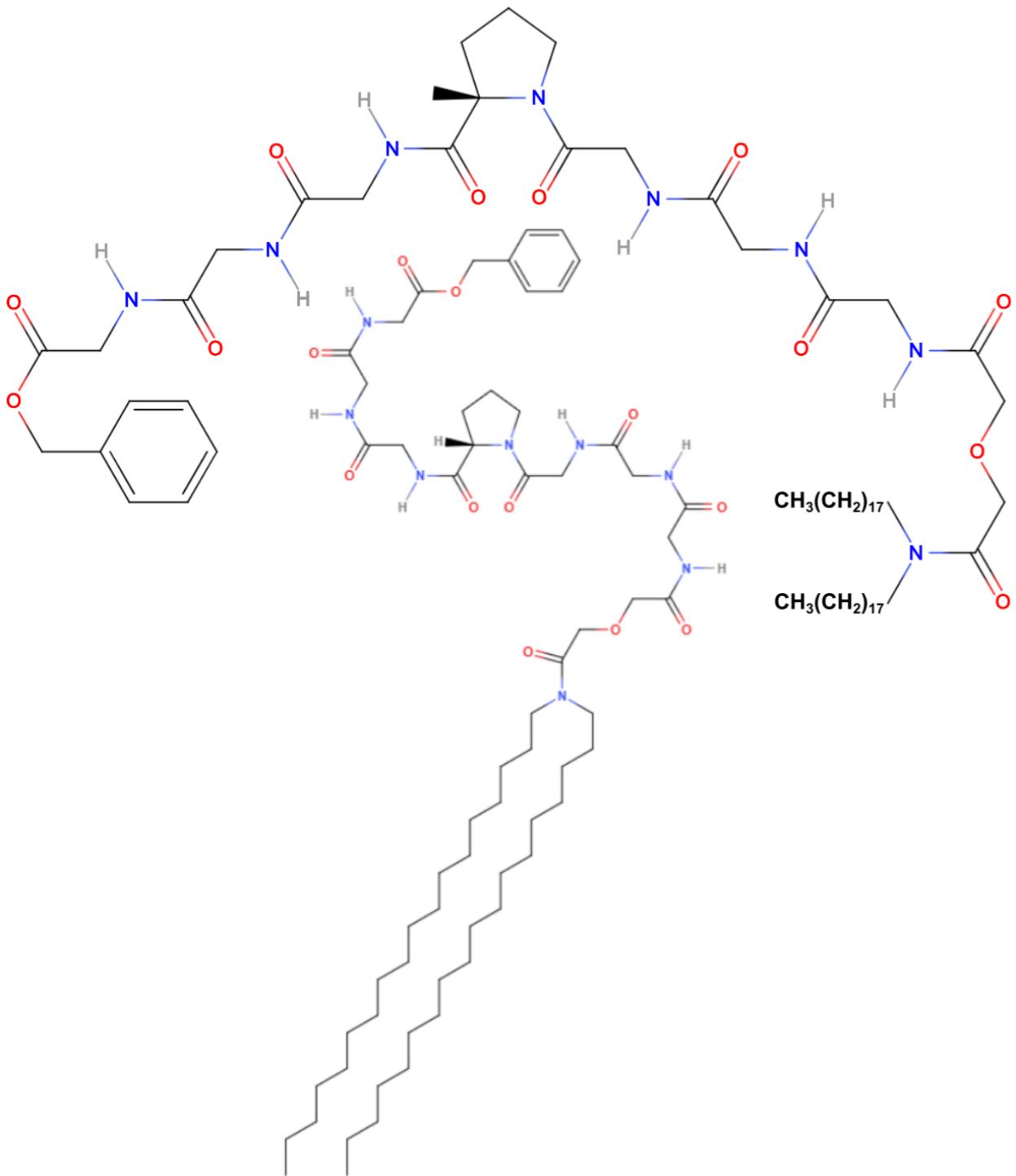


Fig. S1: SCMT Molecule: SCMT molecules are designed to be lipid-like with tails like glycerol derived lipids linked to a heptapeptide section placed between the tails and various headgroup types, a phenyl fragment in this case. Variations of SCMTs include different tail lengths, tail saturation, tail cis-trans geometry, amino acid sequences and headgroup types. The GLY-GLY-GLY-PRO-GLY-GLY-GLY ((C₁₈H₃₇)₂NCOCH₂OCH₂CO-GGGPGGG-OCH₂Ph) sequence shown here is generally considered the most active.

Pore size (water molecules) vs. Simulation time

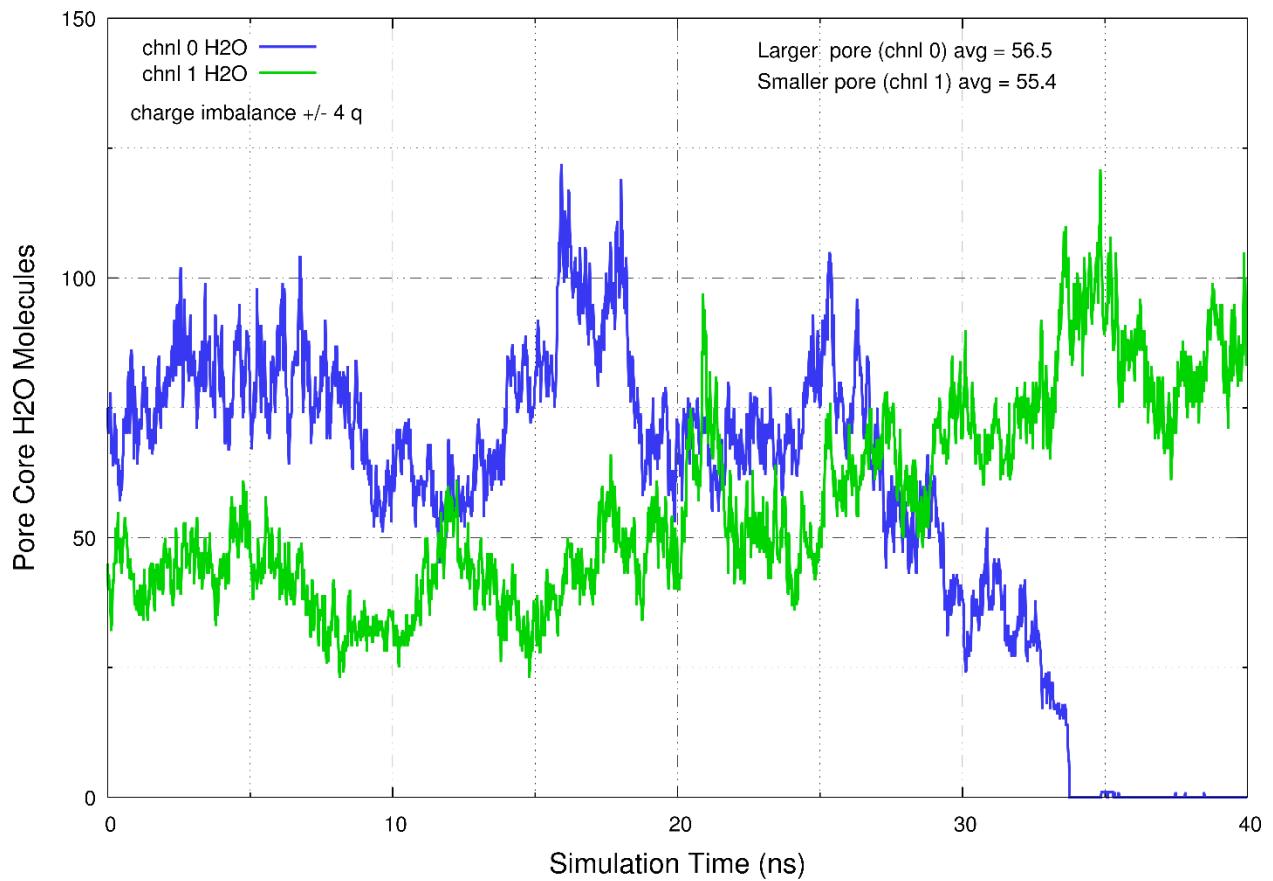


Fig. S2: History of pore core water molecules in upper and lower membranes during one 40 ns simulation. The method of measuring these values on a frame by frame basis is shown and explained below.

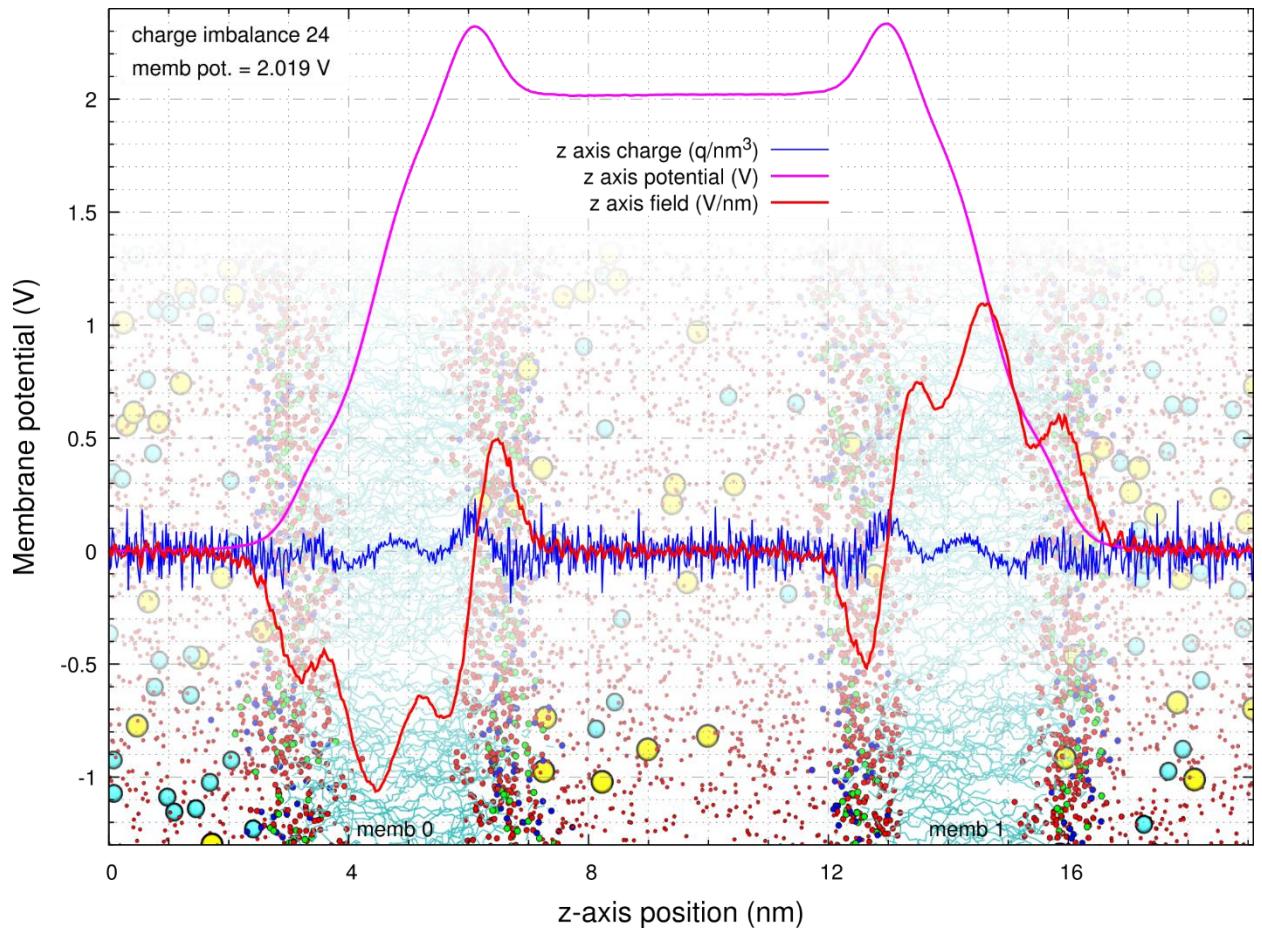


Fig. S3: Details of electrical parameters for pre-pore stage of large (4X) system. To equate these details to the membrane structures, a background image of an accurately scaled frame from the trajectory used for the GROMACS potential analysis was added. Since the simulation conformation at this stage is uniform over the x-y plane, the resulting analysis data is valid. Lipid headgroup atoms are same colors as **Fig. 1**. Water oxygens are rendered as small red dots.

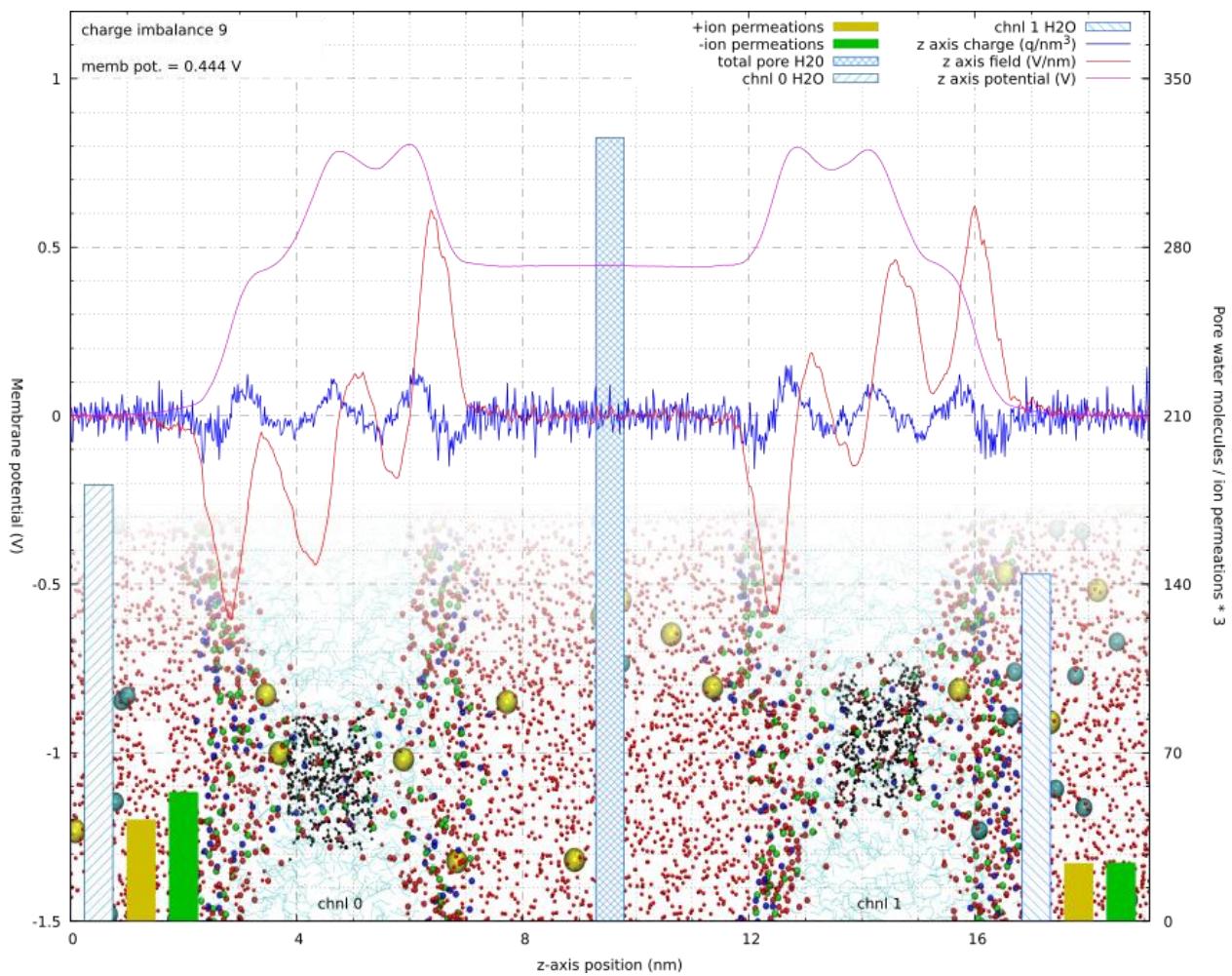


Fig. S4: Results of GROMACS potential analysis for a 40 ns production run combined with pore size and permeation event details (see legend). Core water molecules are rendered as black CPK. A high-fidelity pore development model will require an analysis tool to elucidate the true details of the three-dimensional non-uniform electric field in this conformation.

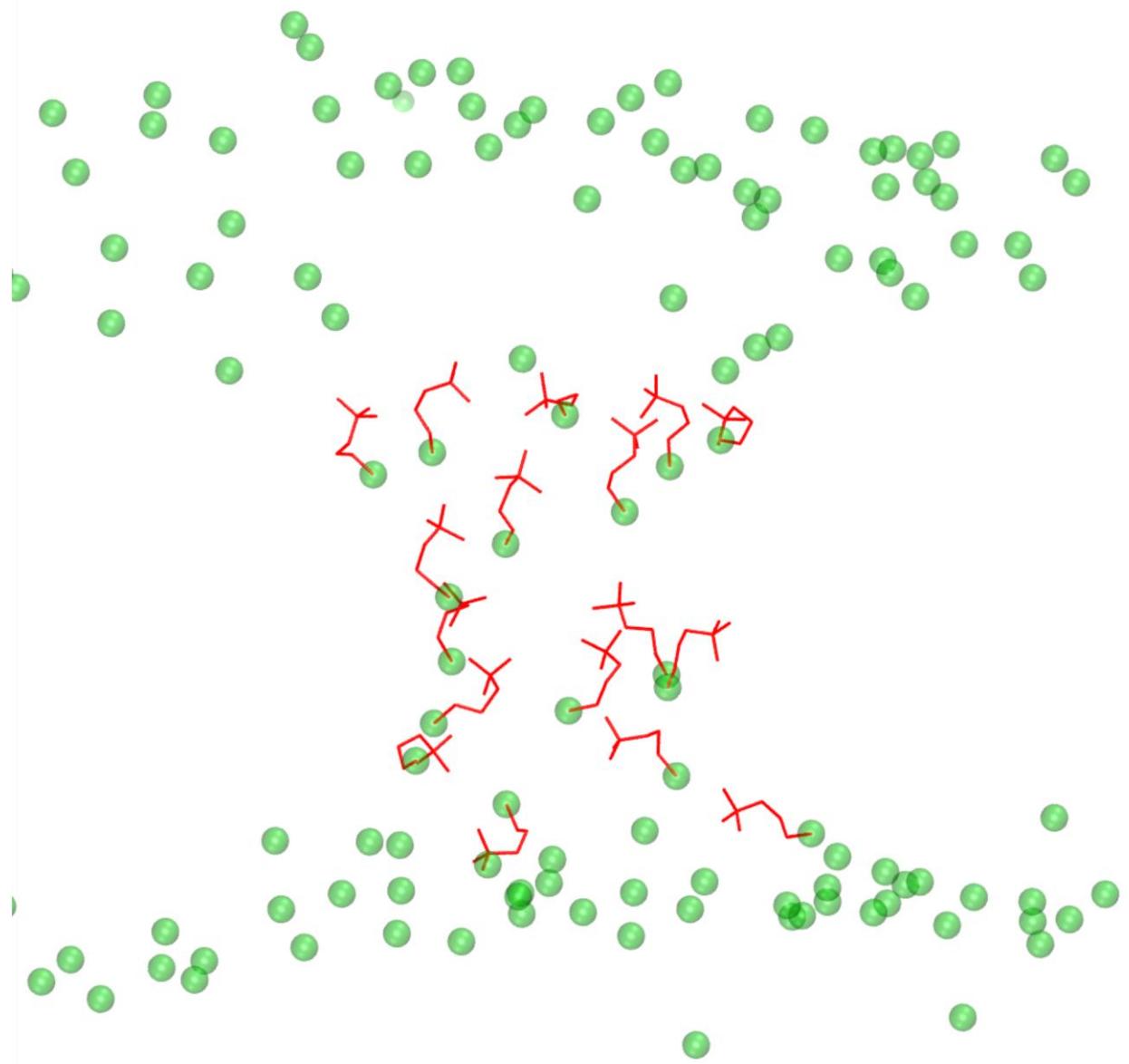


Fig. S5: Vertical alignment of choline moieties extending into center of pore from the phosphorous atoms of the DOPC headgroups in the direction of the concentrated electric flux, i.e. higher electric field. Orographic projection orthogonal to z-axis. Choline moieties are shown if the phosphorous atom is in the “core” of the pore, i.e. +/- 0.75 nm from center of membrane along the z-axis.

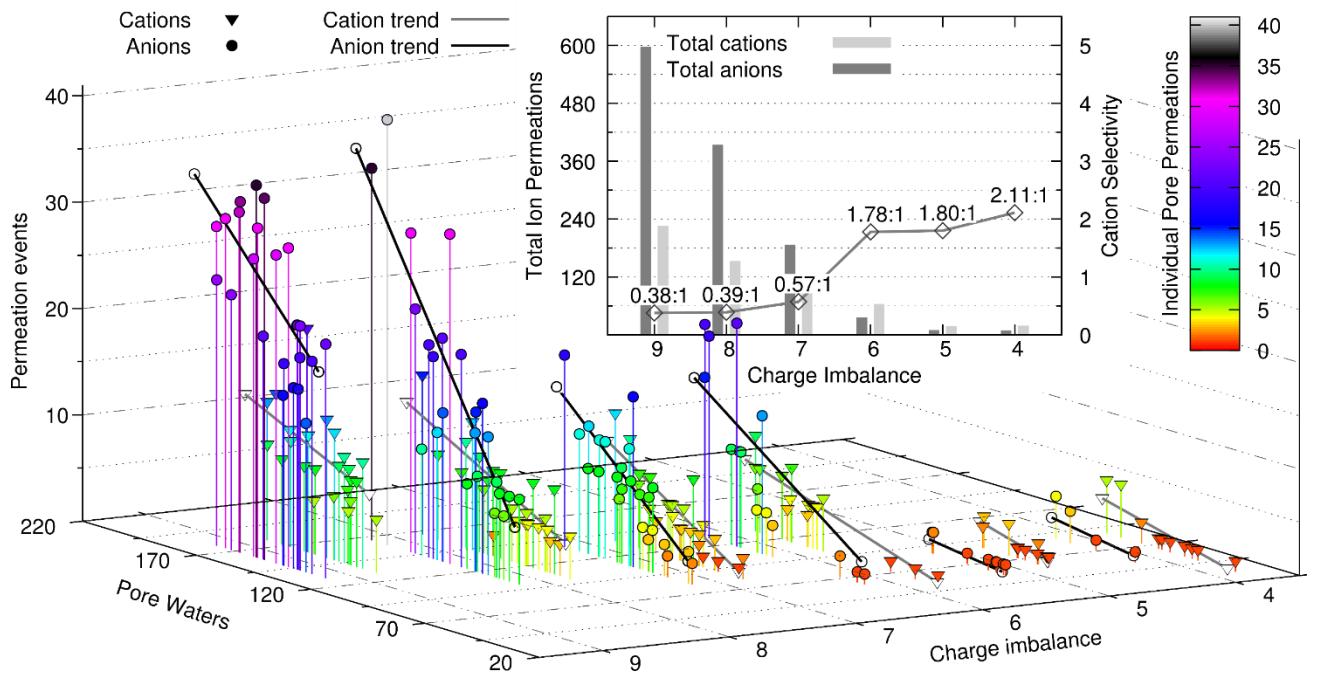


Fig. S6: Ion permeation rates vs. pore size and charge imbalance for pores containing SCMTRs. Data about each charge imbalance (integer values) are split left and right to separate permeation of anions and cations. Note relatively constant trend of cation diffusion vs. pore size compared to steadily increasing trend of anion diffusion.

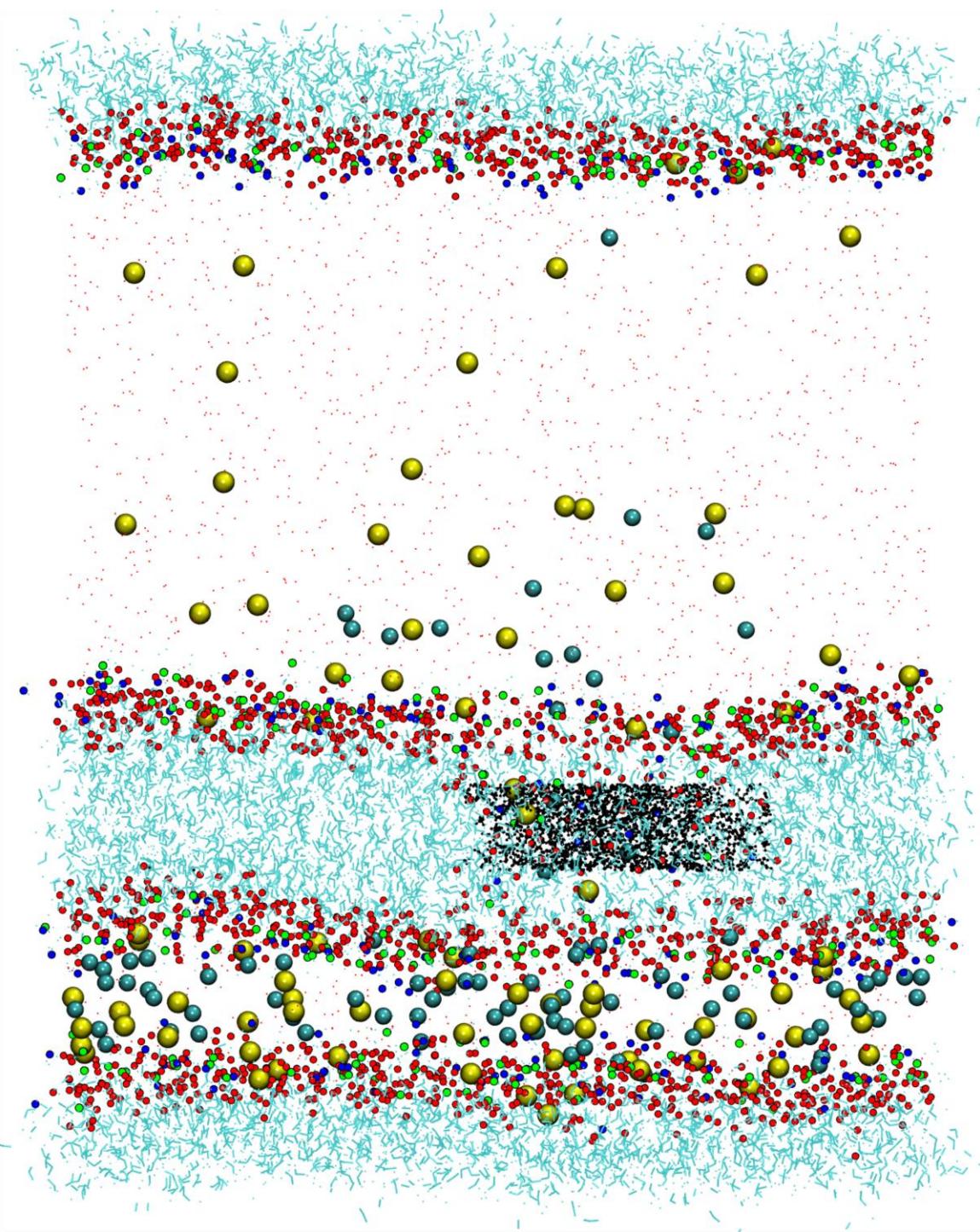


Fig. S7: Large system after 125 ns. Note distribution of chlorine ions (cyan) in upper bath. This is the extreme of the asymmetric development of the system. Disparity of data between this system and the smaller is inescapable until the larger system can be simulated in a fashion to mitigate or eradicate this asymmetric development. Similar size pores in each membrane may be sufficient.

Production run .MDP file including CompEL parameters

```
integrator          = md
dt                 = 0.002
nsteps             = 20000000 ; 40 ns
nstlog              = 1000
nstxout             = 0
nstxout-compressed = 1000
nstvout             = 0
nstfout             = 0
nstcalcenergy      = 100
nstenergy            = 1000
;
cutoff-scheme      = Verlet
nstlist              = 20
rlist                = 1.2
coulombtype         = pme
rcoulomb             = 1.2
vdwtype              = Cut-off
vdw-modifier        = Force-switch
rvdw_switch          = 1.0
rvdw                = 1.2
;
tcoupl               = Nose-Hoover
tc_grps              = tmem bmem SOL_ION
tau_t                = 1.0 1.0 1.0
ref_t                = 303.15 303.15 303.15
;
pcoupl               = Parrinello-Rahman
pcoupltype           = semiisotropic
tau_p                = 5.0
compressibility       = 4.5e-5 4.5e-5
ref_p                = 1.0 1.0
;
constraints           = h-bonds
constraint_algorithm = LINCS
continuation          = no
gen_vel                = yes
gen_temp               = 303.15
gen_seed               = -1
;
nstcomm               = 100
comm_mode              = linear
comm_grps              = upper lower
;
;;;;;;;;;  CompEL Section  ;;;;;;;;
```

```

;

swapcoords = Z

swap-frequency = 100
split-group0 = tmem
split-group1 = bmem
massw-split0 = no
massw-split1 = no
solvent-group = TIP3
coupl-steps = 10
iontypes = 2
iontype0-name = SOD
iontype0-in-A = -1      ; requested number of Na ions in
compartment A
iontype0-in-B = -1      ; requested number of Na ions in
compartment B
iontype1-name = CLA
iontype1-in-A = -1      ; -1 means: use the number of ions
iontype1-in-B = -1      ;           as found at time step 0
bulk-offsetA = 0.0
bulk-offsetB = 0.0
cyl0-r    = 5.0
cyl0-up   = 0.75
cyl0-down = 0.75
cyl1-r    = 5.0
cyl1-up   = 0.75
cyl1-down = 0.75
threshold = 1

```

Center of mass motion removal

The double membrane box geometry introduces one artifact completely unrelated to the CompEL algorithm having to do with center of mass motion removal. When three center of mass motion removal groups were selected (one for each membrane and one for the solvent and ions), GROMACS seemed to compensate for motion in one membrane in the x-y plane by imparting equal and opposite motion to the other membrane, possibly creating un-natural shear at the membrane-water boundary. To mitigate this effect, the top membrane was grouped with all the solvent and ions to produce the first motion removal group leaving only the bottom membrane to become the second group.

Production run history plots

In post processing, pore core water plots were produced for all 144 production runs, permitting a quick visual check for any post processing anomalies. For each charge imbalance, a representative set of the plots for the two membrane pores (channels 0 & 1) are shown below. (Fig.'s S8-S13)

The core water plot (purple) shows the water molecule count in the central 1.5 nm of the pore. The throat (abbr. thrt & t) water plot (evergreen) shows the water molecule count in the central 3.0 nm of the pore. Note that the throat water molecules are plotted at half scale. To accurately read the water molecules value represented by the evergreen trace, the value from the left-hand scale must be doubled.

The key will also include the number of ion permeations if there are any. Ion permeation events times are indicated by the vertical red and green lines at the bottom of the graph. The green lines (which are plotted after the red lines) are plotted at half height so that an underlying red line will not be totally obscured. Multiple red or green lines have occasionally obscured another of the same color, consequently, a few plots will have a larger number of ions shown in the key than the corresponding vertical lines.

Below the key are the simulation charge imbalance followed by the average pore and throat water molecule counts. Following this are two sets of numbers, first for the average pore and throat Cation permeation times and second set is the same pair for Anions. These represent the average, total time each ion spends in the throat and pore regions for each permeation during the simulation. The units are ns/permeation.

Of the first four plots, notice the two where the purple trace decays at a point to a value of zero. These illustrate the data from simulations where a membrane pore has totally disintegrated.

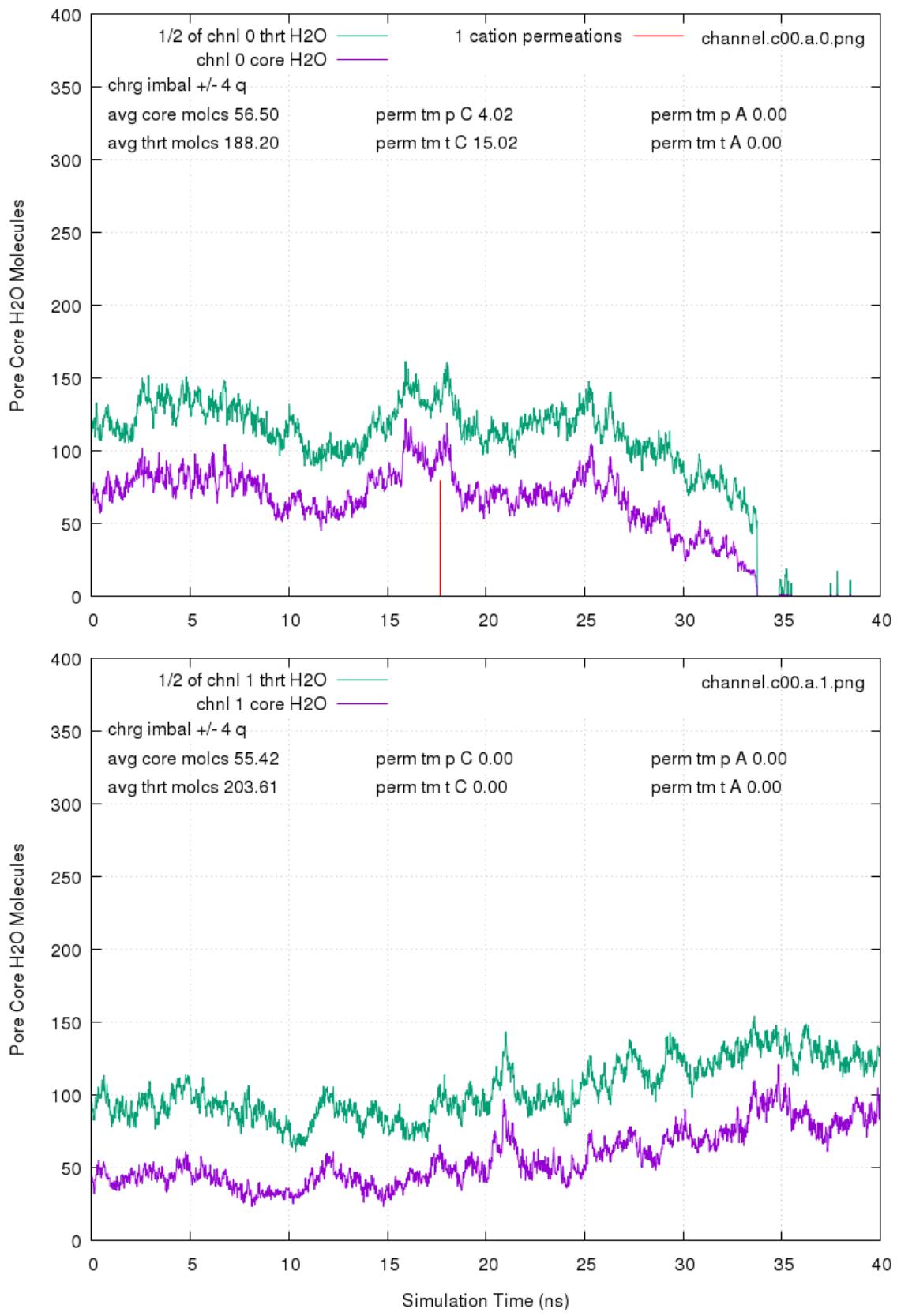


Fig. S8: Plots for production run history with +/- 4 e⁻ charge imbalance.

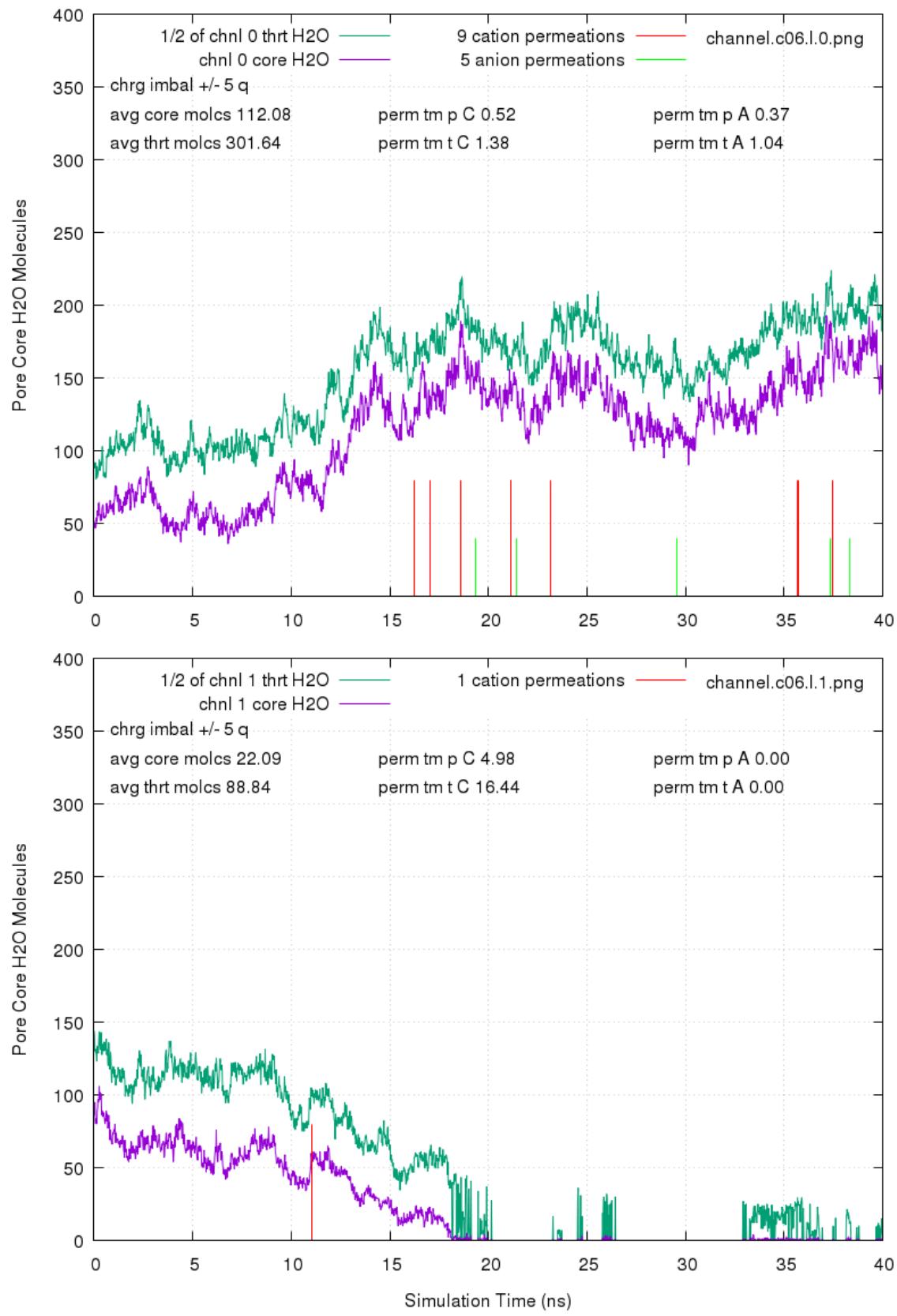


Fig. S9: Plots for production run history with +/- 5 e⁻ charge imbalance.

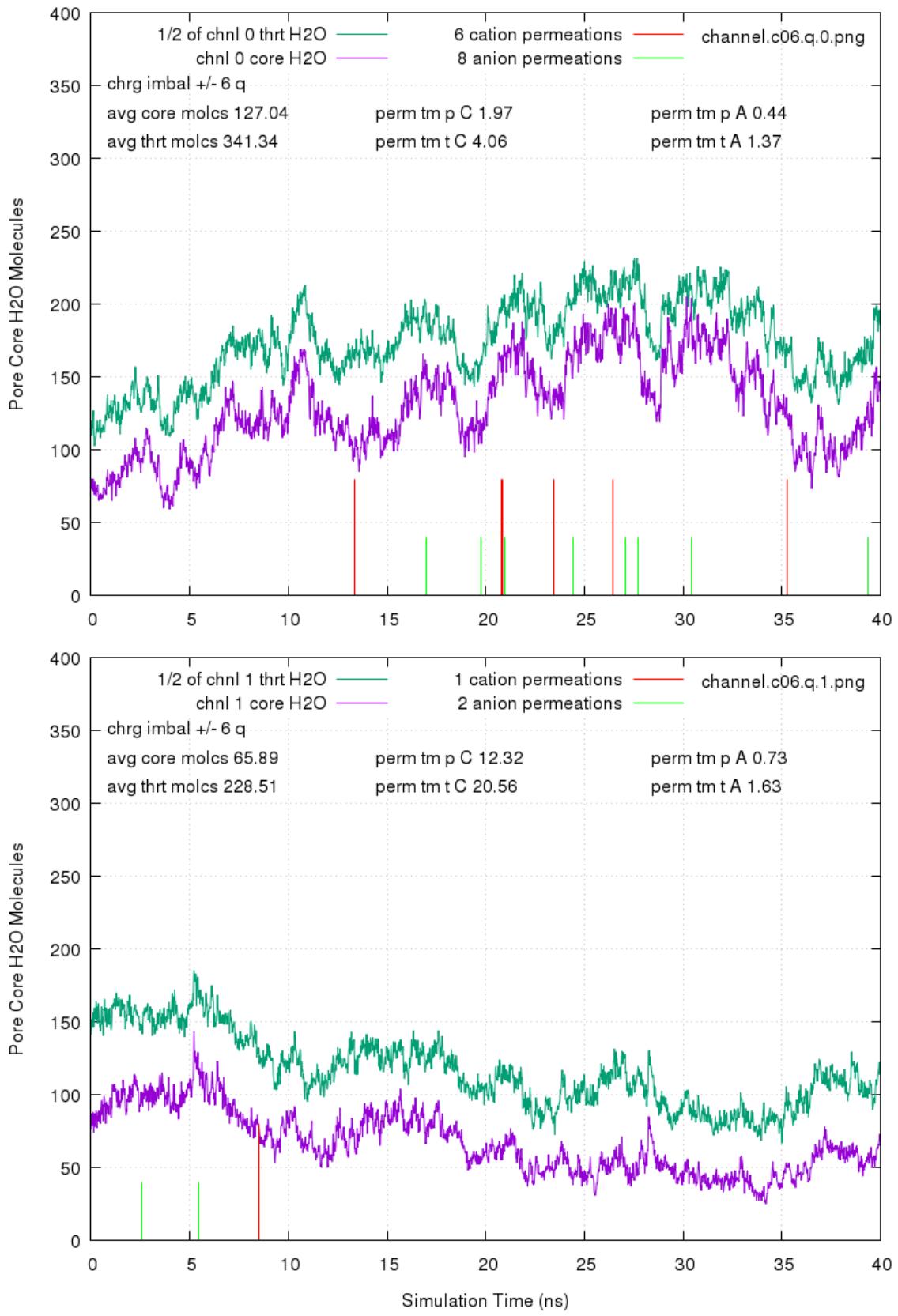


Fig. S10: Plots for production run history with +/- 6 e⁻ charge imbalance.

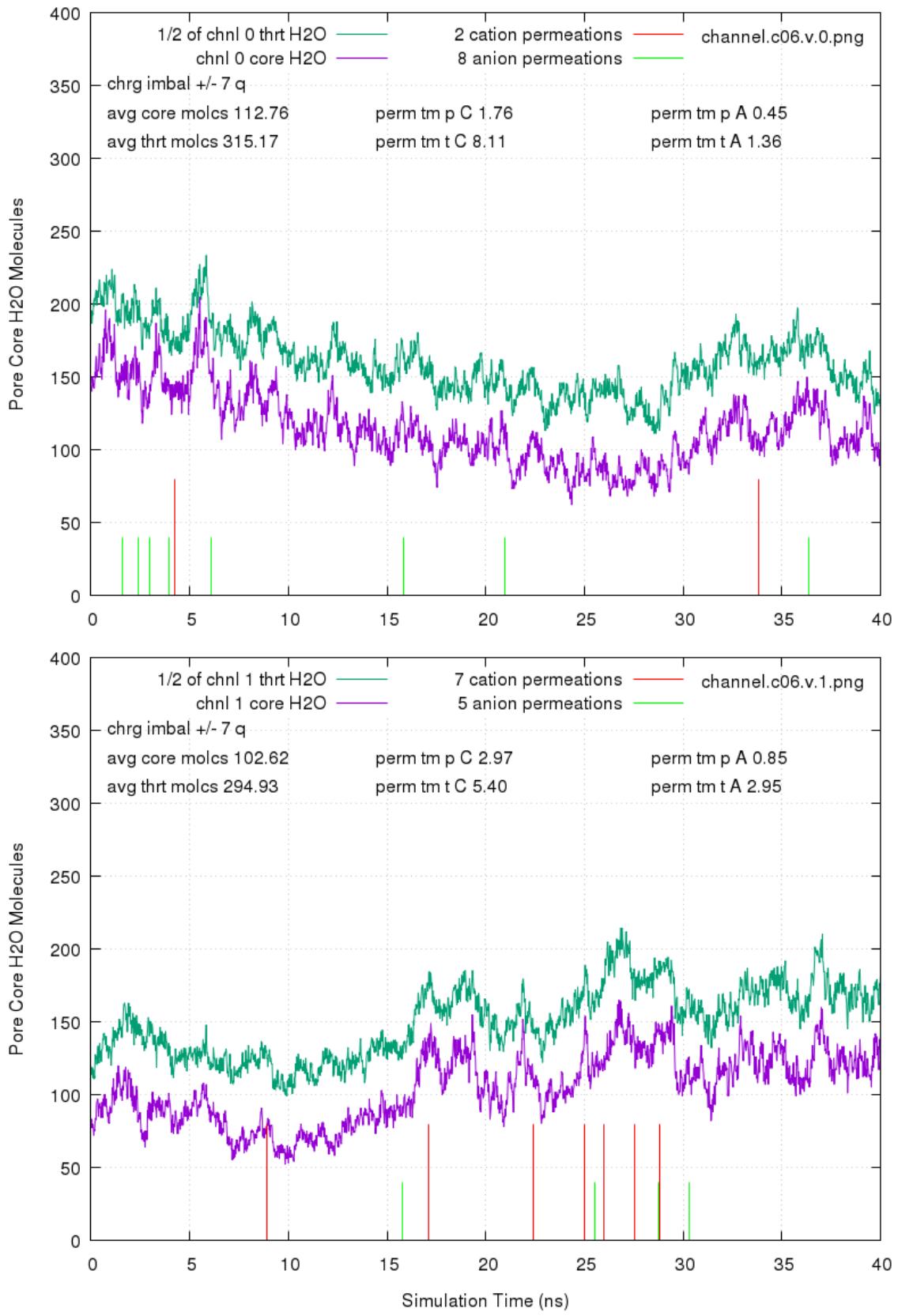


Fig. S11: Plots for production run history with +/- 7 e⁻ charge imbalance.

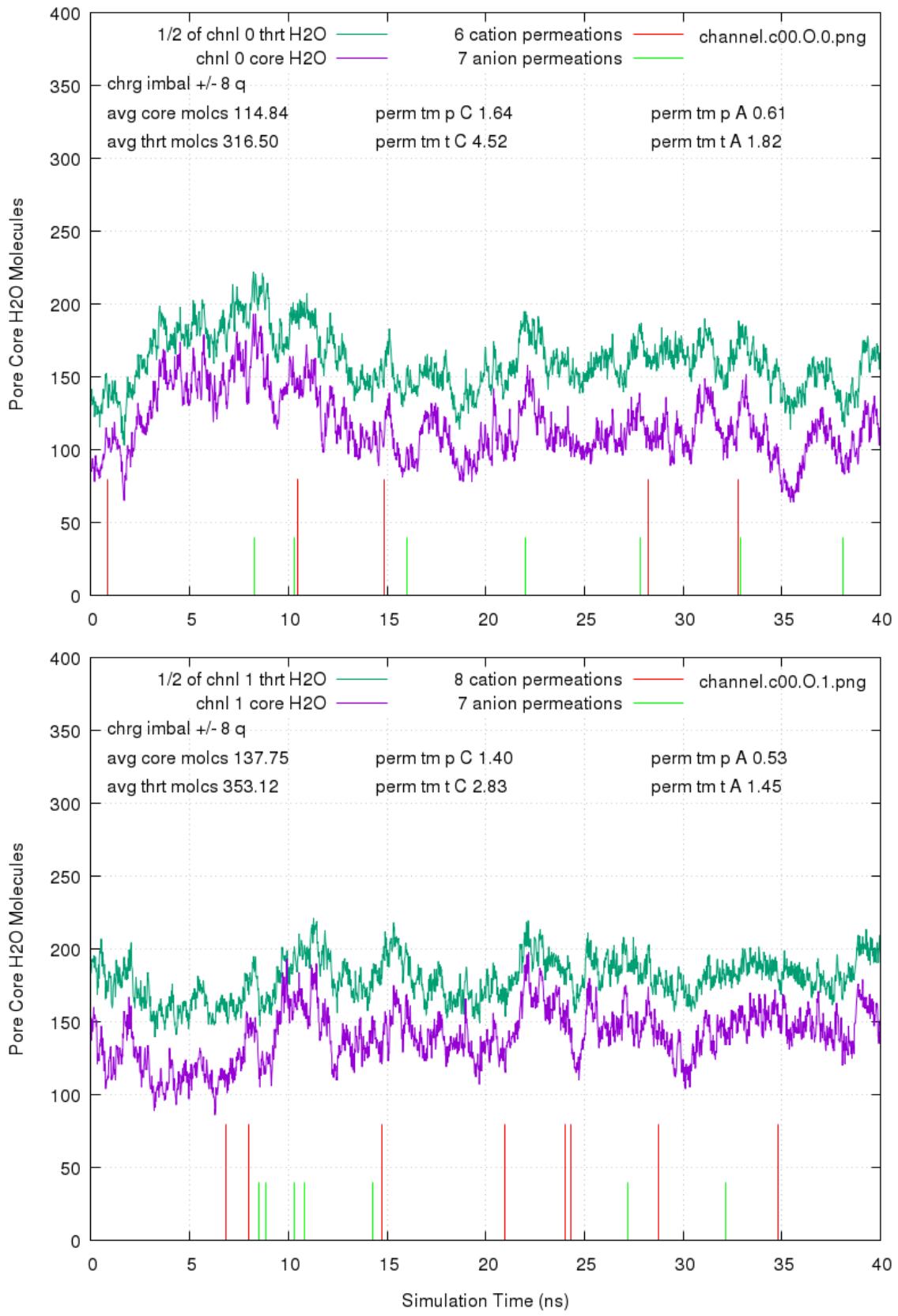


Fig. S12: Plots for production run history with +/- 8 e⁻ charge imbalance.

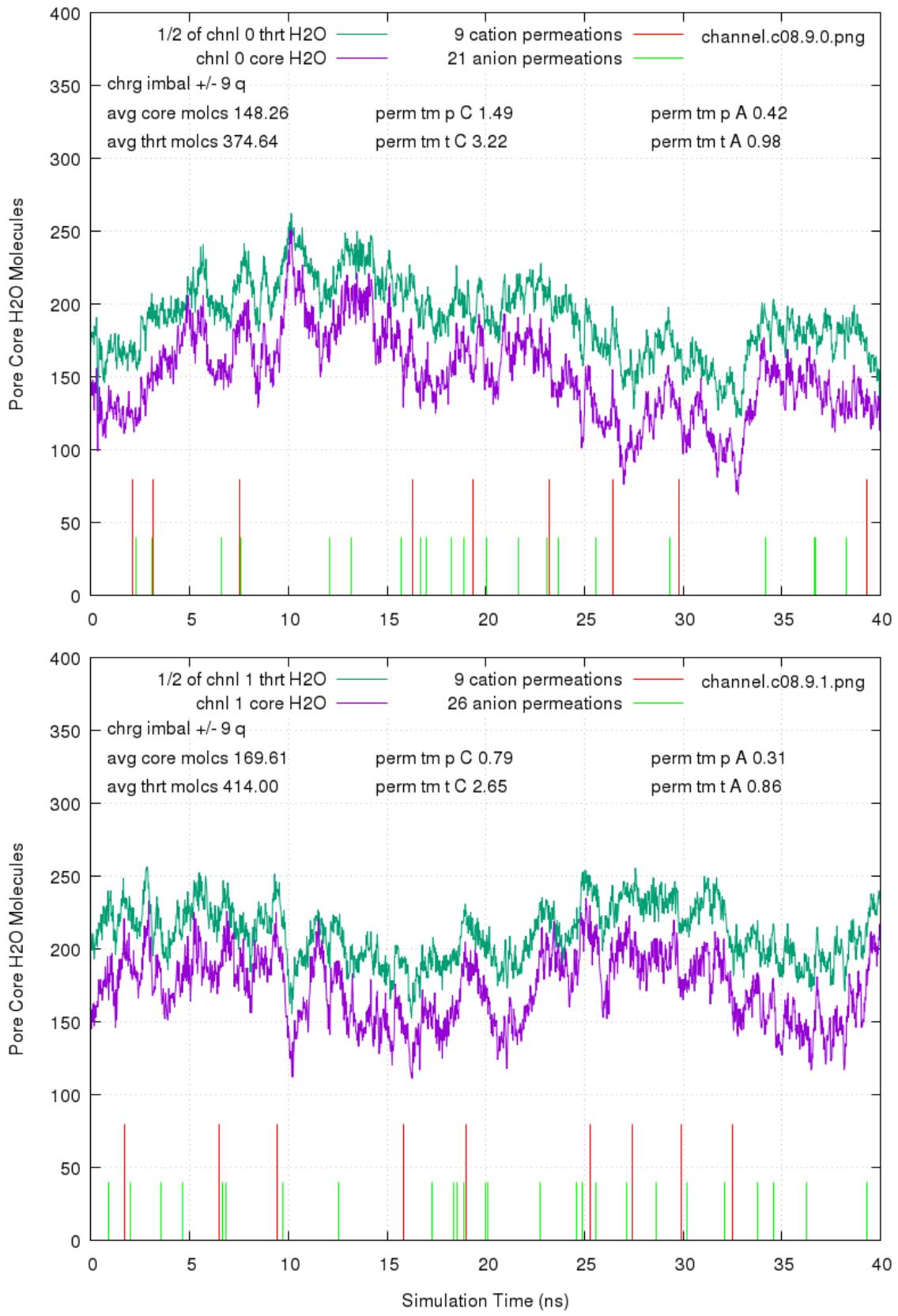


Fig. S13: Plots for production run history with +/- 9 e⁻ charge imbalance.

Post processing issues

If pores drift across periodic boundaries during a simulation, determining the coordinates of the exact center of the pore in every frame becomes problematic. There are instances where the GROMACS post processing utilities are unable to do so and will return a value of NaN (not a number). This must be corrected before any reliable counting of pore water molecules can be conducted. The GROMACS *trjconv* utility has a periodic boundary processing option -cluster that may be able to handle every situation but takes considerably more trouble to implement and time to perform.

For the problem of drifting pores, a much easier and more efficient method of correcting this problem was developed and at its state of development handles completely automatically about 98% of the problem trajectories. Simple manual tweaking of the translation values fixes the other 2%. It requires the use of three scripts over four steps, all of which have been automated. Visual inspection of the simulation trajectory history plot (which is performed anyway), will reveal those trajectories that need manual tweaking.

Links for Github access to two of the scripts are listed below (the third just plots the data). The first is named *select*. It counts water molecules and dumps trajectory coordinates. The second is named *shift* which uses data produced by a *select* routine to performs the necessary translations to keep pores from crossing periodic boundaries. To determine if a top membrane pore needs adjustment, the steps are:

- 1) Use script *select* to tabulate the pore center coordinates throughout the simulation into a file called twatc.xvg. The command is: source select twatc

- 2) Use script *shift* to translate the trajectory, if necessary, creating a file called trajout0.xtc. The source trajectory file for this operation is called trajout.xtc. If the translation is necessary and file trajout0.xtc is created, all subsequent post processing operations for channel 0 data will preferentially select trajout0.xtc over file trajout.xtc. The original file, trajout.xtc may not need any translation to extract data for channel 1 and must be retained. (Since data extraction is an activity unique to each pore, the same processes are applied to channel 1, and if necessary, will produce a preferentially selected in subsequent steps file trajout1.xtc.)
- 3) Use script *select* to tabulate the water molecules in the top (channel 0) pore on a frame by frame basis into file twat.xvg. The command is: source select twat. If a pore still comes too close to the periodic boundaries, some of the values in file twat.xvg will be NaN.
- 4) Plot the water molecules in the pore over the duration of the simulation. If plotted in gnuplot with a line, the NaN values will cause the line to jump from the last good value down to the y-axis zero value until the next real number occurs in the file. These are immediately obvious on visual inspection and comprise the roughly 2% of the trajectories that need some manual tweaking. The manual tweaking process is described later.

Listing of scripts

The first script *select* has measuring and counting routines. They are selected by the first and only parameter passed to the script. The parameter choices are present in the script as the case options. For adjusting the channel 0 trajectory frames, only the twatc and twat options are needed. The scripts are written in bash (bourne again shell).

Script: [*https://github.com/JPatrickBrian/Redstone-Engineering/blob/master/select..*](https://github.com/JPatrickBrian/Redstone-Engineering/blob/master/select..)

The second script examines file twatc.xvg (and bwatc.xvg for channel 1) to determine if/by how much a trajectory file needs to be translated to keep a pore off the periodic boundaries. It does so by creating histograms of the x and y coordinates of the center of the pore throughout the simulation and attempts to position the center of the most significant gaps in the histogram coordinates on the simulation periodic boundaries.

Script: [*https://github.com/JPatrickBrian/Redstone-Engineering/blob/master/shift*](https://github.com/JPatrickBrian/Redstone-Engineering/blob/master/shift).

Manually tweaking a trajectory translation

If a pore water molecule count file (twat.xvg) still contains nan values after being translated, it requires manual tweaking. This occurs when the pore migrates over the entire x or y dimension (or nearly so) during the simulation. The image below (**Fig. S14**) shows an example of how this migration of the pore center over the periodic boundaries compromises the counting of pore water molecules. Note the data dropout after 33 ns is due to pore collapse and not conflict with periodic boundaries.

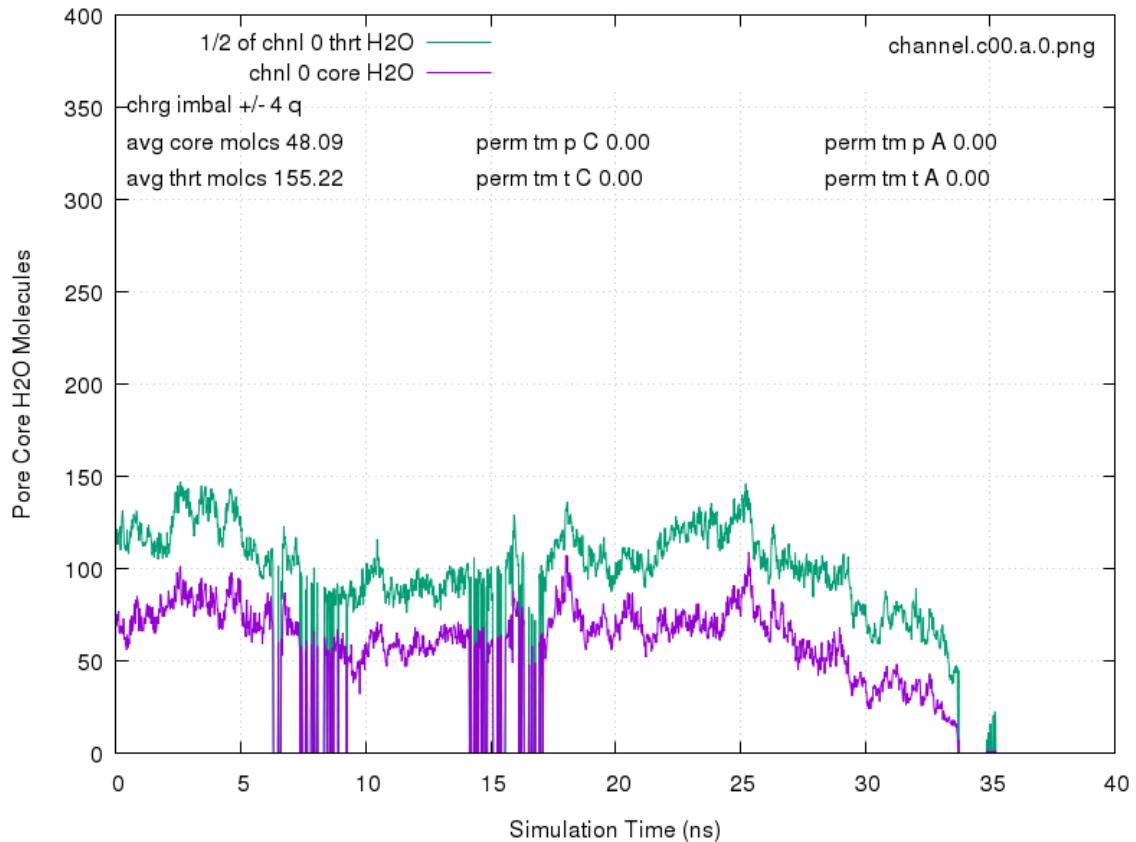


Fig. S14: Signature pattern of water molecule counts when the geometric center of a pore in a frame of data has an indeterminate (NaN) value.

The two following commands produce the histogram data for the center of the pore over the *x* and *y* dimensions. Following that is the histogram data from the first command

(x -axis). Ideally, the smallest counts (left column) at the top and bottom of the list are consistent with a unimodal distribution (this example satisfies that). The other beneficial distribution of the data is nearly equal size gaps between 0 and the first coordinate (right column) and between the last coordinate and the maximum axis dimension (this example is NOT like that since the maximum axis coordinate is about 7.3). If this example had a problem, the fix would likely be to translate the frames \sim (-2.0 nm) along the x axis. Most commonly, a trajectory translation only needs to be tweaked just a few angstroms. Occasionally, the histogram can appear to have two gaps and the solution is to translate the frames for the other gap to line up with the periodic boundaries.

```
FR=`sed -n '/      0.000/{=;q}' twatc.xvg`;tail -n +$FR twatc.xvg | awk
'{printf "%1.1f\n", $2}' | sort -k1,1n | uniq -c | grep -v nan
FR=`sed -n '/      0.000/{=;q}' twatc.xvg`;tail -n +$FR twatc.xvg | awk
'{printf "%1.1f\n", $3}' | sort -k1,1n | uniq -c | grep -v nan
```

```
1 4.0
9 4.1
14 4.2
43 4.3
107 4.4
99 4.5
90 4.6
62 4.7
95 4.8
107 4.9
150 5.0
124 5.1
73 5.2
52 5.3
64 5.4
69 5.5
33 5.6
43 5.7
25 5.8
28 5.9
31 6.0
48 6.1
77 6.2
104 6.3
94 6.4
99 6.5
82 6.6
67 6.7
59 6.8
46 6.9
6 7.0
```

Complex query explanation

Below is a copy of the query used to count the water molecules present in the core of a membrane pore on a frame by frame basis despite bulk motions of the system caused by other factors. Note that a pore must not drift across a periodic boundary in a trajectory submitted to this query. Trajectories where that is not the case must have other preprocessing to correct that before processing with this step. Only the -select command string is explained as the other parameters are sufficiently covered in the standard GROMACS documentation. Although ostensibly the syntax of the select string is also explained, without an example of a non-trivial application of its capabilities, realizing its full power may require a proficiency with structured query language statements uncommon beyond the domain of database experts.

```
gmx select -f traj.trr -s topol.tpr -os bwater.xvg -rmpbc -pbc -select
'
lhgrps = resname "DOPC" and (name "P.*" or name "N.*" or name "O." or name
"O[12].");
zm      = z of cog of lhgrps;
wcog    = cog of resname TIP3 and ( (z+.45) > zm and (z-.45) < zm );
xw      = x of wcog;
yw      = y of wcog;
zw      = z of wcog;
radius  = ((x-xw)^2+(y-yw)^2)^0.5;
lcog    = cog of lhgrps and ( (z+3) > zw and (z-3) < zw ) and radius < 1.5;
zc      = z of lcog;
dyn_mol_cog of resname TIP3 and ( (z+0.75) > zc and (z-0.75) < zc ) and radius
< 1.5
'
```

The last parameter of the select command is the -select flag selection command string and is enclosed in single quotes to prevent the operating system shell from interpreting and processing any text between the quotes. Explaining the statement line by line:

```
lhgrps = resname "DOPC" and (name "P.*" or name "N.*"
or name "O." or name "O[12].");
```

`lhgrps` stands for *lipid head groups*. Here it exists as what GROMACS calls a *variable* and in this syntax, it is equated to an *expression*.

Variables are just names used to help us remember what they represent, because when used later, they become a shorthand method of writing the much longer and more complicated *expression*. Every line beginning with an “=” sign (the first 9 of the selection string) are simply equating *variables* to *expressions*. Each *expression* is a short and relatively simple use of an element of the selection syntax described at

<http://manual.gromacs.org/documentation/2018/onlinehelp/selections.html>.

In fact, this entire string only executes one single element of the selection syntax, which is specified on the very last line. By using *variables* with functionally descriptive monikers, a much simpler written statement causes the invocation of a much more complex fully detailed statement that precisely articulates every condition and limitation the candidate water molecules must satisfy to be included in the result. The problem non-database experts encounter is that the fully detailed statement is so seemingly complicated as to appear to be unreadable. Even database experts find it daunting and would use variables as well because it increases readability, maintainability and statement writing efficiency.

If this seems arbitrarily and unnecessarily complex, it is because structured query language statements, all types including select, must be executed on databases one at a time in manner that would be analogous to a run-on sentence, paragraph and chapter if need because there can be one and only one period at the end of the instruction. They cannot be “split” into two sentences. If a statement hasn’t finished detailing every aspect of the

selection criteria into one sentence, it must continue until it has. Technically, this requirement comes from the necessity for databases to maintain something called “transactional integrity” which is beyond the scope of this explanation. And our trajectory file which our data comes from is not a database.

Nevertheless, it is very common for select commands to access and merge data from files that are not relational database files with data from files that are. To maintain functional compatibility between databases created by different companies, structured query language statements must conform to a standard known as ODBC. ODBC stands for open database connectivity. Whether the GROMACS software has been developed with some future interactivity in mind or not, and whether it is ODBC compliant or not is not clear to us. But making statement mistakes when trying to fashion complex queries give zero feedback about the nature or location of the problem other than “syntax error”. To deal with this at the user end, it means a way is needed to understandably create a complex selection criterion, in a fashion that can be simply read and understood. One method is to use *variables*.

The select element syntax allows the selection of atoms, molecules and residues using three different types of *expressions*, numerical, atom and position. Think of it like this: Correctly written *expressions* define a specific subset of *atoms*, sets or subsets of specific *atoms* are located at a specific set of *positions* and each *position* in that set is comprised of three x,y and z axis coordinate *numerical* values. Note there are expressions that can define sets or subsets of atoms and numerical values as well. By appropriate use of applicable properties of each of these types of sets, or subsets of them, as suits the purpose, a variety of data element types can be extracted from a trajectory file.

In the first selection string line, *variable* `lhgrps` is equated to an *atom_expression*. Henceforth, `lhgrps` becomes a *variable* representing a specific set of atoms. Which ones? For starters, any atom whose parent residue name is “DOPC” but with an additional stipulation. The additional stipulation is added with the “and” and the statement between the following parentheses. Between the “()”, the new criteria are added that relates to each atom’s name, but by using a scripting convention called regular expressions, used here because they allow the creation of “a complex selection criterion, in a fashion that can be simply read and understood”. With regular expressions, that can only happen after learning the conventions and syntax for writing and implementing them. The reality is that without a decent working knowledge of regular expressions, the capacity to implement the FULL power of dynamic select statements will remain substantially out of reach. The scope of regular expression capabilities is enormous, is fully documented elsewhere, and is vastly beyond the scope of this explanation. However, the statement used in this script will be explained.

```
(name "P.*" or name "N.*"
```

Selects any atom whose name starts with the letter P or N, irrespective of the total name length.

```
or name "O."
```

Selects any atom whose name starts with the letter O and has a total name length of exactly 2 characters.

```
or name "O[12].");
```

Selects any atom whose name starts with the letter O, whose second name character is the numeral 1 or 2 and has a total name length of exactly 3 characters. The semicolon at the end of the line denotes that the variable *expression* ends here.

Given a fully labelled drawing of a DOPC molecule and the explanation above, one has the information to be able to determine the number of atoms in the set represented by the variable `lhgrps` for a single molecule.

```
zm = z of cog of lhgrps;
```

This is a *numerical* expression. The set of atoms represented by `lhgrps` has a center of geometry (as opposed to its center of mass). The variable `zm` represents the numerical value of the z-axis coordinate of this center of geometry.

```
wcog = cog of resname TIP3 and ((z+.45)>zm and (z-.45)<zm);
```

This is a *position* expression. TIP3 water molecules whose z-axis center of geometry coordinate value is within +/- 0.45 nm of the value of `zm` are a subset of all TIP3P water molecules. The variable `wcog` represents the *position* of the center of geometry of this subset of atoms.

```
xw      = x of wcog;
yw      = y of wcog;
zw      = z of wcog;
```

These are *numerical* expressions. The variables `xw`, `yw` and `zw` are given the values of the x,y and z axis coordinates of the *position* represented by `wcog`.

```
radius = ((x-xw)^2+(y-yw)^2)^0.5;
```

This is a *numerical* expression and should be self-explanatory. The interesting thing is that the *expression* is not linked to any specific *position* or sets of *positions* without which a logical source for the values of x and y will be undefined. This linking can be established in the statements where the *expression radius* is invoked to act upon a *position* or set of *positions*. This capability is elegant and powerful, and the full implications of this ability require some rumination to fully appreciate.

```
lcog = cog of lhgrps and ((z+3)>zw and (z-3)<zw) and  
radius < 1.5;
```

This is a *position* expression. The set of atoms represented by `lhgrps` whose z-axis coordinate value is within +/- 3 nm of the value of variable `zw` and are closer horizontally to the x,y coordinates of position `wcog` than 2 nm are a subset of `lhgrps`. `lcog` represents the position of the center of geometry of this subset. Due to its placement in the *expression*, the variable `radius` gets the x and y coordinates it needs from the *positions* of the *atoms* in the set represented by `lhgrps`.

```
zc = z of lcog;
```

Numerical expression.

Now all the variables needed have been defined to be able to write the statement in a simplified fashion to extract the data sought, specifically the number of water molecules in this frame that satisfy the selection criteria. So far, the actions of the variable expressions have been explained, but not the geometrical significance of these expressions in relation to the biostructure in which it will count a very specific subset of the water molecules.

Initially the z coordinate of the center of the membrane of interest is needed. `1hgrps` defines the set of atoms used to do this and `zm` gets this from the collective center of geometry of this set. `wcog` gets the center of geometry of water molecules present in a thin horizontal slice centered on `zm`. In normal cases, this should be very close to the center of the pore, particularly on the x,y plane. Because larger membranes exhibit waves and undulations, pores move up and down with these undulations. To move with these undulations, we need the z coordinate of the center of geometry of the lipid headgroups in the immediate vicinity of the pore. This will allow the query to keep track of the pore's center in the z direction as it moves up and down with the waves and undulations. Variables `lcog` and `zc` achieve this aim.

With this foundation laid, the GROMACS *select* statement can be executed. In the statement options, option flag `-os` is specified which writes a file recording the number of entities that satisfy the selection criteria in each frame, in this instance the number of water molecules in the core of the pore. The statement that brings this about is the final line of the select string.

```
dyn_mol_cog of resname TIP3 and ((z+0.75)>zc and (z-0.75)<zc) and radius < 1.5
```

This is a position expression. The `dyn_mol_cog` specifies to select center of geometry positions (`_cog`), for molecules (`_mol`), on a frame by frame basis (`dyn`). The number of those positions that satisfy the selection criteria will be recorded into the file specified after the `-os` option flag. The selection criteria are all water molecules (`resname TIP3`) whose elevations are within +/- 0.75 nm of the z coordinate of the

center of the membrane in the immediate vicinity of the pore and whose horizontal distance from that x,y center are less than 1.5 nm.

The presence of the `dyn` modifier at the beginning of the statement causes all the dependent variables it references to also be evaluated on a frame by frame basis. The result is a completely accurate count of the water molecules that satisfy the criteria, unaffected by the bulk motions of the system.

It is one thing to read a description of how and why a statement achieves its stated purpose. It is another to acquire enough of the skill to be able to write the needed statements to do novel data extraction from other trajectories to address the next research analysis problem that crops up. In our experience, that has only developed through a commitment to learn the necessary skills. Hopefully this example has demonstrated the value of the available tools and enough insight to make others' attempts at learning them successful.

SECTION 2

Post-Processing and Visualization

Electric Fields and Computational X-Rays

ABSTRACT

In section 1, links are provided to two post-processing algorithms posted on Github essential to extracting the results of that paper from the simulation trajectories. The actions of those routines are fully documented in the paper and the source code and a full understanding of their role is required to fully understand the additional processing described in this section required to map the three-dimensional, non-uniform electric fields. These additional routines are fully described and documented in part 1 of this section. As described in the introduction, while attempting to validate the integrity of the data files along the processing path for the electric field maps, computational x-rays were inadvertently created. Computational x-rays, unlike normal x-rays, not only reveal structure, but composition as well. To produce them requires two additional steps, one in

post-processing to calculate number densities throughout the simulation volume (essentially composition), and a second to convert the data to a format that can properly visualize the resulting information. These additional steps along with some discussion of the future capabilities and applications of computational x-rays are included in part 2 of this section.

PART 1 – MAPPING ELECTRIC FLUX FROM SIMULATION TRAJECTORIES

These scripts and processes were developed over a two-year period in support of my doctoral research. Their development was an evolutionary process. None of this work was imagined or planned in the beginning. Unexpected results were encountered, and processes were developed to better measure and verify the initial indications. This process is nowhere close to being complete. There is much more work to do to better understand what has been discovered, and to bring more utility to the work that has been completed so far. It is beyond the scope of this document to explain everything that is taking place. The purpose of this document is to outline the sequence of steps required to implement these tools on simulation trajectories to produce the resulting analysis. The first step to understanding this work would be to carefully read the supporting information in section 1. This will give an introduction of the scripts select. and shift. These enable a very accurate count of water molecules in a membrane pore on a frame by frame basis. They also prepare the trajectories for the following processing.

The title of my dissertation proposal was "Mapping three-dimensional interactions between biomolecules and electric fields". To do this, I needed to create trajectories that had all extraneous motion relative to the biostructure I wanted to study the effect of the

non-linear electric fields on, and have that biostructure perfectly centered in the middle of the simulation box. I think the best way to demonstrate how this is accomplished is to show the sequence of commands to produce the final images to be generated. I'll intersperse comments as needed. All the routines are written in BASH, (bourne again shell) calling standard Unix or GNU utilities. **Fig. XX** is a screen capture of the beginning of my Github repository where a full listing of every routine can be examined or downloaded. Its web address is:

<https://github.com/JpatrickBrian/Redstone-Engineering>

Branch: master		New pull request	Create new file	Upload files	Find file	Clone or download
JPatrickBrian Update METHODS Latest commit cb3e5f5 5 hours ago						
3dchrg4A.awk	Update 3dchrg4A.awk	21 days ago				
3dchrg4C.awk	Add files via upload	24 days ago				
3dchrg4H.awk	Add files via upload	24 days ago				
3dchrg4L.awk	Add files via upload	24 days ago				
3dchrg4N.awk	Add files via upload	24 days ago				
3dchrg4P.awk	Add files via upload	24 days ago				

Fig. 00: Screen capture of top of home page of my Github repository for making code freely available to the research community at large.

These following steps run the initial simulation, then do a quick (cheap computationally) analysis to create some plots that can be reviewed to make sure everything looks OK. For simplicity and completeness every command including the generation of the initial trajectory files is shown

```

# runs simulation
gmx mdrun -cpi
# makes molecules whole with respect to periodic boundaries
gmx trjconv -f traj_comp.xtc -o trajout.xtc -pbc mol -skip 10 <
    <(echo -e "System\n")
# crudely, but quickly counts water molecules in top membrane
gmx select -f trajout.xtc -n index.ndx -s topol.tpr -os
    waters.xvg -select 'dyn_mol_com of resname TIP3
        and ((z+.75) > (z of com of (group DOPC and name P))
        and (z-.75) < (z of com of (group DOPC and name P)))'
# crudely, but quickly counts water molecules in bottom membrane
gmx select -f trajout.xtc -n index.ndx -s topol.tpr -os
    watersb.xvg -select 'dyn_mol_com of resname TIP3
        and ((z+.75) > (z of com of (group D2PC and name P))
        and (z-.75) < (z of com of (group D2PC and name P)))'
# tabulates totals of ion crossing events for each pore
source ..../jobs/cross.
# plots the initial results, validation of no extreme problems
gnuplot -c ..../jobs/cross
#
# The next steps make very accurate counts of membrane pore water
molecules.
source ..../jobs/select. twatc
source ..../jobs/select. bwatc
source ..../jobs/shift. run
source ..../jobs/count.

```

Now for each membrane pore (each of two membranes in the sim have one pore), a trajectory is produced in which the center of the pore is translated to the exact center of the simulation box. From these trajectories, a set of coordinates for every atom in every frame of the simulation is created (coord.xvg). This file is used by script 3dchrg4A.awk to produce the composite electric charge map of the simulation volume elements. Finally, this file is processed by Bfinteg8.awk to produce the electric field and potential maps. In this example, channel 0 (top) is being processed. #!NOTE: Since my work has focused at this point on the analysis of only one of the two pores in any single simulation, the file names would conflict if both pores were analyzed simultaneously.

```
# centers a pore in the exact center of the simulation box,
one parameter is passed, 0 - center top pore, 1 - center
```

```

bottom pore
source ../jobs/center. 0

```

This is a rather complex routine, so we will list its details. The comments (following #) in blue are not contained in the listings but added here for clarity of the dissertation.

```

#!/bin/bash -l
#=====
# center.      ### script for centering pores    ###
#           ### in trajectories                  ###
# parameters
# $1           channel to center
#   options: 0 - channel 0 in the top membrane
#           1 - channel 1 in the bottom membrane
#_____
```

```

case ${1} in
# this section handles the top channel
0)
# trajout.xtc is the original trajectory file, if no
shifting was required to count the water molecules, we can
use that file. If shifting was required for the top
channel, that will exist as trajout0.xtc. Pick accordingly
if [ -s trajout0.xtc ]; then
  tfyle="trajout0.xtc"
else
  tfyle="trajout.xtc"
fi
# if the box dimensions file doesn't exist, create it
if ! ( [ -a boxcntr.xvg ] ); then
  source ../jobs/boxdims.
fi
# load the box center dimensions
declare -a boxcntr
boxcntr=($(cat boxcntr.xvg))
# we must make sure the pore center coordinates file
matches the most recent version of the trajectory file
if [ "${tfyle}" == "trajout0.xtc" -a ${tfyle} -nt
"twatc.xvg" ]; then
  source ../jobs/roll. twatc.xvg
  source ../jobs/select. twatc
fi
# we need to regenerate the pore center coordinates, do so
if ! ( [ -s twatc.xvg ] ); then source ../jobs/select.
twatc; fi

```

```

# calculate the translation vectors for every frame
    s ..../jobs/trans. ${boxcntr[0]} ${boxcntr[1]}
${boxcntr[2]} twatc.xvg
# translate the frames
    gmx trjconv -f ${tfyle} -o /local/scratch/temp.gro -pbcc
mol -sep -exec 'source ..../jobs/translate.' < <(echo 0)
# if we are about to overwrite the centered trajectory
file, give it a unique index number instead to preserve
geneology
    if [ -s "trajout0c.xtc" ]; then source ..../jobs/roll.
trajout0c.xtc; fi
# concatenate the centered frames into a new trajectory
file
    gmx trjcat -f /local/scratch/tump{0..2000}.xtc -o
trajout0c.xtc
# remove the no longer needed individual trajectory frames
    rm /local/scratch/tump*.xtc
# in the Jaeger lab cluster, we compile our GROMACS to run
in single precision. To perfectly center the pore requires
a second pass identical to the first. The same comments
would apply
    tfyle="trajout0c.xtc"
    source ..../jobs/select. twatc
    s ..../jobs/trans. ${boxcntr[0]} ${boxcntr[1]}
${boxcntr[2]} twatc.xvg
    gmx trjconv -f ${tfyle} -o /local/scratch/temp.gro -pbcc
mol -sep -exec 'source ..../jobs/translate.' < <(echo 0)
    source ..../jobs/roll. trajout0c.xtc
    gmx trjcat -f /local/scratch/tump{0..2000}.xtc -o
trajout0c.xtc
    rm /local/scratch/tump*.xtc
;;
# this section handles the bottom channel. Other than
different file names and parameter options, it is a carbon
copy of the first section. The same comments would apply
1)
    if [ -s trajout1.xtc ]; then
        bfyle="trajout1.xtc"
    else
        bfyle="trajout.xtc"
    fi

    if ! ( [ -a boxcntr.xvg ] ); then
        source ..../jobs/boxdims.
    fi

declare -a boxcntr

```

```

boxcntr=$(cat boxcntr.xvg)

if [ "${bfyle}" == "trajout1.xtc" -a ${bfyle} -nt
"bwatc.xvg" ]; then
    source ../jobs/roll. bwatc.xvg
    source ../jobs/select. bwatc
fi

if ! ( [ -s bwatc.xvg ] ); then source ../jobs/select.
bwatc; fi

source ../jobs/trans. ${boxcntr[0]} ${boxcntr[1]}
${boxcntr[2]} bwatc.xvg

gmx trjconv -f ${bfyle} -o /local/scratch/temp.gro -pbc
mol -sep -exec 'source ../jobs/translate.' <<(echo 0)

if [ -s "trajout1c.xtc" ]; then source ../jobs/roll.
trajout1c.xtc; fi

gmx trjcat -f /local/scratch/tump{0..2000}.xtc -o
trajout1c.xtc

rm /local/scratch/tump*.xtc

bfyle="trajout1c.xtc"
source ../jobs/select. bwatc
source ../jobs/trans. ${boxcntr[0]} ${boxcntr[1]}
${boxcntr[2]} bwatc.xvg
gmx trjconv -f ${bfyle} -o /local/scratch/temp.gro -pbc
mol -sep -exec 'source ../jobs/translate.' <<(echo 0)
source ../jobs/roll. trajout1c.xtc
gmx trjcat -f /local/scratch/tump{0..2000}.xtc -o
trajout1c.xtc
rm /local/scratch/tump*.xtc
;;
*)

echo center : ${1} is not a valid case
;;
esac

```

Now that the trajectory has been centered, a set of coordinates for every atom in every frame of the simulation is created (coord.xvg), no small task since its size is only 61 GB. This file is used by script 3dchrg4A.awk to produce the composite electric charge map

of the simulation volume elements. The other routines 3dcharg4H.awk through 3dchrg4S.awk calculate the same for individual moieties of interest in the system. Finally, these files are processed by Bfintegr8.awk to produce the electric field and potential maps (by GNUPLOT routine plfldmaps). In this example, channel 0 (top) is being processed. #!NOTE: Since my work has focused at this point on the analysis of only one of the two pores in any single simulation, the file names would conflict if both pores were analyzed simultaneously.

```
Gmx trajectory -f trajout0c.xtc -ox coord.xvg -n index.ndx <
<(echo -e 'System\n')
source ../jobs/3dchrg4A.awk 50 chrgs3.xvg chrg5A.xvg 1 4 10
11
source ../jobs/3dchrg4H.awk 50 chrgs3.xvg chrg5H.xvg 1 4 10
11
source ../jobs/3dchrg4N.awk 50 chrgs3.xvg chrg5N.xvg 1 4 10
11
source ../jobs/3dchrg4C.awk 50 chrgs3.xvg chrg5C.xvg 1 4 10
11
source ../jobs/3dchrg4P.awk 50 chrgs3.xvg chrg5P.xvg 1 4 10
11
source ../jobs/3dchrg4W.awk 50 chrgs3.xvg chrg5W.xvg 1 4 10
11
source ../jobs/3dchrg4R.awk 50 chrgs3.xvg chrg5R.xvg 1 4 10
11
source ../jobs/3dchrg4T.awk 50 chrgs3.xvg chrg5T.xvg 1 4 10
11
source ../jobs/3dchrg4Y.awk 50 chrgs3.xvg chrg5Y.xvg 1 4 10
11
source ../jobs/3dchrg4L.awk 50 chrgs3.xvg chrg5L.xvg 1 4 10
11
source ../jobs/3dchrg4S.awk 50 chrgs3.xvg chrg5S.xvg 1 4 10
11
#!NOTE The above 11 lines can be invoked with source lchrgall.
Source ../jobs/Bfintegr8.awk 50 chrg5A.xvg
source ../jobs/Bfintegr8.awk 50 chrg5H.xvg
source ../jobs/Bfintegr8.awk 50 chrg5N.xvg
source ../jobs/Bfintegr8.awk 50 chrg5C.xvg
source ../jobs/Bfintegr8.awk 50 chrg5P.xvg
source ../jobs/Bfintegr8.awk 50 chrg5W.xvg
source ../jobs/Bfintegr8.awk 50 chrg5R.xvg
```

```

source ../jobs/Bfintegr8.awk 50 chrg5T.xvg
source ../jobs/Bfintegr8.awk 50 chrg5Y.xvg
source ../jobs/Bfintegr8.awk 50 chrg5L.xvg
source ../jobs/Bfintegr8.awk 50 chrg5S.xvg
#!NOTE The above 11 invocations of Bfintegr8.awk can be
invoked with source integr8all.
# split apart the chrg5?.xvg files into the electric field
and electric potential sections
for s in A H N P W S C R T L Y; do
    s ../jobs/spltfld. ${ARG2:0:1} ${s}
done
# plot all the electric field and electric potential plots
gnuplot -c plfldmaps
#
# The Net Electric Flux maps show the resulting field and
potential for every atom in the system. The plot with the
fyl* name shows the electric field and the plot with the
fylt* name shows the potential. The other plots show the same
fields and potentials for the indicated moieties in the
system. Some of these are in the range of ten times the
magnitude of the overall field and potential. Each plot is
scaled to reveal all the detail present in the maps for that
particular moiety. The relative values of these scales are
contained in the a[script] array in the gnuplot script
plfldmaps.

```

Note that the 3dchrg4?.awk routines have three or seven parameters passed to them. Every one of the routines on Github have comment section at the beginning that gives the necessary information for their use, including the description of the parameters passed to them, the options for them and their default values. Note that not all parameters can necessarily have a default value specified. All of these from the Github repository are listed in appendix A. Those for the 3dchrg4?.awk and Bfintegr8.awk files are shown here for illustration in **Fig.’s 00-99.**

```

#!/bin/bash -l
=====
# 3dchrg4A.awk ### calculate avg charge in bins of volume for a trajectory #####
#           #####
# parameters
# $1      number of horizontal (x-y) slices to create along z axis (2-?)
# options: the output of this algorithm can be used to calculate non-uniform
#          electric fields or to produce computational x-rays. since the
#          computational cost of calculating electric fields scales as N6, and
#          that algorithm must process all the slices created here, balancing
#          resolution and computational time of both steps is required subject
#          to the requirements and limitations of each system on which it is used.
#          higher resolutions can be produced here if the intention is to
#          produce computational x-rays since they incur negligible scaling
# $2      name of file containing charges of each atom in system
# $3      output filename containing bins of average charge for each system volume
#          element
#
ARG1=${1:-'50'}
ARG2=${2:-'chrgs2.xvg'}
ARG3=${3:-'chrg4.xvg'}

```

Fig. 00: Parameters and explanation on use of routine 3dchrg4A.awk

```

#!/bin/bash -l
=====
# 3dchrg4H.awk ### calculate avg charge in bins of volume for a moiety #####
#           ### H is the lipid atoms between the phosphate group up #####
#           ### to and including the ester groups. #####
# parameters
# $1      number of horizontal slices
# $2      file name, contains the charges for each atom in the .gro file
# $3      name of the output file
# $4      membrane channel to analyze
# options 0 - top membrane, 1 - bottom membrane
# $5      number of extra water molecules in simulation
# #! NOTE there should have always been 11480 H2O molecules in each
#           simulation. Somewhere, somehow, another variant with 11484 H2O
#           molecules crept in. We have to know which we're dealing with.
# $6      number of ions in alpha bath
# $7      number of ions in beta bath
#
ARG1=${1:-'50'}
ARG2=${2:-'chrgs2.xvg'}
ARG3=${3:-'chrg4H.xvg'}
ARG4=${4:-'0'}
ARG5=${5:-'0'}
ARG6=${6:-'11'}
ARG7=${7:-'11'}

```

Fig. 00: Parameters and explanation on use of routine 3dchrg4H.awk

```

#!/usr/bin bash -l
=====
# Bfintegr8.awk ### performs brute force calculation of ####
#           ### electric field at location of each bin ####
#           ### in simulation from charge in each bin ####
# parameters
# $1    number of bins in z direction
# $2    charge file to integrate
# $3    output file name - outputs 2 sections, 1st is field on
#           just that particular bin
#           2nd is total field, i.e. accumulated
#           field from pb to this posn
#   #!NOTE - the computational cost of this script scales as N6
#           of parameter $1. it gets very expensive very quick
#

```

ARG1=\${1:-'50'}

Fig. 99: Parameters and explanation on use of routine *Bfintegr8.awk*

Before looking at the results we should briefly consider their significance by considering what analysis and visualization can be done without them. Membrane potentials certainly play a critical role most especially in the conduction of nerve signals to and from the brain as well as inducing pore formation as practiced in electroporation when even larger electric fields are experienced. The calculations are done by the GROMACS utility potential. However this utility assumes planar symmetry of charge distribution along any axis it calculates with respect to. This is completely understandable, since analytical solutions of Gauss's law only exist for conformations where the charge distribution is highly symmetric. The figures on the following two pages show the results of these calculations for a membrane with no pore, thereby possessing charge distribution symmetry in the X-Y plane, making field and potential calculations along the Z axis possible and relatively straightforward.

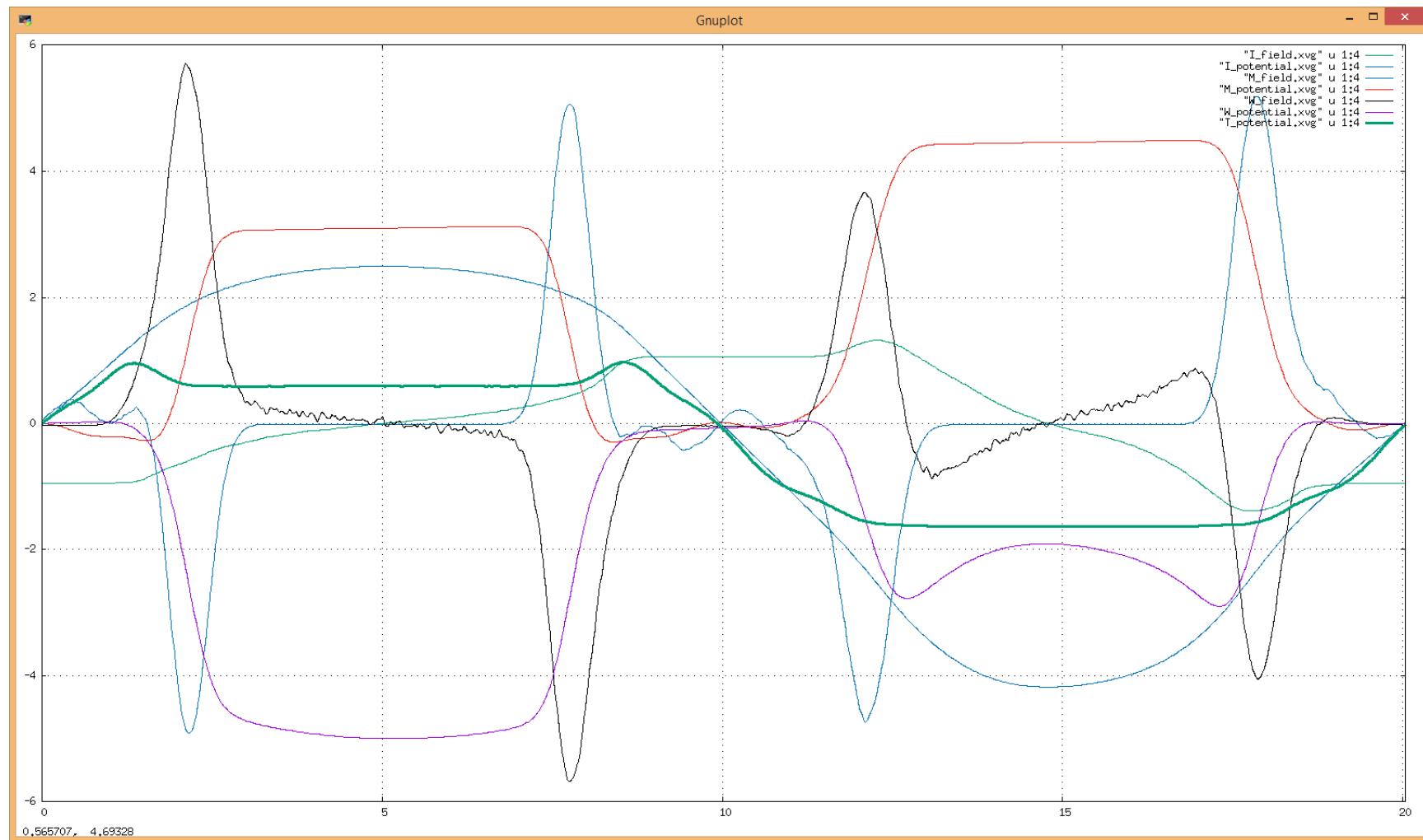


Fig. 00: Composite electric field and potential plot for centered membrane with no pore. Thick green trace shows overall potential. Y-scale is volts, X-scale is nm. The membrane potential is the difference in the overall potential between 5 and 15 nm; ~2.0V. Note the overall electric field magnitude is a fraction of the fields present due to individual moieties. This significance is discussed in the text.

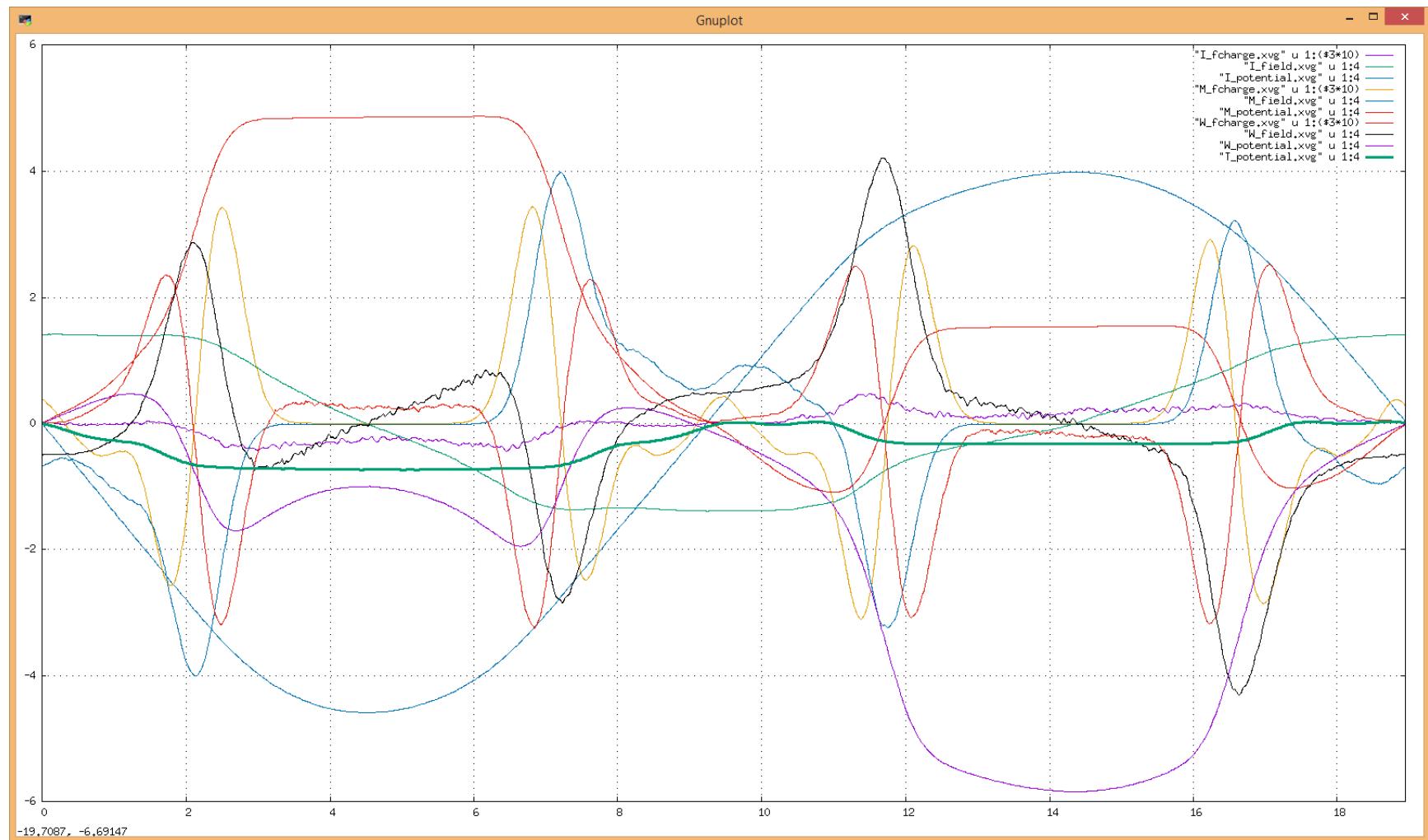


Fig. 00: Composite electric field and potential plot for centered membrane with with pore. The data in the previous figure and this was generated by the GROMACS potential utility. The first figure is valid because of the system symmetry along the axis analyzed. This figure is not because the pore formation disturbs the planar symmetry the potential utility assumes is present to perform the analysis.

In the first figure, which is an accurate visualization, a membrane centered in the simulation box at the 10 nm position is experiencing an ~2.0V transmembrane potential, plotted as the thick green trace. The other six traces show the electric field and potential values for the ions, membrane lipids and water molecules. Note that the individual moieties are exerting and interacting with fields and potentials many times larger than the overall field. These strong coulombic interactions are the forces that give the membrane its amazing self assembling, self healing properties and give insight to what energy potential is available to drive the incorporation of membrane proteins. A pair of the interactions that are easily distinguished are the sharp black and cyan traces that show the opposite orientations of the lipid and water dipoles.

The second plot shows a similar set of traces as well as the charge distributions giving rise to those fields and potentials for a membrane containing a pore. This plot, though interesting and informative is not valid. Note for instance that the indicated transmembrane potential, indicated by the difference in value of the thick green trace on either side of the central membrane indicates a transmembrane potential of only ~0.45V. Both plots were from trajectories that had equal charge imbalances, charge imbalances that the CompEL method maintains during a simulation. How can they have such different transmembrane potentials, or at least on the surface, such different indications? The answer is that in the first figure all the electric field vectors are parallel to the z-axis, the one along whom the fields and potentials are being evaluated. In the second they are not all parallel to the z-axis. Some are parallel, some are orthogonal, and there is a distribution of vectors of fields of every angle between these two. They are non-uniform and a two dimensional plot cannot possibly convey (visualize) the entire reality of what is taking place.

Furthermore, the biomolecules and non-uniform electric fields are not each acting independently of one another. Because many of the molecules involved either have permanent dipoles or are zwitterionic as are the DOPC molecules, they are mutually involved in interactions between field directions and structure conformations that are inseparably reacting and adjusting to assume lowest free energy arrangements, except briefly when overcoming energy barriers in the free energy surface profile.

Now that we can fully appreciate the unique nature of the three dimensional fields that have been calculated we can review and discuss them. Fig. 00 shows the electric field

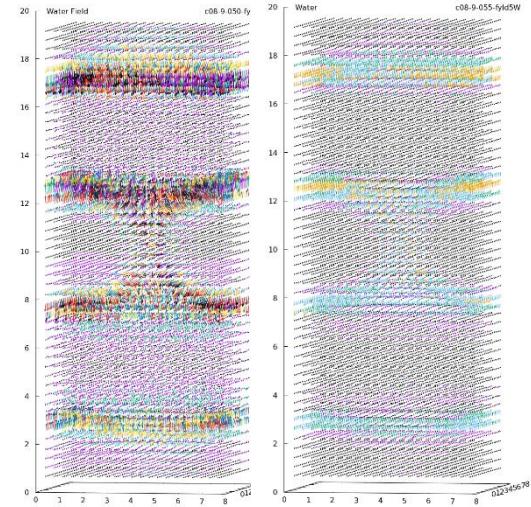


Fig. 00: Comparison of electric field map of water moiety at two different scales.

map for the water moiety plotted at two different scales. These are the lowest resolution plots that have been calculated containing 50 vertical layers. These work well when considering the overall picture within the simulation box. The left plot shows the full dynamic range of the vector magnitudes for the water moiety only. The right plot shows the same vectors plotted at the scale of the moiety with the largest vector magnitudes. So the left plot type shows all the detail of the vectors of its own moiety and the right plot type shows a representation that can be compared to other moieties that are all plotted at a consistent scale. Fig 11 on the next page shows this water plot at the consistent scale next to the plot for the choline moieties at the same scale. It clearly portrays the difference in the relative strength or contribution of the individual molecule parts to the overall field

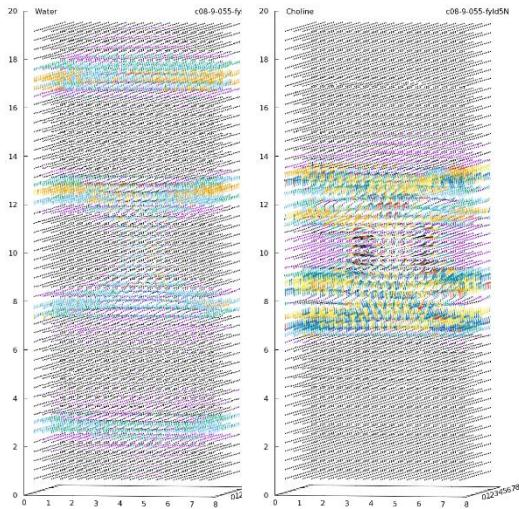


Fig. 00: Comparison of electric field map of water and choline moieties at same scale.

simulation volume as well. Electric potential is a scalar quantity and it is measure of potential difference only, since a universal reference of zero electrical potential does not exist. The best way to visualize such data is with equipotential surfaces and isolines. Currently I have only plotted the data in GNUPLOT with vectors, so the plots do not contain iso-surfaces, and the iso-lines (the vector arrows) have different lengths, a meaningless dimension in this context. The iso-lines show the directions that are orthogonal to the equipotential surface at that point, and the color represents the magnitude, but not the sign of the potential.

Since the simulation volume is periodic, any periodic path (line integral) through that volume must return to its starting point with a net potential change of zero. To keep the post-processing math as simple as possible all the data is biased (valid, since potentials are only relative to one another to begin with) to make the potentials at the boundary origins equal to zero. Notice in figures 00 and 00 that the potentials all have a value of zero volts at the extents of the z-axis. The point is that with my present set of graphing utilities, I have

sub-optimal potential maps that could be misunderstood and misinterpreted. In light that clarification we can risk showing some of the rather dramatic appearing results. From right to left are the: net electric flux potential, sodium potential, chlorine potential and water potential.

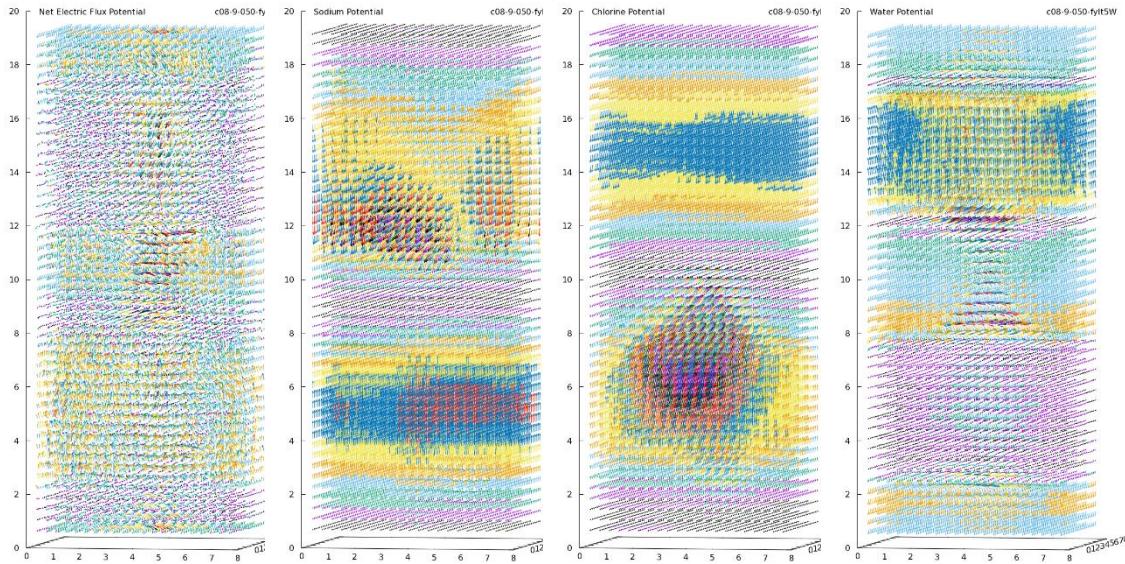


Fig. 00: Electric potential maps of the net electric flux, sodium ions, chlorine ions and water dipoles. Same scale for all plots.

One aspect these plots do correctly convey is the relative magnitude of the driving forces at work and how their individual contributions to equilibrium are somewhat masked when all summed together. An important insight is to realize the magnitude of the energy potentials available in cell membranes that could be exploited in drug design or treatment scenarios like electroporation are represented by the right three plots, not the one on the left.

Much work remains to begin to put together a more comprehensive understanding of the synergy designed into these interactions of molecules and electric fields, and what insights may come with this more detailed view of the ways nature's modulation of the

operations of these biostructures work. Initially better visualization tools that can simultaneously combine these field and potential maps with the matter of the biomolecules comprising them will be a top priority. Fortuitously and unexpectedly, another new capability fell out of this work that can provide unprecedented physical detail of these biostructures, an imaging technique we think we have justifiably titled Computational X-Rays.

PART 2 – COMPUTATIONAL X-RAYS

As described in the introduction starting on page one, my excitement over the first computational x-ray (Comp-XR) was not for the image itself, but the realization of the potential of what further information could be gleaned from the stabilized, centered biostructure trajectories. In fact, to this point, I haven't even shown a "real" Comp-XR yet. The figures shown in the introduction were only color-coded images of the charge distribution. By the end of part 2, I hope you will be glimpsing the enormity of the insights hiding in that 61 GB data file of the position of every atom in what is only one-tenth of the frames generated by the simulation.

Early in my research, because of the post-processing times required, I adopted the practice of initially checking the fidelity of simulations by doing analysis or creating movies from every tenth (recorded) frame of a simulation. Some of my simulations have calculated over a billion frames, but that number is dictated by the simulation time step size, which in the majority of instances is 2 fs, and at the extreme never more than 5 fs. It is typical for us to only record a frame every 100 to 5000 steps, depending on what we are studying. In the simulation campaign for our COMPEL models, we were recording 20000

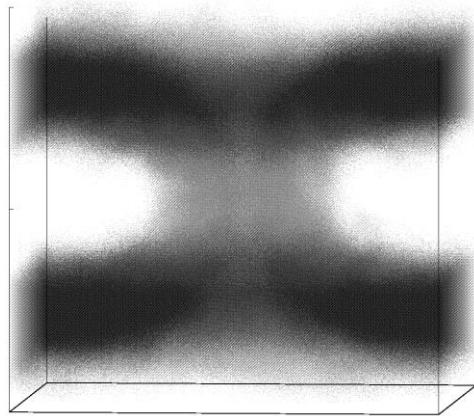
frames per simulation, but initial analysis was in practice only applied to every tenth frame. The first fake Comp-XR came from this initial data of every tenth frame.

To produce a real Comp-XR, the same voxel space as implemented for electric field calculation is used, but the trajectory data is used to compute number densities, not just of atoms, or atom types but what I will call atom sub-species, and with the power of the GROMACS select utility selection syntax at my disposal, the power to define atom sub-species types is limited only by my lack of imagination. In fact, I haven't pointed it out yet, but this is precisely what is taking place in the AWK routines for calculating charges for the individual moieties (sub-species) listed in part one. If you were paying close attention, you noticed that only three parameters are passed to the overall electric flux charge calculation routine, 3dchrg4A.awk, while there are seven parameters passed to all the rest. These additional parameters enable the routines to calculate the indexes of the atoms of the different sub-species that I had wanted to investigate the roles of.

By creating this mechanism, I had created a method to create a Comp-XR. All I needed to do was substitute a file for each atom that assigned it a value of one instead of the fractional charge of each atom. The result would be a map of the volume space in terms of number densities. If I could represent the density as the opaqueness of transparent pixels that would cumulatively add when they overlapped, I could generate an unprecedented image of the biostructure. In the introduction, the figures I showed of these steps were still just plots of the charges, and I was using choline, phosphate and ester moieties because they have significant charge. I was striving for the transparency behavior so I could see how the charge density was varying throughout the volume to validate the integrity of the

processing of the data to that point. While **Fig. 4** was plotting out, the possibilities of what I have just described kind of exploded in my head.

At this point I still hadn't created the first electric field map, I was still hoping to graduate in December, but I felt I had to flesh out for myself some demonstration of what I was envisioning so others could understand what I was talking about. In total, this took over a month, during which time I had to abandon the notion of finishing by December with this new development. As quickly as I could, I did the work to create the first real Comp-XR. To the degree, that our MD models model reality, Comp-XR's are real



scientific images, not pictures, plots, renderings or artists impressions. To meet that standard, there was some output normalization details to work through that will have to be described when I publish the first paper on this work. **Fig. 00** was the first image to come out of that work. The throat of

Fig. 00: First "real" Comp-XR, although not without problems.

the pore visible in the center was so exciting to see, because it was of a real "inaccessible to direct observation" structure, and I emailed to everyone on my committee. The next day I uncovered that the aspect ratio was incorrect and that I was essentially "double exposing" (I found a code flaw) atoms that made the central throat look smaller than it was. Eventually I satisfied myself the results met the standards of a scientific image and I generated sets of them for pores with and without SCMTRs for charge imbalances one through nine for select moiety combinations. All these images are in the Github repository for review. The next two pages, **Fig.'s 00 and 00** show two of these for pores with associated SCMTR's.

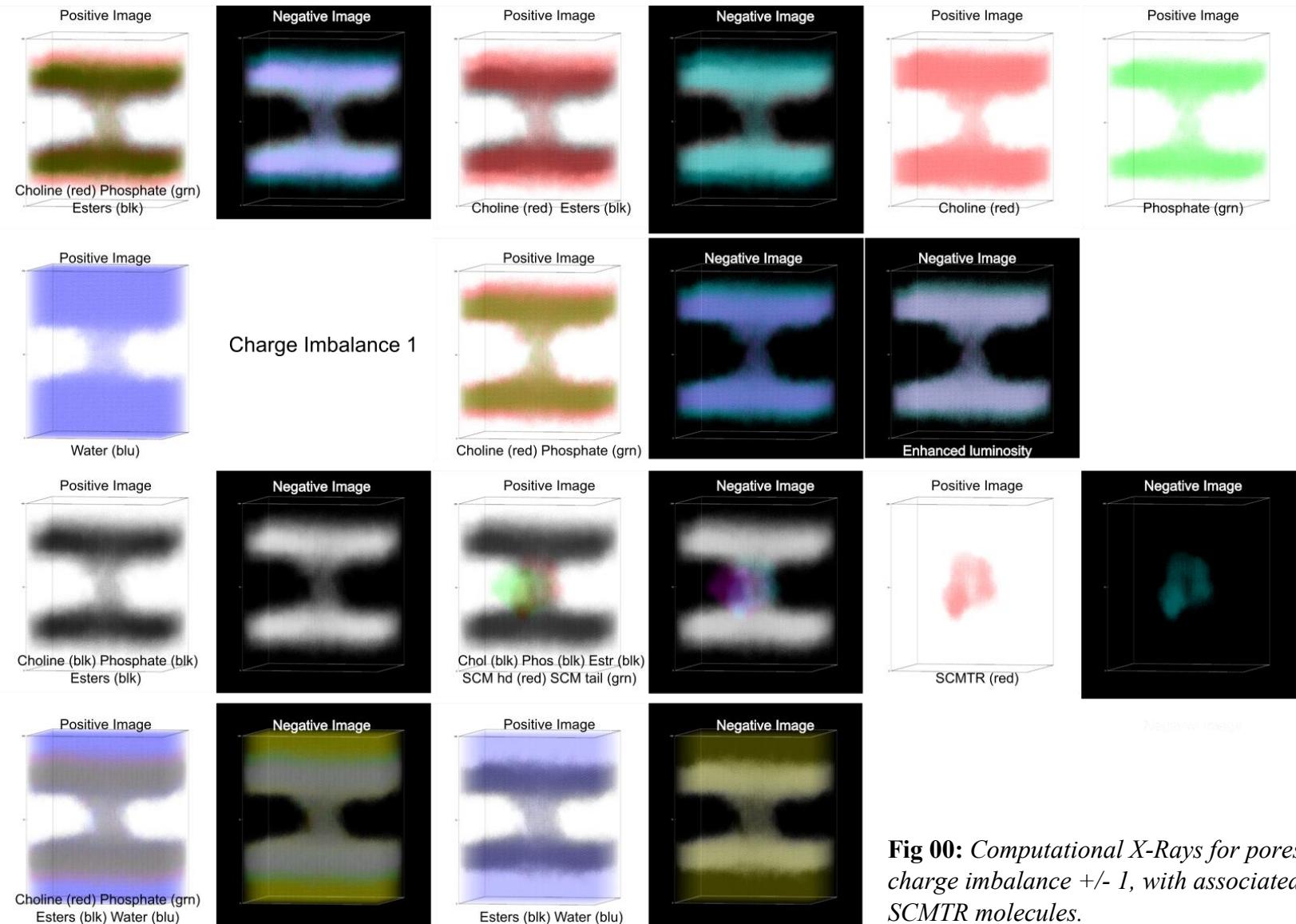


Fig 00: Computational X-Rays for pores at charge imbalance +/- 1, with associated SCMTM molecules.

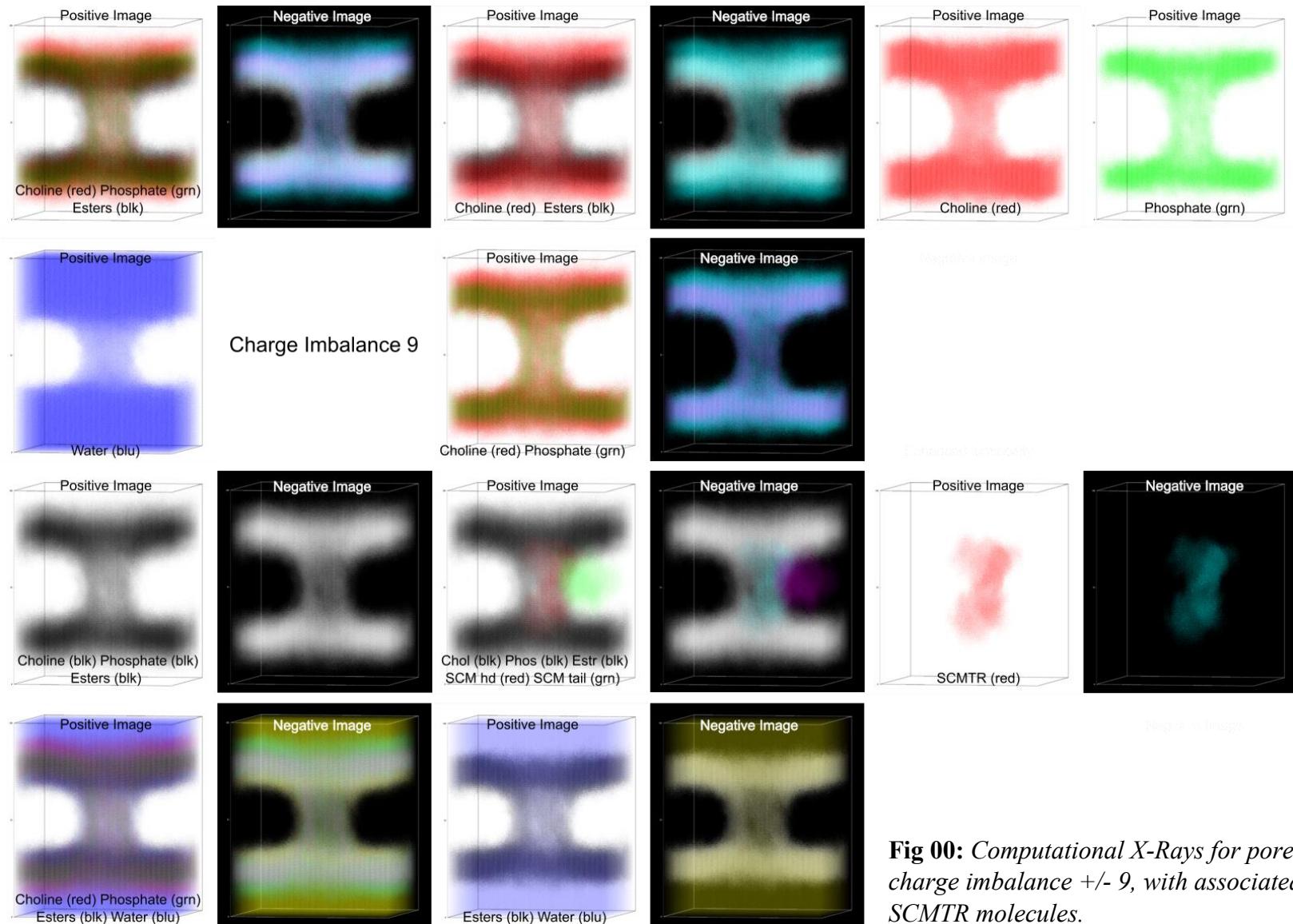


Fig 00: Computational X-Rays for pores at charge imbalance +/- 9, with associated SCMTM molecules.

The images I generated like this using GNUPLOT, although interesting and informative, still suffered from the issue of non-additive opacity nor could I render iso-surfaces. I needed to create an image that would genuinely hint at the potential of the Comp-XRs. To do this I wrote a routine (chrg2pdb.awk) to convert the number density information into a format that could be understood by VMD, which has two methods of rendering surfaces. **Fig. 00** was the first to be produced by this method and begins to reveal the level of detail that can be produced by computational x-rays.

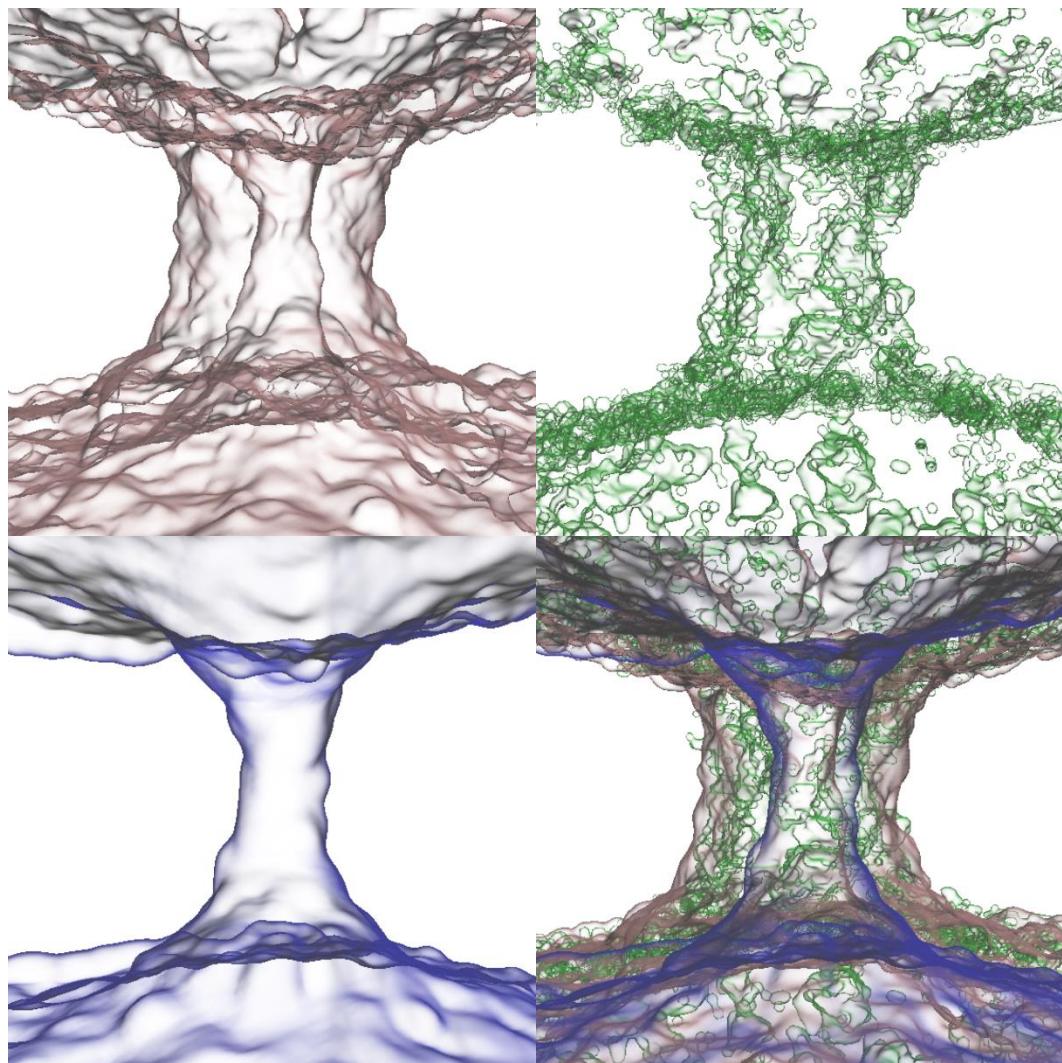


Fig. 00: Computational X-Ray of pore rendered in VMD. Red – Choline, Green – Phosphate, Blue – Water.

By rendering in VMD, it is now possible to make use of the number densities that have been calculated. This requires some explanation. The original routines developed for calculating the electric field maps must do so for every volume element in the volume space. By applying the same routine to every atom sub-species, after normalization, a map of the volume space is created that represents the fraction of the simulation that a volume element is populated by that sub-species of atom. If the density fraction for a volume element for a sub-species like choline is 0.08, that is equivalent to saying the molar composition of choline at that position is 8% The first two images in **Fig 00** are exactly what the surfaces are representing, the boundaries at which the molar composition of choline and phosphate are 8% or greater. Notice the difference in character of how the two moieties are distributed. These and other Comp-XRs answered a long-standing question in my mind, how the zwitterionic elements of DOPC interacted, because the cartoons typically used to illustrate membrane surfaces in research papers couldn't be the reality.

The blue surface represents water, but at a molar composition of 75%. Within the throat of the pore as well as above and below, the molar water composition is 75% or greater. With appropriate manipulation of the composition values we render as a surface, much more detailed information of the pore structure in terms of conformation and composition can be extracted. For example, it can quickly be demonstrated that the DOPC lipid phosphates form clumps in a narrow band surrounded by a halo of the oppositely charged cholines. As improved as these images are from the GNUPLOT variety, effectively conveying the full three-dimension structure with motion or by stereoscopic means improve insights for research as much or even more. For this purpose, a short clip of Fig. 00. in motion revolving about the y-axis was inserted as **Fig. 00**. If viewing this

document in electronic form simply click the Movie 1 image below. Otherwise, it can be viewed at <https://www.youtube.com/watch?v=6LboGlx40ng>.

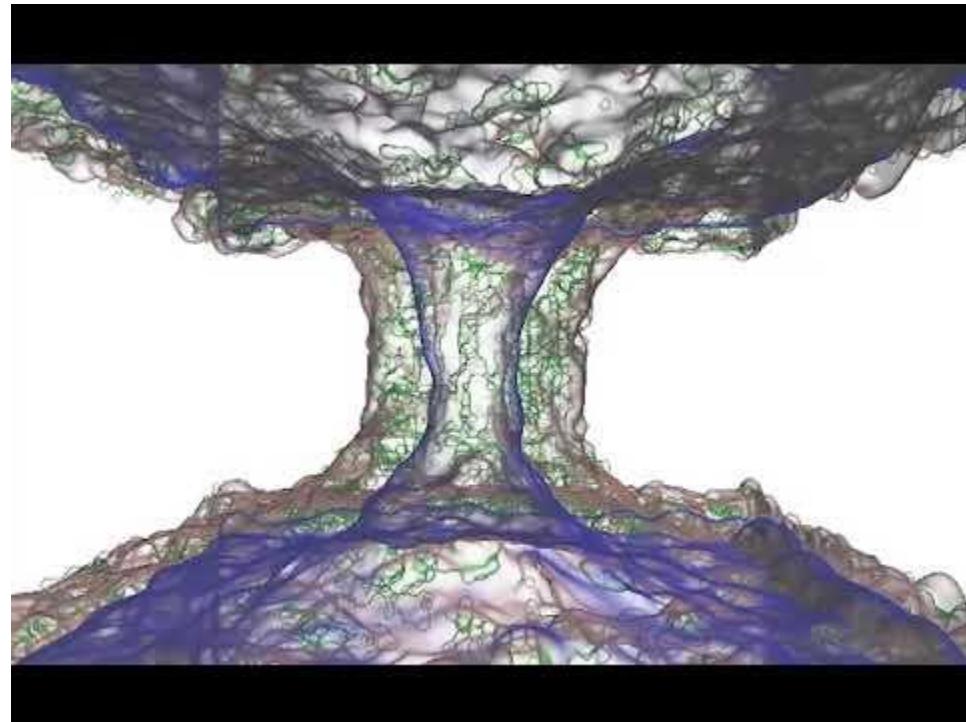


Fig. 00: Short clip of composite Fig 00 in rotation about y-axis.

As good as the three-dimension aspects are in **Fig. 00**, the steps and time required are not always justified in the day to day work of research. When less time is warranted, fields requiring such visualizations like crystallography have long relied on stereoscopic projections. Recently, a top of the line means for viewing such projections, virtual reality (VR) headgear has dropped dramatically in price due to video game driven demand. An inexpensive technique for producing such images are anaglyphs, Red/blue anaglyphs only require a pair of red/blue light filters over the eyes to create the illusion of depth, although the color palette available producing sharp contrast between each eye is much diminished from the full spectrum. Nevertheless, they simulate depth in still images well

even with these constraints. Intrigued about the halo structure of choline atoms formed around the lipid phosphate atoms revealed in the first good Comp-XRs, and for use to demonstrate the imaging made possible by the technique, I produced the video clip in Fig. 00 which is a red/blue anaglyph, which is the reason it may appear slightly fuzzier than Fig. 00 if viewed without red/blue eyeglass lenses.

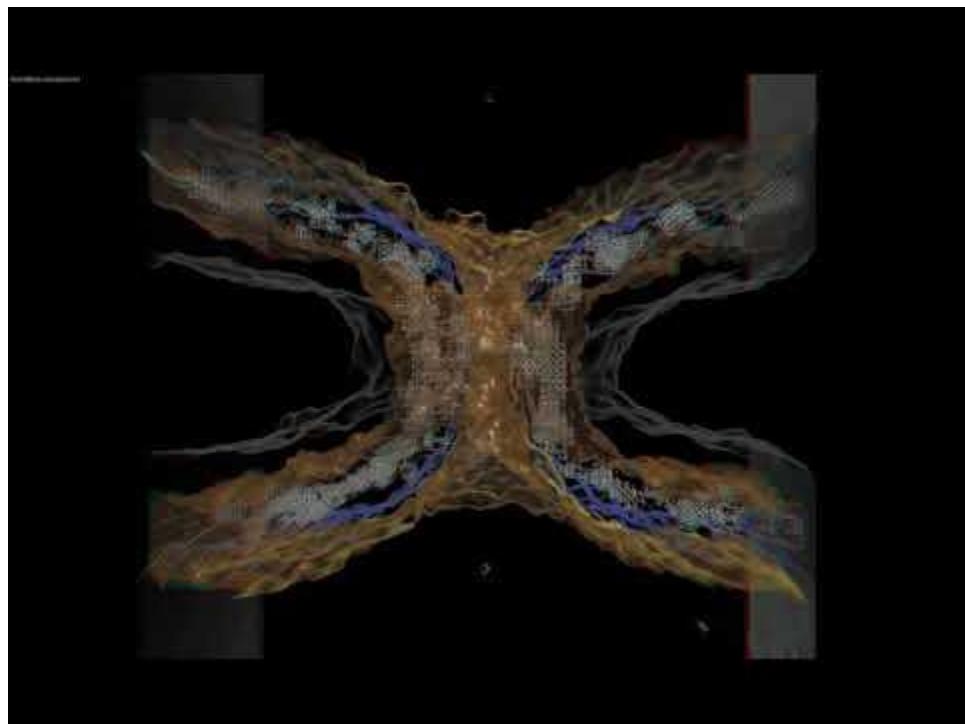


Fig. 00: Short clip of composite Fig 00 in rotation about y-axis.

This clip has more surfaces than Fig. 00. One of them is a 5% water molar composition surface noticeable in the hydrophobic region of the membrane. This surface is being created by water molecules hydrogen bonding with the ester oxygens present between the lipid headgroups and the alkyl tails. The esters are not visualized in this image. Also notice the phosphate clumps rendered as beads in white and how they are surrounded by the halo of the choline atoms. This clip concludes the Comp-XR work done to date.

SECTION 3

Effects of Salt, Water, and Protein Force Fields
on Protein Folding Thermodynamics in
Molecular Simulations

SECTION 4

Nitrogenase Fancier Title

APPENDIX A

Listings of Programs

CURRICULUM VITA

NAME: J. Patrick Brian, P.E.

ADDRESS: Department of Chemical Engineering
216 Eastern Parkway
University of Louisville
Ernst Hall, Rm 207
Louisville, KY 40202

DOB: Louisville, Kentucky – February 1, 1957

EDUCATION,

TRAINING

& EXPERIENCE: B.S., Chemical Engineering
University of Kentucky, 1976-80

BFGoodrich/Westlake Chemical Corporation
Calvert City, Kentucky 1980-2012

Professional Engineering License, KY #14222
Chemical Engineering, 1985

Professional Engineering License, KY #14222
Electrical Engineering, 1988

Ph.D., Chemical Engineering
University of Louisville, 2014-2020

PROFESSIONAL SOCIETIES: Omega Chi Epsilon
Kentucky Society of Professional Engineers
National Society of Professional Engineers

PUBLICATIONS: Controlling the Product Syngas H₂:CO Ratio through Pulsed-Bias
Electrochemical Reduction of CO₂ on Copper
ACS CATALYSIS - July 1, 2016

New trends in the development of heterogeneous catalysts for
electrochemical CO₂ reduction
CATALYSIS TODAY - July 15, 2016

Reduced SnO₂ Porous Nanowires with a High Density of Grain
Boundaries as Catalysts for Efficient Electrochemical CO₂-into-
HCOOH Conversion

ANGEWANDTE CHEMIE-INTERNATIONAL EDITION

February 23, 2017

Electrophysiological measurements reveal that a succinyl linker
enhances performance of the synthetic chloride channel SCMTR.
Chemical communications (Cambridge, England) - April 20, 2018

Heterogeneously catalyzed two-step cascade electrochemical
reduction of CO₂ to ethanol

Electrochimica Acta - April 30, 2018