



University College London

Ensemble-Based Coarse-Grained Molecular Dynamics Simulations of
Multifunctional DNA Nanopores

by

Katya Ahmad

A thesis submitted in partial fulfilment for
the degree of Doctor of Philosophy in Chemistry

April 2021

Declaration of Authorship

I, Katya Ahmad confirm that the work and data presented in this thesis titled “Ensemble-Based Coarse-Grained Molecular Dynamics Simulations of Multifunctional DNA Nanopores” are my own. This work was done wholly while in candidature for a research degree at UCL. Where information has been derived from other sources, and work has been carried out by collaborators, I confirm that this has been indicated in the thesis. Where I have quoted or discussed the published work of others, this has been clearly stated and appropriately cited.

Signed:

Date: 01/04/2021

Abstract

Transmembrane pores are highly specialised nano-devices, with intrinsic specificity and gate-keeping properties that can be exploited in the field of nanobiotechnology. Recently, DNA-origami inspired transmembrane pores with tailorable surface chemistry and programmable dimensions have been rationally designed in an effort to overcome the limitations of protein-based membrane pores such as their fixed lumen size and limited structural repertoire. [1][2] Ongoing experimental research into the potential applications of triethylene glycol-cholesterol DNA nanopores (DNPs) has been fruitful, with a particular emphasis on drug delivery and biosensing. [3]

In this thesis, I describe an ensemble-based coarse-grained MD protocol devised to probe the interactions between bilayer lipids and DNPs, and to determine the effect of membrane encapsulation and salt concentration on the dynamics, structure and conductance of these nanopores. Furthermore, I aim to elucidate the mechanisms by which DNPs mediate translocation of small molecules across lipid bilayers, and the energetics associated with these mechanisms with constant-velocity steered MD and umbrella sampling simulations. I have found that the DNP has no distinct lumen in bulk solution, where it adopts a bloated, amorphous structure with strained and constricted termini regardless of the salt conditions, with significant kinking and fraying of helices. However, salt conditions have a profound effect on the structure of a DNP as it spans a planar lipid bilayer, where it assumes a barrel-like structure with a defined lumen. Sites of constriction in the lumen of the membrane-spanning DNP present a significant barrier to translocation of fluorophores bearing dense negative charges.

Impact Statement

The design of the DNA nanopores studied in this thesis has been under gradual refinement since 2015, and different functionalities (e.g. ligand gating and thermal gating) have been achieved through the introduction of subtle modifications to the archetypal cholesterol anchored six-helix bundle explored here. The published experimental studies that preceded and precipitated this computational study mostly focused on showcasing the successful implementation of gating functionality in these nanopores, highlighting their potential utility in controlled vesicle-based drug delivery systems. Until now, the structural and dynamical properties of the nanopores have remained unknown, and many aspects of the nanopores' behaviour in the experimental conditions used in the aforementioned studies were poorly understood – particularly their response to different salt conditions and their interactions with lipid bilayers. The initial binding of DNA nanopores to the surface of lipid bilayers is known to occur rapidly, however, the membrane insertion kinetics are very slow, especially in solutions of lower ionic strength. Fluorophore translocation through vesicle-encapsulated nanopores is highly dependent on the net charge of the fluorophore, and the nanopore is seemingly able to discriminate between negatively charged fluorophores with very similar structural motifs based on the magnitude of their charge. Additionally, these nanopores exhibit a preference for curved bilayers over planar bilayers, and recent TEM microscopy data have suggested that these nanopores are able to remodel lipid bilayers in certain conditions.[4]

The work done for this thesis has provided some putative explanations for these observations and has identified features of the nanopore that may give rise to this behaviour. These findings should inform future designs with improved insertion kinetics and orientational stability within planar bilayers and identify the kind of molecular cargo that these nanopores are best suited to transport.

Finally, this thesis describes a novel application of an ensemble-based coarse-grained molecular dynamics (CG-MD) protocol to extract robust and reproducible data concerning the conformational dynamics, ion sensitivity, current-voltage relationships and translocational mechanisms of a highly heterogeneous DNA origami/membrane system. With this protocol, I was able to produce high-quality equilibrium molecular models of the DNA nanopore that showed very good agreement with cryo-EM density maps produced by our collaborators at Birkbeck College, University of London. This work describes the first successful structural characterisation of a DNA nanopore using a hybrid CG-MD / cryo-EM approach.

Acknowledgements

First and foremost, I owe a debt of sincere gratitude to my supervisors Professor Peter Coveney and Professor Stefan Howorka for giving me the opportunity to pursue a scientific career and providing me with unwavering support throughout my PhD. I am grateful to Stefan Howorka for his excellent mentoring and training, which has boosted my confidence and nurtured my professional development. I am forever indebted to Peter Coveney for the time, resources and belief that he has invested in me and my work, and I would like to thank him wholeheartedly for his patience and understanding when times were tough, and for his relentless encouragement throughout. I would also like to extend my gratitude to Daria Kieczka for the hard work and long hours she put into her Masters' project, the results of which assisted me with the validation of the work described in this thesis. Likewise, I am also thankful to Dr. Abid Javid and Professor Elena Orlova for their contributions to the validation of this work.

I would like to thank my friends and colleagues at the Centre for Computational Science, Alex Gheorghiu, David Wright, Emily Lumley, James Suter, Jon McCullough, Robbie Sinclair and Robin Richardson for all of the invigorating scientific and technical discussions, topical debates and laughs we shared over the years. This PhD would be much more difficult and much less enjoyable without them. I would also like to offer my sincere gratitude to the entire Howorka group for their support, camaraderie and scientific insights. I will never forget all the great times I shared with Conor Lanphere, Dr. Daniel Offenbartl-Stiegert and Helena Philpott. Thank you for your friendship, and for making the fourth-floor office of the CIB feel like a second home.

Finally, I am perpetually indebted to my family: Mum, Dad and Rami for their steadfast belief in me. They always did their best to make sure I never wanted for anything. Thank you to my best and oldest friends Jessica, Luigi and Richard for keeping me grounded and coaxing me out of my work-induced funks.

Contents

Declaration of Authorship.....	1
Abstract.....	2
Impact Statement.....	3-4
Acknowledgements.....	5
List of Figures.....	9-14
List of Tables.....	15
List of Abbreviations.....	16-17
1. Introduction.....	19-49
1.1 Transmembrane Pores	21
1.2 Engineered and Synthetic Pores	21-25
1.3 DNA Nanopores	26-28
1.4 Hydrophobic Mismatch	28-30
1.5 Experimental Methods in Nanopore Research	30-35
1.5.1 Investigating Transport Properties of Nanopores	30
1.5.2 Structural Characterisation of Nanopores	33
1.6 Design, Synthesis and Applications of DNA nanopores	36-40
1.7 Molecular Dynamics Simulations of DNA nanopores	41-49
2. Molecular Dynamics.....	50-72
2.1 Introduction to Molecular Dynamics Simulations	50-54
2.2 Force Fields	54-65
2.2.1 All-Atom Force Fields	54
2.2.1.1 The CHARMM36 Additive Force Field	55
2.2.1.2 The CHARMM General Force Field	57
2.2.2 Coarse-Grained Force Fields	58
2.2.2.1 The MARTINI Force Field	59
2.2.2.2 Parameterisation of the MARTINI DNA Force Field	63
2.2.2.3 Mapping AA Models to MARTINI	65
2.3 Integration Algorithms	65-66
2.4 Thermodynamic Ensembles	66-68

2.5 Temperature and Pressure Coupling	69-70
2.6 Uncertainty Quantification in Molecular Dynamics Simulations	70-71
2.7 Aims	71-72
3. Coarse-Grained Simulations of DNA Nanopores	
in Aqueous Salt Solutions.....	73-95
3.1 Model Building	73-76
3.2 Simulation Details and Analytical Methods	74-81
3.3 Benchmarking and Computational Resources	81
3.4 Results and Discussion	81-91
3.5 Further Validation of CG Simulation Data	91-102
3.5.1 All-Atom Simulations of the Solvated DNP in 0.3 M NaCl	92
3.5.2 Cryo-EM Models	92
3.5.3 Validation of CG Simulation Data	95
3.5.3.1 Pore Dimensions and Structural Fluctuations	96
3.5.3.2 Helix Kink Angles and Percentage of Base-Pair Breakage	100
3.6 Conclusions	103-104
4. Coarse-grained Simulations of Lipid Bilayer-Spanning	
DNA Nanopores.....	105-127
4.1 Model Building	106-107
4.2 Simulation Details and Analytical Methods	107-109
4.3 Benchmarking and Computational Resources	110
4.4 Results and Discussion	110-125
4.5 Conclusions	126-127
5. Computational Electrophysiology Simulations of Lipid	
Bilayer-Spanning DNA Nanopores.....	128-137
5.1 The CompEl Protocol	128-130
5.2 Simulation Details and Analytical Methods	130-132
5.3 Computational Resources	132
5.4 Results and Discussion	133-137
5.5 Conclusions	137

6. Constant-Velocity Steered Molecular Dynamics Simulations of Fluorophore Translocation through a Lipid Bilayer-Spanning DNA Nanopore.....	138-171
6.1 Simulation Details and Analytical Methods	141-144
6.1.1 The Pull Code	141
6.1.2 Computing the Potential of Mean Force	142
6.1.3 Umbrella Sampling	143
6.1.4 The Jarzynski Equality	144
6.2 Model Building	146-152
6.2.1 Parameterisation of All-Atom Fluorophore Models	146
6.2.2 Parameterisation of Coarse-Grained Fluorophore Models	150
6.2.3 Assembly of Models for Pulling Simulations	152
6.3 Computational Resources	155
6.4 Results and Discussion	155-169
6.4.1 Transport PMFs for SRB	158
6.4.2 Transport PMFs for SRB	163
6.4.3 Transport PMFs for ATTO 655	166
6.5 Conclusions	170-171
7. Summary and Future Outlook.....	172-178
7.1 Stiff Elastic Network vs Soft Elastic Network for MARTINI DNA	174
7.2 Energetics of Insertion and Expulsion of DNA Nanopores	175
7.3 CompEl and Applied Electric Field Simulations	175
7.4 Refinement of cv-SMD protocol for Fluorophore Translocation	176
7.5 Final Remarks	177
8. Bibliography.....	178-187
Appendices.....	188-201

List of Figures

Figure 1.1: Structures of alamethicin and α HL nanopores.	21
Figure 1.2: Genetically engineered α -haemolysin nanopores designed for stochastic sensing of different analytes.	22
Figure 1.3: Cross-sectional rendering of an engineered α HL nanopore translocating a strand of poly-cysteine DNA across a lipid membrane under an applied voltage. . .	23
Figure 1.4: An engineered α HL pore with a DNA polymerase enzyme coupled the extracellular cap domain.	24
Figure 1.5: A silicon nitride/carbon based solid-state nanopore sensor used for DNA sequencing.	25
Figure 1.6: Schematic showing hybridisation of DNA origami objects, and accompanying TEM images.	26
Figure 1.7: 2D caDNAno schematic for a 6-helix DNA origami nanopore, and its' 3D representation.	27
Figure 1.8: Examples of negative and positive hydrophobic mismatch between trimeric proteins and a lipid bilayer.	30
Figure 1.9: A simplified representation of a chip-based apparatus used for planar lipid bilayer electrophysiology experiments.	31
Figure 1.10: A simplified representation of a blockade experiment based on a modified α HL pore to which an analyte can selectively bind.	32
Figure 1.11: Illustration of a dye efflux assay performed on a gated DNA nanopore embedded in a vesicle containing self-quenching fluorophores.	33
Figure 1.12: AFM images of mutant α -haemolysin nanopores on mica and a lipid bilayer.	34
Figure 1.13: High-resolution cryo-EM structures of the <i>E. Coli</i> Cytolysin A dodecamer.	35

Figure 1.14: DNA origami nanopore designed by Bell <i>et al.</i>	36
Figure 1.15: Schematic diagram illustrating the structure of the DNA nanopore designed by the Simmel group.	37
Figure 1.16: 3D representation and caDNAno schematic of a six-helix bundle designed by the Howorka group.	37
Figure 1.17: Examples of different hydrophobic anchors used in DNA nanopore designs by Burns <i>et al.</i> , with 2D maps.	39
Figure 1.18: Ligand-gated nanopore by Burns <i>et al.</i>	40
Figure 1.19: Equilibrated all-atom model of the ethyl-PPT modified DNA nanopore designed by Yoo <i>et al.</i>	41
Figure 1.20: Starting geometry of the DNT1 model and the final geometry after 170 ns of dynamics, simulated by Maingi <i>et al.</i>	42
Figure 1.21: TTP-tagged DNA strand spanning a DphPC bilayer studied by Göpflich <i>et al.</i> , showing the movement of ions through the DNA/bilayer interface. .43	
Figure 1.22: All-atom model of the TTP-DNA/lipid system with graphical representations of DNA/lipid densities.	44
Figure 1.23: DNA porin designed by the Aksimentiev and Keyser groups, shown alongside the local densities of lipids and DNA.	45
Figure 1.24: Snapshots from AA simulations of a DNA nanopore spanning a bilayer performed by Maingi <i>et al.</i>	46
Figure 1.25: Minimised structure of the membrane-spanning 6-helix nanotube, and simulation snapshots taken at various stages of the simulations by Joshi <i>et al.</i>	48
Figure 1.26: Simulation data and snapshots illustrating the influence of increasing NaCl concentration on the rigidity of a DNA nanotube studied by Joshi <i>et al.</i>	49

Figure 2.1: Illustrations of intramolecular potential energy terms	53
Figure 2.2: MARTINI models of water	61
Figure 2.3: Atom-to-bead mapping used in the MARTINI DNA force-field, alongside the ball-and-stick representation of the elastic DNA network	63
Figure 2.4: Partition free energies of nucleobases for a water-chloroform system and a water-octanol system.	64
Figure 2.5: Proposed insertion mechanism for the TEG-C DNA nanopore.	72
Figure 3.1: CaDNA schematic of the DNP, alongside an illustration of the idealised six helix bundle structure.	74
Figure 3.2: 3D renderings of the DNP in its artificial starting configuration, and after equilibration	75
Figure 3.3: AA to CG mapping for the TEG-cholesterol anchor, with MARTINI bead assignments.	76
Figure 3.4: Splitting the DNP into eight distinct pore regions, illustrated on a schematic of Helix 2.	78
Figure 3.5: Illustration of the index groups used for the calculation of average pore height and width.	79
Figure 3.6: Ensemble average plots illustrating the convergence of pore dimensions and the associated bootstrapped standard errors in the solvated DNP simulation ensembles.	81
Figure 3.7: Solvated DNP dimensions in two different salt conditions.	83
Figure 3.8: Representative structures of the DNP exhibiting partial fraying of the <i>ds</i> DNA duplexes.	84
Figure 3.9: Structural fluctuations of a solvated DNP represented by RMSF-per- residue plots.	86

Figure 3.10: Average total dwell time of TEG-cholesterol anchors within DNA grooves.	88
Figure 3.11: Plot of the RMSD of TEG-cholesterol anchors vs time in the both salt concentrations.	88
Figure 3.12: Comparison of the average RMSF of each helix between the two salt concentrations.	90
Figure 3.13: Fluctuation of base pairs (% BPB) as a function of time for the 0.3 M NaCl simulations and 1.0 M NaCl simulations.	91
Figure 3.14: Results of the flexible fitting of cryo-EM electron density maps to CG MD derived atomic models of the DNP.	94
Figure 3.15: Rendered images of highly populated DNP clusters obtained from three different datasets.	97
Figure 3.16: RMSF-per-residue plot generated from the ensemble of 15 x 30 ns AA simulations of the solvated DNP in 0.3 M NaCl	99
Figure 4.1: Ensemble average plots illustrating the convergence of the average bilayer thickness and average pore height.	109
Figure 4.2: Plot showing the evolution of the average pore height with time for each of the 15 individual replicas.	111
Figure 4.3: Snapshots taken from a simulation in which the DNP departs from the bilayer in a multi-step process.	112
Figure 4.4: Simulation snapshots illustrating the formation of a toroidal pore around the DNP	114
Figure 4.5: Analysis of the influence of salt concentration and DNP binding mode on the ensemble averaged bilayer properties.	115
Figure 4.6: Representative structures and RMSF-per residue plots for the membrane-spanning DNP in both salt concentrations.	117

Figure 4.7: Histograms and overlaid estimated probability density functions representing the distribution of system potential energies.	118
Figure 4.8: Simulation snapshots and average width plots illustrating the formation of toroidal pores within the POPC membrane and compression of the DNP.	119
Figure 4.9: Radial distribution functions for the DNP/POPC systems.	121
Figure 4.10: Variation of the membrane-spanning DNP dimensions in the two different salt conditions.	122
Figure 4.11: Helix kink angles calculated for helices 3, 4 and 5 in the 1.0 M NaCl DNP/membrane model.	124
Figure 4.12: Averaged lumen width profile of the DNP in 1.0 M NaCl.	125
Figure 5.1: Illustration of a double-bilayer system built for CompEl simulations. .	129
Figure 5.2: The pre-equilibrated double bilayer/DNP system.	132
Figure 5.3: Ensemble average plots illustrating the convergence of the average conductance	134
Figure 5.4: Series of current-voltage plots illustrating the decline of the average conductance at higher voltages.	135
Figure 5.5: Ion flux vs time plot illustrating cation selectivity of the DNP.	136
Figure 6.1: Skeletal representations of the three fluorophores.	139
Figure 6.2: Illustration of the ‘unlocking’ of the ligand-gated nanopore designed by Burns et al alongside kinetic release traces for CF and SRB.	140
Figure 6.3: Setting up umbrella sampling windows from a pulling simulation. . . .	144
Figure 6.4: QM optimised geometries of the three fluorophores.	150

Figure 6.5: AA atom to MARTINI CG pseudo-atom mapping schemes used for parameterisation of the fluorophores with PyCGTool.	151
Figure 6.6: Tukey boxplots comparing the total R_g of the CHARMM and AMBER derived CG fluorophore models.	152
Figure 6.7: Simplified illustration of the idealised TL and IF pathways.	156
Figure 6.8: Averaged force profiles extracted from cv-SMD simulations of fluorophore translocation.	157
Figure 6.9: PMFs representing the 2D free energy landscapes associated with the translocation of SRB via the TL and IF pathways.	160
Figure 6.10: Schematic illustrating the tendency for the fluorophore molecule (green triangle) to deviate from the idealised IF translocation pathway.	162
Figure 6.11: PMFs representing the 2D free energy landscapes associated with the translocation of CF via the TL pathway and IF pathways.	165
Figure 6.12: PMFs calculated for translocation of ATTO 655 via the TL and IF pathways.	167
Figure 6.13: US simulation snapshots showing the orientation adopted by the ATTO 655 fluorophore as it clears the DNP lumen in the TL pathway.	168
Figure 6.14: US simulation snapshots illustrating the important role of flexion of the ATTO 655 molecule as it enters the DNP/membrane interface.	169

List of Tables

Table 2.1: Common ensembles used in MD simulations.	67
Table 3.1: Comparison of the calculated pore dimensions.	84
Table 3.2: Percentage of broken base pairs within the DNP.	87
Table 3.3: The relationship between the RMSD cutoff used for clustering CG trajectories and the number of clusters obtained.	95
Table 3.4: Comparison of DNP dimensions obtained from AA simulations in 0.3 M NaCl, CG simulations in 0.3 M NaCl and cryo-EM microscopy simulations conducted in 12 nM MgCl ₂	98
Table 3.5: Comparison of the percentage of base pair breakage calculated for the CG and AA solvated DNP systems in 0.3 M NaCl.	101
Table 4.1: Comparison of the calculated pore dimensions of the membrane/DNP models at two different salt concentrations.	123
Table 4.2: Comparison of the distances between adjacent helices in a membrane-spanning DNP at the two different salt concentrations.	124
Table 6.1: Distribution of trans-lumen and interfacial translocation events for each ensemble of cv-SMD replicas, represented as the percentage of replicas featuring the given translocation pathway.	156

List of Abbreviations

Biochemistry terms

AFM	atomic force microscopy
α HL	alpha-haemolysin
BPB	base-pair breakage
C	cholesterol
CF	carboxyfluorescein
cryo-EM	cryogenic electron microscopy
DNA	deoxyribonucleic acid
DphPC	diphytanoylphosphatidylcholine
<i>ds</i> DNA	double-stranded DNA
DOPC	dipalmitoylphosphatidylcholine
DNT	DNA nanotube
FRET	fluorescence resonance energy transfer
GUV	giant unilamellar vesicle
DNP	DNA nanopore
DNP-C	closed state DNA nanopore
DNP-O	open state DNA nanopore
OmpF	outer membrane protein F
PEG	poly-ethylene glycol
poly(A)	poly-adenine
poly(C)	poly-cytosine
poly(T)	poly-thymine
POPC	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
PPT	phosphorothioate
PSM	polymer supported membrane
RNA	ribonucleic acid
<i>ss</i> DNA	single-stranded DNA
SUV	small unilamellar vesicle
SRB	sulpho-rhodamine B
TEG	triethylene glycol
TEM	transmission electron microscopy
TPP	tetraphenylporphyrin

Computational terms

AA	all-atom
AMBER	Assisted Model Building with Energy Refinement
CG	coarse-grained
CompEl	computational electrophysiology
cv-SMD	constant-velocity steered molecular dynamics
CHARMM	Chemistry at Harvard Macromolecular Mechanics
CGenFF	CHARMM general forcefield
DFT	density functional theory
GROMACS	GRONingen MACHine for Chemical Simulations
HF	Hartree-Fock
HPC	high performance computing
JSON	javascript object notation
IF	interfacial
MARTINI	MARrink's Toolkit INItiative
MC	Monte Carlo
MD	molecular dynamics
MM	molecular mechanics
NAMD	NANoscale Molecular Dynamics
PDB	protein data bank
PME	particle mesh Ewald
PMF	potential of mean force
py	Python
QM	quantum mechanics
R_g	radius of gyration
RHF	restricted Hartree-Fock
RMSD	root mean squared deviation
RMSF	root mean squared fluctuation
TIP3	transferable intermolecular potential with 3 points
TL	trans-lumen
VMD	visual molecular dynamics
US	umbrella sampling
USW	umbrella sampling window
WHAM	weighted histogram analysis method

This thesis is dedicated to all the schoolgirls who have ever cried over their maths homework, and to my parents. We all tried our best.

Chapter 1

Introduction

Here, I introduce the subject of transmembrane pores and the principles behind using DNA to construct synthetic analogues of these pores and discuss their current and potential future applications in science, and the experimental techniques that are commonly associated with their synthesis and characterisation. DNA nanotechnology is an emergent branch of nanotechnology, which seeks to exploit the structural and chemical properties of DNA to build modular, programmable structures for application in drug delivery, stochastic sensing, engineering and industrial chemical synthesis.[5] The foundation of DNA nanotechnology rests on the seminal work by Rothemund[6], which describes a one-pot, bottom-up method for designing and fabricating highly complex 2D and 3D structures from single strands of DNA, known as “DNA-origami”. While the size, complexity and potential utility of published DNA origami structures continues to grow, robust and industrially viable technologies based on these structures are still in the conceptual phase, for reasons that will be discussed later in this introduction.

Gállego and co-workers[7] recently designed and synthesised thiolated DNA origami patches for nanolithography of gold surfaces with sub-10 nm precision – a objective that had not been attained with other state-of-the-art nanolithography methods. This method could in principle be adapted to other materials by altering the functionalisation of the DNA origami “stamps” to create metamaterials with applications in nanoelectronics and photonics. DNA origami nanostructures have also been utilised as components in complex multi-enzyme nanoreactors, whereby the nanostructure may act as a platform for the co-assembly of coupled enzymes, such as the DNA-tile based nanoreactor for the glucose oxidase/horseradish peroxidase enzyme pair designed by Fu et al., and the modular DNA-nanotubule based nanoreactor for the same enzyme pair developed by Linko et al. [8] Cross linking amino-acids of the protein surface with short, thiolated DNA strands introduces the “anchors” on the enzymes, which then hybridise with complimentary strands on the DNA origami nanostructure. Application of this methodology to extended cascades of

enzymes and cofactors has the potential to revolutionise the production of industrial biomaterials and biochemicals, working towards sustainable high-yield flow-based biomanufacturing with minimal waste.

Efforts toward the application of DNA nanotechnology in the spheres of drug delivery and stochastic sensing are centred around a particular structural motif – the “DNA nanopore”, which is the focus of the work described in this thesis. Designed to mimic biological protein pores, these DNA nanopores can be incorporated into lipid, polymer or solid-state membranes, where they facilitate the passage of small-molecule drugs or analytes. Sensing functionality can be introduced through chemical modifications to the DNA backbone atoms inside the lumen of a membrane-spanning DNA nanopore, and functionality for vesicular drug-delivery systems can be achieved with the introduction of ligand, temperature or voltage activated DNA “gates” made from located at the termini of a vesicle-spanning DNA nanopore. [1]

Although the sequence-specific base-pairing and folding of helices in a DNA nanopore is predictable, working with DNA nanopores in the laboratory presents many challenges, due to their intrinsic flexibility and high charge-density. It is reasonably simple to design a DNA nanopore with the desired shape and dimensions, but it is difficult to predict how it will interact with its environment, and how these interactions will affect their functionality. In addition, the narrow application field associated with the experimental techniques that have been used to study synthetic DNA nanopores interacting with membranes thus far (e.g., single channel current recordings, fluorescence microscopy, tracking and localisation microscopy) presents a barrier to the advancement of the field. While these techniques have successfully shown that these nanopores can associate with bilayers and trigger poration events, it is not clear whether or not the nanopores adopt a transmembrane configuration with a persistent lumen in a manner that is consistent with biological protein pores such as α -haemolysin. In order to assess the practical utility of DNA nanopores in drug delivery and stochastic sensing technologies, we must understand the key functional features of the transmembrane pores that they are designed to mimic.

1.1 Transmembrane Pores

Transmembrane proteins make up a huge class of highly specialised biomolecules that are designed to carry out a multitude of vital cell processes in prokaryotes and eukaryotes. A vast majority of transmembrane proteins function as membrane channels, which carefully regulate the uptake and release of ions and hydrophilic molecules, and also help coordinate cell-signalling processes. These naturally occurring channels are sophisticated architectures of pore-forming proteins which fold in a very specific manner, to ensure stability within the membrane and the appropriate balance of interactions between the internal lumen of the pore and its cargo, so that translocation through the lumen is quick and effective. Transmembrane proteins vary in size and structure, but the pores that they form within the membrane are typically of nanometer size and are referred to as ‘nanopores’ in the field of nanobiotechnology.

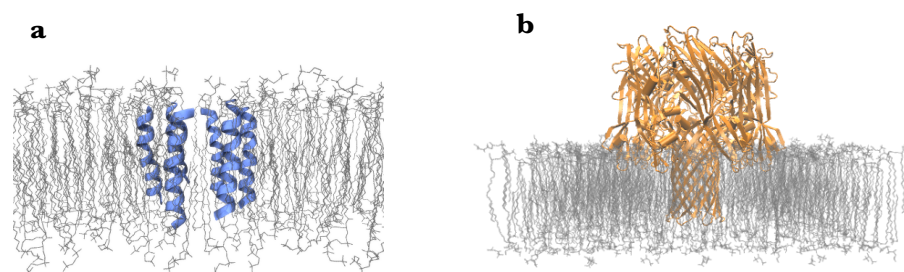


Figure 1.1: (a) Diagram of an alamethicin nanopore [1] from the fungus *Trichoderma Veridie* embedded in a lipid bilayer. (b) An α -haemolysin (α HL) nanopore from the species *Staphylococcus Aureus*. The alamethicin pore is a classic example of an α -helical pore-forming peptide, which can form multiple stable oligomers. [8] The bilayers shown in these figures are composed of unequilibrated lipids and are therefore not representative of the orientation adopted by lipid bilayers upon association with the protein pores.

1.2 Engineered and Synthetic Nanopores

Nanopores are of particular interest to scientists for their potential application in single molecule detection, DNA sequencing and drug delivery. Their small internal lumen size, modifiability, and ability to insert easily into membranes make them ideal stochastic sensors. Stochastic sensing with engineered pores is an extremely precise method of single-molecule detection – allowing scientists to study individual binding

events in depth, investigate binding kinetics, and detect specific analytes at ultra-low concentrations. Nanopores can be genetically engineered to bind a single analyte, or a range of analytes by introducing a number of different binding sites. [9] Biological pores can be modified both the genetic and chemical level for stochastic sensing applications. Binding sites can also be introduced in a very specific manner with covalent or non-covalent modifications of specific amino acids inside the pore lumen, ensuring optimal interactions with the analyte. This has been demonstrated by the Bayley group[10] with α -haemolysin, which was successfully modified with covalently-attached biotin ligands designed to bind with streptavidin proteins, and with non-covalently attached β -cyclodextrins to detect a range of small molecules.

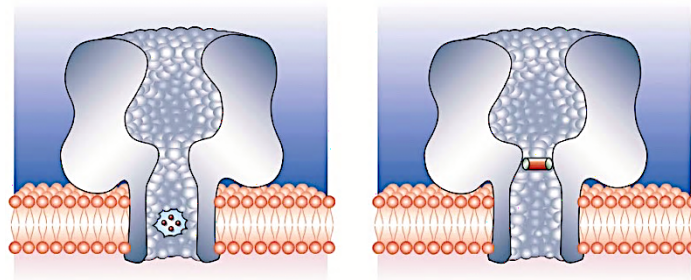


Figure 1.2: Genetically engineered α -haemolysin nanopores designed for stochastic sensing of different analytes. The nanopore on the left features a binding site for metal ions within the lumen of the membrane spanning region. The nanopore on the right contains a binding site for organic molecules. Figure adapted from Bayley *et al.* [9]

Bacterial α -haemolysin, fungal alamethicin and bacterial gramicidins were among the first biological nanopores to be engineered for stochastic sensing applications. Soon after these nanopores proved their efficacy as small molecule detectors, scientists began exploring the possibility of using them for DNA sequencing. The idea was first proposed in a paper by Deamer and Akeson [11] published in 2000, in which they successfully translocated RNA and DNA homopolymer strands through an α -haemolysin pore embedded in a membrane. Each homopolymer strand produced a unique blockade trace during single-channel current recording experiments, and the authors were able to detect the transition from a 30-nucleotide block of poly-adenine to a 70-nucleotide block of poly-cytosine within a single RNA strand. However, single-nucleotide resolution was not possible due to the high speed of DNA/RNA translocation. [11]

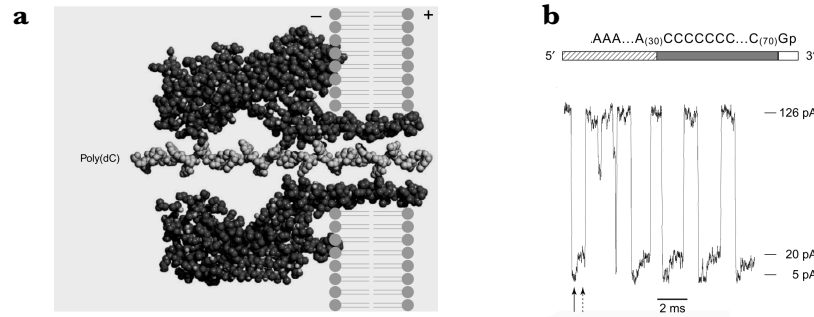


Figure 1.3: (a) Cross-sectional rendering of an engineered α -haemolysin nanopore translocating a strand of poly-cysteine DNA across a lipid membrane under an applied voltage. (b) Single-channel current trace showing the distinct blockade events associated with the translocation of the poly-A and poly-C portions of a single RNA strand, which impede the flow of current as they pass through the pore lumen. However, the current is not blocked completely so a residual current remains, and the magnitude of this residual current is specific to the nucleotide that induces the blockade. The passage of a poly-A section of DNA through the pore coincides with a characteristic drop in current from 126 to 20 pA, while the translocation of poly-C sections is associated with a 5 pA residual current. Adapted from Deamer *et al.* [11]

Nanopore DNA sequencing technology has evolved significantly since then – DNA can now be sequenced with single-nucleotide precision, on the go with a handheld nanopore-based MinION device produced by Oxford Nanopore Technologies.[12] The device features an α -haemolysin pore with a DNA polymerase enzyme attached to the “cap” portion of the pore, which processes the “sense” strand and feeds it through the α -haemolysin pore at a much slower rate than freely translocating DNA, allowing individual bases to be detected. [13]

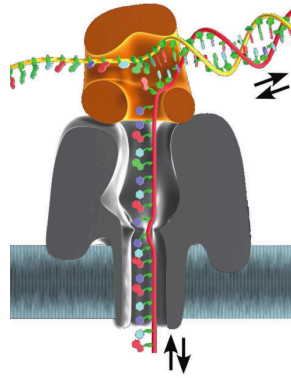


Figure 1.4: An engineered α -haemolysin pore with a DNA polymerase enzyme coupled the extracellular cap domain, which slows the rate of DNA translocation through the sensing element within the α -haemolysin lumen. Figure adapted from Schneider *et al.* [14]

The vast majority of the work towards redesigning biological pores for such applications has been directed towards α -haemolysin, mostly because its structure-function relationships are well understood and thus site-directed mutagenesis with either natural or unnatural amino acids can be executed without affecting the pores' channel forming abilities or geometry.[15] While covalent or non-covalent attachment of protein-binding ligands broadens the analyte scope of α -haemolysin in stochastic sensing applications, these methods are of little use when designing nanopores to mediate the translocation of larger molecules through a lipid bilayer. As a result of the narrow lumen (which ranges between 2.7 nm in the cap region and ~ 2 nm in the transmembrane region) genetically engineered α -haemolysin pores are only capable of transporting small organic molecules and ssDNA.

Single-molecule detection of larger analytes such as human and bovine thrombins using *Salmonella typhi* derived Cytotoxic cytolysin A (ClyA) pores modified with DNA aptamers has been achieved in work by Soskine *et al.*[16] These modified ClyA pores were shown to be capable of translocating proteins up to 40 kDa through its 3.8 nm diameter transmembrane lumen. A follow-up study[17] by the same group made use of a directed evolution approach to yield three different ClyA oligomers (12-mer, 13-mer and 14-mer) with lumen diameters ranging from 3.3 nm to 4.2 nm. In principle, larger protein pores such as twin-arginine translocases and SecA translocases found in bacterial secretion systems. However, the introduction of stimuli-responsive gating elements for controlled release of substances from vesicle-based drug delivery systems

(through ligand activated gates, photo-caging or temperature-gating) has not yet been attempted. [1] Redesigning biological protein nanopores to fit a set of desired specifications (specific length, lumen diameter, lumen chemistry, hydrophobicity, gating functionality) is a lengthy iterative process, whether it be through site-specific mutagenesis or directed evolution, modifications with aptamers or site-specific ligands or any combination of these labour-intensive techniques. Much of the literature on engineered protein nanopores focuses on tuning one key feature of the nanopore, such as lumen diameter[16]–[19], or the electrostatic/chemical properties of the lumen[9], [10], [20], [21], as tuning multiple properties simultaneously is extremely strenuous. This makes *de novo* design of nanopore from biological protein pores very difficult, so interest in synthetic alternatives such as solid-state nanopores and DNA nanopores has grown in recent years as they allow for a greater degree of tunability with less effort.

Solid-state nanopores usually consist of an ultra-thin silicon nitride/graphene-based membrane, in which holes of a chosen diameter are drilled with electron beams. The diameter size, pore length and pore geometry can be tailored to accommodate the class of analytes under investigation, and the lumen walls can be chemically modified to an extent. However, the quantity, loci and specificity of these modifications are much more difficult to control than in protein nanopores. It is also relatively difficult to fabricate solid-state membranes that are thin enough to achieve single-nucleotide precision for DNA sequencing. In order to distinguish between different nucleotides, the membrane thickness must correspond to the length of a single nucleotide (6 Å), as each nucleotide must occupy the nanopore fully and singularly to exert its characteristic current blockade. [9][22].

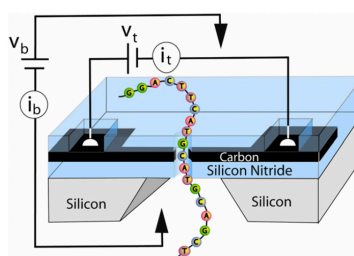


Figure 1.5: A silicon nitride/carbon based solid-state nanopore sensor, used for DNA sequencing. The transverse current (i_t) between the two carbon electrodes indicated by white semi-circles is modulated by the passage of individual nucleotides. Adapted from Spinney *et al.* [23]

1.3 DNA Nanopores

Ideally, we would like to harness all of the best features of engineered biological pores (the specificity, chemical modifiability, and ease of assembly) together with the durability and size-tuneability of solid-state nanopores to create *de novo* nanopores which fulfil the specific criteria required for the intended application. [22] DNA is an excellent candidate for the design of such nanopores, as the phosphate groups on its backbone are easy to modify with simple nucleophilic substitution reactions, and the specificity of its base pairing makes the assembly of DNA nanostructures predictable. Invoking the principles of DNA origami, a series of complementary scaffold and staple strands can be rationally designed to fold into interlacing DNA duplexes, which all hybridise spontaneously to form complex three-dimensional structures. [24] The first DNA origami structures were being synthesised in the early 2000's, and were published in the groundbreaking 2006 Nature paper by Rothemund. [6] These structures were relatively simple, 2-dimensional shapes designed using the “raster-filling” approach, where a long single strand of viral DNA is folded into the desired shape, and the unpaired bases are filled with smaller staple strands.

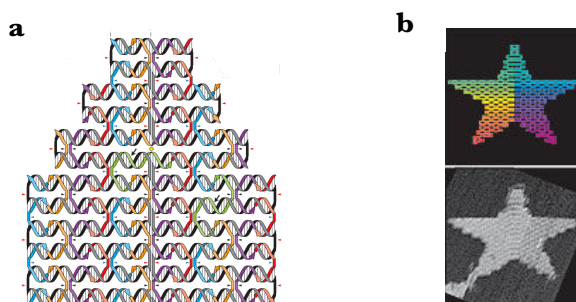


Figure 1.6: (a) Schematic showing how staple strands (coloured) hybridize with a large scaffold strand (black) to form a network of interlacing DNA double helices. (b) TEM images of two-dimensional star-shaped DNA origami structure designed and synthesised by Rothemund. Figures adapted from the publication by Rothemund. [6]

DNA nanopores can be designed according to the same protocol, using a specialised software package called caDNAno. They can be designed to have a cylindrical, conical or even cuboidal shape, and the diameter of the pore lumen can range anywhere from 2 nm to around 20 nm. [1][3][14][15] Hydrophobic groups can be introduced quite easily onto the outward-facing surface of the pore by making chemical modifications

to the backbone phosphate groups or to the strand termini. [15][16] The addition of hydrophobic anchors facilitates the insertion of the densely charged DNA nanopore into the strongly hydrophobic membrane environment and allows it to reside within the membrane for an extended period of time. A number of different anchors have been trialled in experiments, ranging from small ethyl groups to large porphyrin rings.

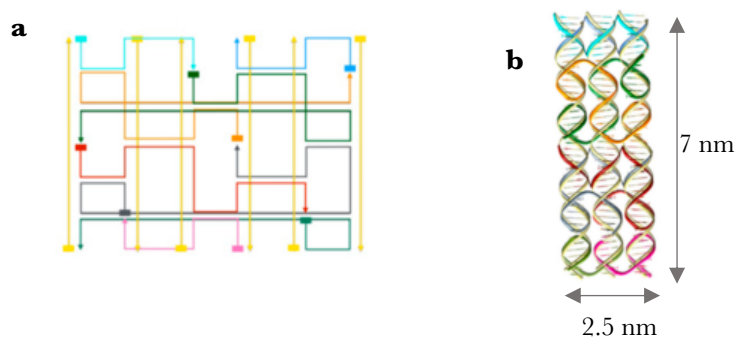


Figure 1.7: (a) 2D caDNAno schematic showing the placement of DNA strands across the 6 hexagonally arranged helices that make up a 6-helix DNA origami nanopore. The 5' ends are denoted with a square, and the 3' ends with a triangle. Two of the longer 'staple' strands (green) are incorporated into all 6 duplexes, and by doing so they bind the helices together. (b) 3D representation of the caDNAno schematic. Figures adapted from Maingi *et al.* [28]

The advantages of DNA nanopores over their solid state and biological counterparts are numerous. Their tailorability, ability to span biological membranes combined with their sensing capabilities makes them more widely applicable than solid-state nanopores. Targeted drug delivery is one groundbreaking potential application that is currently under investigation. By incorporating ligand-gated or photosensitive DNA nanopores into vesicles containing bioactive compounds, the release of drugs/bio-imaging agents can be localised and controlled, providing a precise method of treatment with limited side effects. [3] The residual negative charges on the surface of anchored DNA pores gives them the ability to modulate the structure of planar membranes, as described in a study by Birkholz *et al.*[4] which reported observations of curved protrusions in DOPC polymer-supported membranes exposed to cholesterol-anchored DNA nanopores. The deformation of the membrane was a result of re-orientation of membrane lipids such that the positive charged choline moieties on the DOPC lipid face the DNA backbone, in an attempt to alleviate the hydrophobic

mismatch (discussed in Section 1.4) between the membrane and the nanopore. Given their ease of assembly and functionalisation, they could be engineered to target specific membranes, and even specific cells, to regulate biological functions like apoptosis and cell signalling. [2][1] As interesting as this novel branch of DNA nanobiotechnology is, it is not without its limitations. While hydrophobic anchors make insertion of DNA nanopores into membranes possible, the process is neither quick nor easy.[3],[4] Some of the smaller DNA nanopore designs based on the 6-helix bundle motif (discussed in Section 1.5 and shown in Figure 1.16) are not as structurally robust as protein nanopores due to the intrinsic flexibility of dsDNA helices, but full structural characterisation of these small nanopores has not yet been achieved with experimental techniques. There are considerable kinetic and thermodynamic barriers that must be overcome in order for the nanopores to adopt an ideal transmembrane geometry, and it is difficult to quantify these barriers experimentally. Furthermore, it is difficult to ascertain how and why the association of DNA nanopores with the membrane influences the orientation of the bilayer lipids to the extent that deformations are observed, and determine if/how membrane encapsulation influences the structure of the DNA nanopore.[4]

Molecular dynamics methods can be used to probe the interactions between DNA nanopores and lipid bilayers in detail, and to calculate the macroscopic properties of these systems, allowing us gain insight into the nature of the DNA nanopore/lipid interface.[1]

1.4 Hydrophobic Mismatch

One of the chief factors that dictates whether or not a nanopore will assume a stable trans-membrane conformation is the degree of ‘hydrophobic mismatch’ between the membrane and the membrane spanning region of pore i.e. the mismatch between the thickness of the hydrophobic bilayer and the length of the hydrophobic region of the pore. To avoid the energy penalty associated with this phenomenon, most protein pores feature a belt of hydrophobic residues in its membrane-spanning region to facilitate fast and stable insertion. [30][31]

From a thermodynamic perspective, it would be ideal if the length of the hydrophobic membrane spanning region matched the thickness of the bilayer. However, in nature,

this is not always the case, and as a result we observe a phenomenon known as “hydrophobic matching” in membrane biophysics. This is a process by which membrane lipids re-organise and/or adjust their chain thickness to match the thickness of the transmembrane region of the pore. Alternatively, the transmembrane pore could tilt, twist its backbone, form aggregates or alter its conformation/orientation within the membrane in an effort to reduce contact between its hydrophilic surfaces and the membrane. [30][29] Hydrophobic mismatch is said to be ‘positive’ when the hydrophobic length of the membrane spanning structure is larger than the hydrophobic width of the lipid bilayer, and ‘negative’ in the reverse case. [31] Hydrophobic matching is a complex thermodynamic phenomenon, which can be understood through the use of Monte Carlo (MC) and molecular dynamics (MD) methods. MD simulations can be performed on nanopore/membrane systems to extract information on the energetics associated with pore formation and lipid reorganisation, and to investigate the interactions between the nanopores and the bilayer lipids in detail. [32]–[35] The large-scale MD study of hydrophobic matching mechanisms employed by various KALP proteins published by Kandåsamy and Larson[31] in 2008 provides an excellent example of this. By systematically altering the KALP protein chain lengths and the lipid chain lengths in the bilayer, they authors built a series of nanopore-membrane systems, each representing a different mismatch condition. The simulations (each between 50 ns and 200 ns long) showed that in positive mismatch conditions, there is a strong tendency for the protein pore to tilt violently within the membrane to alleviate the mismatch, whereas the negative mismatch condition triggers the lipids to reorganise in such a way that causes the bilayer to curve while the pore remains relatively static, with lysine residues “snorkelling” towards the polar lipid headgroups. [31] While there are several published MD studies on DNA nanostructures interacting with lipid bilayers, the specific hydrophobic matching behaviours of these nanostructures have not been investigated as rigorously as those of membrane proteins, and hence they are less well understood. These findings of these studies will be discussed in Section 1.7.

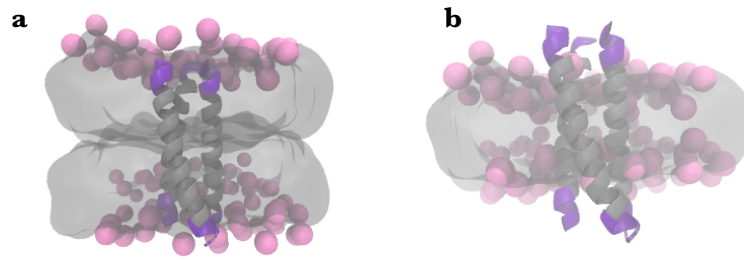


Figure 1.8: (a) Example of negative hydrophobic mismatch between a trimeric protein and a lipid bilayer. The hydrophobic regions of the bilayer and the protein are coloured grey, and the hydrophilic regions are coloured pink/purple. (b) Example of positive hydrophobic mismatch.

1.5 Experimental Methods in Nanopore Research

There is a well-established retinue of techniques used in nanopore research, which give scientists information about the structure, conformation, permeability, conductance and translocational properties of a nanopore. While these methods are powerful in their own right, their results are often used in conjunction with simulation data to build up a more detailed, comprehensive picture of the dynamical properties of biological (and synthetic) nanopores at the smallest timescales, and the effect that these properties have on their function.

1.5.1 Investigating Translocational Properties of Nanopores

The translocational properties of nanopores are usually studied both quantitatively and qualitatively, using a combination of electrophysiological and microscopic techniques. Single-channel current recording experiments (a specific type of electrophysiological technique) enable the measurement of the ionic conductance of a single nanopore, and also observe single-molecule translocation events.

The planar lipid bilayer technique is the most commonly employed single-channel electrophysiological method in nanopore research, with the aim of characterising the conductance of the nanopore. The apparatus consists of a chamber divided into two compartments (*cis*- and *trans*-) holding an electrolytic solution, which are connected by a small aperture. The *cis*- compartment mimics the ‘outside’ of the cell, and the *trans*-

compartment mimics the inside of the cell. Lipid solution is applied to the aperture and a planar lipid bilayer subsequently forms, in which the nanopore is embedded for analysis. The aperture size is tailored to the size of the pore being analysed, to ensure that the conductance is being measured for one single nanopore. When a voltage is applied, ions are translocated through the nanopore exclusively (as the membrane is non-conductive) to either the *cis*- or the *trans*- side, depending on their charge, and the amplitude of the transmembrane current is recorded. The magnitude of the current is recorded at a series of different voltages, and a current-voltage (I/V) relationship is then plotted. The conductance of a nanopore (the degree to which it can conduct electric current, the reciprocal of resistance) is then calculated by taking the gradient of the current-voltage relationship.[36]

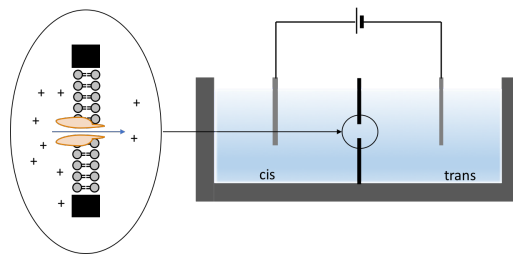


Figure 1.9: A simplified representation of a chip-based apparatus used for planar lipid bilayer electrophysiology experiments. The lipid solution forms a bilayer within the aperture in the partition between the *cis* and *trans* chambers, and the nanopore inserts into the bilayer. When a potential difference is applied, a current flows through the nanopore, carried by the ions translocating through the pore lumen.

Planar lipid bilayer techniques can also be used to differentiate between the closed and open states of nanopores that exhibit gating behaviour, such as the voltage-gated ion channels that are found in neurons. Altering the conditions of the experiment and subsequently measuring the conductance of the pore allows for the investigation of the effect that voltage, pH and ionic strength have on the gating properties of the nanopore. They are also employed heavily in blockade experiments, which are used to study the binding or translocation of molecules to engineered nanopores used for sensing. [36]

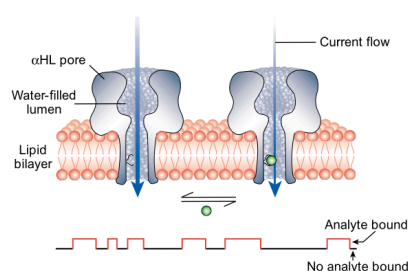


Figure 1.10: A simplified representation of a blockade experiment based on a modified α HL pore to which an analyte (green ball) can bind selectively. When a potential difference is applied, current flows through the pore, as the membrane is non-conductive. The current is perturbed whenever there is a binding event, and a signal specific to the particular analyte is observed on the trace. Image adapted from Bayley and Cremer *et al.* [9]

Confocal fluorescence microscopy is a versatile optical method that can be adopted in both a qualitative and a quantitative capacity. In nanopore research, it is often used with FRET pairs (which act as ‘spectroscopic rulers’) or simple self-quenching fluorescent tags to directly visualise and confirm proper pore insertion and confirmation within a membrane/vesicle and to determine the kinetics of insertion. When used in dye influx/efflux assays, it can be used to monitor the translocation rate of a fluorescent dye through the nanopores and establish influx/efflux kinetics. [37]

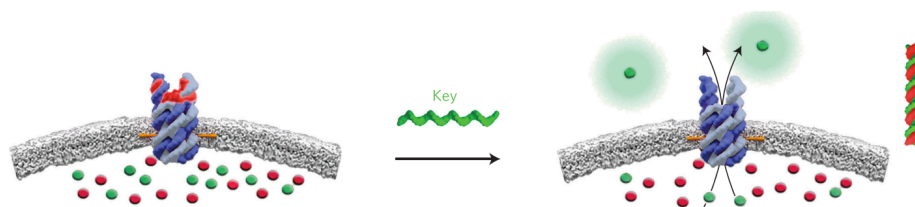


Figure 1.11: Illustration of a dye efflux assay performed on a ligand-gated DNA nanopore embedded in a vesicle containing self-quenching fluorophores carboxy-fluorescein (red) and sulpho-rhodamine (green). The figure illustrates the ‘unlocking’ mechanism of the ligand-gated nanopore and subsequent release of the fluorophore molecules from inside a vesicle. The green “key” is a short strand of *ss*DNA with a sequence that is complementary to the red “lock” strand – an *ss*DNA strand that ties together two overhanging loops of *ss*DNA located on one of the pore termini in the nanopores closed state by forming tentative base pairs with them. When exposed to the key strand, the lock strand dissociates from the nanopore to form a stable dsDNA double helix with the key strand. When the nanopores are opened, the fluorophores diffuse through the nanopore from concentration inside the vesicle; where emission is low due to self-quenching, to an area of low concentration outside the vesicle, where the fluorophores are highly visible under the confocal microscope. Adapted from Burns *et al.* [38]

1.5.2 Structural Characterisation of Nanopores

Due to their very small size, it is difficult to closely analyse the structural properties of nanopores using experimental methods. That being said, there some techniques that have been successful at providing basic information on the shape, dimensions, assembly status and lumen diameter of various biological and synthetic nanopores. Transmission electron microscopy (TEM), atomic force microscopy (AFM), and cryo-electron microscopy (cryo-EM) are three imaging techniques that are capable of achieving the sub-nanometer resolution required for examining nanopores. Of these three techniques, AFM is by far the most commonly employed method for directly visualising nanopores that have been embedded or fabricated within solid supports (e.g. mica) or membrane environments. AFM is an extremely high-resolution microscopic technique, which allows 3-dimensional imaging of sample surfaces at atomic resolutions using a physical probe that scans the sample and outputs

topographic information, allowing researchers to take direct measurements of pore dimensions *in situ*. High-resolution AFM typically uses a thin silicon nitride tip (~ 2 nm radius at the tip apex) attached to cantilever controlled by piezoelectric drivers to scan the surface of the sample, in the presence of a laser, which acts as an ‘optical lever’. The forces (attractive and repulsive) between the tip and the surface are measured, along with the deflections of the laser beam off the cantilever, and this information is processed into a high-resolution 3D image. [39]

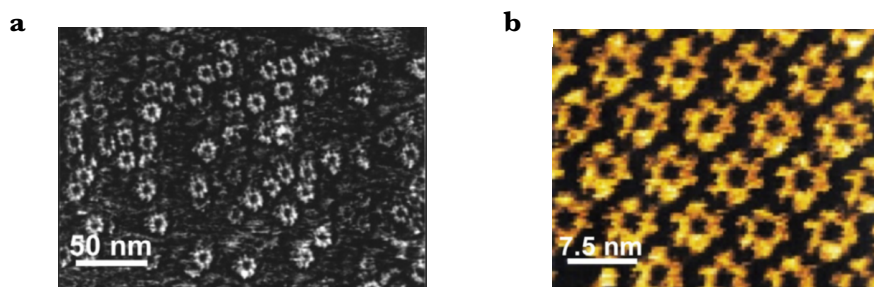


Figure 1.12: (a) AFM image of mutant α -haemolysin nanopores on mica. (b) AFM image of α -haemolysin on a lipid bilayer. Images taken from Yilmaz and Kobayashi. [40]

More recently, cutting-edge cryo-EM methods capable of achieving near-atomic resolution have been brought to the fore of biomolecular imaging techniques. The basic principle behind these methods is centred on the analysis of specially prepared flash-frozen specimens using either traditional transmission electron microscopy, electron tomography, electron crystallography, or any combination of these techniques. Direct electron irradiation is highly damaging to organic matter, causing breakage of covalent bonds and subsequent loss of secondary, tertiary and quaternary structure, so specimens must be shielded from the harmful effects in some way. Flash-freezing samples at liquid nitrogen temperatures coats them in a protective glass-like layer of ice, permitting the use of higher electron doses than those that are routinely used in traditional TEM. Higher electron doses increase the signal-to-noise ratio during the imaging process, and the end result is an ultra-high resolution, artefact-free 3D reconstruction of the biological structure. Alternatively, averaging across multiple images captured under cryogenic conditions with lower electron doses can be used to yield reconstructions of similar resolution. [41]

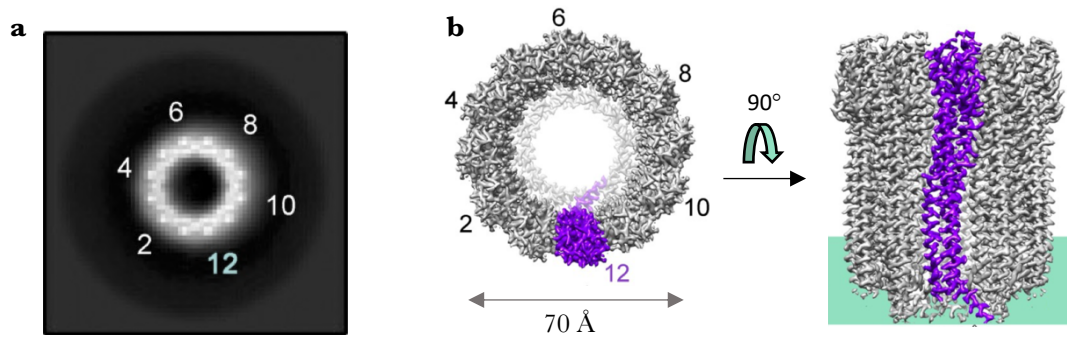


Figure 1.13: High-resolution cryo-EM structures of the *E. Coli* Cytolysin A dodecamer **(a)** Class-averaged 2D image of the dodecamer (birds eye view). **(b)** Complete 3D electron density maps of the dodecamer, reconstructed at a resolution of 2.8 Å. Figures adapted from Peng *et al.* [42]

There are other, more indirect methods that can be employed to investigate structural properties of nanopores. Differential polymer exclusion is one such method; it uses various poly-ethylene glycols (PEGs) of different molecular weights to determine the size of the channel lumen. The nanopore is embedded in a planar lipid bilayer within an electrophysiology chamber (Figure 1.9), and a transmembrane voltage is applied. Separate experiments are then carried out, where a solution containing a specific PEG is added to the *cis*- side of the chamber, and the current trace is monitored for signal perturbations. PEGs with a hydrodynamic diameter smaller than the pore lumen will enter the pore and cause blockade events, whereas larger PEGs will be unable to enter the pore and the signal will be unperturbed. By plotting the blockade level against the hydrodynamic diameter of the PEG molecules, the size of the lumen can be inferred. [43][44]

1.6 Design, Synthesis and Application of DNA Nanopores

Research into the applications of DNA nanopores is still in its youth, but the advances and observations described in the recent literature have highlighted a number of potential applications. Bell *et al* published the first of many studies detailing a successful application of DNA origami nanopores in early 2012. They designed a large conical DNA channel from an 8634 nucleotide-long scaffold strand, combined with 142 staple strands, which they embedded into various differently sized solid-state SiN nanopores. Single channel current demonstrated that ions could diffuse across the hybrid pore freely with an average conductance of ~ 80 nS. The hybrid SiN/DNA pore was also shown to be capable of detecting λ -DNA: short DNA strands harvested from the lambda virus. [45]

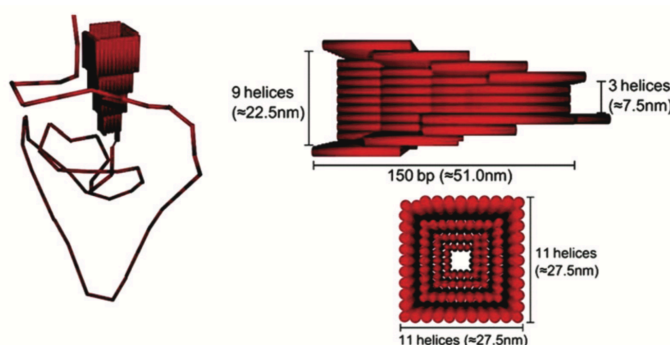


Figure 1.14: DNA origami nanopore designed by Bell *et al*. Double helices are represented as rods, and the overhanging dsDNA strand acts as a thread to guide the nanopore into the SiN pore. Figure adapted from Bell *et al*. [45]

To assess the practicality of standalone DNA nanopores in contrast to their protein counterparts, the translocation of ions through the pores in biological media was investigated using chip-based planar bilayer methods. The Simmel group designed a slightly shorter cylindrical DNA channel inspired by the structure of α HL, featuring a large 54-helix ‘cap’ and a narrow 6-helix transmembrane domain.

A total of 26 cholesterol-modified oligonucleotides were used to introduce hydrophobic anchors into the structure, and these hybridised to ‘adaptor’ strands of ssDNA hanging off the bottom of the pre-folded honeycomb cap structure. [46] Using relatively low concentrations, the group were able to successfully insert the synthetic DNA channels into differently sized POPC SUV’s and planar DphPC lipid bilayers, which they were

unable to do without cholesterol anchors due to the hydrophobic mismatch between the The reported conductance of the DNA channel was ~ 0.87 nS; slightly less than that of a typical protein pore in the same conditions (~ 1 nS) and slightly higher than the calculated value; an observation that was attributed to lateral conductivity of small ions in the cap region, which appeared to have leaky walls. [46]

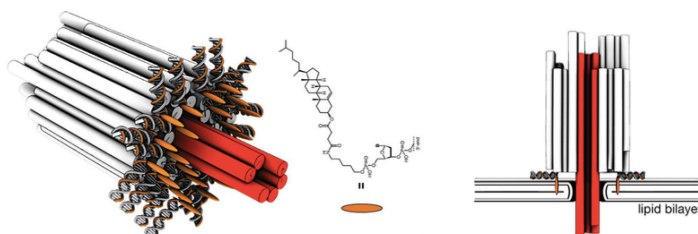


Figure 1.15: Schematic diagram illustrating the structure of the DNA nanopore designed by the Simmel group. Orange strands within the helices represent modified ssDNA strands with cholesterol moieties (orange ellipsoids) attached to the 3' end. Adapted from Langecker *et al.* [46]

These early studies on DNA nanopores focussed on their potential applicability in sensing and sequencing, and the nanopores used in these studies were quite large, bulky, and cumbersome to fold. In 2013, Burns, Stultz and Howorka began looking beyond this *status quo ante*, designing relatively simple but elegant DNA nanopores with a broader range of applications. One of their earlier designs consisted of a rationally designed six-helix bundle with a hydrophobic belt around the membrane-spanning surface comprised of 77 neutral ethyl mercapto groups. [24]

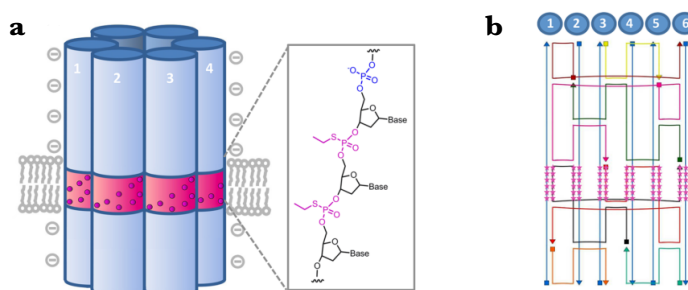


Figure 1.16: (a) 3D representation of the six-helix bundle designed by the Howorka group, with the hydrophobic belt shown in magenta. Commercially available DNA strands with phosphorothioate (PPT) modifications were sourced for the synthesis of the nanopore, and these were reacted with ethyl iodide to yield ethyl-phosphorothioate groups. (b) The corresponding 2-dimensional caDNAno schematic of the nanopore. Figures adapted from Burns *et al.* [24]

The dimensions of the pore were measured using AFM experiments, which gave a length of 21.3 ± 4.0 nm and a width of 10.3 ± 1.9 nm (for pores adsorbed on mica). The authors considered the measurements to be in good agreement with the theoretical dimensions of 15nm by 5.5nm, when accounting for tip convolution artefacts. The artefacts correspond to a broadening in the dimensions of the features when the AFM tip radius is similar to or larger than that of the feature. The discrepancy between the predicted dimensions and the AFM measurements, and large errors associated with the AFM measurements highlighted the that the structural properties of these DNA nanopores are likely to highly sensitive to experimental and analytical conditions.

Single channel current recordings in DphPC revealed a linear current-voltage characteristic matching those described in the previous literature on DNA nanopores, and the group reported an average conductance of 0.395 ± 0.097 nS. It was noted that the nanopores stayed intact and sound within the membrane at voltages up to 200 mV, but at higher voltages the pore exhibited gating properties, as suggested by the transition from Ohmic to non-Ohmic behaviour in the current-voltage relationship past 100 mV – usually indicative of a transition from an open state to a closed state. At applied voltages above 200 mV, the nanopores had a tendency to pop out of the membrane. [24] Since the initial 2013 design, there have been many iterations of the 6-helix bundle DNA nanopore. The Howorka group have experimented with many alternative arrangements of oligonucleotide strands and used several different types of hydrophobic anchor. These nanopore designs had similar overall dimensions and conductance values but exhibited a variety of functional behaviors on account of the structural differences in their frameworks.

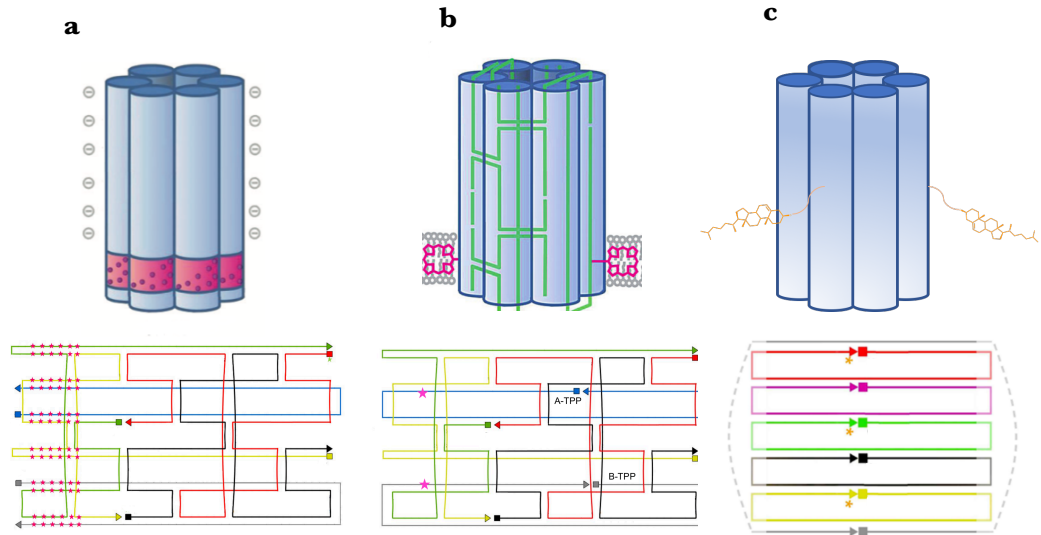


Figure 1.17: Examples of different hydrophobic anchors used in DNA nanopore designs by Burns *et al.*, with 2D maps below. **(a)** Cytotoxic ethyl-PPT anchored DNA nanopore. Adapted from Burns *et al.* [47] **(b)** Tetraphenylporphyrin-anchored DNA nanopore. Adapted from Burns *et al.* [27] **(c)** A shorter, TEG-cholesterol anchored DNA nanopore. [3][29]

A 2014 study by Burns, Howorka and co-workers explored the possibility of using ethyl-PPT DNA nanopores (Figure 1.17a) as therapeutic agents. After incubating the ethyl-modified DNA nanopores with cervical cancer cells for 24 hours at a concentration of 60 mg mL^{-1} , they saw a 20% decrease in cell viability compared to controls. While the study didn't provide much evidence to suggest that the DNA nanopores could specifically target cancer cells, it proved that DNA nanopores with hydrophobic anchors are capable of penetrating biological membranes in a manner similar to pore-forming toxins such as α HL and gramicidin. [47]

Different variations of the six-helix bundle design have been used to develop DNA nanopores with unique gating properties. The TTP-anchored nanopore shown in Figure 1.16(b) exhibited voltage-gating characteristics; adopting a high-conductance state at low voltages with an average measured conductance of $\sim 1.6 \text{ nS}$. [43] The TEG-cholesterol anchored nanopore shown in Figure 1.16(c) was used as the foundation for a ligand-gated DNA nanopore developed by Burns *et al.* and published in 2016, and a temperature-gated equivalent published three years later. The locking mechanism of the former was irreversible, meaning that the transition between the closed state and the open state could be achieved only once, as the application of the

ligand “key” removed the locking mechanism from the DNA nanopore completely. The temperature-gated pore was designed with reversible locking and unlocking in mind, and the authors successfully demonstrated that transitions between the open and closed state could be controlled through subtle increases and decreases in temperature.

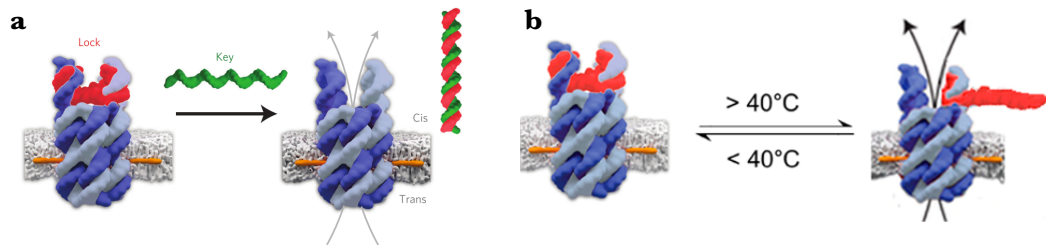


Figure 1.18: (a) Ligand-gated nanopore by Burns *et al* [38]. The figure illustrates the ‘unlocking’ of the nanopore and subsequent release of cargo from inside a vesicle. The “key” is a short strand of DNA, which hybridizes to the “lock” strand. The lock strand is an ssDNA strand that ties together two overhanging ssDNA strands located on one of the pore termini in the nanopores closed state by forming tentative base pairs with them. When exposed to the key strand, the lock strand dissociates from the nanopore to form a stable dsDNA double helix with the key strand. (b) Temperature-gated nanopore designed and studied by Arnott and Howorka [48] based on the modified locking DNA nanopore in **a**, where the lock stand is covalently bound to the body of the DNA nanopore, allowing reversible temperature-induced transitions between the open and closed states to occur. Figures adapted from Burns *et al* [38] and Arnott and Howorka [48].

1.7 Molecular Dynamics Simulations of DNA Nanopores

DNA nanobiotechnology is still an emerging field and hence computational literature on the subject is fairly sparse. The first molecular dynamics simulations of DNA origami nanopores were performed in 2015; on hand-built all-atom structures that had not yet been synthesised by experimentalists, that were based on the 6-helix bundle motif popularised by Howorka et al. The Aksimentiev group [49] were among the first to model DNA nanopores, and their first simulation study focused 6-helix bundle nanopore based on the 6-helix bundle designed published by Burns et al [24] in 2013. Ethyl modifications (72 in total, the same number as the original design by Burns et al) were added to the backbone phosphates around the central circumference of the barrel using the CHARMM software package. Parameters for the ethyl-phosphate groups were derived from the CHARMM General Force-field (CGenFF), which were prepared for simulation using the NAMD MD code, implementing the CHARMM36 force field.[49]

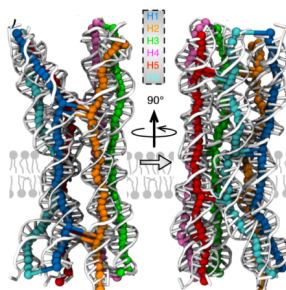


Figure 1.19: All-atom model of the ethyl-PPT modified DNA nanopore designed by Yoo *et al*, after 70 ns of equilibration time. Figures adapted from Yoo *et al*. [49]

The goal of their production simulations (of which there was 6 replicas) was to explore the local structural fluctuations of the DNA nanopore within a membrane with atomic precision, and to determine which factors influence the flow of ions/conductance mechanism of the pore. The base pairs located at the termini of the nanopore had the highest RMSF values – suggesting that these regions are conformationally non-rigid and may exhibit gating behavior. They ran additional production simulations to explore the relationship between membrane tension and nanopore conductivity and found that membrane compression (high tension conditions) caused the nanopore to expand, increasing pore conductance and strengthening the integrity of the DNA-lipid interface. Density flow maps revealed that ions moved almost exclusively through the lumen at lower voltages, and that higher voltages caused the DNA/lipid interface to break down, causing lateral diffusion of ions.[49]

In the same year Maingi, Sansom and co-workers[28] conducted a similar investigation to that of Aksimentiev *et al*, but with a specific focus on the lateral permeability and gating properties of DNA nanopores in solution, in the absence of a membrane. The models were based on a design by Burns *et al* [24] (Figure 1.15) but the hydrophobic anchors were omitted in this study. Four starting configurations were selected, and these models were submerged in octahedral solvent boxes, with a KCl concentration of 1M. Three of the models were simulated using the AMBER12 MD engine together with the AMBERff99+parmbasc0 force-field, and the remaining model was simulated using GROMACS alongside the CHARMM36 force-field.

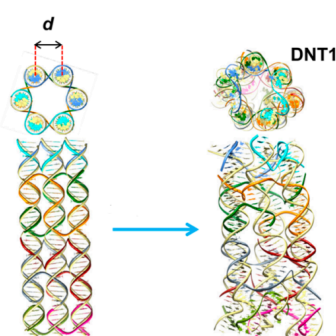


Figure 1.20: Starting geometry of the DNT1 model (left), and the final geometry after 170 ns of dynamics (right). Figure adapted from Maingi *et al*. [28]

Production simulation times ranged between 115 ns (AMBER) and 500 ns (GROMACS). The low RMSF values at in the midsection of the nanopore indicated the pore have a structurally stable lumen, whereas the pore termini had high RMSF values, indicating gating characteristics. The mean square displacements of water and ions were used to calculate diffusion coefficients down the lumen, along the principal pore axis. The values reported were slightly lower than the diffusion coefficients calculated for the bulk solution, and visual inspection revealed that this was due to the closure of one of the terminal “gates” due to narrowing of the pore lumen. Further visual inspection showed that there was a tendency for K^+ ions to reside in the “walls” of the nanopore, between the adjacent DNA helices, and for some lateral diffusion of ions (perpendicular to the principal pore axis) from the interior of the pore lumen to the outer surface of the nanopore. The average pore radius for the four models was $\sim 0.8\text{nm}$; close to the expected value of 1 nm.[28]

The following two simulation-based studies on DNA nanopores took a departure from the six-helix bundle framework and considered two opposite ends of the spectrum. The first was based on a nanopore that consisted of a single DNA duplex, and the second featured a large DNA-origami porin, similar to the design put forward by the Keyser group in 2012. The first study used both computational and experimental data to demonstrate that a single 19-base pair duplex bearing 3 porphyrin-based membrane anchors (Fig. 1.21) was able to form a pore within a DphPC bilayer through which ions could diffuse freely.[26]

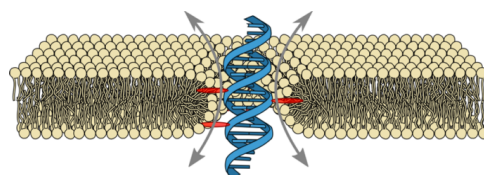


Figure 1.21: Cartoon representation of the TTP-tagged DNA strand spanning a DphPC bilayer, showing the movement of ions through the DNA/bilayer interface. Figure adapted from Göpfrich *et al.*[26]

The DNA/membrane model was solvated in a box containing an aqueous solution of KCl (1M) and was equilibrated in two steps using the NAMD2 simulation code, with CHARMM36 forcefield. Force field parameterisation of the TTP anchors was done with the assistance of the CGenFF web-server. For the production phase of the experiment, trajectories were collected for a single model over ~ 600 ns of simulation time under a simulated external electric field. Local number densities and ionic currents were calculated at different voltages, and these values were used to construct current-voltage plots. They reported an average simulated conductance of 95 pS, which was corroborated by their experimental value of ~ 100 pS. Their ionic flux magnitudes indicated that ions moved mainly through the DNA, with some diffusion of ions between the DNA helix and the bilayer.[26]

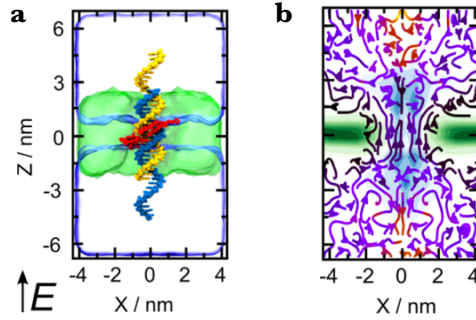


Figure 1.22:(a) All-atom model of the TTP-DNA/lipid system before equilibration. (b) Local densities of lipids (green) and DNA (blue), and ionic currents (streamlines). Arrows on streamlines indicate direction of the current flux, which corresponds to the positive membrane bias. Black lines correspond to low ionic current flux, and red lines correspond to high ionic current flux. Figures were adapted from Göpfrich *et al.* [26]

Months later, the Keyser and Aksimentiev groups published their second hybrid computational/experimental study. The large-conductance, conical DNA-origami porin described in their paper [50] had the widest pore lumen ever reported for a DNA nanopore, and the largest overall dimensions too. The nanopore featured 19 cholesterol anchors, which were introduced in the structure using the technique previously described in Figure 1.14. Fluorescent confocal spectroscopy experiments confirmed that this number of cholesterol anchors was necessary to induce full permeation of the nanopores into vesicular membranes. The model-building and simulation protocols used for the computational part of the study were similar to those used in the previous study by Göpfrich *et al* [26]. The results of the all-atom production simulations (~ 20 ns at ± 100 mV and ~ 50 ns at ± 30 mV) were used to construct conductance histograms, from which an average conductance of 46.6 nS was calculated. The number density plots (Fig. 1.23c) and ionic flux calculations indicated that $\sim 80\%$ of ion flow occurred through the pore lumen, while the remaining 20% of ion diffusion occurred through a gap at the DNA/bilayer interface. The authors go on to address the large variance in their conductance data, which they attribute to insufficient sampling; a result of the relatively short simulation timescales they employ.

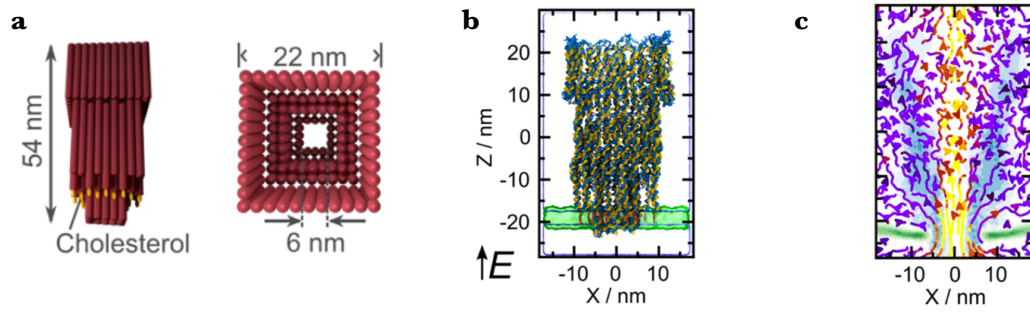


Figure 1.23: (a) Simplified schematic illustrating the structure of the DNA porin designed by the Aksimentiev and Keyser groups. (b) All-atom model of the DNA porin, before equilibration. (c). Local densities of lipids (green) and DNA (blue), and ionic currents (streamlines). Colour codes are the same as those used in Figure 2.7. Figures adapted from Göpfrich *et al* [50].

One of the most detailed simulation-based studies on DNA nanopores to date was published in mid-2017 by Maingi, Sansom and co-workers.[25] The focus of the study was on specific interactions between the nanopore and the membrane lipids, and the effect that they have on the stability of membrane-spanning DNA nanopores. To maximise simulation timescales and sampling while maintaining precision, a combination of coarse-grained (CG) and all-atom (AA) simulation techniques were used. Coarse grained models were built and simulated for up to 1 μ s, before being converted to their respective AA representations to be further simulated for up to 500 ns. The subject of these simulations was the ethyl-PPT anchored DNA nanopore featured in 2013 publication by Burns and Howorka [3], shown in Figure 1.15.

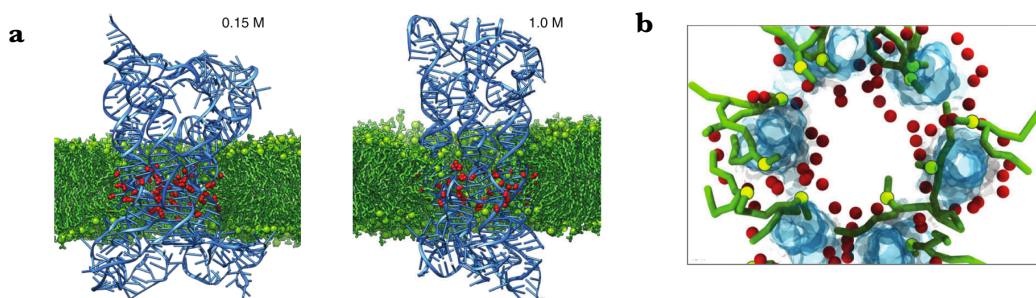


Figure 1.24: (a) Snapshots from AA simulations of the ethyl-PPT anchored DNA nanopore spanning the bilayer, illustrating the effect of salt concentration (KCl) on the structure of the pore. (b) Snapshot of the same simulation, illustrating the interactions of individual POPC lipid chains (green) with the major and minor grooves in the DNA. The red spheres represent the hydrophobic ethyl-PPT anchors, which were covalently attached to the phosphate groups located on the midsection of the DNA nanopore. Figures taken from Maingi *et al.*[25]

Although previous MD studies [16],[74] suggested the presence of an ion-permeable gap in the DNA/membrane interface, close inspection of lipid dynamics over extended simulation times ($\sim 1 \mu\text{s}$) and across 6 replicas revealed that the lipids did in fact pack around this particular nanopore, with the fatty acid chains settling within the major and minor grooves of the helix. This may have been facilitated by the hydrophobic ethyl-PPT groups, or this may be unique to the DNA/POPC system; as the characteristics of POPC bilayers are markedly different from those of DphPC bilayers due to the nature of their fatty acid chains.[3], [51] One important point these studies on DNA/membrane systems fail to address is the reproducibility of their results. They do not offer any justification for the duration of their simulations, or the number of simulations performed, and hence they cannot comment on the precision of their results. Without a robust uncertainty quantification protocol, one cannot determine whether or not the sampling of the phase space in a set of simulations is adequate and should exercise caution when drawing conclusions from this data.

A similar MD study on the interactions between DNA nanopores and lipid bilayers carried out by Joshi and Maiti was published soon after. The nanopore under investigation was a stackable hexagonal nanotube with “sticky ends” designed and synthesised by Wang *et al* in 2012. [52] The authors built an AA model of a single-unit nanotube embedded in a POPC bilayer patch, with no hydrophobic moieties attached to the nanotube. The authors explored the influence of different salt conditions on the dynamics, conductivity and stability of the membrane-spanning nanotube, and characterised the perturbations experienced by the membrane lipids as a result of their interactions with the nanotube. Five different salt conditions were explored: no salt, 0.5 M NaCl, 0.5 M KCl, 1.0 M NaCl and 1.0 M KCl. The time-evolution of the average RMSD of the nanotube along with average lipid tilt angles, area-per-lipid and ion diffusion coefficients for all four salt conditions were extracted from one-off AA equilibrium simulation trajectories, which were run for 205 ns apiece using the AMBER MD code with AMBER parameters.[53] Computational current-voltage relationships for each salt condition were derived from sets of simulations (five simulations, at transmembrane voltages: 10, 20, 50 100 and 200 mV) performed under a constant applied electric field.

The authors make clear and distinctive comparisons of conductivity and RMSD between different salt conditions, however they do not comment on the errors associated with calculated values, making qualitative assessments rather than quantitative measurements or predictions. As the simulated nanotube/membrane models did not feature any hydrophobic anchors, the authors expected the hydrophobic mismatch between the lipid bilayer and the negatively charged DNA nanotube to be extreme in all cases. In all salt conditions, they observed a large degree of lipid reorientation around the outer surface of the nanotube, with the headgroups curving towards the outer surface of the nanotube to form a toroidal lipid pore.[54]

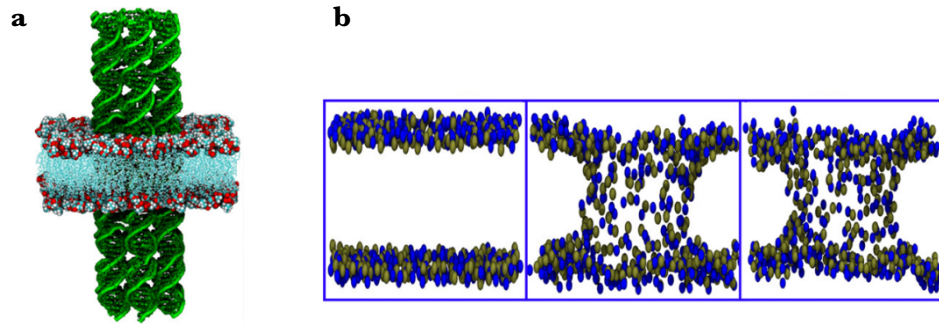


Figure 1.25: (a) Minimised structure of the membrane-spanning 6-helix nanotube. The nanotube consists of a single unit, where the helices are 74 base-pairs long (~ 24 nm along the helical axis). (b) Simulation snapshots showing the orientation of the lipid headgroups at the start of the simulation (left), after 15 ns of simulation time, the point at which the bilayer equilibrates and forms a toroid around the DNA nanopores (middle), and after the full 205 ns of simulation time (right). Figures adapted from Joshi and Maiti.[54]

Higher salt concentrations led to an increase in the rigidity of the nanotube by decreasing the inter-helix repulsion, which led to a stabilisation of the pore lumen that in turn enhanced the average conductance in the high salt (1.0 M) electric field simulations. The Na^+ was shown to have a higher affinity for the negatively charged DNA backbone, and as a result these cations had a greater influence on the rigidity of the nanotube than the K^+ cations. The simulated average conductance of the nanotube in NaCl was lower than that of KCl – another consequence of the higher binding affinity of the Na^+ . The calculated conductances were reported as a range (4.31 to 20.64 nS) and validated against a range of previously published experimental conductances for various DNA nanopores[24], [46], [49], [50], [55], [56], with the authors concluding that their calculated range overlaps well with the range of experimental values. Finally, the authors performed an extended one-off simulation in the no-salt condition (run for 0.5 μs), to establish whether or not the nanotube retained its transmembrane configuration at longer timescales. [54] As this conclusion was based on a single trajectory - which is not representative of the ensemble, the reliability of this result is indeterminate (see Chapter 2.4).

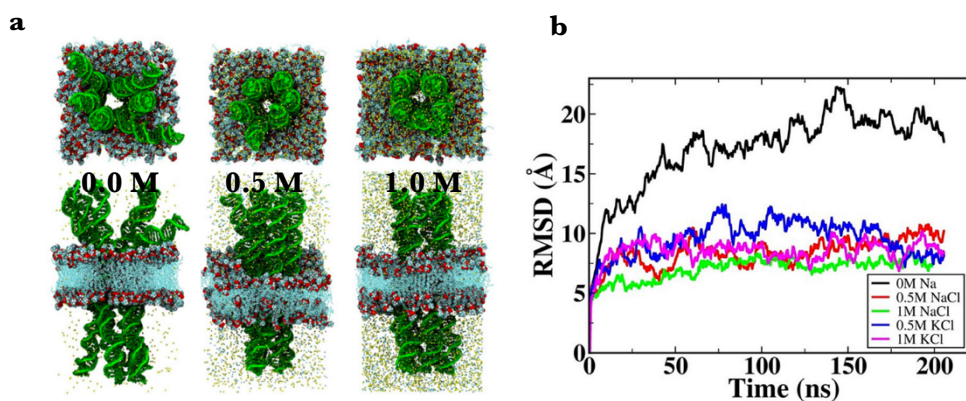


Figure 1.26: (a) Simulation snapshots illustrating the influence of increasing NaCl concentration on the rigidity of the nanotube. (b) RMSD plot with respect to the minimised structure (Fig 28.a) for all five salt conditions. The average RMSD in the absence of salt (black line) was markedly higher than it was in the other four salt conditions, which is indicative of the stabilizing effect of monovalent cations on the structure of the nanotube. The lowest RMSD/ highest degree of conformational stability was observed with 1.0 M NaCl. Figures adapted from Joshi and Maiti.[54]

Joshi and coworkers expanded on this study in 2019, and published another paper[57] focusing on the characterization of the rigidity of the nanotube in these different salt conditions (in the absence of a membrane) by calculating the stretch modulus and bending angles of the individual helices. They use these calculated values to provide further evidence of the conclusions reported in their previous paper.[54] Although errors are reported, they appear to use the same “one-off” simulation protocol as they did in their previous study, and they do not comment on the derivation of their errors.

Chapter 2

Molecular Dynamics

MD simulations are the tool of choice for investigating the configurational and thermodynamic properties of complex biological systems. They allow us to gain insight into microscopic and mesoscopic properties of complex systems that are difficult to analyse experimentally. Unlike MC simulations, MD simulations allow us to monitor the evolution of the properties of the system over time. For time-dependant processes, such as the insertion of a biological nanopore into a membrane, MD simulations give a good representation of the configurations that different components of a system are likely to adopt at each interval within the process. On the other hand, MC simulations generate a probability distribution of likely configurations given a set of conditions using random repeated sampling; making them ideal for the study of disordered systems. A downside of MD is that within a single simulation, the system can become trapped in low energy minima due to the deterministic nature of simulations. Fortunately, this can be overcome by using an ensemble-based protocol, which involves running multiple replicas of a simulation; each with different starting conditions, and calculating ensemble averages for the macroscopic properties of interest.[58]

2.1 Introduction to Molecular Dynamics Simulations

Scientists use molecular dynamics simulations to examine conformational changes in biomolecules, determine structure-function relationships and probe the kinetics and thermodynamics of biological processes such as ligand binding, protein folding and bilayer assembly. By integrating Newton's equations of motion for all the atoms within a system simultaneously, over a series of discrete time-steps – we are able to observe how the positions, velocities and various other qualities of individual atoms evolve over time by collecting a trajectory.[40]-[41]

In MD simulations, nuclei are treated as classical particles, therefore the motions of a multi-atomic system can be represented in the following fashion:

$$m_i \frac{d^2 \mathbf{r}_i}{dt^2} = \mathbf{F}_i(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$$

$$i = 1, 2, 3, \dots, N$$
(2.1)

where the second derivative of \mathbf{r}_i with respect to the time-step (dt) is the acceleration of particle i . Dynamics simulations require initial velocities for each atom, and an interaction potential/potential energy function must also be defined. An approximation of the forces within a van der Waals system can be derived from the Lennard-Jones (LJ) potential:

$$U_{LJ}(\mathbf{r}_{ij}) = 4\varepsilon \left[\left(\frac{\sigma}{\mathbf{r}_{ij}} \right)^{12} - \left(\frac{\sigma}{\mathbf{r}_{ij}} \right)^6 \right]$$
(2.2)

where σ is the distance at which interatomic/intermolecular potential is zero, ε is the potential well depth (the depth of which depends on degree of attraction between the particles), and \mathbf{r}_{ij} is the distance between particles i and j . The coulomb electrostatic potential between two charged particles i and j (usually molecules in MD simulations) is described by the following expression:

$$U_E(\mathbf{r}_{ij}) = \frac{q_i \cdot q_j}{4\pi\varepsilon_0 \mathbf{r}_{ij}}$$
(2.3)

where q_i and q_j are the charges of the particles, ε_0 is the permittivity of the free space, and \mathbf{r}_{ij} is the distance between the two particles. For most biomolecular simulation systems, which often consist of hundreds of thousands of particles, the calculation of the non-bonded interactions (vdW and Coulomb forces) is the most computationally demanding part of the simulation, as these forces arise between each individual pair of atoms. Hence, the calculation of vdW and Coulomb interactions for each atom (i) and its neighbours must be limited to inter-atomic distances between zero and a user-specified cutoff value, where all atoms falling outside of the cutoff distance from atom i are ignored.

This simplification does not significantly impinge on the accuracy of the calculated vdW interactions, which decay to zero rapidly as the distance \mathbf{r}_{ij} increases. However, this is not the case for Coulomb potentials, which have a much longer interaction range and therefore a cutoff cannot be applied in the same manner as is done with vdW interactions without causing truncation errors.[61]

Several electrostatics schemes have been developed to strike a balance between accuracy and computational efficiency, and these tend to fall into one of two categories: Ewald summation techniques[62] and modified cutoff-based methods. Ewald summation methods treat the electrostatic potential as the sum of two terms; one short-range, which rapidly decays as the distance \mathbf{r}_{ij} approaches infinity in real space, and a long-range term with a sum that converges in Fourier space. In the widely implemented particle mesh Ewald (PME) method, the short-range electrostatic interactions are summed for all atom-pairs within a certain cut-off distance. The calculation of long-range interactions is done by assigning the individual point charges to a grid by interpolation, which is Fourier transformed before the potential energies at the grid points are summed. The use of PME algorithms achieves excellent numerical accuracy whilst scaling at $N \ln(N)$ calculations for a system of N particles.[63]

The reaction-field (RF) method[45]-[46] is a popular modified cutoff-based technique that provides a good alternative to Ewald summation, producing results that are generally consistent with the latter technique.[64] In the RF method, a “cavity” around each atom/molecule with a user-defined cut-off radius is defined, in which Coulomb interactions are treated explicitly, and the region outside of this cavity is assumed to be a dielectric continuum (with a dielectric constant ϵ_{RF}). Each cavity is polarised by its interactions with the atoms/molecules inside the cavity. This polarisation of the media gives rise to an electric field (the reaction field) within each cavity. The electrostatic potential energies for each cavity are summed across the entire system to yield a total electrostatic potential at each timestep. It is worth noting that while this method is computationally efficient, the combination of explicit and continuum treatment of electrostatics can have some negative consequences for heterogenous systems with charged surfaces, such as bilayer/nanopore systems in solution. In these systems, there are sharp discontinuities in the dielectric constant that occur over relatively short distances (e.g., dielectric constant in the hydrophobic length

of the bilayer is around twenty times higher than the dielectric constant of bulk water surrounding the bilayer), so assuming a uniform dielectric constant can lead to inaccuracies in the calculation of electrostatic potential in some parts of the system.[66]

The intramolecular (bonded) terms in the potential energy function of a biomolecular system are made up of three key components: bond stretching, angle bending, and torsion (bond rotation).

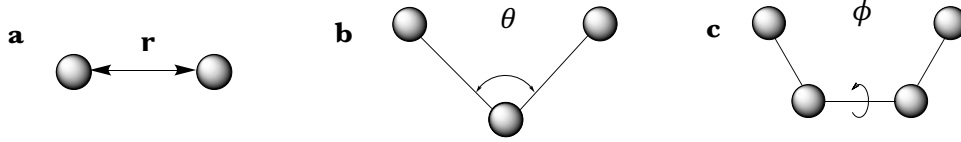


Figure 2.1: Illustrations of the three components of the intramolecular potential energy terms. **(a)** The bond stretching interaction **(b)** The angle bending interaction **(c)** The torsional interaction.

The sum of these interactions is summarised by the following expression:

$$U(\mathbf{r}_n) = \sum_{bonds} \frac{k^r_{ij}}{2} (\mathbf{r}_{ij} - \mathbf{r}_{eq})^2 + \sum_{angles} \frac{k^\theta_{ijk}}{2} (\theta_{ijk} - \theta_{eq})^2 + \sum_{torsions} \frac{k^\phi_{ijkl}}{2} (1 + \cos(n\phi_{ijkl} - \gamma)) \quad (2.4)$$

where k^r , k^θ and k^ϕ are spring, angle and torsion constants respectively, and \mathbf{r}_{eq} , θ_{eq} are equilibrium values. The torsional term $(n\phi_{ijkl} - \gamma)$ is a function of the torsion angle ϕ_{ijkl} , the number of minima and maxima between 0 and 2π (n) and the phase γ . The total potential energy function for a biomolecular system is comprised of intermolecular terms similar to those described in equations 2.2 and 2.3, summed together with the intramolecular terms shown in 2.4, and takes the following general form:

$$U_{total} = U_{bond\ stretching} + U_{angle\ bending} + U_{torsion} + U_{coulomb} + U_{vdw} \quad (2.5)$$

In MD packages, the exact form of this total potential energy function must be defined, along with complementary set of optimised parameters i.e. values for the various constants k^r , k^θ , σ , etc for all combinations of atom type in each class of biomolecule. The potential energy function and parameter set is referred to as a force-field.[58]–[60]

2.2 Force Fields

Biomolecular MD simulations can be performed over a range of resolution scales. The resolution scale is chosen in accordance with the process being studied and the available resources. In all-atom (AA) forcefields, the potential energy function and parameter set describe the interactions between individual atoms. In coarse-grained (CG) force fields, small groups of atoms are represented as ‘beads’, and the potential energy function and parameters describe the interactions between these beads.

2.2.1 All-Atom Force Fields

There are a number of AA forcefields that are applicable to simulations involving DNA. Of these forcefields, CHARMM and AMBER are the most widely used in the literature. The functional forms of the potential energy terms are usually very similar, but there are differences in the way that individual torsional, VdW and coulomb parameters are derived. Force field parameters for individual classes of biomolecules are usually derived using a combination of quantum mechanical (QM) methods, spectroscopic data, and experimental target data such as water/octanol partition free energies, interaction energies and macroscopic quantities of the bulk material e.g. densities and melting points. Macroscopic variables of interest from these simulations are compared to experimental data, and a range of validity is established. [58] Force fields are continuously under improvement, and there have been a number of studies comparing the results of one force-field against another, and against experimental data for various different DNA systems. [67]–[69] The CHARMM36 forcefield was chosen for the study presented in this thesis because it is implemented in both NAMD and GROMACS, and it allows novel residues to be parameterised with relative ease using the CHARMM General Force-field parameter set.

2.2.1.1 The CHARMM36 Additive Force Field

The earliest incarnation of the CHARMM force-field was delivered as part of the CHARMM Biomolecular Simulation Program [45]-[46], which has been in continuous development since the 1980s. The parameter table used in the program listed parameters for proteins and nucleic acids for use within the program exclusively. Eventually, this common parameter table developed into a set of separate CHARMM force fields for proteins, nucleic acids, lipids and small drug-like molecules, which were made available for implementation within other MD engines over the years. The CHARMM36 force field [68] (version 36) is an extensive set of parameters designed to describe highly heterogenous systems by uniting all of these separate forcefields, and is consistently undergoing development. The form of the potential energy function used in the CHARMM36 force field is given as:

$$\begin{aligned}
 U_{total} = & \sum_{\text{bonds}} k^r_{ij} (\mathbf{r}_{ij} - \mathbf{r}_{eq})^2 + \sum_{\text{angles}} k^\theta_{ijk} (\theta_{ijk} - \theta_{eq})^2 \\
 & + \sum_{\text{Urey-Bradley}} k^{UB}_{ik} (\mathbf{S}_{ik} - \mathbf{S}_{eq})^2 \\
 & + \sum_{\text{dihedrals}} k^\phi_{ijkl} (1 + \cos(n\phi_{ijkl} - \gamma)) \\
 & + \sum_{\text{impropers}} k^\omega_{ijkl} (\omega_{ijkl} - \omega_{eq})^2 \\
 & + \sum_{\text{non-bonded pairs}} \left\{ 4\epsilon \left[\left(\frac{\sigma}{\mathbf{r}_{ij}} \right)^{12} - \left(\frac{\sigma}{\mathbf{r}_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0 \mathbf{r}_{ij}} \right\} \\
 & + \sum_{\text{residues}} U_{\text{CMAP}}(\varphi, \psi)
 \end{aligned} \tag{2.6}$$

The first set of summed bond, angle and dihedral terms are explained in Eq. 2.4, and the non-bonded pair term is a combination of the LJ and Coulomb potentials (Eq. 2.2 and 2.3). The Urey-Bradley term is a quadratic function of the distance \mathbf{S}_{ik} between atoms i and k , which are separated by atom j ($i - j - k$), multiplied by the force constant of an imaginary spring between atoms i and k . It is used in addition to the

bond and angle terms in special cases, where it is needed to properly describe certain out-of-plane motions in order to successfully replicate the vibrational spectrum of the molecule – which is often a target during parameterisation of bonded terms within the CHARMM force-field. [46]-[47] The introduction of the spline-based numerical correction term for protein chains (CMAP) came as part of an improvement to the CHARMM22 protein parameters. It was developed through the use of *ab initio* QM calculations and structure-based potentials of mean force, to correct systematic errors in the description of protein backbone by the potential energy function, which would result in discrepancies (around 1 kcal/mol) between computed and experimental values for the free energy of hydration for each amino acid. [71] As force fields may differ in their functional forms or in their parameterisation schemes, it is generally not a good idea to mix parameters from different force fields.

The bonded parameters in the CHARMM force-field are usually based on equilibrium bond lengths and angles obtained from electron diffraction, X-ray diffraction and vibrational spectrum data for a selection of gas-phase and crystalline model compounds. The parameterisation scheme for intermolecular parameters is an iterative one; initial values for each initial parameter within the potential energy function are calculated using QM and MM methods for each model compound and then adjusted (either manually or automatically) until agreement with a varied spread of experimental data pertaining to the model compound is reached, over the course of many MD simulations. The earliest parameterisation of β -DNA within CHARMM heavily on data taken from crystallographic structures and QM calculations, and over the years it became apparent that the β II conformation of DNA was under-represented in simulations, which resulted in inaccurate calculated DNA-protein interactions that did not line up with experimental observations. The variation of phosphodiester torsions results in the emergence of two equilibrium β -DNA conformations (I and II), which have profound effects on the twist, roll and base-pair displacement of DNA – all factors that influence the binding of DNA to proteins. NMR data provides crucial information concerning the equilibrium populations of BI and BII in different solvent conditions, so these data became an important target in later iterations of the CHARMM DNA forcefields (CHARMM36), where the phosphodiester torsions were tuned to align with these observations.[68] The more experimental data available for validation, the more accurate the force field.

2.2.1.2 *The CHARMM General Force Field*

The CHARMM General force field is an organic force field developed to be compatible with the CHARMM36 biomolecular forcefield, to facilitate biomolecular simulations involving small drug-like organic molecules interacting with larger biomolecules e.g. drugs binding to protein receptors, or polymers interacting with lipid bilayers. Organic force field parameters are typically derived using different sets of experimental target data from those used for the derivation of biomolecular parameters, and the functional forms of these two classes of force field are usually dissimilar. CGenFF uses the same functional form as CHARMM36 and was parameterised with an emphasis of reproducing the vibrational spectra and crystal structures for a large set of model compounds that sufficiently describe the majority of chemical space and relies heavily on QM calculations. The CGenFF program can then be used to assign CGenFF parameters to input molecules by analogy, and the resulting parameters can be validated with QM calculations if necessary. Speed and simplicity are important features of its parameterisation scheme; this allows the user to expand the force field to a large number of new compounds with relative ease. [73]

The CGenFF program consists of a series of algorithms responsible for bond perception, atom typing, derivation of partial charges, and bonded parameter assignment, combined with penalty scoring. The CGenFF force-field contains a library of non-bonded and bonded parameters covering a broad range of chemical functional groups, which are assigned to the input molecule/fragment by analogy. In the first stage of parameter assignment, a decision tree is used to categorise the atoms of the input molecule/fragment into atom types. Each combination of adjacent atom types has a set of nonbonded parameters associated with them, and these parameters are assigned automatically during atom typing. The program then assigns appropriate bonded parameters from the force field by analogy. Partial charges then are assigned using a bond-charge increment scheme, starting with an initial assignment of a charge to an atom based on the identity of the atom. Charge increments associated with each bond, angle or dihedral the atom participates in are then added or subtracted to each atom's charge to give the final partial charge. Each bond parameter and partial charge assignment is given a penalty score to quantify the dissimilarity between the assigned

atom type and the actual atom type. Models generated by CGenFF will have a combined penalty score associated with them; a low score indicates that model is a good approximation of the molecule, and a high score advises that QM validation and further refinement of parameters are required.[74], [75]

2.2.2 *Coarse-Grained Force Fields*

The use of CG forcefields decreases resolution and degrees of freedom in a dynamic system by lowering the number of individual particles and flattening the potential energy landscape. This allows a larger time-step to be used, and so CG biomolecular simulations are much faster and more affordable than their AA counterparts, making microsecond-scale or ‘slow’ dynamics much more accessible with CG force fields. They are also used to overcome size issues, where the system is too large and complex to be simulated and sampled sufficient with AA methods. and small motions and individual interactions between atoms can be neglected e.g. simulations of entire viral capsids[76] or photosynthetic membranes in chloroplasts[77]. Additionally, the reduced cost of CG simulations force fields allows us to perform ensembles of simulations on average-sized biomolecular systems, providing more precise and reproducible results.

Parameterisation of CG force fields can be achieved using three different approaches; a bottom-up’, ‘top-down’ or mixed approach. Bottom-up approaches use structure based coarse-graining methods, in which parameters are extracted from reference AA simulations using inverse Monte-Carlo, force matching or iterative Boltzmann inversion methods.[78] Top-down approaches parameterize by calibration against experimental thermodynamic data, with a specific emphasis on reproducing specific thermodynamic properties. Mixed approaches use a combination of both bottom-up and top-down methods. [79]

There are many biomolecular CG force fields in use today. Some are used widely for a range of biomolecular simulation problems, such as the extremely popular MARTINI force field – which was designed to be applied to highly heterogenous systems, with a mapping scheme that allows the user to generate custom parameters for chemically modified biomolecules or small molecules. Some force fields are

parameterised and mapped to solve very specialised problems, or to analyse specific systems where general CG force fields such as MARTINI may not provide the right level of detail. The PaLaCe [80] force field is a pseudo coarse-grained forcefield designed to model the spontaneous folding and force-induced conformational changes of proteins, parameterised using conformational probability distributions generated from a large database of protein crystal structures. It uses a two-tier mapping scheme in which each residue is represented by three CG pseudo-atoms during calculation of non-bonded interactions, meanwhile the individual atoms of the side-chains and backbone are treated explicitly for the calculation of bonded interactions. This tiered approach that utilises different levels of granularity in one simulation is common to many force fields and is often referred to as multi-scale modelling.[81] For the CG studies described in this thesis, only a single level of granularity was used for all components of each simulation system, which were modelled with the MARTINI force field.

2.2.2.1 *The MARTINI Force Field*

The MARTINI forcefield is the most widely used CG force-field for biomolecular MD simulations. It features parameters for proteins, DNA, carbohydrates and lipids, making it ideal for simulating heterogenous systems. The earliest iteration of the force-field was designed for simulations of solvated lipids and used a standard 4-1 heavy atom-to-bead mapping. Each bead was designed to model a different chemical ‘building blocks’ with different charges, masses and thermodynamic properties. The intrinsic properties and nonbonded interactions of the beads were parameterised using a top-down approach, with a focus on successfully replicating partition free energies between water and organic solvents for a broad range of molecules. The CG bonded parameters (force constants and equilibrium values) were derived using a bottom-up approach, with manual tuning of the CG bonded terms until a satisfactory overlap between the CG bond/angle/dihedral distributions and the corresponding AA distributions was established. During this process, care was taken to ensure that the CG bond length/angle/dihedral values did not extend beyond the range of the AA distributions. [82]–[84] This is similar to the methodology used in the parameterisation of the GROMOS all-atom force-field [85], and the coarse-grained OxDNA model.

In the original MARTINI force-field, the beads were classified according to four ‘types’: polar (P), non-polar (N), apolar (C), and charged (Q). The bonded terms between chemically connected beads are modelled by weak harmonic potentials almost identical in form to those described in equation 2.4.[82] Non-bonded interactions between beads are described by the Lennard-Jones potential (2.2), and electrostatic interactions are modelled by the standard Coulomb potential (2.3). LJ interactions between chemically bonded beads (nearest neighbours) are typically excluded, but interactions between second nearest neighbours are not. The parameters associated with LJ and coulomb interactions of the various bead types are described by an interaction matrix, with different ‘levels’ of interaction: attractive (I), semi-attractive (II), intermediate (III) semi-repulsive (IV) and repulsive (V). Each level of attraction is defined by a different set of values for σ (distance at which potential is zero), ϵ (potential well depth in the LJ potential) and ϵ_0 (permittivity in the coulomb potential).[45]-[46] The introduction of special bead types (S and T prefixed) with reduced LJ parameters in later iterations of the force field facilitated the correct modelling of small ring structures. These reduced parameters dampen the repulsive forces between covalently bound beads, allowing them to come together more closely – resulting in more stable and compact models.

The modelling of water and ions makes many approximations. As these particles make up the vast majority of most CG biological systems, and interactions between biological polymers such as DNA and proteins are critical to their structure and function, it is important to understand how these approximations may affect the behaviour of the simulated system. MARTINI ions consist of a single charged (Q-type) interaction site, which represents the ion and the first hydration shell. As all ions are represented by the same bead type, there is no distinction between ions of different sizes. The standard MARTINI bead has a diameter of 5.2 Å, which is close to the diameter of a hydrated sodium ion (~ 5 Å, as determined by Monte Carlo simulations of water/ion systems[86]), thus the MARTINI model provides a faithful representation of sodium ions, but less so for smaller or larger ions. Under- or over-estimation of ion diameter may lead to an under- or over-representation of ion coordination to chelating biomolecules such as DNA.

The current version of MARTINI provides two water models, the original non-polarisable model[83] and the polarisable water (PW) model that was developed in 2010.[87] The original water model mapped four water molecules to one MARTINI water bead, which consisted of a single interaction site thus neglecting dipoles, opting for implicit electrostatic screening and a uniform dielectric environment. As the collective motions of individual dipoles in water generate powerful cooperative forces that drive protein folding, lipid membrane formation, DNA hybridisation and fraying, this simplification can lead to a gross misrepresentation of the of biomolecular systems. This necessitated the development of the PW model, which uses the same mapping (4 water molecules to one bead) but models the bead with three interactions sites (Fig 2.2) – one neutral, one with a charge of $+q$ and another with a charge of $-q$. These charged particles can rotate around the central interaction site in response to interactions with other charged particles within the system, altering the dipole momentum of each bead as they do so. Hence, the orientational polarisability of bulk water is accounted for with the PW model.[87]

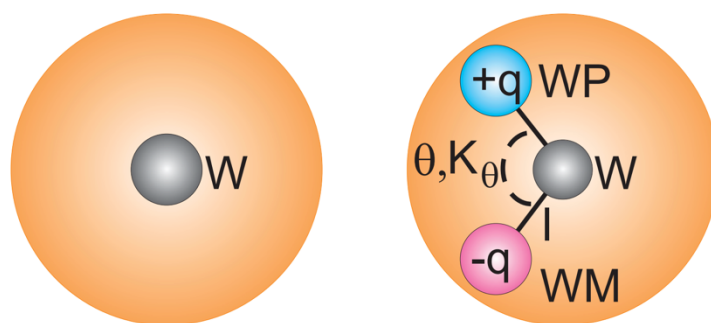


Figure 2.2: MARTINI models of water. The bead on the left illustrates the non-polarisable water bead, modelled by a single neutral interaction site, and the bead on the right shows the polarisable variant, modelled by three interaction sites. The distance l between the central W site and the charged WP/WM sites, the equilibrium angle θ and force constant K_θ are constrained to control the rotation of WP and WM particles and control the distribution of the dipole momenta. Image adapted from Yesylevskyy et al. [87]

While the PW model succeeded in facilitating the CG simulation of processes such as electroporation and solvent transport across ion channels with similar accuracy to AA simulations, it was established that the model does not sufficiently sample the range of dipole moments available to water molecules when compared to AA force-fields [87]. However, the motivation of the MARTINI model is to mirror the bulk properties of the system in question, and the PW model does this to an extent. The freezing temperature of polarizable MARTINI water is still too high (280-285 K), and surface tension (30 mN/m) too low. Also, the implicit hydration shell in CG ions results in a short-range repulsion between positive-negative ions pairs, but this is compensated for by increasing the magnitude of q . Nevertheless, the model reproduces the overall dielectric constant of water at 300K well, and the temperature dependence of these properties is also reflective of what is observed experimentally. The hydration and partitioning free energies of the various CG bead types in a water/hexadecane system and a water/DPPC bilayer system have been investigated extensively and discussed in the 2010 MARTINI paper on the PW model. The authors demonstrated that the model provides the level of electrostatic screening required to realistically describe the movement of charged particles from water to a low dielectric medium (membrane interior) at a level comparable to atomistic models.[87] It is worth noting here that atomistic water/ion models are not without their limitations. It is unclear how accurately polarisable AA water models can reproduce the distribution of dipole moments, and because of this a dielectric continuum is applied outside of the cutoff distance.

As a result, the long-range electrostatic interactions between ions and highly charged molecules are masked, and an artificial clustering of ions is sometimes observed. [87]–[89] The MARTINI force-field has been successfully employed to model the interactions between different classes of biomolecule, in a range of biological systems. In CG MD studies of proteins, carbohydrates and polymers embedded in lipid bilayers, the simulation data acquired with the use of the MARTINI force-field usually correlate very well with experimental data.[90]–[92] There have been very few MARTINI CG studies on a DNA/lipid systems to date[25], and those that do exist do not discuss the reproducibility or uncertainty of their results. Hence, we cannot have complete confidence in the reliability of the MARTINI force-field in such systems, so caution

must be taken when interpreting the data produced by these simulations, and appropriate validation of results against experimental data/atomistic data must be performed.

2.2.2.2 Parameterisation of the MARTINI DNA Force Field

The extension of the MARTINI forcefield to DNA necessitated the introduction of new ‘tiny’ beads (T-prefix) to correctly model the nucleobases and facilitate base-stacking, and these typically have a 2-1/3-1 atom-to-bead mapping, and interact with each-other with a reduced LJ parameter ($\sigma = 0.32$) while interacting with standard and S-type beads of Each tiny bead involved in hydrogen bonding was given a unique bead assignment, and their properties were tuned in a way that ensured that the energy of their interactions with each-other closely resembled the energy of the hydrogen bonds between those bases. To ensure the persistence length and double-helix structure of *ds*DNA is preserved throughout simulation, an elastic network of restraints was constructed – illustrated below. The elastic restraints can be thought of as flexible bonds that link each pseudo-atom to all surrounding pseudo-atoms within a certain cutoff distance. In the “stiff” elastic network, these restraints are applied to all beads within each nucleotide, which limits the flexibility of the base pairs and the backbone, whereas the “soft” elastic network allows for more flexibility in the motions of bases as these beads are not restrained.

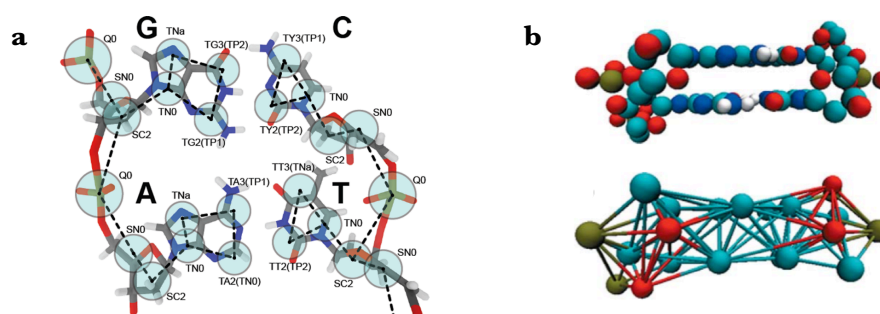


Figure 2.3: (a) Atom-to-bead mapping used in the MARTINI DNA force-field. Adapted from Uusitalo et al.[84] (b) Ball-and-stick representation of the stiff elastic network of restraints used to maintain the structure of base pairs in *ds*DNA. The elastic restraints are flexible bonds that link each pseudo-atom to all surrounding pseudo-atoms within a certain cutoff distance. Adapted from Khalid et al.[93]

The longest step in the parameterisation of the DNA model was the tuning of the non-bonded parameters. The parameters were tested by calculating free energies of partitioning (water to octanol, water to chloroform and water to lipid bilayer) for each CG nucleobase and comparing the CG values to both experimental and atomistic values.

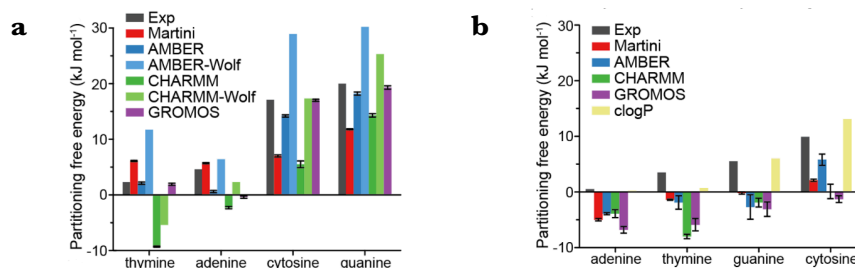


Figure 2.4: Partition free energies of nucleobases for (a) water-chloroform system and (b) water-octanol system. n-Octanol is used as a crude membrane-lipid mimic. Adapted from Uusitalo et al.[84]

While the MARTINI DNA model does not completely replicate the experimental partitioning free energies of the nucleobases, Figure 21 shows that the absolute error associated with the MARTINI CG partition free energies are comparable to those associated with the partition free energies yielded by AA forcefields. The water/membrane partition free energies for the MARTINI CG and various AA forcefields were all in excellent agreement with each other. Additional potential mean force (PMF) profiles were constructed for base-base interactions in water, to the energy landscapes of base stacking and base-pairing interactions. The calculated CG base-pairing and base-stacking energies were compared to their AA counterparts, and the base-stacking CG PMFs closely resembled the AA PMFs. On average, the CG base-pairing free energy values were lower, but they followed the same order as the AA energy values. While this is a considerable limitation, it is an inevitable effect of the granularity of the CG model – as increasing the base-pairing interactions would adversely affect the base-stacking energies. Overall, the MARTINI DNA model is robust enough to conserve the structure of dsDNA, and the top-down parameterisation of the beads with the use of water-lipid partitioning data makes this model suitable for modelling of DNA origami structures in the presence of lipid membranes.[79], [84]

2.2.2.3 Mapping AA Models to MARTINI

The MARTINI model allows CG systems to be built by mapping existing biomolecular AA structures to their coarse-grained representation using an auxiliary programme called *martinize.py*. [94] The program uses standardised mapping schemes for water, ions, lipids, proteins, DNA (Figure 2.2), allowing heterogenous to be built easily in an automated fashion. If a simulation system features non-standard residues or organic molecules, additional steps must be taken to generate MARTINI CG parameters for these custom system components. MARTINI parameterisation tools such as PyCGTool [95] provide an elegant solution; the user feeds in an atom-to-bead mapping scheme and a bonding scheme describing the connectivity of the beads, and once this information has been loaded the program parses the input reference AA simulation trajectories. Non-bonded parameters are assigned according to the mapping and bonding information provided by the user, and bonded parameters are extracted from the reference AA simulations, making use of the Boltzmann inversion method to derive coarse-grained bonded potentials. The output is a GROMACS compatible structure and topology file, ready for incorporation into larger simulation systems. The *insane.py* [96] script is another auxiliary tool commonly employed for the construction of heterogenous MARTINI systems – it allows the user to solvate their custom CG molecules and proteins/nucleic acid structures in water (standard MARTINI water model or the polarizable water model), and/or insert CG macromolecules into pre-equilibrated CG lipid bilayers membrane simulations.

2.3 Integration Algorithms

The most frequently employed algorithms for integrating the equations of motion are all based on the Verlet integration algorithm. It uses randomly assigned initial velocities generated from the Maxwell-Boltzmann distribution together with the force-field parameters to calculate the force on each particle, according to equation 2.6. The force is then used to calculate updated positions and velocities at the next timestep (Δt). [58], [78]

$$\mathbf{a}_i = \frac{d^2 \mathbf{r}_i}{dt^2} = \frac{\mathbf{F}_i}{m_i} = -\frac{1}{m_i} \frac{\partial U}{\partial \mathbf{r}_i} \tag{2.6}$$

There is no explicit velocity term within the Verlet algorithm, which leads to a loss of precision, and there is also a tendency towards truncation errors. To address these limitations, the leap-frog integration algorithm was developed. It is a variation of the Verlet scheme, and is used as the default integrator in the GROMACS MD code.[58], [97]–[99]

$$\mathbf{r}_{i+1} = \mathbf{r}_i + \mathbf{v}_{i+\frac{1}{2}}(\Delta t) \tag{2.7}$$

$$\mathbf{v}_{i+\frac{1}{2}} = \mathbf{v}_{i-\frac{1}{2}} + \mathbf{a}_i(\Delta t) \tag{2.8}$$

where \mathbf{r}_{i+1} is the position of particle i at time $t + \Delta t$, $\mathbf{v}_{i+\frac{1}{2}}$ is the velocity of particle i at time $t + \frac{1}{2}\Delta t$ and $\mathbf{v}_{i-\frac{1}{2}}$ is the velocity of particle i at time $t - \frac{1}{2}\Delta t$. One drawback of this algorithm is that the position and velocity of a particle at a particular time-step cannot be calculated at the same time. The NAMD engine makes use of the velocity-Verlet algorithm, which updates positions and velocities in the same step without loss of precision. [100] It is executed in three steps, which are summarised by the following two relationships:

$$\mathbf{r}_{i+1} = \mathbf{r}_i + \mathbf{v}_i(\Delta t) + \frac{1}{2}\mathbf{a}_i(\Delta t)^2 \tag{2.9}$$

$$\mathbf{v}_{i+1} = \mathbf{v}_i + \frac{1}{2}\{\mathbf{a}_i + \mathbf{a}_{i+1}\}(\Delta t) \tag{2.10}$$

2.4 Thermodynamic Ensembles

At present, biomolecular simulations are only capable of reproducing the behaviour of biomolecular systems at time and length scales in the nano/micro regime, purely through the calculation of microscopic quantities i.e. positions and velocities of individual atoms. From this data, we can extract information about the macroscopic behaviour of the system through the use of statistical mechanics.

The idea of the ‘ensemble’ in statistical mechanics is central to the correct interpretation and understanding of MD simulation data. An ensemble can be thought of as a collection of all the possible microscopic states of a system that give rise to the

same macroscopic or thermodynamic state. The microscopic state of a system is defined by the atomic positions and momenta, and these quantities can be considered as coordinates on a $6N$ -dimensional space known as a phase space. The macroscopic state of the system is defined by three mechanical properties of the system that can be controlled independently. This is usually some combination of: the number of particles (N), the volume (V), pressure (p), temperature (T) and internal energy (E). By keeping three of these variables constant, we define the ‘experimental conditions’ or ‘phase’ of the biomolecular simulation, and we define the ensemble. [61]-[62]

Ensemble Name	Fixed Variable	Conditions
Microcanonical	NVE	Closed system in vacuum
Canonical	NVT	Closed system in heat bath
Isothermal/Isobaric	NPT	Open system in heat bath

Table 2.1: Common ensembles used in MD simulations.[101]

An MD simulation will produce data that corresponds to different points on the phase space as a function of time, all within the same ensemble. For example, each replica within the NVT ensemble will have a different configuration, but they will all have identical values for N , V and T . The macroscopic variables of interest i.e. free energy, interatomic distances etc are as calculated as ensemble averages, according to the following general relationship:

$$S_{macro} = \langle S \rangle_{ensemble} = \frac{1}{N_e} \left(\sum_{i=1}^{N_e} S_i \right) \quad (2.11)$$

where S is the variable of interest, N_e is the number of sampled configurations of the system within the ensemble, and S_i is the value of S for the i th configuration.[85], [103] For each configuration, S_i is a function of the positions and momenta of the particles in the system. As it is not realistic to calculate the ensemble average of a variable as the average over *all* sampled configurations, we classify these configurations as being in a particular ‘state’ and calculate the probability ρ of the state from the number of configurations that exist in the particular state. Hence, the following expression

can be derived from equation 2.11:

$$S_{macro} = \langle S \rangle_{ensemble} = \left(\sum_{i=1}^{N_s} \rho_i S_i \right) \quad (2.12)$$

One important point to note is that for non-kinetic systems, the ensemble average of a variable is the same as its time average:

$$\langle S \rangle_{ensemble} = \langle S \rangle_{time} \quad (2.13)$$

where $\langle S \rangle_{time}$ is the time averaged over infinity. This is known as the ergodic hypothesis, and it is only truly valid for systems already in equilibrium. In the context of biomolecular MD simulations, calculating all the possible states of the system at a single point in time is not possible, and neither is simulating at an infinite timescale. For a complex biomolecular system approaching equilibrium (e.g. a protein undergoing conformational changes, or a ligand binding to a protein), the evolution of the trajectory is sensitive to the initial conditions. While such a system can be considered ergodic, this sensitivity to the starting conditions is ignored in the ergodic hypothesis. Two or more trajectories with the same starting configuration but different initial velocities may diverge exponentially with time, and this has been proved in numerous ensemble-based MD studies.[90], [104]–[107] Hence, it is inappropriate to assume that the results from a single trajectory are truly representative of the ensemble if it simulated for ‘long enough’. One must establish an appropriate simulation length; where the macroscopic variables of interest are permitted to evolve and approach equilibrium, and an appropriate ensemble size consisting of multiple ‘replicas’ with different initial conditions. An ensemble of sufficient size and duration will give a converged and therefore reproducible ensemble-averaged value for the macroscopic property of interest, with minimal errors. Ensemble-based protocols have been employed in several biomolecular MD studies for a broad range of problems, from calculating binding affinities of drugs to predicting inter-helix distances of G protein-coupled receptors in membranes. [90], [107]

2.5 Temperature and Pressure Coupling

The data collected from MD simulations of biomolecular/biological systems are only meaningful if the simulation conditions are reflective of real-life experimental conditions. Hence, it is common practice to run production simulations in either the canonical or isothermal/isobaric ensemble, and these require temperature and pressure control through the use of thermostats and barostats. Thermostats work by ensuring that the average temperature is equal to the desired temperature, whilst still allowing the total kinetic energy to fluctuate. More specifically, the rate of change of the particle velocities is controlled by modifying Newtons second law. This is often done through the use of the Langevin equation:

$$\frac{d^2 \mathbf{r}_i}{dt^2} = \frac{\mathbf{F}_i(t)}{m_i} - \gamma_i(t) \frac{d^2 \mathbf{r}_i}{dt^2} + \frac{\mathbf{R}_i(t)}{m_i} \quad (2.14)$$

where $\gamma_i(t)$ is the specified frictional damping coefficient and $\mathbf{R}_i(t)$ is the stochastic force experienced by particle i at time t , which is randomly assigned at each timestep.[59], [97] The Langevin thermostat is commonly implemented in the NAMD [100] molecular dynamics engine. The Berendsen weak-coupling scheme [108] uses velocity-rescaling to fix the average temperature to the desired value, and this forms the basis of the Berendsen thermostat used in MD engines such as NAMD and GROMACS. At each time-step, the algorithm multiplies the particle velocities by the following scale factor:

$$\lambda = 1 + \frac{\Delta t}{\tau} \left(\frac{T_0}{T(t)} - 1 \right) \quad (2.15)$$

where τ is the time-coupling parameter, which dictates the extent of temperature fluctuation. For larger systems, this thermostat produces roughly the same temperature fluctuations as what would be expected in the canonical ensemble, and hence it is often used for equilibration.[60], [109] The velocity-rescaling thermostat implemented in GROMACS is a modified version of the Berendsen thermostat, which adds an additional stochastic term to the kinetic energy distribution of the system, allowing accurate sampling of the canonical ensemble. [63][110] Just as thermostats couple the

system to a heat bath, barostats couple the system to a ‘pressure bath’. The pressure bath is introduced by modifying the equations of motion to include an additional degree of freedom. The pressure (P) is related to the box volume (V) and the inner virial (Ξ) of the system according to the Clausius virial theorem:

$$P = \frac{2}{3V}(E_k - \Xi) \tag{2.16}$$

where E_k is the kinetic energy of the system and the inner virial is defined by pairwise inter-particle distances and forces. [111] The Berendsen pressure-coupling algorithm[108] (supported by NAMD and GROMACS) rescales the box vectors and distances between individual particle coordinates at each time-step with a scaling matrix, which alters the virial, allowing the average pressure to be controlled throughout the simulation. This results in a first-order relaxation of the pressure towards to the desired pressure. As the Berendsen barostat doesn’t allow the box shape to change, and it cannot accurately sample the NPT ensemble, so it is often used for equilibration rather than production simulations. The Parrinello-Rahman barostat [112] is often used for production simulations, as it does yield a true NPT ensemble. It treats the box vectors individually, adding an additional degree of freedom, which allows the shape and size of the box to change throughout the simulation.[63]

2.6 Uncertainty Quantification in Molecular Dynamics Simulations

As discussed previously in Chapter 2.4, macroscopic properties of a simulated system are calculated as ensemble averages of instantaneous microstates – which may vary significantly over time, and across multiple replicas. Due to the deterministic nature of MD simulations, the initial velocities that are assigned to the particles in a system will dictate which microstates the system will occupy over time. At the same time, the temporal evolution of an individual trajectory is stochastic, meaning that neighbouring trajectories with different initial velocities are often divergent. Increasing both the ensemble size (i.e. the number of replicas) and the replica length will maximise the sampling of microstates. Therefore, we must determine a sufficient simulation timescale for each replica, and an appropriate number of replicas (N) to ensure a reproducible result with minimal uncertainty, such that running $N+1$ would not alter the behaviour significantly. When the ensemble-averaged value of a macroscopic

variable and its associated error remains stable despite further increases in the ensemble size, it is said to be converged.

An appropriate statistical protocol is required to quantify the errors associated with observable quantities that have been calculated from a simulated ensemble. I have employed the bootstrap method [82]-[83] for calculating the standard error of the mean, which provides a good estimate of the errors associated with ensemble-averaged macroscopic properties of the simulated systems. This statistical technique involves random sampling with replacement from a set of N data points generated from an ensemble of N simulations; a process which is repeated 10,000 times to generate 10,000 bootstrapped samples, each containing N data points. The mean of each bootstrapped sample is calculated, and the standard deviation of the distribution formed by the means is taken as an estimate for the standard error of the mean (SEM).[107] This statistical method has been utilised successfully in many ensemble-based all-atom MD studies [104]–[106], [115]–[118], where the typical simulation duration is fairly short (4 – 20 ns), but its use in CG studies at longer timescales has not yet been reported. In this work, I have achieved similar success to previous AA studies, in terms of error control and reproducibility.

2.7 Aims

The simulation work described in the following chapters aims to deepen our understanding of the structure, dynamics, membrane interactions and transport properties of the TEG-cholesterol anchored DNA nanopore (henceforth referred to as DNP) that was first published by Burns et al[38] in 2016 (shown in Fig. 1.16c). The design of this nanopore is much simpler than previous generations; it uses simple oligonucleotides of equal length rather than a combination of scaffold/staple strands of different lengths. It is also predicted to be smaller than previous designs, though a robust structural characterisation has not yet been achieved. Binding assays carried out using fluorescence microscopy [38] have shown that they bind preferentially to curved membranes rather than planar membranes, and that stable insertion into planar lipid membranes for single-channel current recordings is difficult to achieve without the assistance of a Triton-X surfactant and a buffer with a salt concentration (either KCl or NaCl) of ~ 1.0 M minimum. Some recently published tracking-and-localisation microscopy/transmission electron microscopy experiments[4] performed at lower salt

concentrations (0.3 M KCl) have shown that these DNPs are able to remodel planar DOPC bilayers to form hybrid DNA-lipid nanotubes. The binding mechanism of the nanopore was also studied with total internal reflection fluorescence and reflectance interference spectroscopy (TIRFS-RIf), and a two-step binding mechanism was proposed, as illustrated in Figure 2.5.

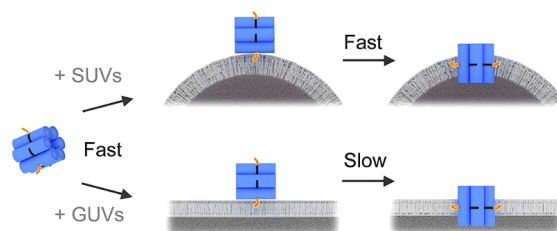


Figure 2.5: Proposed insertion mechanism for the TEG-cholesterol DNA nanopore. A rate constant of $\sim 1.5 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$ was calculated for the initial association of the nanopore with a polymer supported planar lipid DOPC bilayer. Initial tethering to the membrane surface is fast for highly curved membranes (SUV's) and low curvature membranes (GUV's), but the reorganisation/insertion step is markedly slower for low curvature membranes. Figure adapted from Burns et al. [3]

These observations of unpredictable behaviour of DNPs in the presence of lipid bilayers at different salt concentrations cannot be explained without a good understanding of their structural and dynamic properties in different salt conditions, both in bulk solution and in the presence of a lipid bilayer. The simulations described in Chapters 3 and 4 aim to explore these phenomena. Chapters 4 and 5 focus on the transport properties of membrane-spanning DNPs – specifically the conductance of ions and the transport of small-molecule fluorophores across the lipid bilayer.

Chapter 3

Coarse-Grained Simulations of DNA Nanopores in Aqueous Salt Solutions

Firstly, I will discuss the protocols I employed to build and equilibrate the initial all-atom and coarse-grained models of archetypal TEG-cholesterol anchored DNA nanopores (illustrated in Fig. 3.2), which have not been modelled previously. I will also discuss the results yielded from the two ensembles of CG simulations performed at two different salt concentrations (0.3 M NaCl and 1.0 M NaCl, two salt conditions that have been used in experimental studies[4], [38]) in aqueous conditions, in the absence of a bilayer. These simulations give us reliable and reproducible information on the structural properties of solvated DNPs in solution and provide insight into the effect that monovalent salt concentration has on the gross conformation, dynamics and structural integrity of DNP's in these conditions. Macroscopic quantities calculated from the 0.3 M NaCl simulation ensemble are then validated against an ensemble of shorter AA simulations in the same conditions, as well as cryo-EM experiments.

3.1 Model Building

Before any CG models could be built, an all-atom (AA) model of the DNP was constructed. Helices 1-6 (schematic shown in Fig. 3.1, sequences in Appendix 1) were built with the Nucleic Acid Builder module in AMBERTools [53]- [53], and arranged hexagonally with a ~ 2 nm inter-helix spacing using a python script deployed in PyMOL.[119] This arrangement was achieved by positioning the helices 2 nm apart, at an angle of 120° relative to the adjacent helices, and then rotating the helices such that the terminal base pairs of adjacent helices were also 120° apart, corresponding to the interior angle of a hexagon. Assuming 10.5 base pairs per turn, which corresponds to twist angle of 34.29° per base pair in a helix consisting of 21 base pairs, the two terminal base pairs of each helix would have an angle of 0° between them, as $21 \times 34.20^\circ$ gives a helix twist angle of 720° , which is exactly divisible by 360° .

The polyT inter-helix crossovers, TEG-C anchors and the covalent linkages between the TEG-C and the DNP were also made using PyMOL utilities. The TEG-cholesterol anchors and the covalent linkages to the DNP were parameterised within the CHARMM General Force Field (CGenFF V 3.0.1)[73]–[75], with the assistance of the CGenFF program (v.1).[120] Parameters are supplied in Appendix 2. NAMD v2.12 [100] was used for all AA simulations, in conjunction with the CHARMM36 + CGenFF force fields. The nanopore was solvated in a 35 x 35 x 40 nm box of TIP3 water molecules using the VMD *solvate* plugin, and the *autoionize* plugin was used to set the concentration of KCl to 0.3M. [121]

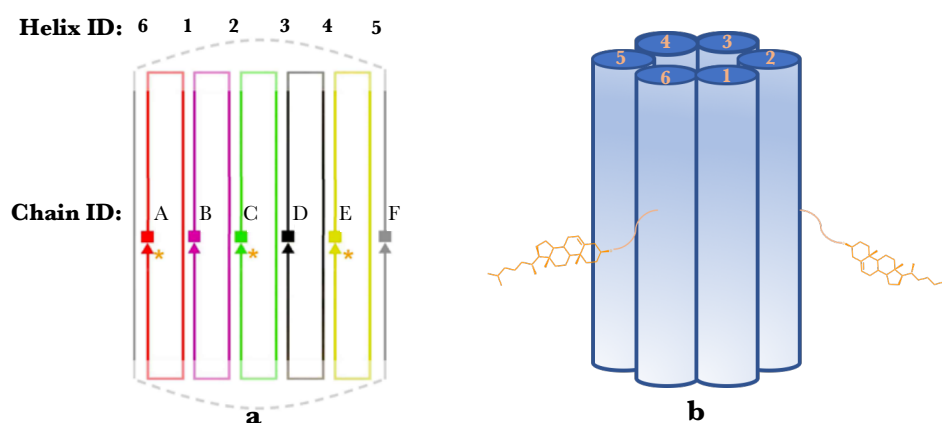


Figure 3.1: (a) CaDNAno schematic of the DNP. Orange stars denote the loci of the TEG-cholesterol anchors, and blurred hairpin loops represent the poly(T) crossovers. Strands A, B and C have TEG-C anchors covalently attached to the 3' ends, which are represented as arrowheads. Hence, the anchors are located on helices 2, 4 and 6. The 5' ends are represented as squares. Figure adapted from Burns *et al.* [3] (b) Illustration of the idealised six helix bundle structure of the TEG-C anchored DNP.

The system was minimised for 10,000 steps (2fs time-step) using the default conjugate-gradient algorithm, to remove the unfavourable interactions and relax strained bonds that originate from the artificial starting structure. The system was then equilibrated for 25 ns with positional restraints, to allow for a slow and stable relaxation of the artificial starting configuration. The force constant (k) of these restraints was gradually reduced throughout the equilibration period, starting with 1000 kJ/mol/nm² for the first nanosecond, 500 kJ/mol/nm² for the second nanosecond, 200 kJ/mol/nm² for

the following 3 nanoseconds, 100 kJ/mol/nm² for the next 10 nanoseconds, 10 kJ/mol/nm² for another 10 nanoseconds, and no restraints for the final 5 nanoseconds. The model was then subjected to another 50 ns of unrestrained dynamics, allowing the temperature and average RMSF of the DNP backbone to reach a steady state. The temperature was set to 300K with the Langevin thermostat, and isotropic pressure control at 1.013 bar was achieved using the Nosé-Hoover Langevin piston method. Electrostatic forces were calculated using the particle mesh Ewald (PME) algorithm, and a 1.2 nm cutoff was used for VdW and electrostatic interactions.

The PME method involves calculation of short-range electrostatic interactions in real space, combined with calculation of the electrostatic potential across the entire unit cell (in Fourier space) by assigning charges to a grid using interpolation. A 1.2 nm shifted cutoff was used for VdW interactions and short-range electrostatic interactions, to allow a slow decay to zero as the distance (\mathbf{r}) increases. [100]

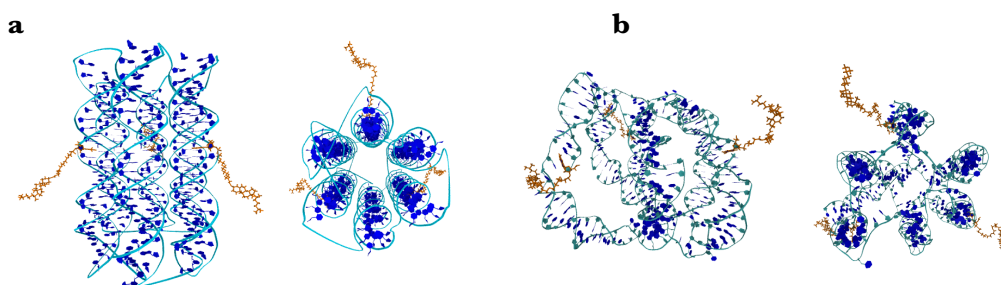


Figure 3.2: (a) 3D rendering (side view and birds eye view) of the DNP in its artificial starting configuration. (b) Snapshot of the DNP after 25 ns of equilibration.

The artificial relaxed-barrel configuration of the DNP used to build the AA DNP/POPC system was converted to its MARTINI[84] coarse-grained representation using the *martinize-dna.py* script. [76] The script automatically generates an elastic network of restraints optimised to maintain the base-pairing, general structure and persistence length of double-stranded β -DNA. The ‘stiff’ elastic network was used for all CG simulations. MARTINI parameters for the TEG-C anchors were generated with the PyCGTool program developed by Graham et al [95], using the AA to CG mapping scheme illustrated in Fig. 3.3a.

Radius of gyration (R_g) data for the CG TEG-C anchors were generated from a set of 6 unrestrained CG simulations of the DNP in solution (using protocols described in Section 3.2), was compared to the (R_g) data taken from 6 unrestrained 30 ns AA reference simulations of the solvated DNP. R_g data provides an indirect measure of the overall conformation adopted by small molecules, making it a convenient metric for validation of the CG anchor parameters against AA data. For each anchor, the percentage difference between the total R_g from the AA reference simulation and the total R_g from the CG simulation was under 5%, indicating that the CG parameters gave a good representation of the dynamic behaviour of the AA TEG-C anchors.

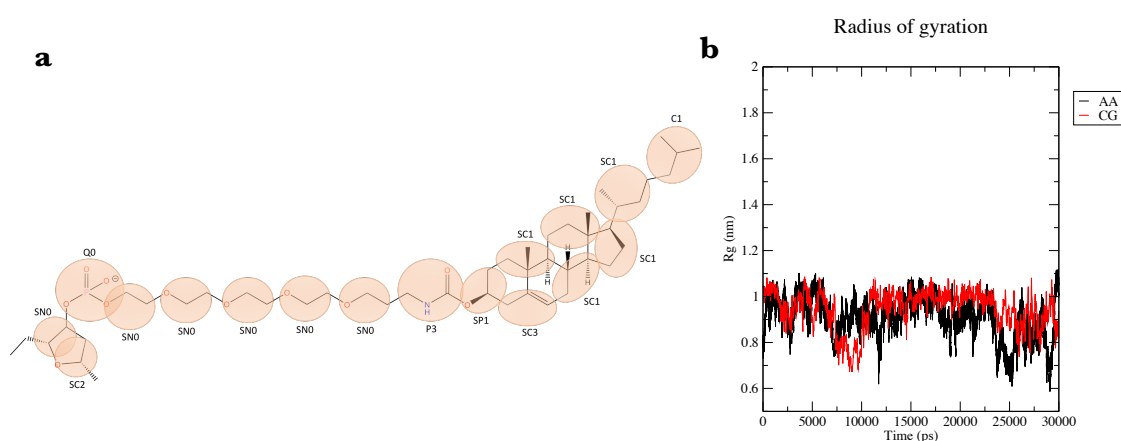


Figure 3.3: (a) AA to CG mapping for the TEG-C anchor, with MARTINI bead assignments. (b) Total radius of gyration vs time for a TEG-C anchor.

To study the influence of aqueous solvents with different ionic strengths on the macroscopic properties and dynamics of the DNP, two CG models were built – one representing low salt conditions and one representing high salt conditions. For the low salt model, the CG DNP structure was centred in a box of polarisable PW MARTINI water molecules with NaCl concentration set to 0.3M using the *insane.py* script.[96] The high salt model was built with the same script, but with NaCl concentration set to 1.0 M NaCl. The two solvated models consisted of $\sim 90,000$ pseudo-atoms and had overall cell dimensions of 18 nm x 18 nm x 18 nm. These dimensions were chosen to ensure that the cholesterol anchors did not interact with each-other across the periodic boundaries, as each anchor is around 5 nm long at full extension and any tilting, breathing motions or lateral movement of the nanopore within the simulation box result in artefacts due to self-interactions.

3.2 Simulation Details and Analytical Methods

All CG simulations were performed with GROMACS v5.1.4. Temperature control was achieved with the use of the velocity-rescaling thermostat developed by Bussi et al[110] with a time constant of 1 ps. Two different barostats were employed for pressure control (at 1.013 bar) during NPT simulation phases; the Berendsen thermostat for early equilibration (with time constant of 2.0 ps) and the Parrinello-Rahman barostat[112] for production/data collection. Two separate temperature and pressure coupling groups were defined – one for the water and ions and one for the DNP and its cholesterol anchors. Replica production simulations were run by generating separate run files with different random velocity seeds. Velocities, energies, forces and coordinates were recorded every 40ps, corresponding to 25 ‘snapshots’ per ns. The maximum distance allowed for bonded interactions with domain decomposition (rdd) was adjusted throughout the production simulations to facilitate the use of a 5-fs timestep, and typically remained in the range of 1.5 -1.9 nm. The standard 1.1 nm cut-off was used for electrostatic and VdW interactions.

Once the solvated DNP models were built, they were subjected to 5,000 steps of minimization using the steepest-descent algorithm before equilibration. A 2 fs timestep was used for the first 10 ns of equilibration time, as the initial relaxation of system involves drastic conformational changes which result in large forces being calculated at each time step. For first phase of the equilibration (in the NVT ensemble), strong positional restraints were applied to the beads of the DNP. The force constants of the restraints were gradually reduced to zero over the course of ~4 ns. The temperature was held at 300K throughout the equilibration. The next 16 ns of equilibration was performed without restraints in the NPT ensemble with the Berendsen barostat[108], after which timestep was increased to 5 fs for a further 10 ns of equilibration. For the final phase of equilibration (lasting 50 ns) the Parrinello-Rahman barostat[112] was employed with a time coupling constant of 12 ps. Production simulations were then run for 500 ns apiece, with a 5 fs timestep and with temperature and pressure held at 300K and 1.01 bar, respectively.

The average pore height was calculated as the arithmetic mean of the 6 helix lengths. Specific pairs of oppositely-facing DNA backbone beads at the termini of each double helix (6 in total) were selected, and the distance between these pairs was calculated using the `gmx_distance` tool in GROMACS.[70] [82] To facilitate the calculation of root-mean-squared fluctuation (RMSF), pore width and pore lumen profiles, each helix composing the DNP was split into eight distinct regions, illustrated below. The outer pore width for each region was calculated by taking the average of the maximum distance between opposing helices within the relevant regions — of which there are three (Figure 3.5). The lumen width for each region was approximated in a similar fashion, by taking the average of the minimum distance between opposing helices. The RMSF for each region was calculated using the `gmx_rmsf` tool.[63]

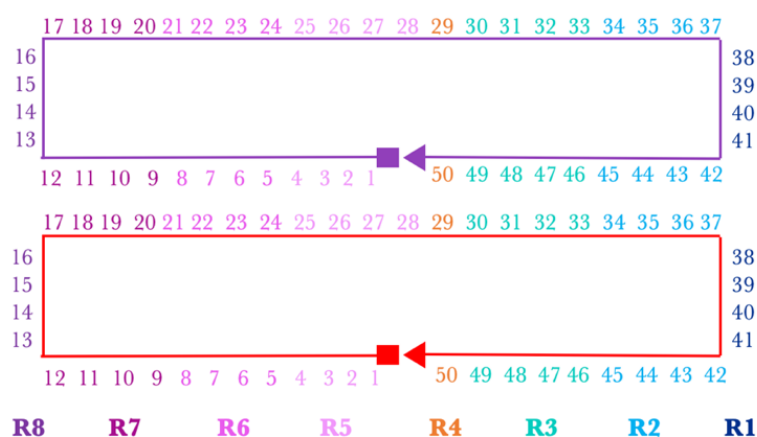


Figure 3.4: Splitting the DNP into eight distinct pore regions, illustrated on a schematic of helix B, (comprised of strands 1 and 2). The pore termini are comprised of the unpaired poly(T) crossovers (four nucleobases per strand) and singular base-pairs at the strand termini; three of which bear the TEG-cholesterol anchors, make up R4. The remaining regions are each composed of four consecutive base-pairs.

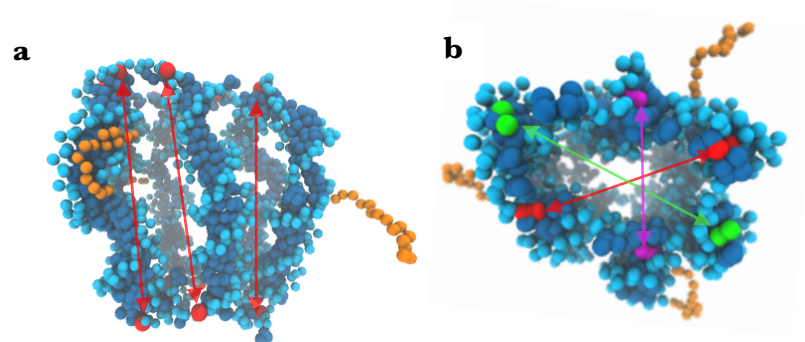


Figure 3.5: (a) Illustration of the index groups used for the calculation of average pore height. Three of the six bead pairs used for the helix length calculations are shown in red. (b) The three sets of opposing helices used to calculate average pore widths and lumen widths at each pore region (R1-R8), which were used to generate pore width and lumen width profiles.

The percentage of base pair breakage was calculated in VMD with a custom Tcl script, which loops through all frames within each trajectory and counts the intact hydrogen bonds between base pairs within the DNP. An intact hydrogen bond is defined by the same geometric criteria in most analysis packages, such as the VMD HBond plugin[120] and MDAnalysis[121]. The distance between the hydrogen bond acceptor and donor (in adenine-thymine and guanine-cytosine base pairs) must be less than or equal to 0.3 nm, and the angle between the donor and acceptor must be greater than or equal to 140° . The script then subtracts the number of intact base pairs from the total number of base pairs possible in the given selection (helix, region or entire DNP) to yield the number of broken base pairs, and this is reported as a percentage of the total number of base pairs in the selection. These distance and angle criteria work well in all-atom systems where the hydrogen bonds occur between explicit hydrogen and oxygen/nitrogen atoms, but are less suitable for MARTINI CG systems, where the hydrogen-bond donor and acceptors are modelled by larger pseudo-atoms. Hence, the cutoff distance of 0.3 nm for classification of hydrogen bonds is too small for MARTINI DNA systems, so I used a cutoff of 0.5 nm as this enabled better discrimination between broken and intact base pairs. The equivalent calculation of base pair breakage for the AA ensemble described in Section 3.5 used a 0.3 nm cut-off and an angle of 140° .

Average dwell times for cholesterol anchors within the grooves of the DNA helices was approximated by calculating the minimum distance between the centre-of-mass of the cholesterol portion of the TEG-cholesterol anchors and the side-chain pseudo-atoms of the helices as a function of time for each simulation. The timesteps at which the minimum distance was between 0.3 nm and 0.6 nm (the typical range of interaction distance for vdW forces) were counted, and the average gross dwell time was calculated by summing these timesteps.

Helix kink angles were calculated using the *gmx bundle* analysis tool. Prior to analysis, the region containing the kink angle centre was identified for each helix. In all six helices, the centre of the kink angle originated at the cusp of R4 and R5 in all replica simulations. Pseudo-atom indices of the R4/R5 region were defined as the “kink centre” group, from which the two vectors emerge to make the angle. Pseudo-atoms making up region R2 was defined as the “kink-top” group, and R5 was defined as the “kink-bottom” group. During the kink angle analysis, *gmx bundle* fits vectors around the kink angle centre using the defined index groups and calculates the angles between them at each timestep.

The first phase of the study for the solvated DNP simulation systems was dedicated to establishing an appropriate ensemble size. By monitoring the decay of the bootstrapped standard error associated with the average pore height and width, while increasing the replica number and length, I have determined that 12 replicas run for *at least* 500 ns apiece are required to provide robust, reproducible results for the aqueous DNP models, as the averaged pore height and pore width did not change significantly after 12 replicas. A total of 15 replicas were performed for each ensemble, to facilitate a thorough analysis of the convergence of macroscopic properties of interest. The ensemble average plots shown in Figure 3.6 overleaf show how the bootstrapped standard errors decay rapidly when increasing the number of replicas before levelling off after around 10 replicas – suggesting that this is the minimum number of replicas of this duration required for fully converged ensemble averages.

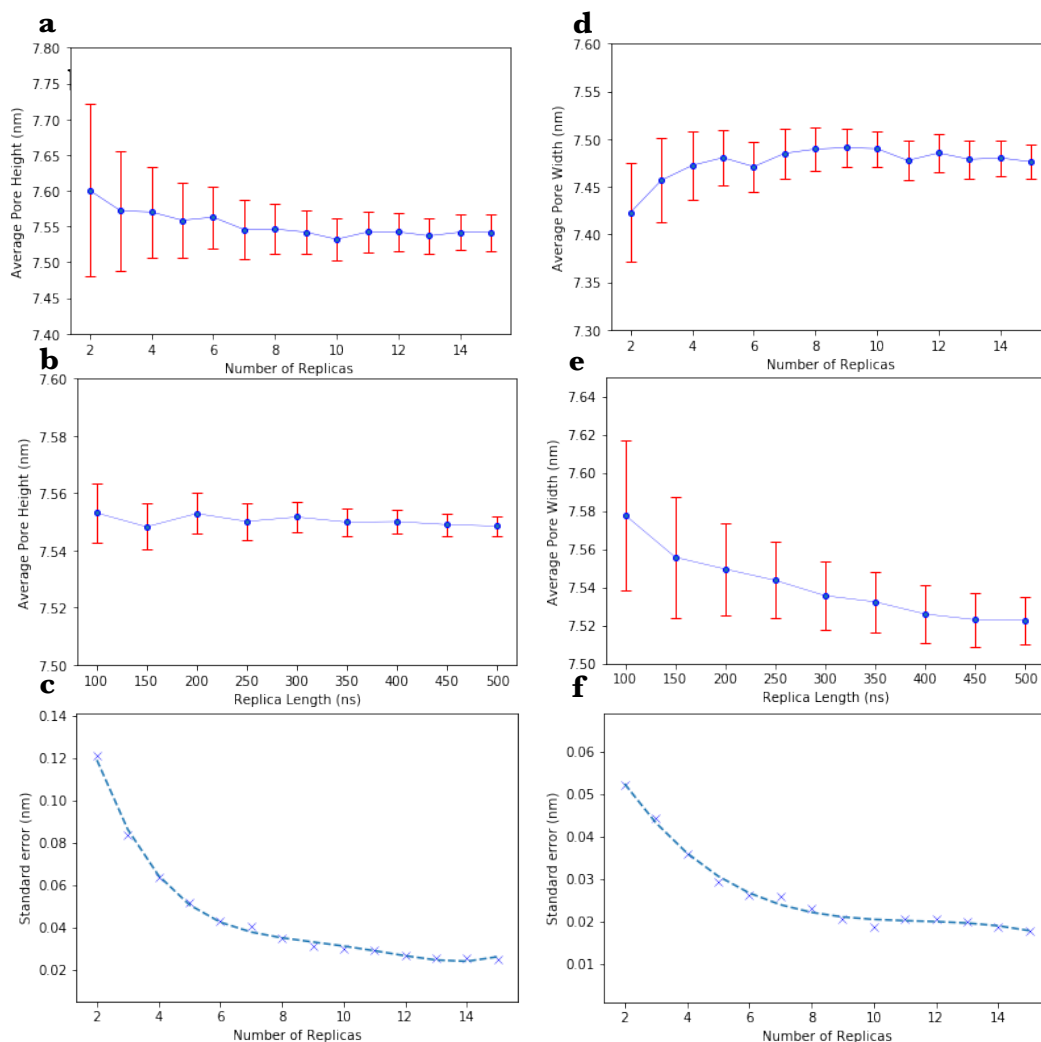


Figure 3.6: Ensemble average plots illustrating the convergence of pore dimensions and the associated bootstrapped standard errors in the solvated DNP simulation ensembles. Error bars correspond to the bootstrapped standard error. **(a)** Convergence of the average pore height as a function of replica number for the 1.0 M NaCl model demonstrates that 10 replicas are sufficient to ensure reproducible results for these systems. **(b)** Convergence of the average pore height with simulation time for the 1.0 M NaCl model shows that 500ns is an appropriate duration for the full set of 15 replicas. **(c)** Decay and subsequent plateau of the standard error associated with the pore height for the 1.0 M NaCl model suggests that 10 replicas were sufficient for control over errors. **(d)** Convergence of the average pore width as a function of replica number for the 0.3 M NaCl model. **(e)** Convergence of the average pore width as a function of simulation time. **(f)** Decay of the bootstrapped standard error associated with the average pore width in the 0.3 M NaCl model.

3.3 Benchmarking and Computational Resources

In order to approximate the level of computing resources that would be required for these lengthy simulations, strong scaling tests were performed for one of the CG solvated DNP simulation systems (running with GROMACS) on two HPC platforms; the UK national supercomputer ARCHER[123] and the Dutch national supercomputer Cartesius. [124] For an optimal balance between speed and cost, I chose to run each individual simulation on 360 CPU cores, as the simulation box was too small to be divided amongst more cores than this. The thirty 500 ns - long simulations of the aqueous DNP models were completed on ARCHER, and these consumed 325,000 core hours in total. The computational resources used for this work were provided by the CompBioMed Centre of Excellence.[125]

3.4 Results and Discussion

Here, I interpret and compare the results of the solvated DNP simulations. In the absence of a membrane, the DNP is an anisotropic globular structure without a distinct pore lumen. It adopts a highly bloated conformation in solution, with regions of constriction at the pore termini - where the inter-duplex poly(T) crossovers are located. Increasing the salt concentration from 0.3 M to 1.0 M had little effect on the average pore dimensions; in both cases the average pore height remained stable at around 7.6 nm, and the pore width at around 7.5 nm. Minimum pore width values were calculated for both models (Table 3.1 and Figure 3.7), and in both sets of simulations, the minimum values were clustered at the pore termini; specifically, the R8 and R1 regions.

The effect of increasing the salt concentration on the average width of the DNP in solution is very modest due its intrinsic asymmetry and structural lability, meaning that a constriction in one dimension causes an expansion in another dimension. For this reason, I calculated three widths for each region – corresponding to the distance between three sets of opposing helices (Figure 3.5b). The calculated average pore dimensions at 0.3 M (Table 3.1) are in fairly good agreement with the predicted dimensions (9 nm by 5 nm), and with experimental values reported from atomic force microscopy (AFM) experiments performed in 0.3 M KCl buffer solution. It should be noted that the predicted values do not account for the effects of the solvent or salt

concentration, and that AFM measurements of these delicate DNA nanostructures are prone to tip convolution effects, as was discussed previously in Section 1.4.2.[39] In both sets of simulations, we notice a profound and persistent difference in the minimum widths of the two termini of the DNP. The minimum width at the R8 terminus is ~ 1 nm narrower than the minimum width at the R1 terminus, giving rise to a fixed polarity in pore/lumen width down the central pore axis. This asymmetry may play a key role in the translocational properties of the DNP (discussed in Chapter 6).

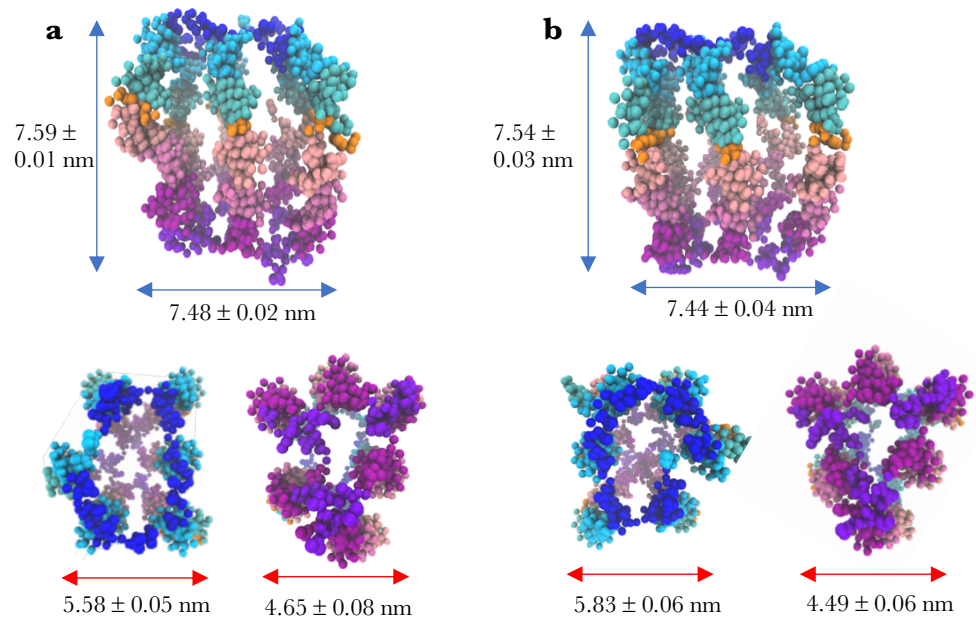


Figure 3.7: Solvated DNP dimensions in two different salt conditions. **(a)** Average height and width of the DNP in 0.3 M NaCl (indicated by the blue arrows, shown above), and the widths of the constrictions at the termini, which correspond to the mean of the calculated minimum pore widths at R1 and R8 (indicated by red arrows, shown below). **(b)** Average height and width of the DNP in 1.0 M NaCl, and the widths of the constrictions at the termini.

	0.3 M NaCl	1.0 M NaCl	AFM
Average Pore Height (nm)	7.59 ± 0.02	7.54 ± 0.06	9.0 ± 1.5
Average Pore Width (nm)	7.48 ± 0.04	7.44 ± 0.08	5.1 ± 1.1
Average Pore Width R1 (nm)	6.59 ± 0.04	6.54 ± 0.08	n/a
Average Pore Width R8 (nm)	6.11 ± 0.02	6.06 ± 0.02	n/a
Minimum Pore Width R1 (nm)	5.58 ± 0.10	5.83 ± 0.12	n/a
Minimum Pore Width R8 (nm)	4.65 ± 0.16	4.49 ± 0.12	n/a

Table 3.1: Comparison of the calculated pore dimensions of the solvated DNP models. Experimentally derived average dimensions are presented for validation. All calculated values correspond to the mean \pm the bootstrapped 95% confidence interval obtained from an ensemble of 15 trajectories.

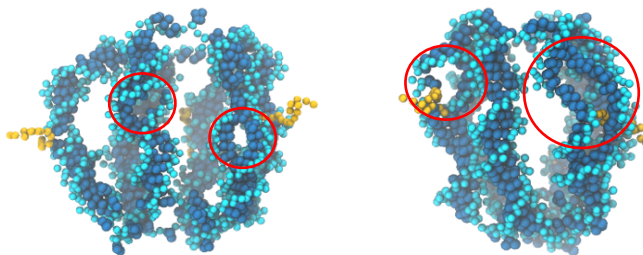


Figure 3.8: Representative structures of the DNP in 0.3 M NaCl (left) and 1.0 M NaCl (right) exhibiting partial fraying of the *ds*DNA duplexes.

While the overall pore dimensions are more or less unchanged upon increasing the NaCl concentration from 0.3 M to 1.0 M, the RMSF-per-residue plots (Figure 3.9) reveal a subtle change in the structural fluctuations of the pore, particularly in the R1-R4 region where the duplexes are shorter (9 base-pairs long). In the low-salt simulations, the degree of fluctuation is relatively uniform throughout the entire length of the nanopore, with some sparse clustering of extreme values (greater than the median 0.3 nm) in the central region of the nanopore. In the high-salt simulations, there is a pronounced reduction in the extent of fluctuation in the R1 region. We also observe a higher incidence of extreme values in the upper-central region (R2-R4) of the nanopore, and a lower incidence of extreme values in the lower-central region (R4-R6), suggesting that the longer duplex sections located below the nicks are stabilised at the expense of the shorter duplex sections, which are located above the TEG-cholesterol anchors.

The extreme duplex fraying within this region observed in the 1.0 M NaCl simulations (illustrated in Figure 3.8) supports this hypothesis, and the relatively high percentage of broken base pairs (% BPB) shown in Table 3.2 in the R1-R4 region (47.3 ± 1.52 %) provides further evidence of helix destabilization in this region at 1.0 M NaCl.

It is important to note that shorter helix sections (R1-R4) of helix 2, 4 and 6 are covalently connected to the TEG-cholesterol anchors, which are highly mobile within the solution. The RMSD vs time plots shown in Figure 3.10 indicate that the average RMSD of the cholesterol anchors evolves sinusoidally with time, and the magnitude of their average RMSD is consistently higher than the RMSF of the main body of the DNP. The forceful motions of the connected cholesterol anchors are likely to be a contributing factor the helix fraying exhibited in the shorter duplex sections. In addition, the freely-moving hydrophobic TEG-cholesterol anchors frequently make close Van-der-Waals contacts with the hydrophobic grooves within the DNA, which may disrupt the hydrogen bonding between base-pairs. In order to identify the pore region that the cholesterol anchors interacted with most frequently, the average total duration of (non-consecutive) simulation time in which the distance between the anchors and the grooves of the R1-R4 region was less than 0.6 nm was calculated, and the same was done for the R5-R8 region. The 0.6 nm cut-off is analogous to the upper boundary of the range of Van-der-Waals interaction distances. At both salt concentrations, the dwell time of the anchors within the R1-R4 grooves is significantly higher than the dwell time within the R5-R8 grooves (Figure 3.10). The difference in the dwell times in the 1.0 M ensemble is significantly more pronounced than it is in the 0.3 M ensemble; the duration for which the cholesterol anchors occupy R1-R4 is more than twice the duration they occupy R5-R8, indicating a strong preference for the shorter helix sections, which may further contribute to their fraying and increased RMSF-per-residue values.

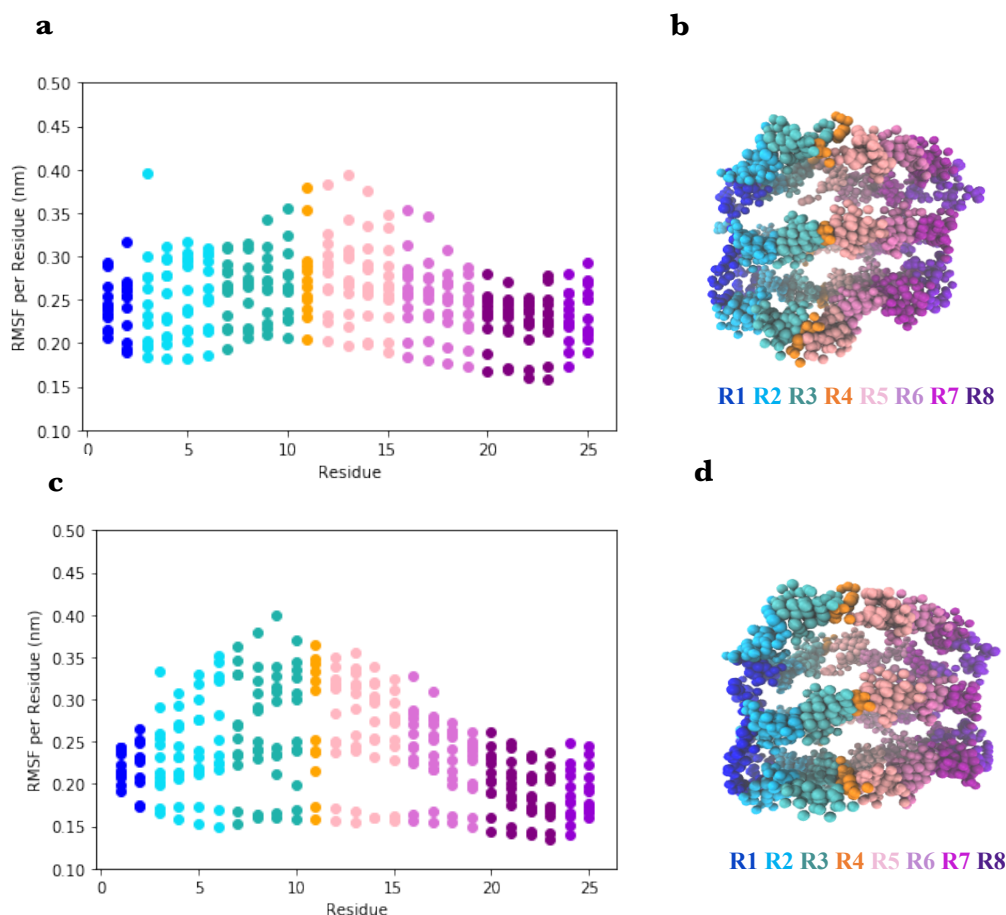


Figure 3.9: Structural fluctuations of a solvated DNP. **(a)** Root mean square fluctuation (RMSF) per residue plot for the DNP in 0.3M NaCl, generated from an ensemble of 15 x 500 ns trajectories. The median RMSF value is 0.25 nm. **(b)** Average structure of the DNP in 0.3 M NaCl (anchors omitted). The length of each DNA duplex corresponds to 21 base pairs (not including the poly(T) crossovers), which have been grouped and colour coded according to their position along the pore axis. The helix sections located above the nicks (R2-R4) are 9 base-pairs long, whereas the helix sections below the nicks (R5-R7) are constituted from 12 base-pairs. The orange pseudo-atoms represent nucleobases that are directly bound/adjacent to TEG-cholesterol anchors. **(c)** RMSF per residue plot for the DNP in 1.0M NaCl. The median RMSF value is 0.23 nm. **(d)** Average structure of the DNP in 1.0 M NaCl.

Further % BPB and RMSF analyses were performed to determine which individual helices display more mobility in general, which ones are particularly prone to fraying in either salt condition and how this might influence the mechanical properties of the DNP. Surprisingly, only three out of the six helices had over 50% of their base pairs intact at either salt concentration, and the identity of these helices is highly dependent on the concentration of NaCl (Table 3.2). The conformation of the DNA throughout the pore clearly deviates from the expected β -DNA structure in the presence of NaCl, across a range of concentrations, with helices 1, 3 and 6 experiencing the most disruption at the lower salt concentration. It is worth noting here that the MARTINI force field for DNA has been shown to underestimate the free energy associated with base-pairing[84], so the absolute values of the percentage of base-pair breakage reported here are likely to be exaggerated. This is discussed further in Section 3.5.3.

	0.3 M NaCl	1.0 M NaCl
R1-R8	39.0 ± 2.48	42.1 ± 1.29
R1-R4	30.2 ± 3.40	47.3 ± 1.52
R5-R8	45.7 ± 2.08	38.2 ± 1.71
Helix 1	60.4 ± 2.95	4.80 ± 1.15
Helix 2	5.58 ± 1.66	86.6 ± 0.46
Helix 3	67.8 ± 1.75	12.8 ± 0.21
Helix 4	16.4 ± 2.00	54.3 ± 0.80
Helix 5	44.0 ± 3.31	40.8 ± 0.83
Helix 6	53.6 ± 3.37	82.0 ± 0.94

Table 3.2: Percentage of broken base pairs within the DNP. At the higher salt concentration, the helix fraying is more prominent in the shorter (9 base-pairs long) helix sections located above the nicks (R1-R4), whereas the longer helix sections (12 base-pairs long) below the nicks (R5-R8) experience more fraying at the lower salt concentration. Errors are reported as the bootstrapped 95 % confidence interval.

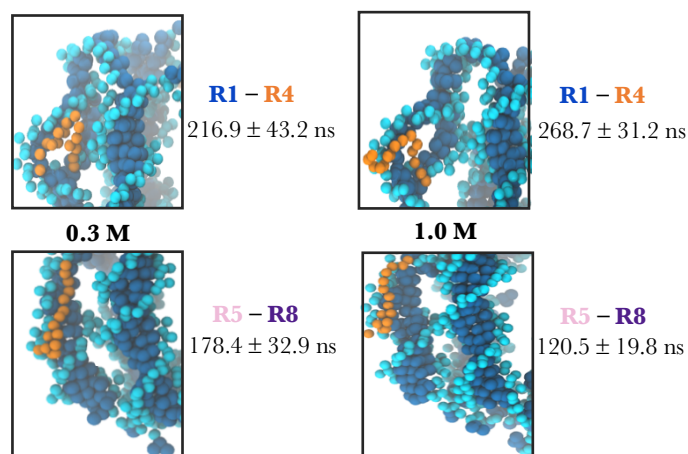


Figure 3.10: Average total dwell time of TEG-cholesterol anchors within the DNA grooves of the R1-R4 region, and the R5-R8 region for both salt conditions. The higher salt concentration appears to increase the propensity for the cholesterol anchors to reside within the R1-R4 grooves rather than the R5-R8, which coincides with an increase in %BPB in the former region a decrease in %BPB in the latter. Errors are reported as the bootstrapped 95 % confidence interval.

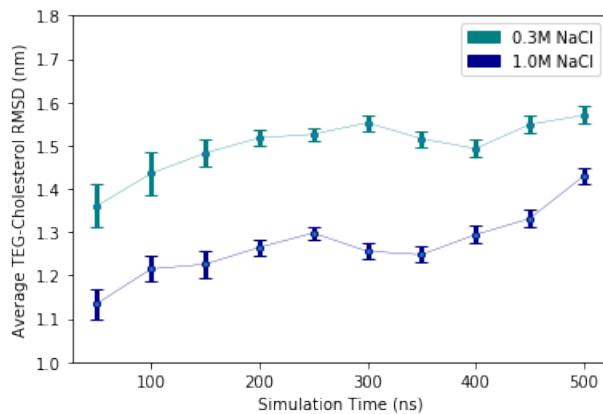


Figure 3.11: Plot of the RMSD of TEG-cholesterol anchors vs time in both salt concentrations. Error bars represent the bootstrapped standard error associated with each RMSD value.

The RMSF-per helix plot shown in Figure 3.12 reveals a fixed asymmetry in the dynamics of the helices in 0.3 M NaCl; where one “half” of the 6-helix bundle (helices 1, 2 and 3) are significantly less mobile than the other half (helices 4, 5 and 6), as evidenced by their lower RMSF values (Figure 3.12a). The higher mobility of helices 4, 5 and 6 suggests that it may be these helices that modulate the dimensions and dynamics of the pore lumen at this salt concentration, as they are capable of moving closer together to restrict the lumen, or further apart to expand it. In 1.0 M NaCl, the distribution of mobile helices is more evenly distributed, with helices 2, 4 and 6 experiencing the highest RMSF. It may be that lumen width is modulated by symmetrical contractions of these helices towards the DNPs centre of mass at higher salt concentration. It is highly likely that relatively high mobility of the higher-RMSF helices in both salt conditions are the result of lateral/axial motions primarily, and base-pair breakage to a lesser extent.

The base-pair breakage data shown in Table 3.2 illustrates the profound effect of salt concentration on the % BPB of each helix, with helices 2, 4 and 6 experiencing dramatic increases in base pair breakage, while base-pair breakage in helices 1, 3 and 5 decreases substantially. Overall, the average % BPB of the higher RMSF helices in 0.3 M NaCl (4, 5 and 6) decreases in response to increasing ionic strength, and an increase in base-pair breakage occurs for the lower RMSF helices. However, the average RMSF of all helices but helix 2 decreases by ~ 0.2 nm in response to the increasing ionic strength, so there is no clear correlation between increasing % BPB and increasing RMSF. Tracking the evolution of broken base pairs within the DNP as a function of time, we see the % BPB increases very gradually over time at both salt concentrations, indicating that base pairs have a tendency to slowly drift apart over time, rather than spontaneously breaking and reforming over the course of the simulations. Hence, we can attribute the majority of the individual helix mobility to large-scale lateral and axial motions, rather than helix fraying alone.

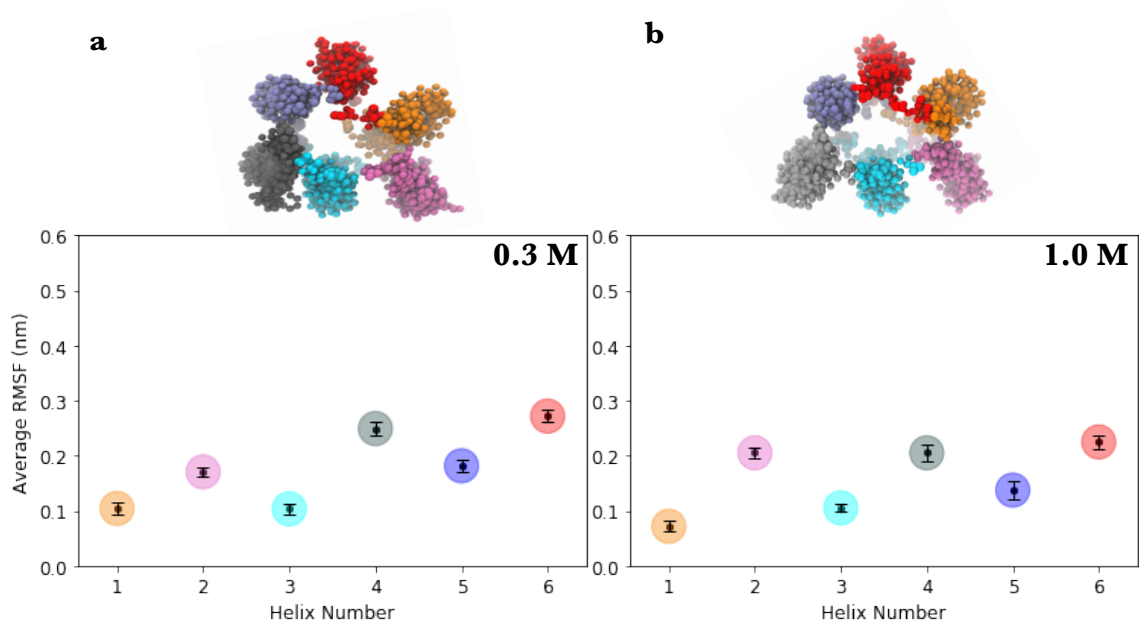


Figure 3.12: Comparison of the average RMSF of each helix between the two salt concentrations, presented with colour-coded top-view images of the averaged DNP structures. The average RMSF of the helix is taken as the mean of the RMSF of all CG pseudo-atoms within the helix. **(a)** RMSF per helix for the 0.3 M NaCl system. The lower RMSF of helices 1, 2 and 3 compared to helices 4, 5 and 6 indicates that one half of the pore is more structurally rigid than the other at this concentration of NaCl. **(b)** RMSF per helix for the 1.0 M NaCl system. Helices 2, 4 and 6 exhibit the highest average RMSF, and also the largest extent of base-pair breakage. The RMSF of the helices is generally lower by ~ 0.2 nm at the higher salt concentration, with the exception of helix 2, which experiences higher fluctuation in high salt conditions.

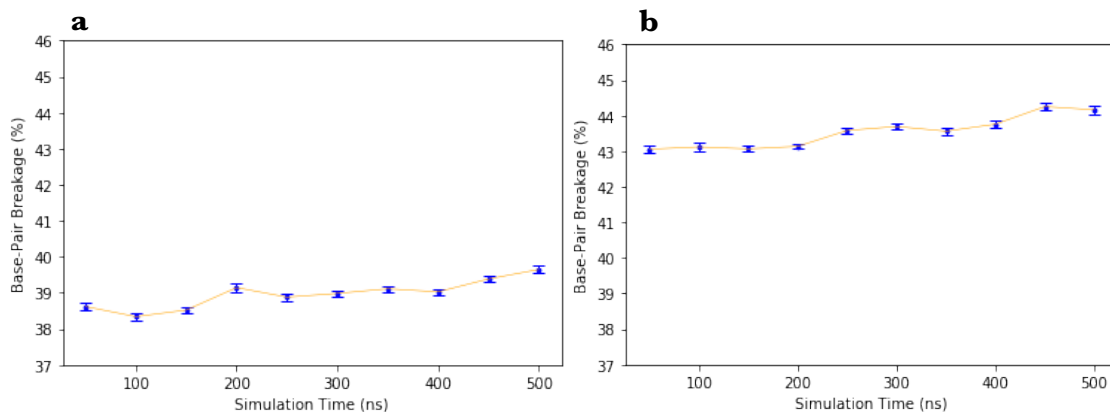


Figure 3.13: Fluctuation of base pairs (% BPB) as a function of time for **(a)** the 0.3 M NaCl production simulations and **(b)** the 1.0 M NaCl production simulations. The percentage of broken base pairs gradually increases over time in both cases, and at a similar rate. The % BPB differs between the two salt concentrations significantly at $t = 0$, on account of base pair breakage occurring during the equilibration period in both models. Error bars represent the bootstrapped standard error.

3.5 Further Validation of CG Simulation Data

Here, I present some experimental and all-atom MD simulation work on the characterisation of the DNP done by both internal and external collaborators, which validate the results discussed in this chapter. AA simulations of the solvated DNP in 0.3 M NaCl detailed in Section 3.4.1 were performed by Daria Kieczka, a masters' student under my supervision, and the cryo-EM experiments and subsequent analyses were undertaken by Dr. Abid Javed and Professor Elena V. Orlova at Birkbeck, University of London. To facilitate comparisons of CG data to the cryo-EM maps, structures representing the most highly populated clusters obtained from the CG trajectories were back-mapped to their all-atom representation, as the flexible fitting procedure for the construction of cryo-EM images must be performed with all-atom MD models. The clustering of CG trajectories was performed with the GROMOS algorithm, implemented within the *gmx cluster* analysis module in GROMACS.[63], [98]

3.5.1 *All-atom simulations of the solvated DNP in 0.3 M NaCl*

The equilibrated all-atom DNP model that was constructed to serve as the basis for the solvated CG model (shown in Figure 3.2b) was used to initiate the production simulations of the AA solvated DNP model. All equilibration and production simulations were performed with the 2019 version of GROMACS.[63] The system was minimised for 10,000 steps (2fs time-step) using the default conjugate-gradient algorithm, to remove the unfavourable interactions and relax strained bonds that originate from the artificial starting structure. The system was then equilibrated for ~ 15 ns with a network of elastic restraints to reinforce base-pairing and base-stacking interactions. The force constant (k) of these restraints was gradually reduced throughout the equilibration, from $500 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ to zero. The model was then subjected to another 50 ns of unrestrained dynamics. The temperature was set to 300K with the velocity-rescale thermostat, and isotropic pressure control at 1.013 bar was maintained through the use of the Parrinello-Rahman barostat. Electrostatic forces were calculated using the particle mesh Ewald (PME) algorithm, and a 0.9 nm cutoff was used for VdW and electrostatic interactions. A 0.9 nm shifted cutoff was used for VdW interactions and short-range electrostatic interactions. A total of fifteen production simulations were run for 30 ns apiece, and each replica was initiated with a different random velocity. Average pore height and pore width values were calculated from the AA DNP simulations according to the protocol described in Section 3.2, as well as helix kink angles and % BPB. These full set of AA simulations consumed a total of 250,000 CPU hours on Cartesius.[124]

3.5.2 *Cryo-EM Models*

The classification of electron-density maps yielded from cryo-EM experiments conducted in a solution of 0.0012 M MgCl_2 identified a total of six well-populated distinct structural classes (Fig. 3.14). Though the use of a monovalent salt medium matching the salt concentrations used in the CG MD simulations would have provided us with a more directly comparable dataset, this was not ultimately possible as the resolution of the images yielded in these conditions was comparatively poor. It is well documented that Mg^{2+} ions have a profound influence on the structure and stability of the *ds*DNA helix,[126]–[128], and that salt concentration strongly influences the persistence length and flexibility of DNA.[129] Experimental and computational

studies have demonstrated that increasing the concentration of Mg^{2+} ions in the media leads to stabilisation of base-pairing and attenuation of the fluctuations in the phosphate backbone, and Mg^{2+} ions are also capable of inducing β to Z-DNA transitions in helices with high cytosine and guanine content[130], and kinking in helices with more adenine and thymine content[131]. Therefore, we can expect some discrepancies between observations made in cryo-EM and those made in the simulations. Five of these six classes were of high enough resolution for flexible fitting against MD-derived atomic models.

Before the flexible fitting procedure can be initiated, one must establish an appropriate set of back-mapped AA models that are distinguishable at the cryo-EM level of resolution (typically anywhere between 0.2 nm and 2.0 nm) [132], whilst also being representative of the range of conformations present in the CG simulation ensemble. This can be achieved by tailoring the RMSD cut-off used for clustering simulation trajectories, as this directly influences the number of clusters that are generated, and the degree of resolution between these clusters. I tested a range of RMSD cut-off values between 0.3 nm and 0.5 nm, and ultimately selected the clusters obtained with the 0.4 nm RMSD cut-off for the comparisons detailed here, as these clusters yielded the highest correlation with the cryo-EM maps. The CG middle structures of these clusters were then back-mapped to their all-atom representation through the use of the *backward.py* script [133] provided by the developers of the MARTINI force-field. The structures and percentage populations of these twelve clusters are shown in Appendix 3. Three of the twelve back-mapped clustered models correlated extremely well with the six cryo-EM maps during the flexible fitting procedure, with a cross correlation score of > 0.9 in each case (Fig. 3.14). To obtain these correlation scores, first the clustered DNP structure PDB files were converted into SPIDER density maps, and the resolution of the computed EM map was adjusted to match that of the cryo-EM map. The computed EM maps were then docked into cryo-EM electron density maps using the SPIDER software suite[134], according to the method described by Li and Frank.[135] After docking, the flexible fitting procedure is initiated, which uses an external potential derived from the cryo-EM map to drive the atomic positions of the MD-derived molecular model into high density areas in the cryo-EM map while preserving the conformation of the MD-derived model through the application of harmonic restraints to the atomic positions, to avoid over-fitting. Finally, the Pearson's

cross-correlation coefficient between the averaged voxel values of the EM and MD maps is calculated to give a measure of the goodness of fit.[136]

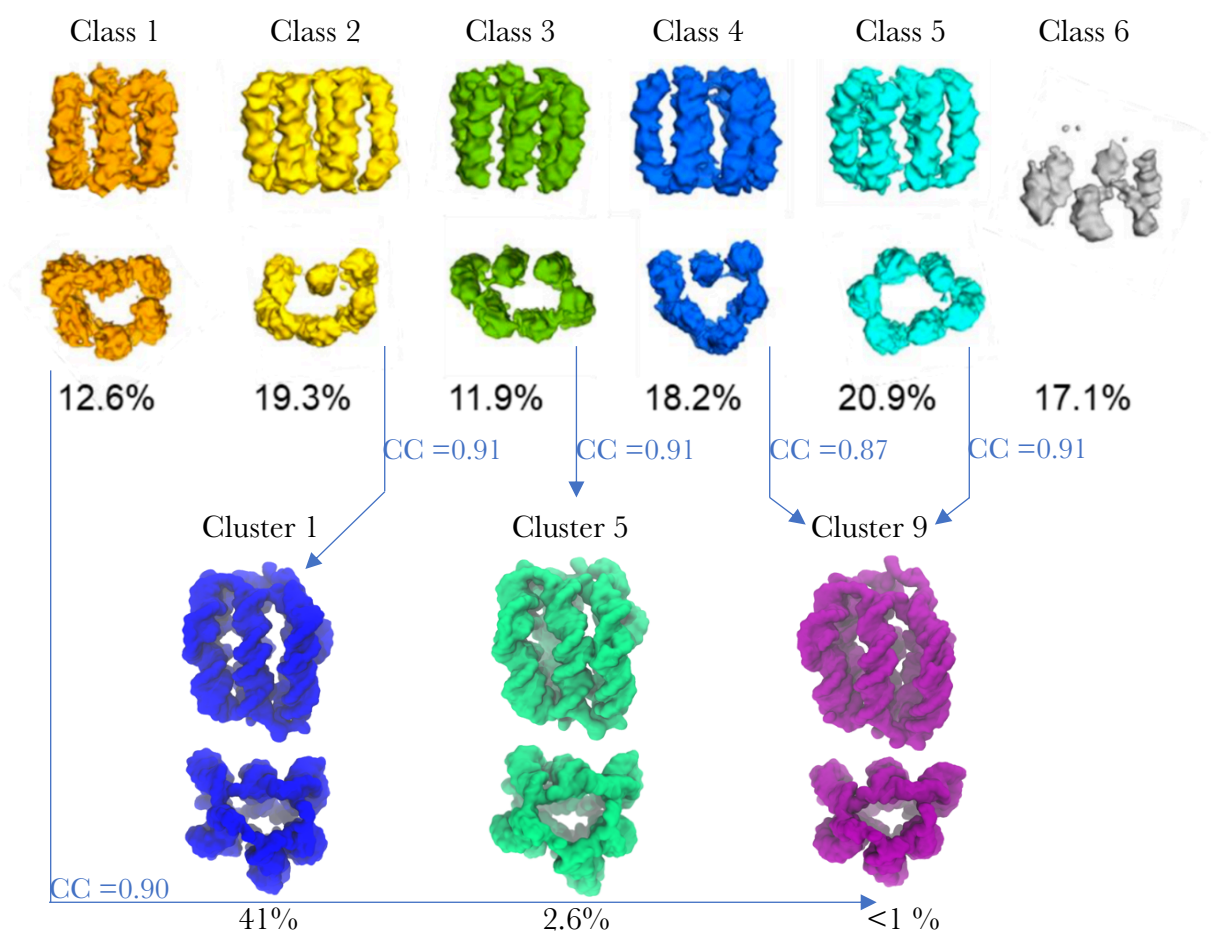


Figure 3.14: Results of the flexible fitting of cryo-EM electron density maps to CG MD derived atomic models of the DNP. The six structural classes obtained from the cryo-EM characterisation of the DNP in 12 mM MgCl_2 solution, and other are shown in the top panel. No detergents or nanodiscs were used in the preparation of cryo-EM samples. The population of these classes is represented as the number of structures belonging to each class taken as a percentage of the total number of structures obtained. The three back-mapped middle structures obtained by clustering the CG trajectories are shown in the bottom panel, along with their population percentages. Blue arrows indicate the strength of the correlation, denoted by correlation coefficient CC, between the cryo-EM maps and the MD models.

RMSD cut-off	CG Clusters
0.30 nm	3
0.35 nm	28
0.40 nm	12
0.45 nm	6
0.50 nm	4

Table 3.3: The relationship between the RMSD cutoff used for clustering CG trajectories and the number of clusters obtained. The set of 12 clusters generated with the 0.4 nm RMSD cut-off were used for flexible fitting to the cryo-EM maps.

3.5.3 Validation of CG simulation data

At the time of writing of this thesis, only a small preliminary set of cryo-EM derived quantities were available for direct comparison to computed quantities. However, the strength of the correlation between the cryo-EM reconstructions and the back-mapped AA models extracted from the CG simulation ensemble is demonstrative of the accuracy and reliability of the CG ensemble-based protocol for the characterisation of key structural features of a solvated DNP. To assess the accuracy of the dynamical behaviour and observed structural features yielded from the CG nanopore model, as well as the overall suitability of the MARTINI force-field for the characterisation of novel and complex DNA nanoarchitectures; direct comparisons were made between CG-derived quantities and their AA equivalents. Specifically, I focused on pore dimensions, % BPB, helix kink angles and RMSF-per residue plots, as these observable quantities taken together capture both the gross structure of the pore and more detailed, site-specific features and motions that are likely to contribute to the pores' functionality.

Overall, we find there is very good agreement between the results of the AA simulations and those of the CG simulations, particularly in relation to the proportions of the pore and areas of strain and/or loss of idealised dsDNA structure, though the CG model does appear to overestimate the absolute dimensions and % BPB.

3.5.2.1 Pore Dimensions and Structural Fluctuations

The side-by-side comparison of the computed AA and CG dimensions presented in Table 3.3 suggests that the MARTINI CG model of the nanopore overestimates the average pore height and width by ~ 0.5 nm, as a consequence of the standardised 4-1 atom-to-bead mapping and the associated loss of resolution.[73],[114] However, the ratio between the height and width in the CG model is nearly identical to the ratio observed in the AA model. In addition, the regions of constriction observed in the AA models correspond to the regions of constriction in the CG models. In both models, the pore is constricted at the termini, with the narrowest constriction occurring at the R8 terminus – suggesting that the overall pore geometry and specifically the fixed polarity of the pore width down the principal axis is successfully reproduced by the MARTINI force-field. This polarity is also observed in the cryo-EM images that were fitted to back-mapped clustered AA models, and this is reflected in the high cross correlation scores associated with the flexible fitting procedure. The slight overestimation of the absolute pore dimensions by the MARTINI force-field is a problem that can be easily overcome by measuring these quantities from representative back-mapped AA structures. The average pore height calculated for the 12 back-mapped AA structures generated from the clustered CG trajectories was 7.01 ± 0.05 ; almost identical to the value obtained from the ensemble of AA simulations. This ability to quickly recover atomic-resolution macroscopic properties from CG simulations whilst taking advantage of greater computational efficiency and longer simulation timescales further proves the power and versatility of CG MD simulation protocols. The efficiency of the CG protocol here is indeed remarkable – the ensemble of 500ns CG simulations required only 65% of the CPU hours that were consumed when running the ensemble of 30ns AA simulations.

As the absolute AA dimensions are inherently more accurate than CG equivalent, these are the dimensions we should use to compare against the cryo-EM values. We notice that the average measured height of the DNP in the cryo-EM dataset is larger than it is *in silico* by ~ 0.6 nm, and the average measured width is only very slightly narrower, though the relatively large uncertainty in the measured width makes this comparison slightly more difficult. The larger pore height in cryo-EM is likely to be a consequence of the relatively low salt concentration used in the cryo-EM experiments.

As the cryo-EM experiments were conducted after the MD data were collected, and the cryo-EM protocol explicitly required an MgCl_2 buffer, it was regrettably not possible to achieve a match between the solvent conditions. Transient electric birefringence experiments have shown that, regardless of the cation valency, increasing the salt concentration from low (~ 0.001 M) to moderate concentrations (~ 0.01 M) causes the persistence length of *ds*DNA to decrease monotonically from the range of 80 - 100 nm down to the limiting persistence length range of 45 - 50 nm. [129][138]. While it cannot be measured directly, it can be assumed that the persistence length of the DNA in the cryo-EM conditions used here is likely to lie in the 80 - 100 nm range. A higher persistence length in cryo-EM conditions would result in a stiffer helix and lower degree of helix kinking, and therefore a longer helix. Helix kinking was indeed observed in the cryo-EM images, and kink angles are currently being measured for the cryo-EM dataset, but cannot be commented on at this time.

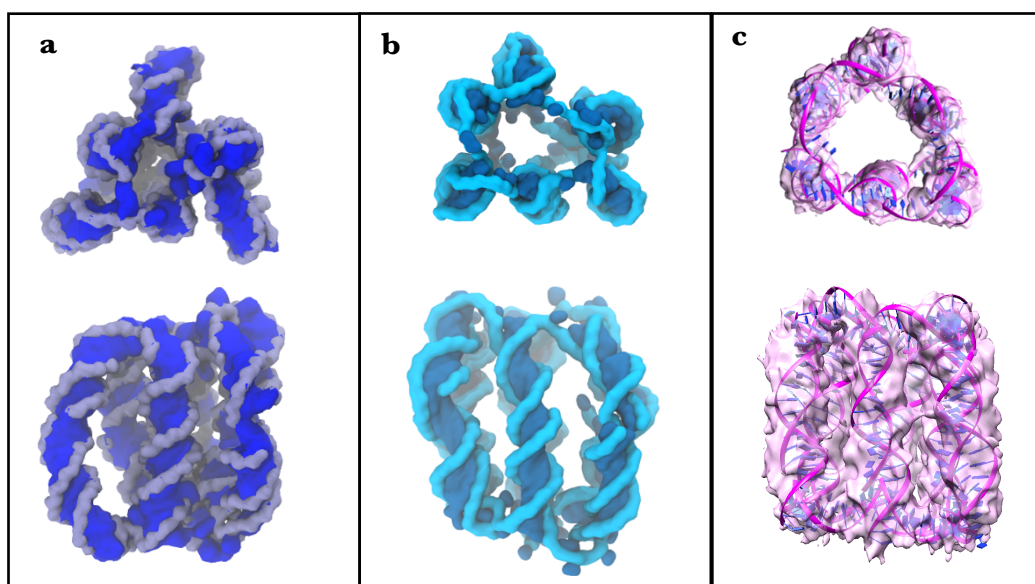


Figure 3.15: Rendered images of highly populated DNP clusters obtained from three different datasets. The top-view images are taken from the R1 terminus. (a) Model of the most populous structure obtained from the AA 0.3 M ensemble. (b) CG model representing the most highly populated cluster obtained from the CG 0.3 M NaCl ensemble (c) Cryo-EM structure of the DNP built by flexible fitting of the back-mapped AA representation of model **b** to the most populated structure observed in cryo-EM (class 5 in Fig. 3.14).

	AA 0.3 M NaCl	CG 0.3 M NaCl	Cryo-EM 12 nM MgCl ₂
Average Pore Height (nm)	7.02 ± 0.04	7.59 ± 0.02	7.60 ± 0.4
Average Pore Width (nm)	7.09 ± 0.06	7.48 ± 0.04	6.85 ± 1.2
Average Pore Width R1 (nm)	5.56 ± 0.14	6.59 ± 0.04	TBC
Average Pore Width R8 (nm)	5.11 ± 0.10	6.11 ± 0.02	TBC
Minimum Pore Width R1 (nm)	4.83 ± 0.14	5.58 ± 0.10	TBC
Minimum Pore Width R8 (nm)	4.58 ± 0.12	4.65 ± 0.16	TBC

Table 3.4: Comparison of DNP dimensions obtained from AA simulations in 0.3 M NaCl, CG simulations in 0.3 M NaCl and cryo-EM microscopy simulations conducted in 12 nM MgCl₂. The higher-resolution AA-derived pore dimensions are taken as the more reliable estimates of the absolute dimensions. The pore height calculated from back-mapped CG clusters (7.01 ± 0.05 nm) overlaps extremely well with the AA pore height.

Fluctuation analysis was performed for the AA simulation ensemble according to the same protocols used for the CG ensemble (Fig. 3.9). The RMSF-per-residue plot constructed from this data provides a convenient means to assess the validity of the characteristic motions and conformational flexibility observed in the CG ensemble, as the plots in Figure 3.9a (CG) and Figure 3.16a (AA) can be compared directly. The AA RMSF-per-residue plot reveals a median RMSF of ~ 0.6 nm, which is over twice the median RMSF seen in the CG equivalent (0.25 nm), which suggests that the conformational flexibility of the DNP is somewhat suppressed in the CG model. This may be the consequence of the use of the stiff elastic network model in the production simulations, which applies flexible restraints to all of the pseudo-atom beads in order to preserve the β -DNA structure (discussed in Section 2.2.2.2). However, the distribution of more “extreme” values (greater than the median) through the pore body is very similar to the distribution of extreme values in the CG plot. In both plots, extreme values are distributed fairly uniformly through the pore body, and the highest RMSF values are observed in the R4/R5 regions, indicating that the breathing motions of the bloated midsection observed in the CG model are also present in the AA model. However, the AA RMSF plot displays an asymmetry in the mobility of the termini that was seemingly not captured by the CG simulations. The mean RMSF of R1 is higher than the mean RMSF of R8 by ~ 0.7 nm, suggesting that the poly(T) crossovers at R8 terminus experience more strain than the poly(T) crossovers at the

R1 terminus. While this is not directly observed in the CG RMSF plot, this is reflected somewhat in the asymmetry in the widths of the terminal constrictions seen in the CG model (and validated by the AA dimensions) presented in Table 3.4. The narrower terminus is likely to be less flexible than the wider terminus, as the two sets of poly(T) crossovers are identical in length and sequence, so the narrowing of the R8 terminus compared to the R1 terminus must be a result of strain imposed by the neighbouring R7 region.

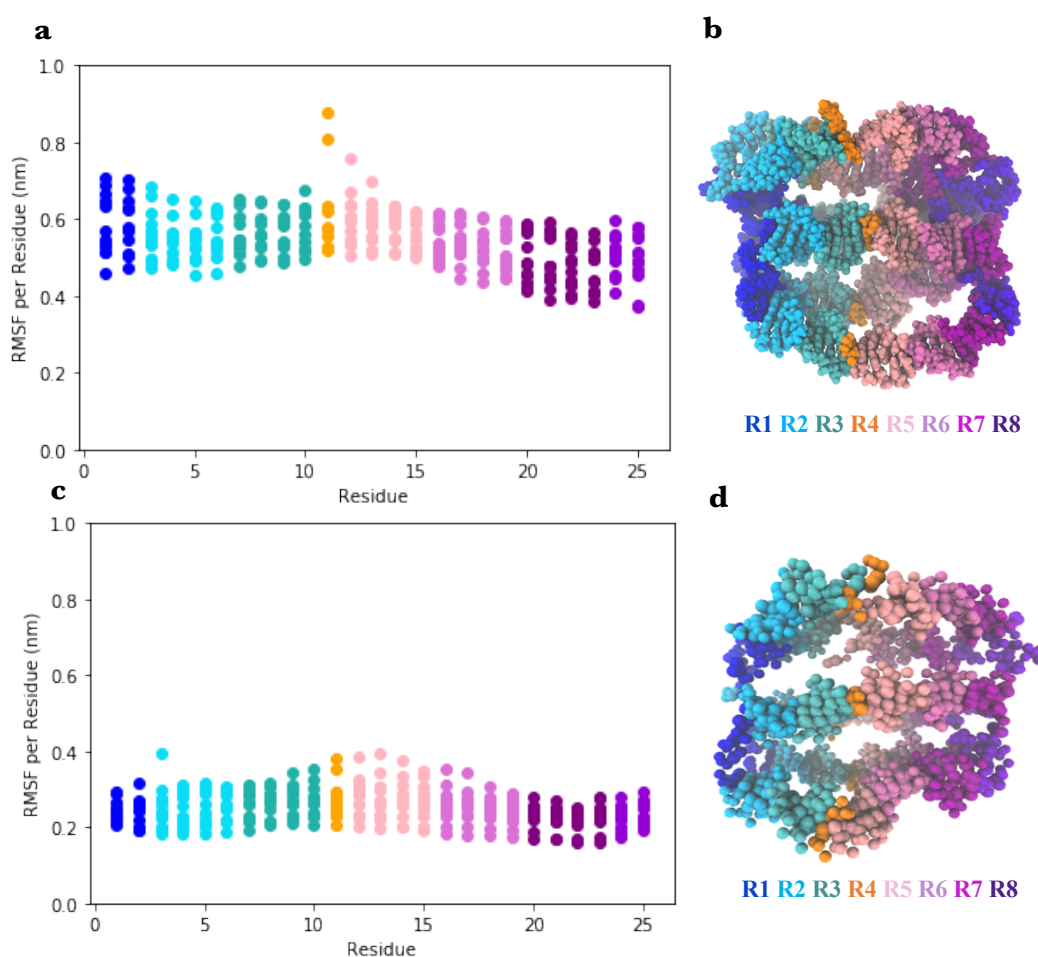


Figure 3.16: (a) RMSF-per-residue plot generated from the ensemble of 15 x 30 ns AA simulations of the solvated DNP in 0.3 M NaCl, illustrating the degree of fluctuations through the pore body from R1 to R8. (b) Representative structure of the AA DNP model in 0.3 M NaCl taken from the most populous cluster of conformations observed in the AA ensemble. (c) RMSF per residue plot for the CG DNP model in 0.3M NaCl taken from Figure 3.9a and readjusted to match the range of y-axis values in the AA equivalent. (d) Average CG structure of the DNP in 0.3M NaCl.

3.5.2.1 Helix Kink Angles and Percentage of Base-Pair Breakage

Here, I compare the % BPB extracted from the CG simulation ensemble in 0.3 M NaCl (presented in Table 3.2) to the % BPB observed in the AA simulation ensemble and relate these quantities to observations of structure loss in specific helices in the cryo-EM structures. The total % BPB throughout the pore (14.8 ± 1.00 %) calculated from the AA ensemble is considerably lower than the value reported for the CG ensemble, which is suggestive of a two-fold to three-fold overestimation of base-pair breakage within the MARTINI CG model. However, the AA data does give credence to some observations made in the BPB analysis – namely the identification of helices that experience breakage of over 50% of the base pairs. In Section 3.4 (Table 3.2), we concluded that helices 1, 3 and 6 exhibit a relatively high % BPB in 0.3 M NaCl. In the AA dataset presented below, we can see that helices 1 and 6 do indeed experience above-average BPB as suggested by the CG model, though helix 3 is below the average. The cryo-EM structures of the DNP in 12 mM MgCl_2 are not of high enough resolution to allow for distinction of frayed helices, however they do show evidence of loss of β -DNA structure in specific areas, namely helices 2, 5 and 6. These helices closely resemble Z-DNA in the images, suggesting that a major conformational change occurs in these helices in MgCl_2 . The transition from β -DNA to Z-DNA involves transient disruptions of base-pairs, and typically occurs on the scale of seconds to minutes.[139] Both the CG and AA models predicted a relatively high percentage of base-pair breakage in helix 6, highlighting the potential for predicting the loci of these transitions *in silico*. However, BPB data from both models did not identify helices 2 and 5 as sites for these potential β -to-Z conformational transitions, which is not completely unexpected given that the time scale for complete base-pair formation and breakage in *ds*DNA helices ranges from several hundreds of nanoseconds to milliseconds and is still a heavily debated topic[140], [141].

Aside from the difference in salt conditions between the MD simulations and the cryo-EM experiments, it should also be noted that that the DNP structure studied by cryo-EM does not feature TEG-cholesterol anchors, which may introduce further sources of discrepancy between the MD and cryo-EM data sets.

	CG % BPB	AA % BPB
R1-R8	39.0 ± 2.48	14.8 ± 1.00
R1-R4	30.2 ± 3.40	14.4 ± 1.52
R5-R8	45.7 ± 2.08	15.0 ± 1.12
Helix 1	60.4 ± 2.95	18.9 ± 1.29
Helix 2	5.58 ± 1.66	10.8 ± 1.60
Helix 3	67.8 ± 1.75	10.4 ± 3.10
Helix 4	16.4 ± 2.00	14.6 ± 0.72
Helix 5	44.0 ± 3.31	12.8 ± 1.22
Helix 6	53.6 ± 3.37	20.0 ± 2.60

Table 3.5: Comparison of the percentage of base pair breakage calculated for the CG and AA solvated DNP systems in 0.3 M NaCl. Overall, % BPB is significantly lower in the AA ensemble than it in the CG ensemble, and the asymmetry in base pair breakage above (R1-R4) and below (R5-R8) the helix nicks that is observed in the CG ensemble is much less pronounced in the AA ensemble. However, both ensembles identify helices 1 and 6 as sites prone to above-average BPB.

Given that the MARTINI force-field is known to underestimate the free energy associated with base pairing in the DNA model, as discussed in the 2015 paper by Uusitalo *et al.*[84] The minima in the potential of mean force (PMF) profiles associated with the interactions between two CG hydrogen-bonding nucleotide bases approaching each-other on a fixed plane are much shallower than in the equivalent AA models simulated with CHARMM or AMBER parameters. On average, the magnitude of the free energy at the global minimum of PMFs calculated for a single base-pair is ~ 1.5 times smaller than that of the CHARMM force-field. Adjustments of the non-bonded hydrogen-bonding pseudo-atom parameters were made in an attempt to correct this misrepresentation of base-pairing, however these adjustments led to a negative impact on representation of base-stacking interactions, which is generally very accurate in the MARTINI forcefield. Fortunately, the agreement between the AA and CG data with respect to calculated helix kink angles is much stronger than for BPB. The AA kink angles calculated for helices 1 – 5 overlap very well with the CG kink angles (both shown in Table 3.6), and both models indicate that helix 4 experiences the most severe kinking, followed closely by helices 2 and 5. The AA and CG-derived kink angles for helix 6 are in reasonable agreement, however the average kink angle in

the CG model lies slightly outside the lower bound of the confidence interval of the AA value. The magnitude of the errors reported for the AA kink angles are generally higher than they are in the CG equivalent, on account of the shorter simulation timescale of the AA simulations, and the higher average RMSF in the AA models, which gives rise to relatively sharp fluctuations in the kink angle as a function of simulation time (plot provided in Appendix 4). The cryo-EM structures reveal significant kinking in all six helices, though verified measurements cannot be provided at this time.

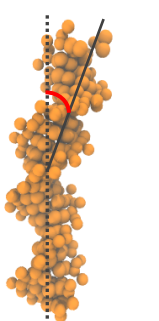
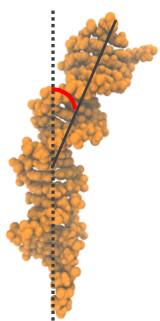
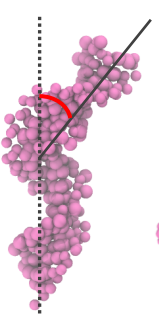
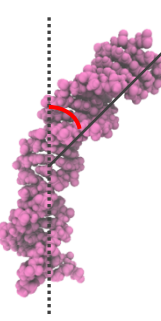
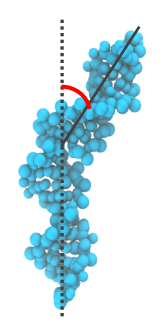
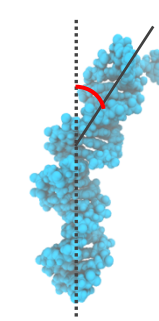
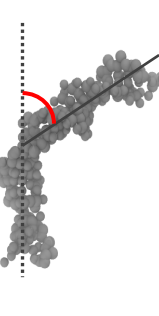
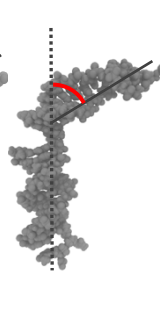
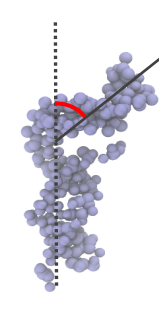
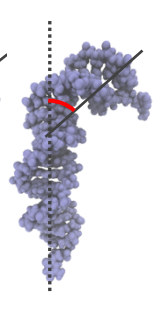
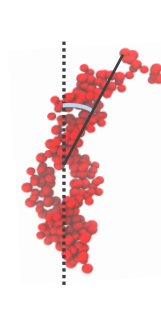
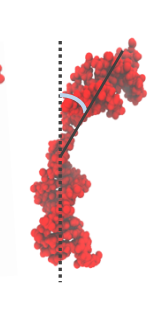
Helix 1		Helix 2		Helix 3	
CG	AA	CG	AA	CG	AA
$40.0 \pm 0.3^\circ$	$35.9 \pm 7.8^\circ$	$62.5 \pm 0.5^\circ$	$70.3 \pm 7.3^\circ$	$50.8 \pm 0.5^\circ$	$47.2 \pm 4.0^\circ$
					
Helix 4		Helix 5		Helix 6	
CG	AA	CG	AA	CG	AA
$67.3 \pm 0.7^\circ$	$78.3 \pm 5.3^\circ$	$62.5 \pm 0.5^\circ$	$60.0 \pm 9.1^\circ$	$31.0 \pm 0.8^\circ$	$38.9 \pm 3.4^\circ$
					

Table 3.6: Helix kink angles calculated for the CG and AA simulation ensembles of a solvated DNP in 0.3 M NaCl. Helix kink angles were calculated using the *gmx bundle* analysis tool, which calculates angles between vectors representing the kink-top and kink-bottom groups. The solid black lines represent the vector of the R1-R4 region (kink-top) in each helix, and dotted lines represent the vector of the R5-R8 region (kink-bottom). In all six helices, the centre of the kink angle originated at the cusp of R4 and R5 in all replica simulations. Pseudo-atom indices of the R4/R5 region were defined as the “kink centre” group, from which the two vectors emerge to make the angle.

3.6 Conclusion

This chapter explored the conformational dynamics, average dimensions and key structural features of a DNP in bulk solution, and the influence that ionic strength has on these properties. Overall, the solvated DNP has no distinct lumen, and is conformationally very flexible, with a significant amount of helix fraying and characteristic breathing motions in the midsection that are facilitated by kinking of the helices at the nick sites. The pore is constricted at the termini, with a fixed polarity in the widths of the termini that is also observed in cryo-EM electron density maps. These observations contrast greatly with the idealised pore structure, which was assumed to be a rigid barrel with a well-defined lumen when it was designed. The poor rate of insertion associated with this pore design may be a consequence of the lack of rigidity and generally globular conformation of this particular design of DNP. It would not be appropriate to compare the structural characteristics of these DNPs to protein pores such as α -haemolysin or ClyA as these proteins exist in a monomeric form in solution, which aggregate and assemble into their multimeric functional form once they come into contact with lipid bilayers. Similarly, we cannot attribute the relatively slow rate of insertion to structural properties by directly comparing the rates of insertion of the DNPs studied here to those of other published DNP designs, as insertion kinetics of other designs have not been published yet.

Through the comparisons made against AA simulation data and cryo-EM experiments, we have seen that the use of the MARTINI CG force-field combined with subsequent back-mapping of clustered structures allows us to accurately predict the dimensions of a solvated DNP, and CG models alone were capable of identifying regions of constriction and strain within the structure. The CG models also identified the helices most susceptible to fraying and kinking, and the kink angles calculated from the 0.3 M NaCl CG simulation ensemble were successfully validated by AA simulations of the DNP in the same conditions. However, the loss of resolution and the invocation of the elastic network model when using the MARTINI force-field does lead to an underestimation of the degree of fluctuations, and in particular the relative flexibility of R1 terminus compared to the R8 terminus, though in theory this could be corrected for by using softer elastic restraints. Over-estimation of the percentage of base pair fraying is likely unavoidable when using the MARTINI model, due to the

difficulty involved in refining non-bonded parameters to preserve realistic base-pairing and base-stacking interactions simultaneously. Nevertheless, the model predicted a relatively high incidence of base-pair breakage in helices 1 and 6, and this observation was validated by AA simulations. The CG simulation ensemble provided highly reproducible results with consistently low errors due to the ease and affordability of accessing longer timescales with the MARTINI model, and the breadth of observations made in these ensembles yielded results that were comparable to cryo-EM experiments – which has not yet been achieved in any CG simulation studies of DNA nanostructures thus far. To make the best use of these findings, the simulation protocol described in this chapter should be applied to other published DNP frameworks, and the kinetics of insertion of these DNPs into planar membranes must also be established, using techniques such as total internal reflection fluorescence and reflectance interference spectroscopy (TIRFS-RIf).[4] Comparing the dynamical behaviour of the simulated DNPs in solution allows for the identification of design motifs (e.g. helix length, helix nicks, terminal polyT crossovers, Holliday junctions) that confer structural rigidity or areas of flexibility to the pore scaffold. It is not yet clear whether or not the conformational flexibility, helix kinking, and helix fraying exhibited by the TEG-C DNPs in solution here detrimental or advantageous to membrane insertion. If simulation data for a range of DNP designs are considered alongside experimentally derived insertion kinetics, the answer to this question could be elucidated, and future designs can be refined to improve membrane encapsulation of DNPs.

Chapter 4

Coarse-Grained Simulations of Lipid-Bilayer Spanning DNA Nanopores

The protocols and results discussed here builds on those discussed previously in Chapter 3. To understand the influence of ionic strength on the functionality of a DNP, we must analyse its' dimensions and conformational dynamics as they span a lipid bilayer, as well as in bulk solution. The concept of hydrophobic mismatch (discussed in Chapter 1) is important when we consider the interactions between DNPs and lipid bilayers. As the DNP possesses negative charges along the entirety of the outer surface, and just three thin and flexible hydrophobic anchors in lieu of a hydrophobic region, we can surmise that the hydrophobic mismatch between the DNP and the membrane will be very large. A greater number of cholesterol anchors on the DNP may in theory reduce the hydrophobic mismatch, however when tested experimentally these DNPs had a tendency to aggregate in solution (results pending publication). The composition of the lipid membrane would also influence the extent of hydrophobic mismatch and the type of hydrophobic matching behaviour that occurs between the DNP and the bilayer, as acyl chain length and cholesterol content control the hydrophobic thickness of the membrane, as well as its rigidity and susceptibility to curvature.[29]

As we are interested in understanding observations made in published experimental studies which make use of homogenous PC lipid bilayers[4], [38], the simulation conditions must be reflective of these conditions. Hence, a homogenous POPC bilayer was selected to model the lipid bilayer as they are used extensively in ongoing electrophysiology experiments. Experimental and computational studies of these bilayers have confirmed that POPC bilayers have a hydrophobic thickness of ~ 2.7 nm[142]. As the DNP is highly hydrophilic, can expect to observe some hydrophobic matching effects in the membrane simulations, whereby the system moves to reduce contact between hydrophobic lipid tails and the negatively charged DNP. This can involve tilting of the pore, reorganisation of lipids, and the adoption of alternative binding modes. Indeed, these effects are observed in all of the DNP membrane simulations, and the findings here indicate that the salt concentration plays a crucial role in the degree of tilting and general mobility of the DNP within the bilayer.

4.1 Model Building

The solvated membrane models were built by inserting the CG DNP into the centre of a pre-equilibrated CG POPC bilayer patch, which was solvated with in a box of polarised MARTINI water molecules with the *insane.py* script.[96] Coarse-grained Na⁺ and Cl⁻ ions were added to set NaCl concentrations of 0.3 M (for the low salt model) and 1.0 M NaCl (for the high salt model), also with the *insane.py* script. The two models consisted of $\sim 100,000$ pseudo-atoms, with overall cell dimensions of 16 nm x 16 nm x 13 nm. A 16nm x 16nm membrane patch was chosen, as this allowed for plenty of space between periodic images of the DNP, thus avoiding self-interactions. The initial configuration of the DNP had an initial unequilibrated width of 4.95 nm and was placed in the middle of the membrane patch, which corresponded to initial distance of 11.05 nm between periodic images of the DNP. Membrane simulations are also highly susceptible to finite size effects, as the periodic boundaries often cancel out undulations that may be experienced by membranes in nature[143][144], which can have amplitudes anywhere between 5 nm and 100 nm. Sampling these undulations is important in these simulations, as hydrophobic matching effects i.e. lipid reorganization and membrane thinning are variables of significant interest in this study. This membrane patch size allows us to sample some of the smaller amplitude undulations, as the centred DNP is surrounded by ~ 5 nm of membrane.

The initial models were subjected to 5,000 steps of minimization using the steepest-descent algorithm. A 2-fs time-step was used for the first 10 ns of equilibration time, as the initial relaxation of system involves drastic conformational changes which result in large forces being calculated at each time step. The first phase of the equilibration was performed in the NVT ensemble, with strong positional restraints applied to the beads of the DNP and the charged lipid head-groups. The force constants of the DNP restraints were gradually reduced to zero over the course of ~ 5 ns of equilibration time, during which the temperature increased gradually from 100K to 300K. The subsequent 6 ns of equilibration was performed in the NPT ensemble with the Berendsen barostat[108], to relax the volume of the system whilst slowly releasing the restraints on the POPC head-groups, to prevent the formation of voids within the bilayer. After the removal of all position restraints, the time-step was increased to 5 fs, and the unrestrained system was subjected to another 20 ns of simulation time with the Berendsen barostat. For the final phase of equilibration (lasting 70 ns) the Parrinello-

Rahman barostat[112] was employed with a time coupling constant of 12 ps. Production simulations were then run for 1 μ s, with temperature and pressure held at 300K and 1.013 bar, respectively with the velocity-rescaling thermostat and the Parrinello-Rahman barostat.

4.2 Simulation Details and Analytical Methods

All CG simulations were performed with GROMACS v5.1.4. Temperature control was achieved using the velocity-rescale thermostat with a time constant of 1 ps. Two different barostats were employed for pressure control (at 1.013 bar) during NPT simulation phases; the Berendsen thermostat[108] for early equilibration (with time constant of 2.0 ps) and the Parrinello-Rahman barostat[112] for production/data collection. Three separate temperature and pressure coupling groups were used; one for the water and ions, one for the POPC bilayer, and one for the DNP (including the cholesterol anchors). Ensemble-based production simulations were run by generating separate run files with different random velocity seeds. Velocities, energies, forces and coordinates were recorded every 40ps, corresponding to 25 ‘snapshots’ per ns. The maximum distance allowed for bonded interactions with domain decomposition (rdd) was adjusted throughout the production simulations to facilitate the use of a 5-fs timestep, and typically remained in the range of 1.5 -1.9 nm. A standard 1.1 nm cut-off was used for electrostatic and VdW interactions, and this further ensures that there are no electrostatic or vdW interactions between periodic images.

The calculation of nanopore height, RMSF per residue, percentage base pair breakage, width profiles and lumen profiles for the DNP/membrane models were performed with the same tools and methodologies described in Chapter 3.2. Helix kink angles were calculated for the middle structures of the most highly populated clusters from each replica using the *gmx bundle* analysis tool. The clusters used for these calculations were obtained through the use of the GROMOS clustering algorithm implemented within *gmx cluster*, with the RMSD cutoff set to 0.3 nm. Bilayer thickness was calculated using the *gmx density* [70],[82] tool which was used to compute the density of the lipid headgroups along the bilayer normal. The distance between the peaks in the bilayer thickness profile provides a good measure of bilayer thickness. [145] Second rank lipid order parameters (P_2) for the bonds within the lipid acyl chains were calculated according to the following equation:

$$P_2 = \frac{1}{2}(3 \cos^2\langle\theta\rangle - 1) \tag{4.1}$$

where θ is the angle between each bond and the bilayer normal (z-axis). Averaging the P_2 values gives us an approximation of the alignment of lipid tails with respect to the z-axis. Good alignment with the z-axis is indicated by P_2 values close to 1, anti-alignment is indicated by P_2 values close to -0.5, and a value close to 0 suggests that the lipids have no preferred orientation and the bilayer is structurally anisotropic. [146]

Average bilayer thickness and average pore height were used as convergence criteria to determine the appropriate simulation length and ensemble size for these simulations. By monitoring the decay of the bootstrapped standard error associated with these two quantities, while increasing the replica number and length, I determined that 15 replicas will yield reproducible results for these models. The ensemble average plots in Figure 4.1 show that the averaged pore height and bilayer thickness in the 1.0 M NaCl simulation ensemble remain more or less unchanged after 10 replicas. However, this was not the case for the 0.3 M NaCl ensemble, where the trajectories split into two subsets with different DNP binding configurations. In one subset, the DNP remained in the membrane spanning configuration throughout, and in the other the DNP migrated out of the bilayer and tethered itself to the membrane surface. Therefore, it was necessary to run a total of 15 replicas for the 0.3 M NaCl ensemble, and the same was done for the 1.0 M NaCl ensemble. A simulation duration of 1 μ s was chosen for both ensembles initially, because this is the timescale on which processes such as lipid reorganisation and bilayer phase transitions occur, according to NMR relaxation experiments. [147] As we are interested in probing the interactions between the DNP and the lipid, we must simulate for a long enough time to ensure that hydrophobic matching effects such as lipid reorganisation are captured, and that the properties of membrane and DNP are converged across the entire ensemble.

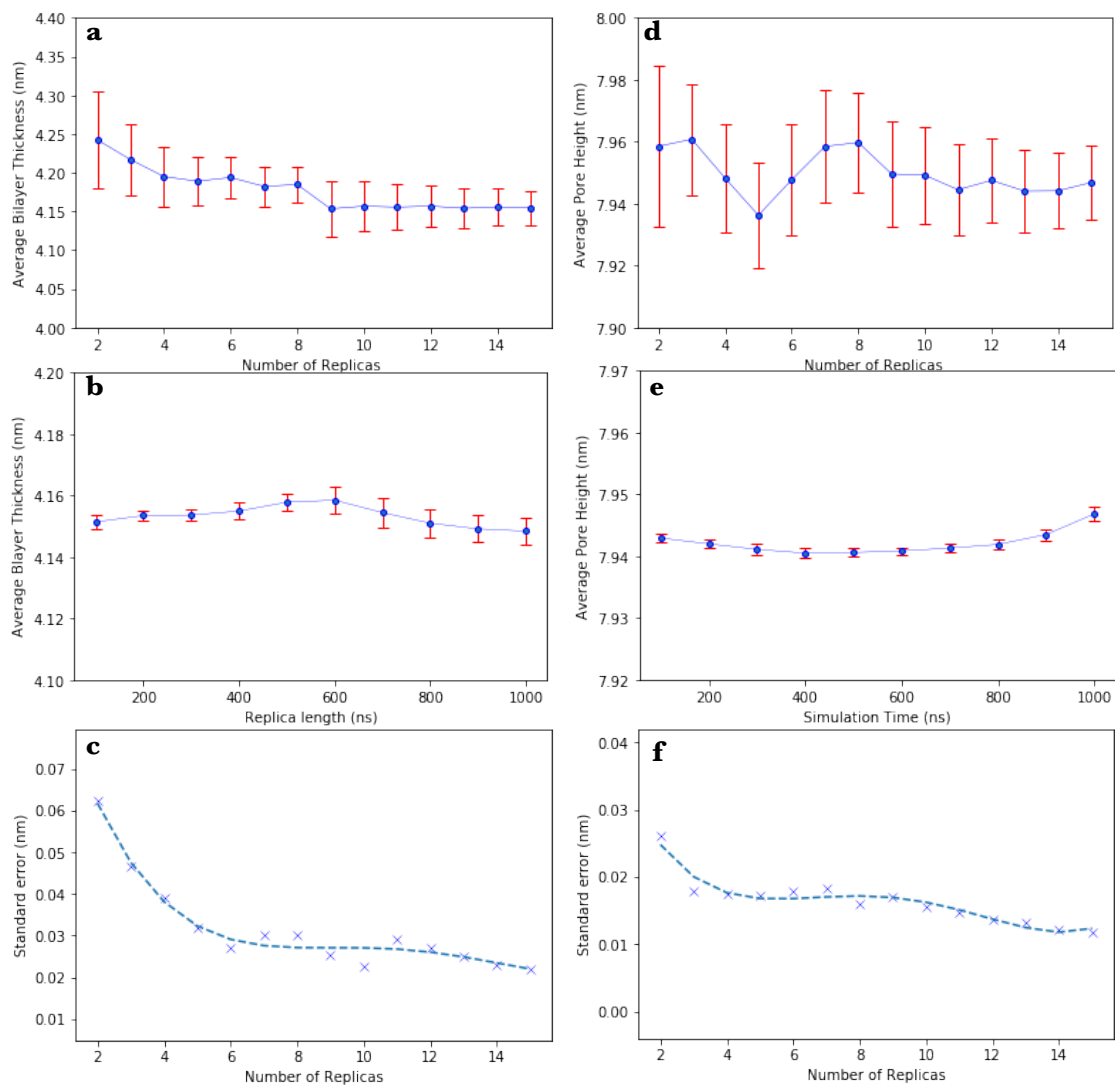


Figure 4.1: Ensemble average plots illustrating the convergence of quantities in the 1.0 M NaCl simulation ensemble. Error bars correspond to the bootstrapped standard error. **(a)** Convergence of the average bilayer thickness as a function of replica number demonstrates that 10 replicas are sufficient to ensure reproducible results, however 15 replicas were performed in total to ensure reproducibility and consistency between the 1.0 M NaCl and 0.3 M NaCl ensembles. **(b)** Convergence of the average pore height with simulation time shows that $1\mu\text{s}$ is an appropriate duration for a set of 15 replicas. **(c)** Decay and subsequent plateau of the standard error associated with bilayer thickness suggests that 15 replicas guarantee fully converged errors. **(d)** Convergence of the average pore height as a function of replica number. **(e)** Evolution of the average pore height as a function of simulation time. The slight increase in pore height at the end of the plot suggests that a longer simulation duration may be prudent. **(f)** Decay of the bootstrapped standard error associated with the average pore height.

4.3 Benchmarking and Computational Resources

In order to approximate the level of computing resources that would be required for these simulations, strong scaling tests were performed for one of the CG DNP/POPC simulation systems (running with GROMACS v5.1.4) on two HPC platforms; the UK national supercomputer ARCHER[123] and the Dutch national supercomputer Cartesius. [124] Each simulation was run on 360 CPU cores, as the simulation box was too small to be divided amongst more cores than this. Half of the thirty 1 μ s - long simulations of the DNP/POPC models were completed on ARCHER, and the other half on Cartesius. These simulations consumed 685,000 core hours in total. The computational resources used for this work were provided by the CompBioMed Centre of Excellence.[125]

4.4 Results and Discussion

The results from the simulations of the DNP as it spanned the POPC bilayer revealed a striking changeability in its long-time behaviour at different ion concentrations. Most remarkably, there was a high tendency for the DNP to pop out of the bilayer and bind to the bilayer surface in the low-salt (0.3 M NaCl) ensemble, but not at all in the high-salt ensemble (1.0 M NaCl). This occurred in 10 out of the 15 trajectories in the low-salt ensemble. In experimental studies, DNPs have been shown to exhibit a strong preference for curved membranes over planar membranes[4], [38], and insertion into planar membranes in similar conditions (0.3 M KCl buffer) has been reported to be very slow. In light of this, single channel current recording experiments with these pores are typically carried out in buffer solutions containing surfactants and relatively high salt concentrations (e.g. 1.0 M KCl) to facilitate the uptake of DNP's into planar membranes. [2], [4], [38] This is possibly indicative of a high energy barrier associated with insertion of the DNP into planar membranes. While we do not have the appropriate simulation data to comment on the thermodynamics associated with membrane insertion, the expulsion of the DNP in the simulations described here suggests that the artificial starting configuration imposed on the membrane-spanning DNP is unfavourable in 0.3 M NaCl. In the affected trajectories, the DNP exits the bilayer in a series of steps, illustrated in Figure 4.3.

We can approximate the time at which the DNP departs the membrane within each simulation by tracking the change in pore height as function of simulation time across all replicas. As the pore tethers to the underside of the membrane, the attractive electrostatic interactions between the negatively charged phosphate backbone and the positively charged choline groups causes DNA helices to elongate, in order to maximise these favourable interactions. The average pore height in the affected replicas increases fairly sharply up until the end of the simulation in each case, indicating that these replicas are exhibiting non-equilibrium behaviour and should therefore be extended in the future. Figure 4.2 shows that the onset of the expulsion event varies considerably between replicas, with the time of onset ranging between 300 ns and 700 ns. To extract reliable information on the dimensions and dynamics of the DNP as it spans the bilayer in the low-salt condition, it was prudent to use the data from the first 300 ns of the 0.3 M ensemble only.

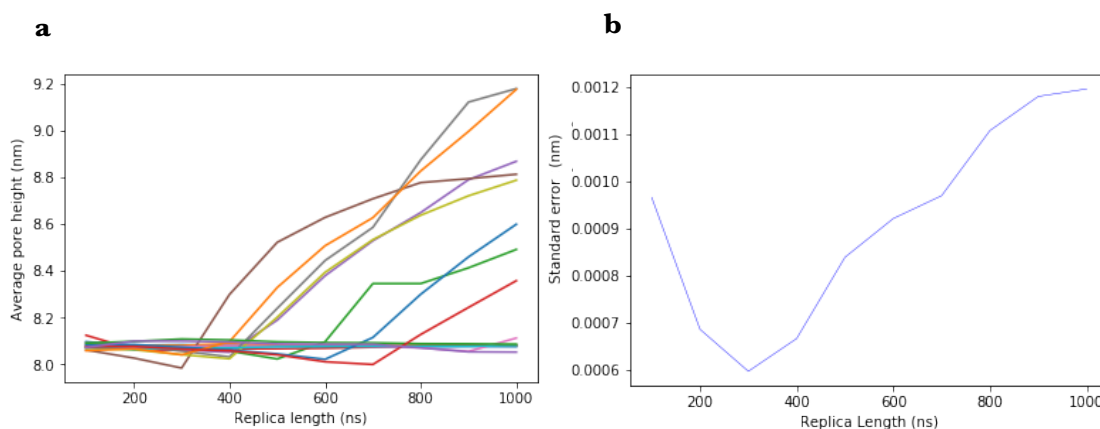


Figure 4.2: (a) Plot showing the evolution of the average pore height with time for each of the 15 individual replicas. The plot-lines showing a sharp increase in the average pore height are representative of the replicas in which the pore tethers to the bilayer surface after ejection. Error bars have been omitted for clarity. (b) Plot of standard error (σ) associated with the average pore height vs replica length, illustrating how the divergence of neighbouring trajectories causes the standard error to increase dramatically after around 300ns.

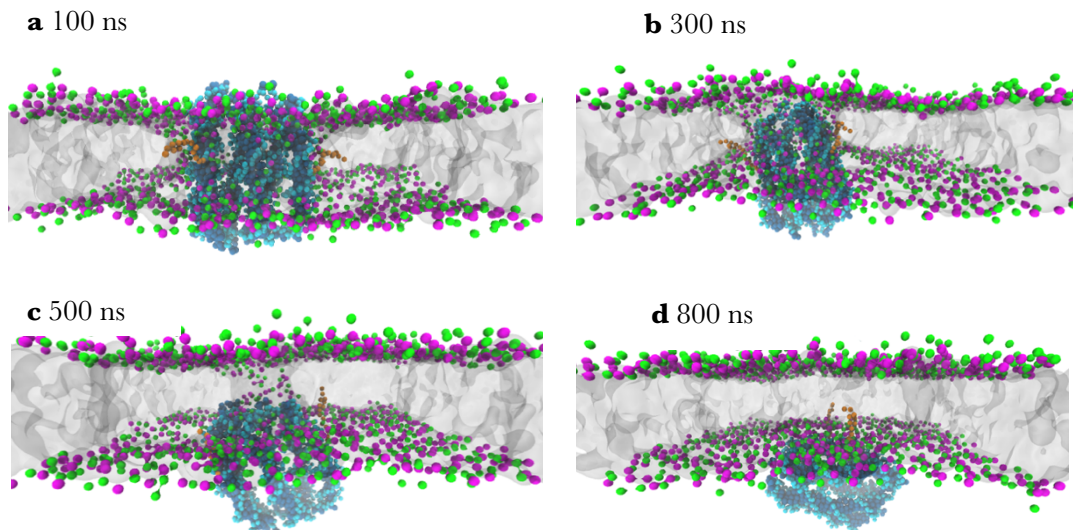


Figure 4.3: Snapshots taken from a simulation in which the DNP departs from the bilayer in a multi-step process. **(a)** The DNP is spanning the membrane normally in the transmembrane binding pose, and the surrounding lipids form a toroidal pore around the DNP by orienting themselves perpendicular to the bilayer normal. The shape of the toroidal pore is shown in detail in Figure 4.4. **(b)** Membrane lipids reorganise, and the DNP descends from the centre of the bilayer. **(c)** The DNP tilts and rotates, moving further out of the membrane. **(d)** The DNA backbone aligns with the POPC headgroups, and the DNP is held on the lower surface of the bilayer by two cholesterol anchors, adopting the side-on binding pose. The positively charged choline groups of the POPC lipids are represented as green spheres, and the negative phosphate moieties are shown as purple spheres.

Previously published experimental studies [4] of the DNP's membrane binding mechanism have alluded to the unpredictability of the DNP in the presence of planar bilayers, due to its propensity to induce curvature and eventually remodel bilayers into lipid nanotubes. Capturing these events using brute force MD is extremely difficult due to the issue of sampling as these processes occur over extensive periods of time (seconds/minutes).[4] To assess the effect of the two observed DNP binding poses (transmembrane vs side-on) on the structural and dynamical properties of the lipid bilayer from this set of simulations, the second-rank lipid chain order parameters, the average area per lipid and the average bilayer thickness was calculated for the membrane-spanning DNP in both salt concentrations, and also for the surface-tethered DNP after expulsion in 0.3 M NaCl. The bilayer thickness is defined as the

averaged distance between the lipid headgroups of the upper membrane leaflet and those of the lower membrane leaflet. Previously published X-ray scattering data[148] and AA MD studies[113][114] have confirmed the average thickness of a pure POPC bilayer to be 3.92 ± 0.01 nm at 303 K. The calculated bilayer thickness taken from the trajectory slices where the DNP is tethered to the bilayer surface (3.980 ± 0.034 nm, see Figure 4.4c) is very close to the published value, though slightly higher. The discrepancy may be a result of coarse-graining – it has been noted that the densities of bilayers in AA and CG models are often not in perfect agreement, and CG membranes are usually thicker than their AA counterparts. [82], [151] Alternatively, the discrepancy could be due to presence of salt ions, which have been reported to cause thickening of pure bilayers at concentrations as low as 0.10 M in x-ray scattering experiments.[152], [153] It is well known that incorporation of cholesterol into bilayers (including POPC) increases both the bilayer thickness and the average lipid order parameter.[149], [150], [154] However, these effects are only observed when the cholesterol is present at percentages of 20 % or above. As there are only 3 cholesterol anchors present in these systems (making up less than 1 % of the bilayer), their influence on the average bilayer properties is likely to be negligible in this case.

When comparing the bilayer properties associated with the two binding poses (transmembrane vs-side on) in 0.3 M NaCl, we can see that the average bilayer thickness decreases by 0.285 ± 0.1 nm when the pore adopts the side-on binding pose after expulsion in 0.3 M NaCl. The penetration of the bilayer by the DNP results in thickening of the POPC bilayer at both salt concentrations (Figures 4.4a/b), as the bilayer thickness values are considerably higher than published value in both cases. Hence, we can infer that the bilayer-spanning DNP induces membrane thickening – a phenomenon that has been observed in x-ray scattering[155] and MD[35], [156] studies of certain protein nanopores, including mammalian aquaporin-0 and a bacterial α -helical stress response protein. For both of these proteins, the hydrophobic length of the protein exceeds the hydrophobic length of the bilayer, and the membrane swells in the z dimension to increase the thickness of its hydrophobic domain in response. In the DNP/membrane system, the hydrophobic thickness of the pore is much smaller than that of the lipid, and the lipids form a toroidal pore around the DNP, so this reasoning is not applicable here.

Toroidal pores form when the headgroups of lipid molecules in close proximity to a transmembrane channel form strong electrostatic interactions with the backbone of the channel. [157] The lipids headgroups form the inner surface of a torus, resulting in a reduction of area-per-lipid of the membrane at the membrane/channel interface. The second-rank order parameters provide some insight on what happens to the orientation of the lipids when the DNP departs from the bilayer in 0.3 M NaCl, and how this may or may not impact the bilayer thickness. The order parameter increases by ~ 0.03 after expulsion, meaning the lipids adjust their orientation slightly to better align themselves parallel to the bilayer normal. The lipid torus that surrounds the DNP features lipids that lie perpendicular to the bilayer normal (corresponding to P_2 values close to -0.5), and therefore the presence of the torus decreases the average lipid order parameter. After the DNP exits the bilayer, the lipid torus is lost and averaged P_2 value decreases. It is also interesting to note that after the expulsion even, a small number of lipids molecules are pulled out of the bilayer and remain nestled within the grooves on the helices (Figure 4.5c). In the 1.0 M NaCl system, the value of P_2 is comparable to that of the system represented in Figure 4.5c (0.3 M, side-on binding), however the bilayer is significantly thicker. This indicates that the reduction in the order parameter that coincides with presence of a lipid torus *does not* have much influence on the bilayer thickness. We also notice that for the transmembrane binding mode, the higher salt concentration causes an increase in the order parameter and a concurrent decrease in the average bilayer thickness, meaning that the usual linear relationship between lipid order parameter and membrane thickness[158] does not apply for these systems.

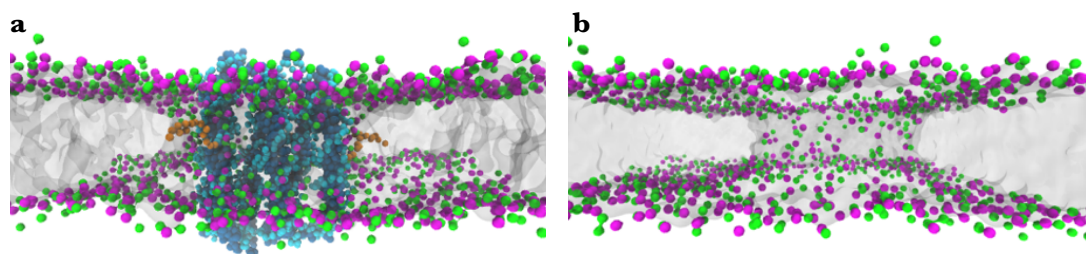


Figure 4.4: Simulation snapshots illustrating the formation of a toroidal pore around the DNP. (a) The positively charged (green) choline moieties of lipids adjacent to the DNP align closely with the negatively charged DNP backbone, causing the adjacent lipids to curve inwards to form a lipid torus, shown clearly in b. (b) The lipid torus with the DNP removed for clarity.

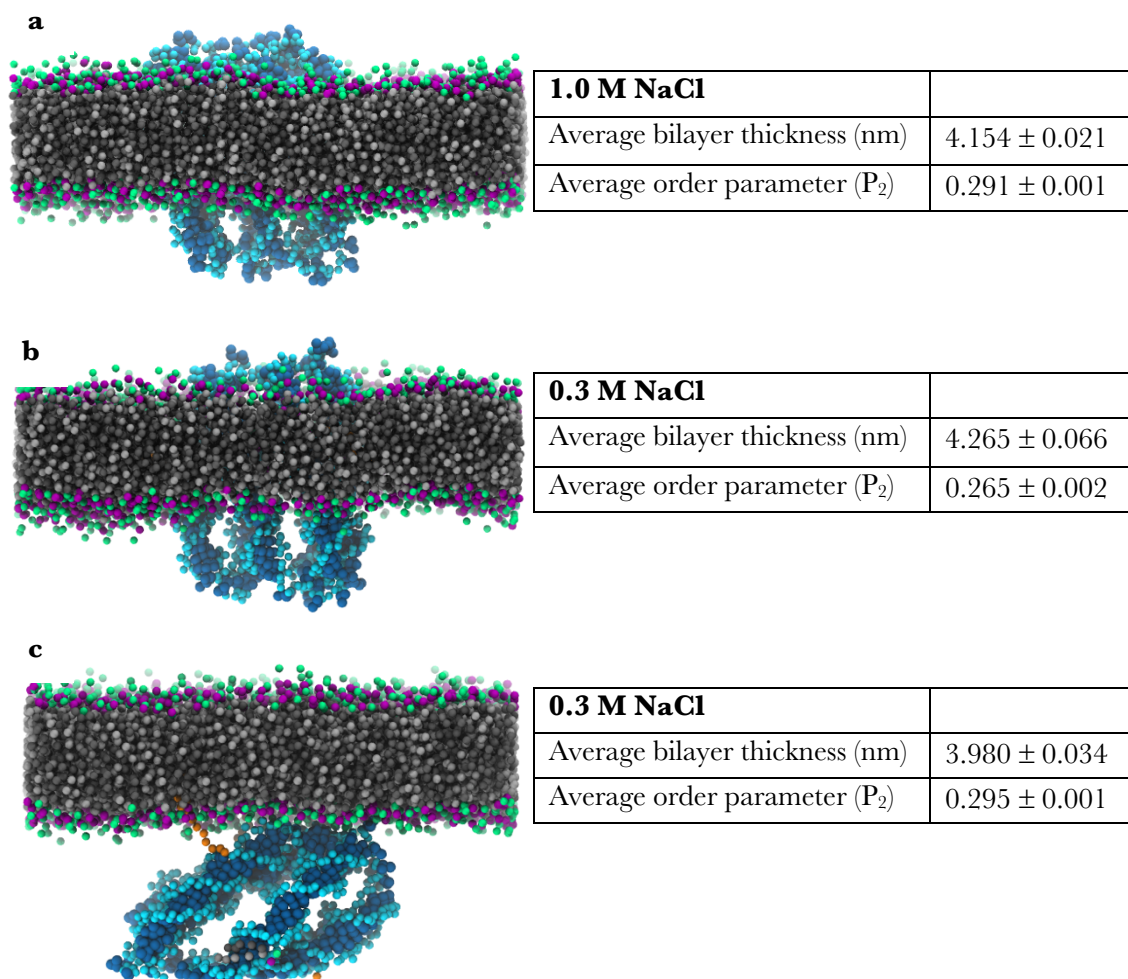


Figure 4.5: Analysis of the influence of salt concentration and DNP binding mode (where applicable) on the ensemble averaged bilayer properties. **(a)** Average bilayer thickness and average second-rank order parameter (P_2) with respect to the bilayer normal for the entire high salt DNP/membrane ensemble, where the DNP remained in the transmembrane configuration throughout each replica. **(b)** Average bilayer thickness and P_2 calculated for the first 300ns of the full low salt DNP/membrane ensemble, where the DNP remained in the transmembrane orientation throughout. **(c)** Average bilayer thickness and P_2 taken from the last 200 ns of the ten trajectories in which the DNP was expelled from the bilayer to adopts a side-on binding configuration.

In each of the 10 trajectories featuring an expulsion event in 0.3 M NaCl, we observe a range of hydrophobic matching effects as the pore exits the bilayer i.e. lipid reorganisation and DNP twisting, tilting and breathing motions. The heightened mobility of the DNP within the membrane in this environment is illustrated in the RMSF per residue plot shown overleaf (Figure 4.6). The DNA backbone is extremely fluctuant; much more so than in the solvated 0.3 M NaCl system (Figure 3.7a). The range of RMSF values in the membrane model is much broader than it is for the solvated model, with much higher upper extremes - around 0.7 nm for the membrane model, compared to 0.4 nm for the solvated model. Furthermore, extreme values are distributed relatively uniformly across the length of the DNP as it spans the membrane, whereas extreme values in the solvated model (0.4 nm) occur mainly within the midsection of the pore, due to the breathing motions of this region.

In the high-salt membrane simulations, we did not observe any signs of the DNP exiting the bilayer at any point, suggesting that the initial transmembrane binding mode is more favourable at higher salt concentrations. The incidence of extreme RMSF values in the high-salt membrane model is very low compared to the low-salt membrane model, and even the most extreme value (0.35 nm) is around 0.1 nm lower than the median value in the low salt RMSF-per-residue plot. Generally, the nucleosides in the membrane-spanning region have RMSF values below 0.2 nm in the high-salt case, indicating that the DNP is far less mobile within the bilayer, and hence less likely to depart. The RMSF-per-residue values for the high-salt membrane model are also lower and more narrowly distributed than those of the solvated high-salt model (Figure 3.7d), suggesting that the midsection of the DNP experiences significant compression and immobilization within the membrane, due to the formation of the toroidal lipid pore (Figure 4.4) around the DNP.

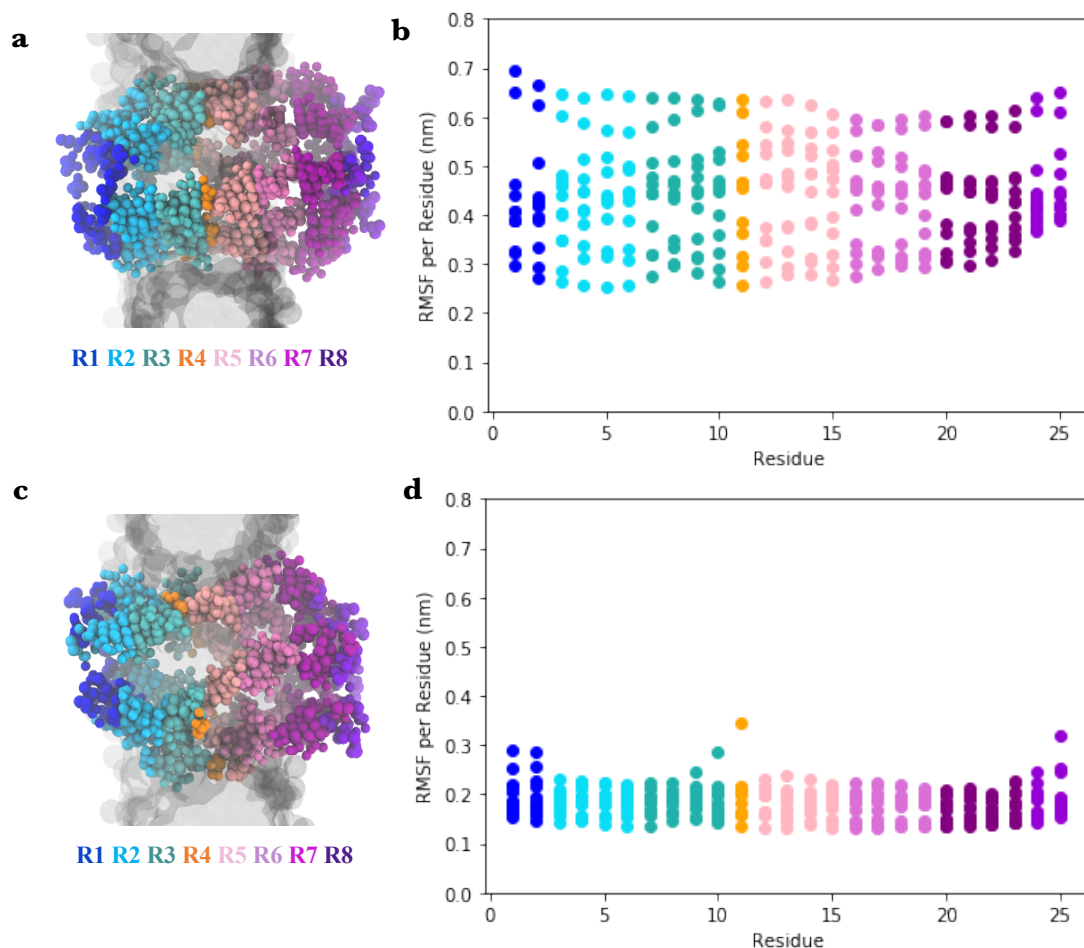


Figure 4.6: (a) Representative structure of the membrane-spanning DNP in 0.3 M NaCl (cholesterol anchors omitted). (b) RMSF per residue plot for the membrane-spanning DNP in 0.3M NaCl, generated from the first 300 ns of the 15 trajectories. The median RMSF value is 0.45 nm. (c) Representative structure of the membrane-spanning DNP in 1.0 M NaCl. (d) RMSF per residue plot for the membrane-spanning DNP in 1.0 M NaCl, generated from the first 300 ns of the trajectories. The median RMSF value is 0.19 nm.

The relative stability of the transmembrane binding mode in 1.0 M NaCl compared to 0.3 M NaCl is also made apparent in the distribution of the total potential energies of these systems, shown in Figure 4.7. The average potential energy of the 0.3M NaCl system (-246×10^5 kcal/mol) is markedly higher than that of the 1.0 M NaCl system (-420×10^5 kcal/mol), suggesting that the additional electrostatic screening proffered by the higher salt concentration makes a large impact on the stability of the DNP/bilayer system. Additionally, the distribution of energies in 1.0 M NaCl resembles a Gaussian

distribution centred around a single global maximum, whereas the 0.3 M NaCl system energies take the form of a bimodal distribution, where there are two separate potential energy wells. The density of the higher energy well on the right-hand side of the plot is approximately 1.5 times greater than the density of the lower energy well, which reflects the ratio of time spent in the transmembrane orientation vs the side-on tethered orientation reasonably well. Given that all fifteen replicas in the 0.3 M NaCl ensemble had the same initial configuration with the DNP spanning the membrane, and the expulsion event in the ten affected trajectories occurred at around 500ns on average, we can estimate that the DNP adopted the transmembrane binding mode for around 66% of the total simulation time (15 μ s) across all replicas. This means the duration of time the DNP spends in the transmembrane orientation within the ensemble is around 1.5 times the duration of time spent in the side-on binding orientation. Hence, it is very likely that the more highly populated, slightly higher energy potential energy well corresponds to the transmembrane binding mode in 0.3 M NaCl.

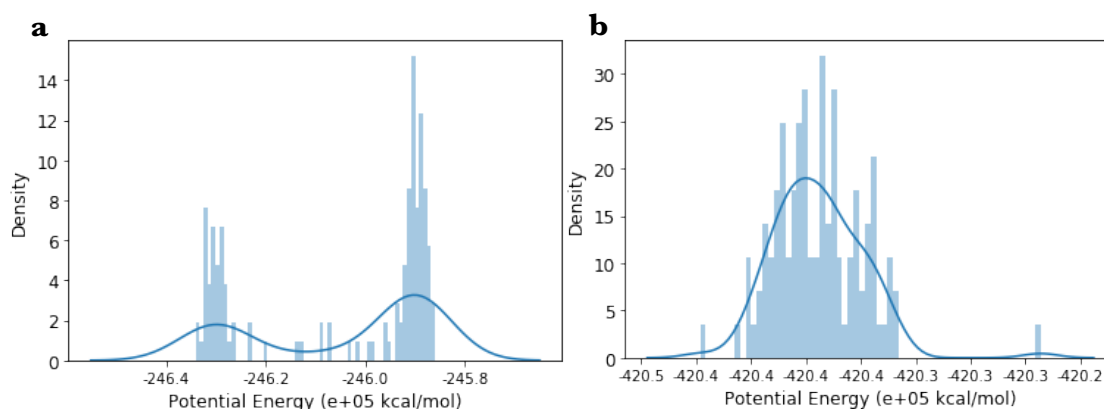


Figure 4.7: Histograms and overlaid estimated probability density functions representing the distribution of the total system potential energy for the membrane spanning nanopore models, calculated from 100ns-long windows across the full ensemble in each case. Probability density functions (smooth curves) were calculated using the kernel density estimator. **(a)** Distribution of potential energies of the 0.3 M NaCl, showing the emergence of two distinct potential energy wells. The larger well on the left-hand side of the plot corresponds potential energies sampled in the simulation windows where the nanopore adopts the less stable (higher energy) transmembrane binding mode. The well on the right represents the lower energy side-on binding mode. **(b)** Distribution of potential energies for the 1.0 M NaCl system, where the nanopore spans the membrane stably throughout the entire ensemble.

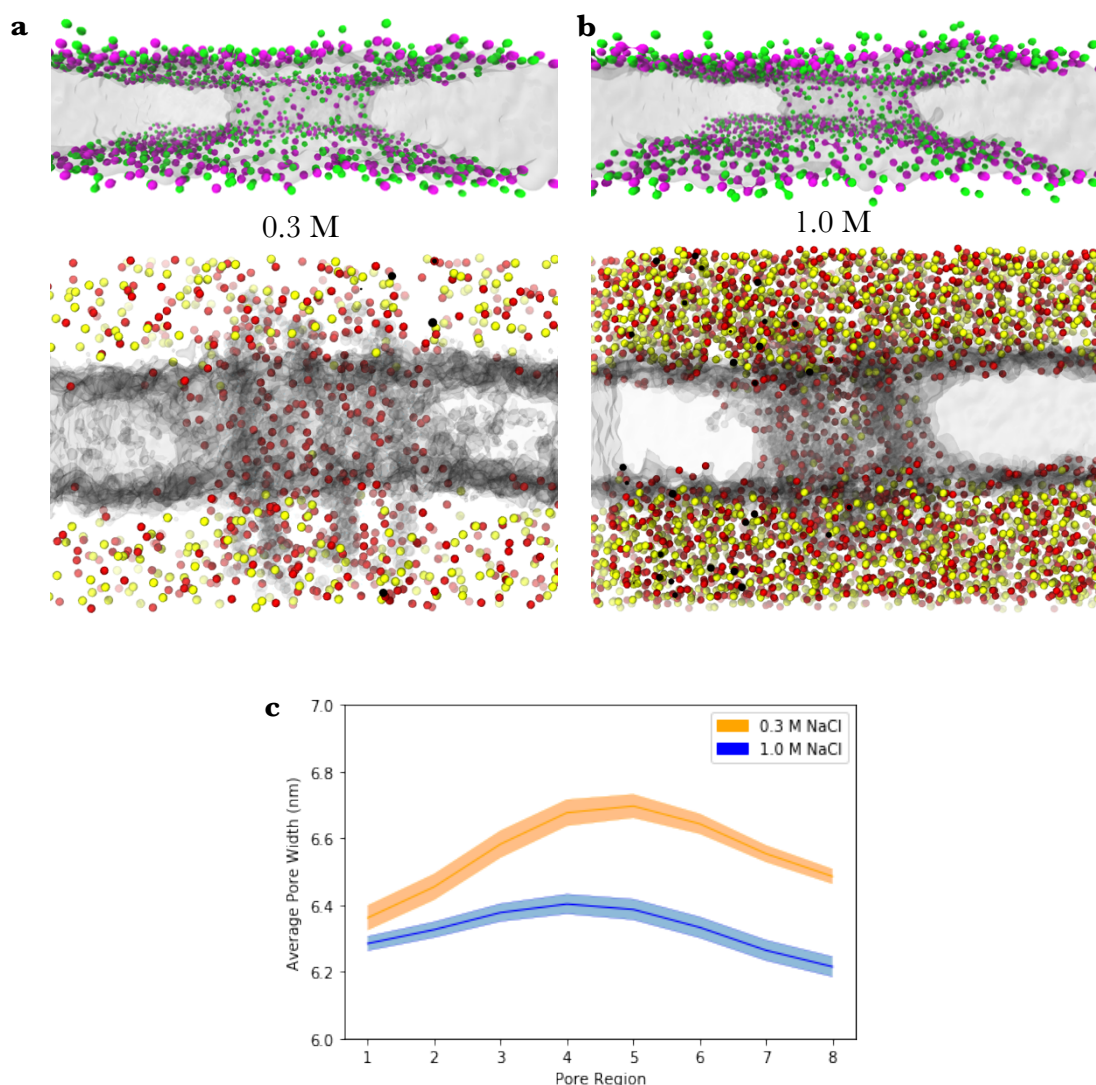


Figure 4.8: Compression of the DNP within the POPC membrane. **(a)** Snapshots taken after 1 μ s of simulation time in 0.3 M NaCl, illustrating the formation of a toroidal pore within the POPC membrane (upper panel, ions and DNP removed for clarity), and the distribution of ions in the system (lower panel). Red spheres represent the sodium ions, and yellow spheres represent chloride ions. Green pseudo-atoms represent the positively charged choline groups, which are electrostatically attracted to the negatively charged backbone of the DNP. Magenta pseudo-atoms represent the neighbouring negatively charged phosphate groups. The hydrophobic region is illustrated in grey. **(b)** Snapshots taken after 1 μ s of simulation time in 1.0 M NaCl, illustrating toroidal pore formation and distribution of ions. **(c)** Plot of average pore width along the pore axis for both ensembles, illustrating the reduction in average pore width in all eight pore regions due to increased pore compression in high-salt conditions.

The additional electrostatic screening in 1.0M NaCl allows the DNP to tolerate the compressive force of the toroidal pore, and therefore the transmembrane binding has a lower potential energy penalty in this medium. The inner edge of the torus is capable of exerting localised pressure to the midsection of the pore, especially when the lipid head-groups and the outer pore surface are oppositely charged, which is the case in this system. In low-salt conditions, compression of a DNA nanopore would be unfavourable, as this would bring densely the negatively charged *ds*DNA backbones together, giving rise to repulsive coulomb forces. Compression in the high-salt model made possible by the enhanced electrostatic screening, which lessens the repulsion between the DNA helices, allowing them to be pushed together to form a stable ellipsoidal lumen. As a result, the average pore width in the high salt condition is 0.24 ± 0.03 nm smaller than it is in the low salt condition (illustrated in Figure 4.10).

The radial distribution functions shown in Figure 4.9 reveals a strong tendency for the sodium ions to aggregate within the pore lumen at both salt concentrations, as the highest peak in sodium ions density occurs immediately prior to the peak DNA density in both distribution functions. However, in the 0.3 M distribution function, the sodium ion density declines steadily beyond 2 nm while the chloride ion density gradually increases beyond this point, indicating that the aggregation of sodium ions in the lumen and around the DNA helices at the DNP/membrane leads to a shortage of sodium ions in bulk solution. Evidence of this can also be seen in Figure 4.8a, where the sodium ions (red spheres) appear concentrated in the DNP lumen and more sparsely distributed in the bulk solution. However, this decline of sodium content in bulk solution (past the 2 nm point) is not seen in the radial distribution function for the 1.0 M NaCl system, and Figure 4.8b illustrates this further. The bulk solvent is rich in both sodium and chloride ions, and chloride ions appear to occupy the DNP lumen in greater numbers than in 0.3 M NaCl. It would follow that the electrostatic potential energies in the more homogenous 1.0 M NaCl solution experience less fluctuation than in 0.3 M NaCl, leading to more stable dynamics. The use of the reaction field method for electrostatics in these simulations makes it difficult to determine exactly how the imbalance between the sodium and chloride ions introduced in the 0.3 M simulation system by the aggregation of sodium in the DNP lumen influences the dielectric electrostatic screening ability of the bulk solvent, because a continuous dielectric constant is assumed outside of the cutoff distance. Further simulations of the 0.3 M

NaCl system with PME electrostatics are needed to elucidate this. Nevertheless, the number density plots show that the density of sodium ions within the pore lumen is 2.2 times greater in 1.0 M NaCl than it is in 0.3 M NaCl. The aggregation of sodium ions in the lumen depletes the bulk in the higher salt concentration to a lesser extent than what is seen at the lower salt concentration; the maximum density is 1.4 times greater than the density at the outer edges of the box in 1.0 M NaCl, and in 0.3 M NaCl the maximum is 2.4 times greater than the density of the outer edges.

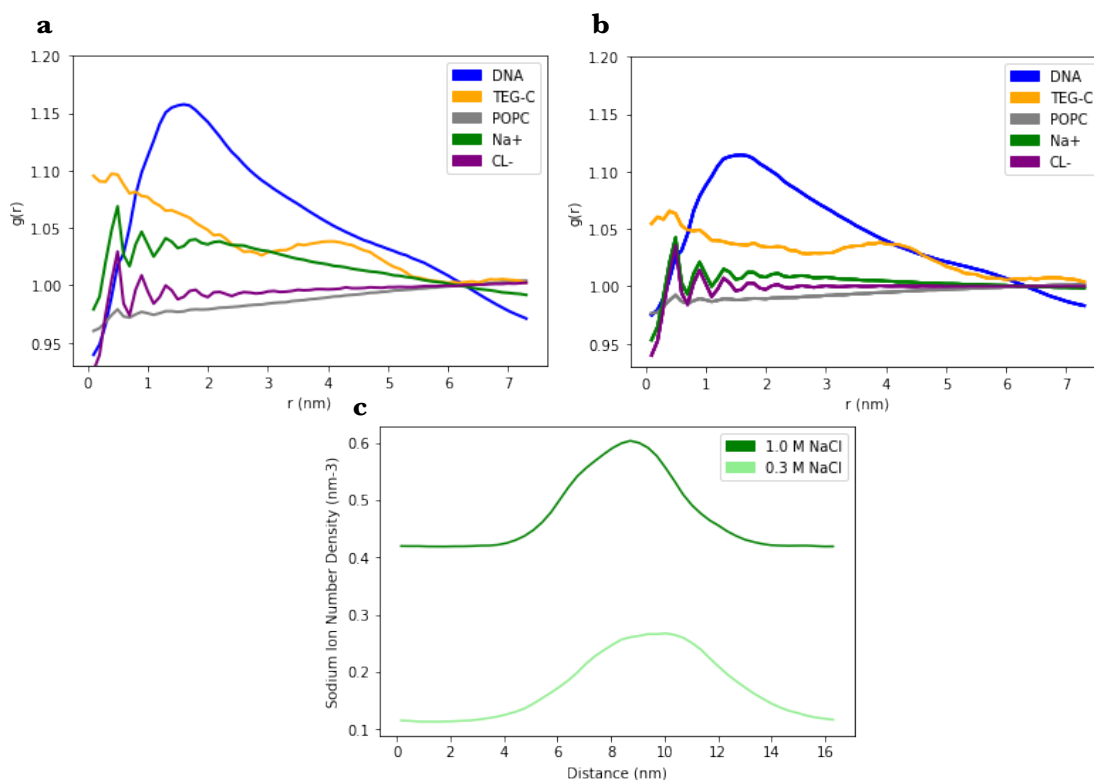


Figure 4.9: Radial distribution functions illustrating the variation of system component densities as a function of distance (r) from the centre of the DNP lumen for **(a)** the first 300 ns of simulation time of the 0.3 M NaCl ensemble and **(b)** the first 300 ns of the 1.0 M NaCl ensemble. The density of the sodium ions within the lumen is higher in 0.3 M NaCl, with a peak at ~ 0.5 nm which tails off gradually as r increases, indicating that their numbers become depleted in the bulk solution, while the density of chloride ions increases as a function of r . **(c)** Averaged number density plots of sodium ions across the simulation box (x dimension), calculated from the frames corresponding to the first 300 ns of the two simulation ensembles. In both plots, the sodium ion density reaches its peak inside the pore lumen, which is centered ~ 10 nm across the box.

The R8 terminus appears to be much more sensitive to the compressive force of the lipid torus than the R1 terminus, and the largest salt-induced decrease in the average pore width appears to occur in the R5 region, located in the pore midsection. This suggests that the R8 terminus may act as an ion sensitive gate, where higher salt concentrations facilitate the translocation of small molecule cargo through the pore lumen. The pore height is 0.16 ± 0.06 nm smaller in the high salt solution than it is in the lower salt condition, indicating that higher ionic strengths cause compression of the DNP in both the lateral and axial dimensions.

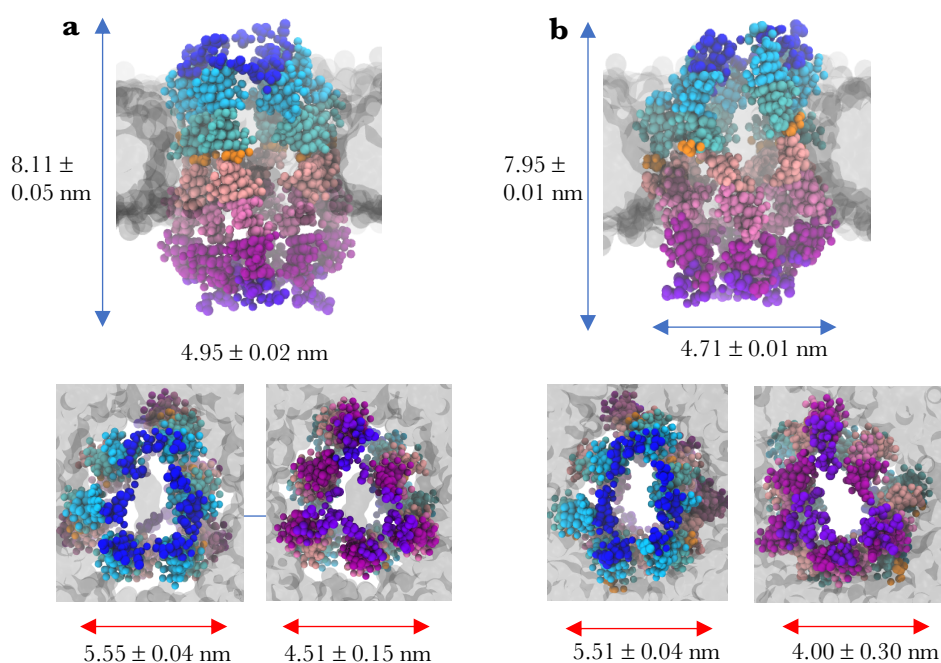


Figure 4.10: Variation of the membrane-spanning DNP dimensions in the two different salt conditions. **(a)** Average height and width of the DNP in 0.3 M NaCl (indicated by the blue arrows, shown above), and the widths of the constrictions at the termini, which correspond to the mean of the calculated minimum pore widths at R1 and R8 (indicated by red arrows, shown below). **(b)** Average height and width of the membrane-spanning DNP in 1.0 M NaCl, and the widths of the constrictions at the termini.

Kinking of the helices in the was observed in the high salt simulations, which was particularly extreme in helices 3, 4 and 5 and most likely a contributing factor to the axial compression and shortening of the DNP. Kink angles were calculated for the affected helices and these are shown in in Figure 4.11. The angles of the kinks in helices 3 and 4 were centred around the R4 region, whereas helix 5, which was the most susceptible to kinking, exhibited large kink angles in the R2 and R4 regions. The kink angle of $\sim 72^\circ$ in the R2 region of helix 5 was severe enough to cause partial unraveling of the helix in this area, with the small, frayed strand section sitting on the surface of the surrounding lipid torus like an anchor.

	Calculated 0.3 M NaCl	Calculated 1.0 M NaCl
Average Pore Height (nm)	8.11 ± 0.05	7.95 ± 0.01
Average Pore Width (nm)	4.95 ± 0.02	4.71 ± 0.01
Average Pore Width at R1 (nm)	5.95 ± 0.02	5.85 ± 0.02
Average Pore Width at R8 (nm)	5.84 ± 0.03	5.97 ± 0.01
Minimum Pore Width at R1 (nm)	5.55 ± 0.04	5.52 ± 0.04
Minimum Pore Width at R8 (nm)	4.51 ± 0.15	4.00 ± 0.30

Table 4.1: Comparison of the calculated pore dimensions of the membrane/DNP models at two different salt concentrations. All calculated values correspond to the mean \pm the standard error obtained from an ensemble of 15 trajectories.

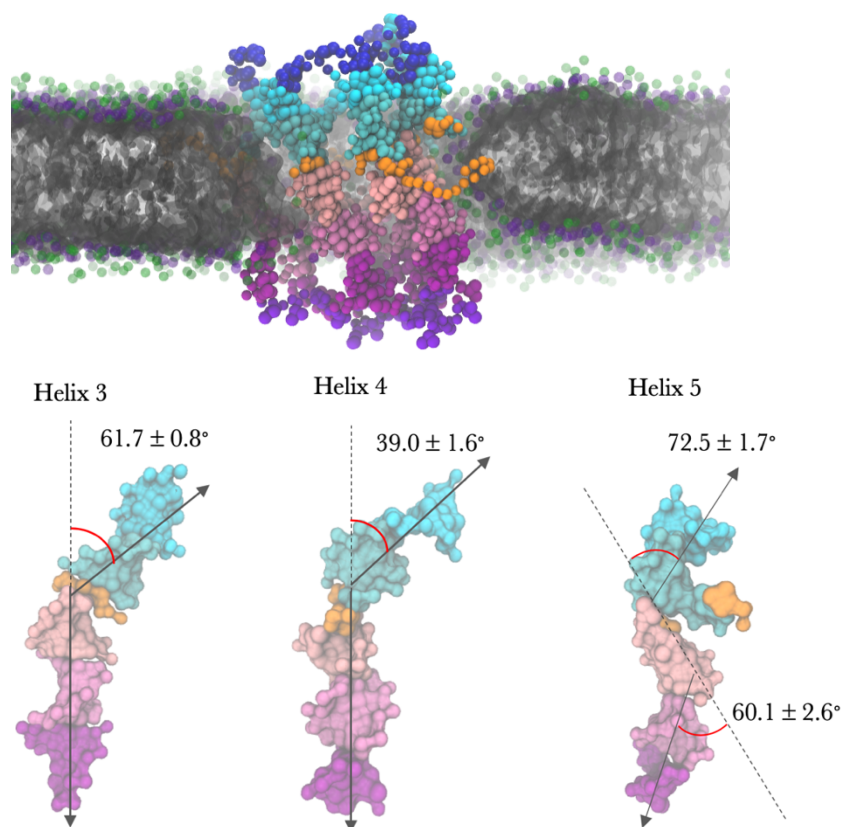


Figure 4.11: Helix kink angles calculated for helices 3, 4 and 5 in the 1.0 M NaCl DNP/membrane model. The order in which the helices appear in the top panel corresponds to their order in the bottom panel. The kink angle of helices may contribute to the decrease in average pore height that occurs upon increasing the NaCl concentration from 0.3 M to 1.0 M.

Mean Inter-Helix Distance (nm)	0.3 M NaCl	1.0 M NaCl
Helix 1 – 2	3.51 ± 0.20	2.56 ± 0.01
Helix 2 – 3	3.80 ± 0.30	2.38 ± 0.02
Helix 3 – 4	2.74 ± 0.09	2.46 ± 0.03
Helix 4 – 5	2.42 ± 0.04	2.14 ± 0.04
Helix 5 – 6	3.53 ± 0.24	2.54 ± 0.03
Helix 6 – 1	3.64 ± 0.27	2.32 ± 0.01
Average	3.26 ± 0.02	2.40 ± 0.10

Table 4.2: Comparison of the distances between adjacent helices in a membrane-spanning DNP at the two different salt concentrations. The marked reduction of the average inter-helix distance upon increasing the salt concentration gives evidence for the stabilising effect of additional electrostatic screening on the membrane-spanning pore.

Further evidence of the higher extent of lateral pore compression in 1.0 M NaCl is given by inter-helix distance data shown in Table 4.2, which clearly illustrates a dramatic reduction in the average inter-helix distance of ~ 0.9 nm upon increasing the salt concentration to 1.0 M NaCl. Additionally, comparison of the lumen width profiles (Figure 4.12) reveals a striking difference in the width and shape of the lumen between the two salt conditions. Both lumen profiles shown noticeable constrictions within the lumen at R2 and R7, due to their close proximity to the strained helix crossovers at R1 and R8. Increasing the salt concentration causes a uniform narrowing and smoothing of the lumen profile, with an additional constriction at R4⁻ where the cholesterol anchors are located. In theory, the hydrophobic cholesterol groups act simply as hydrophobic anchors, allowing the highly charged DNA nanopore span the lipid bilayer. The observation of this lumen constriction at the site of TEG-cholesterol anchoring suggests that the presence of hydrophobic cholesterol anchors, which align themselves parallel to acyl chains of the lipids making up the torus, increasing the force of attraction between the positively charged lipid headgroups and the negatively charged midsection of the DNP, leading to a concurrent increase in pressure exerted on the DNP by the lipid torus.

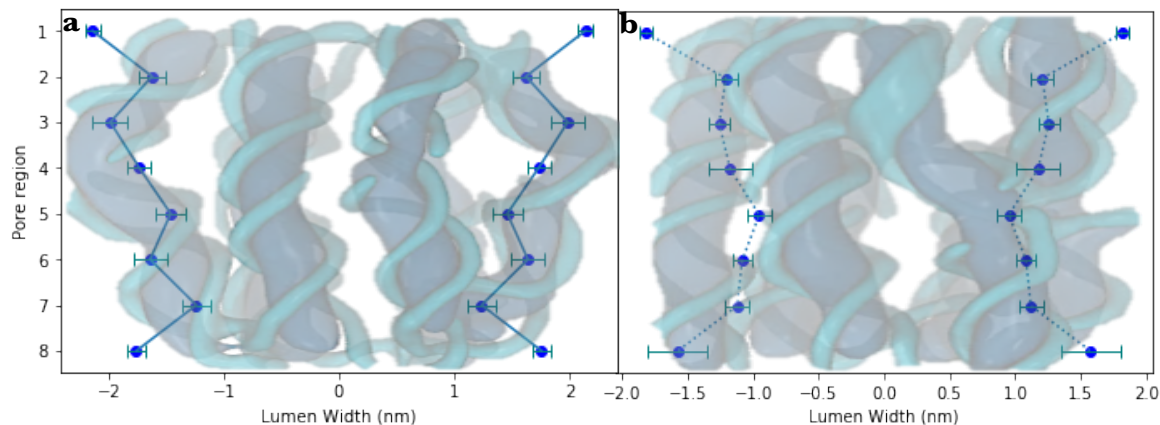


Figure 4.12: (a) Averaged lumen width profile of a solvated DNP in 0.3 M NaCl, generated from an ensemble of fifteen replicas. The lumen width for each region was taken as the mean of three minimum distances measured between opposing DNA helices. The computed average lumen width is 3.39 ± 0.02 nm. (b) Averaged lumen width profile of a membrane-spanning DNP in 1.0 M NaCl. The main constriction within the pore lumen at R4 corresponds to the position of the TEG-cholesterol anchors. The computed average lumen width is 2.21 ± 0.01 nm, which is well in line with the expected lumen width of ~ 2.0 nm confirmed by PEG sizing experiments. The profiles have been symmetrised for ease of visualisation.

4.5 Conclusions

The varied spread of interesting behaviors that were observed in these highly dynamic simulation systems highlights the need for ensemble-based CG protocols for exploratory studies of novel and unpredictable DNA nanoarchitectures such as the TEG-cholesterol anchored DNP studied here. In the low-salt case, a propensity for the DNP to exit the bilayer caused the ensemble to separate into two clusters, revealing two distinct kinds of behavior corresponding to two different membrane binding modes. Capturing these sub-states, which involve biomolecular motions that occur at micro-second timescale with traditional brute-force all-atom MD whilst maintaining close control of errors is not possible without incurring huge costs. If we are to make direct comparisons between the binding modes that are observed in simulation and experiments, thorough elucidation of the free energy difference between the two binding modes is required, alongside simulated binding and insertion kinetics using enhanced sampling simulation techniques e.g., metadynamics. [159], [160]

The conditions in the simulations are of course not completely reflective of the conditions employed in experimental studies of DNPs in the presence of membranes. These studies use significantly larger membrane patches (μm length scale) than the 16 nm x 16 nm simulated POPC patch, KCl rather than NaCl, and slightly alkaline buffers with pH of 8.0. [4], [38] Not much is known about the influence of pH on planar homogenous lipid membranes such as the PC bilayers used in this computational work the published experimental literature, but preliminary fluorescence microscopy-assisted vesicle migration studies by Angelova et al [161] have suggested that higher pH values can decrease the surface tension of DOPC membranes. Due to the simplified representation of ions and water in the MARTINI force-field, it is not possible to exactly recreate the solvent conditions, and the use of reaction-field electrostatics with the polarisable water model must be validated for this system by comparing the results to simulations performed with PME electrostatic treatment. Ideally, this should be done with an AA force-field, as MARTINI was parameterised with short range shifted electrostatic interactions, so long range electrostatics are considered less reliable in MARTINI CG simulations, regardless of the method used for calculating electrostatics.

In these simulations, we observe a strong tendency for sodium ions to accumulate within the pore lumen in both salt conditions, due to the density of negative charges on the DNP interior, and the density of sodium ions within the DNP lumen is of course greater in the higher salt condition due to the abundance of ions. The expulsion of the DNP in 0.3 M NaCl, and the subsequent side-on membrane binding mode suggests that the lower salt concentration does not provide sufficient electrostatic screening between the DNA helices of the DNP in CG models. The inter-helix repulsion within the DNP is apparently too large to withstand the compressive force of the membrane, and therefore it opts to bind to the bilayer side-on instead. A concentration of 1.0 M NaCl appears to stabilise the transmembrane binding mode, and the DNP is compressed within the lipid bilayer to form an ellipsoidal pore, with kinks forming within the DNA helices in response to the compressive force of the lipid torus that forms around it. The simulated pore lumen profile in 1.0 M NaCl is in excellent agreement with the values obtained from PEG-sizing experiments, though more experimental data are needed to validate properties such as helix kink angles and bilayer thickness. The relationship between the lipid order parameters and the bilayer thickness is atypical in the presence of the DNP, and this is difficult to explain without further experimental data (e.g. x-ray scattering), or higher resolution all-atom MD simulations.

Chapter 5

Computational Electrophysiology Simulations of Lipid Bilayer Spanning DNA Nanopores

In this chapter, we explore the conductive properties of a membrane-spanning DNP and assess the relationship between the dynamics of the membrane/DNP system under an applied potential difference and its ability to transport ions across the membrane. The current-voltage relationship and average conductance for this particular DNP have already been established in published experimental literature, primarily through single-channel current recordings (discussed in Chapter 1.5). The DNP exhibits ohmic conductivity under applied voltages below 60 mV, but the average conductance appears to decrease at higher voltages – a phenomenon which has been attributed to conformational changes to the DNP that impede its conductivity. Confirming this hypothesis experimentally would be very challenging, so I have utilised coarse-grained simulation methods to analyse the conductive properties of the membrane-spanning DNP, and any conformational changes that the system experiences as a result of ion transport and the application of a potential difference to the bilayer normal. I have used the computational electrophysiology protocol [162], [163] developed by the creators of GROMACS [98] to derive current-voltage relationships for the membrane-spanning DNP at a range of instantaneous voltages, and to assess the dynamic behavior of the system in response to the resultant voltage applied normal to the bilayer.

5.1 The CompEl Protocol

The computational electrophysiology protocol was developed within GROMACS to address the issues that arise when one attempts to introduce an applied electric field into a simulation system with periodic boundary conditions and fluctuating dielectric properties. Periodic boundary conditions are particularly necessary in membrane simulations to avoid finite size effects (discussed briefly in Chapter 4.1), and one of the consequences of these conditions in the z-direction is that the ionic solution on either side of the bilayer plane will have the same composition. This means that ion concentrations on either side of the bilayer are always equivalent, making it difficult to establish a transmembrane potential difference. [164] The magnitude of the total electric field experienced by the simulation system depends on the side of the

simulation box, as well as the dielectric properties of systems. In simulation systems with fluctuating dielectric properties (such as the dynamics membrane/nanopore systems being studied here), it is difficult to impose a constant electric field as the factor by which net electric field is influenced by the fluctuating dielectric must be corrected for throughout the simulation. The CompEl protocol provides a simple solution to this problem by cloning the system in the z-axis, allowing the user to define two solvent compartments which differ in their ion concentration which introduces a sustained potential difference across the bilayer.

For a typical CompEl simulation, a double-bilayer system is constructed by cloning the nanopore/membrane system in the z-direction, such that two distinct solvent compartments (A and B) are formed, corresponding to the *cis* and *trans* compartments in a traditional chip-based parallel bilayer recording instrument. Virtual cylinders are defined around the two parallel membrane-spanning pores, a point charge difference (Δq) is applied to the system in each simulation, and the net influx of cations and anions into each compartment is recorded at regular intervals.

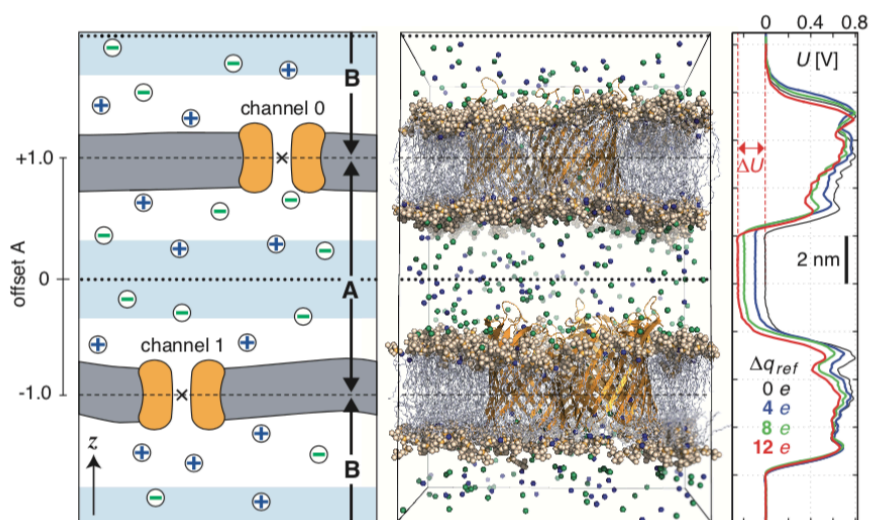


Figure 5.1: Illustration of a double-bilayer system built for CompEl simulations. Ion/water swaps between the two solvent compartments (A and B) are performed after an ion traverses through the pore from one compartment into the other to re-establish a charge imbalance between the two compartments, resulting in a sustained potential difference – denoted by ΔU shown in the plot shown on the right. Image adapted from Kutzner *et al.*[162]

Whenever the ion count in each compartment differs from the reference count (defined by Δq), ions from one compartment are swapped with water molecules from the other compartment to restore the reference count. Once the simulation is complete, the trajectory is split into slices – each slice typically corresponding to 20ns of simulation time. [63], [162] For each trajectory slice, the potential difference (V) is calculated, and the instantaneous current flowing through each pore (I) is calculated according to the following expression:

$$I = \frac{\sum(q_i \times n_i \text{ of swaps})}{2\Delta t} \quad (5.1)$$

where q_i is the charge of the ion, multiplied by the number of ion/water swaps for that ion. The division by two is required as two bilayers are simulated in the CompEl protocol. The computed current and instantaneous voltage for each trajectory slice constitutes a single data point on the computed current-voltage relationship. The gradient of the trend-line yields the average conductance of the nanopore/membrane system.

5.2 Model Building and Simulation Details

To generate an appropriate simulation system for the CompEl simulation ensemble, a new membrane-spanning DNP model was constructed with a smaller membrane patch than the one that was used in the equilibrium simulations discussed in Chapter 4.1. The fully equilibrated DNP model (without membrane) was taken from the 1.0 M NaCl DNP/membrane ensemble and inserted into a 13 nm by 16 nm membrane patch, and solvated in a 13 nm x 16 nm x 16 nm box containing 1.0 M NaCl solution (MARTINI polarisable water model) using the *insane.py*[96] script. The system was then cloned in the z-axis using the *gmx genconf*[98] tool to create a double bilayer/DNP system, consisting of ~90,000 pseudo-atom beads. Each simulation was run for 100 ns, during which ions were counted every 0.2 ps, and ion/water swaps were made between the two solvent compartments to restore the original charge imbalance Δq . The speed of ion diffusion is dictated by the ion diffusion coefficient, which is itself dependant on the identity of the ions, the chemistry of the ion channel, the concentration of the solution and the temperature. In solutions with concentrations between 0.1 and 1 M,

the diffusion coefficients of sodium and chloride ions are reported to be between 1 and 2 nm² ns⁻¹ at room temperate, so we can expect to witness many diffusion events in a simulation system of this size over the course of 100 ns.[165] Kutzner et al[162], [163] achieved accurate predictions for the conductance and ion selectivity of wild-type and mutated bacterial porin PorB with all-atom CompEl simulations in 1.0 M and 0.2 M NaCl solutions with simulation durations of 100 ns, so I chose to run multiple replicas of this duration, increasing the number of replicas until the error associated with the calculated conductance remained constant (Fig. 3.5b). There are currently no published studies of coarse-grained membrane/nanopore systems that make use of the CompEl protocol, so the simplified representation of water along with the underestimation in the size of hydrated chloride ions in the MARTINI force-field may introduce inaccuracies in the ion diffusion coefficient for this system. However, the excellent agreement between the computed and experimental studies indicate that these effects are minor.

For each trajectory frame, the potential difference (V) across the membrane was calculated using the *gmx potential* [63] tool, by summing the charges in the two solvent compartments and taking the second integral of the charge distribution on either side of the bilayer. The system was equilibrated using the same procedure described in Section 4.1. Production simulations were run in the NPT ensemble for a duration of 100 ns at a temperature and pressure of 300K and 1.03 atm, with temperature and pressure control maintained with the velocity-rescale thermostat and the Berendsen barostat. Electrostatics were computed using the PME method, with a cut-off of 1.2 nm. Twenty-five replicas were performed in total; nine with $\Delta q = 0$, eight with $\Delta q = 2$, and eight with $\Delta q = 4$, the charge imbalance values recommended in the CompEl protocol.[162] The transmembrane voltage at each timestep U_t is related to Δq according to the following relationship:

$$U_t = \frac{\Delta q_t}{C} \tag{5.2}$$

where C is the membrane capacitance, which dictates how the membrane voltage responds to changes in current. As the capacitance of the DNP-perforated membrane is not *a priori* known, the transmembrane voltage must be calculated at each time-step

(by integrating the charge distribution). The high frequency of ion/water swaps and relatively small membrane patch caused the instantaneous potential difference to fluctuate sharply throughout each simulation, so some of the trajectory slices were omitted as the potential difference rose beyond the range of voltages that are typically employed in electrophysiology experiments on DNA nanopores (± 100 mV). Around 100 usable data points were extracted from the full set of trajectories, and these were invoked to construct a current-voltage (I/V) curve. Bootstrap error analysis confirmed that twenty-five replicas yielded a converged value for the average conductance G , which was taken as the slope of the line fitted to the I/V curve.

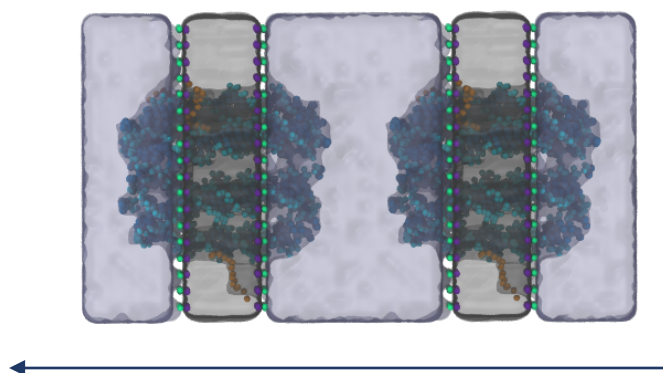


Figure 5.2: The pre-equilibrated double bilayer/DNP system set up for CompEl simulations. The arrow represents the alignment of the simulation box with respect to the z-axis (the vector of ion translocation).

5.3 Computational Resources

All CompEl simulations were performed on the UK national supercomputer ARCHER[123], with each simulation being run on 240 CPU cores, as the simulation box was too small to be divided amongst more cores than this. These simulations consumed 88,000 core hours in total. The computational resources used for this work were provided by the CompBioMed Centre of Excellence.[125]

5.4 Results and Discussion

At the beginning of this investigation, I had to determine an appropriate number of CompEl simulation replicas to ensure a converged result for the calculated average conductance, with a reasonably low standard error. Tracking the evolution of the average conductance (calculated with equation 5.1) and its associated standard error reveals that the ensemble of 20ns windows with instantaneous voltages below 100 mV (a voltage range that is commonly used in experiment) yielded from the 25 simulations was sufficient to provide a reliable value for the computed conductance (1.19 ± 0.49 nS). This value was in good agreement with the experimentally determined conductance of 1.59 ± 0.07 nS. The magnitude of the errors associated with the average conductance is noticeably higher than errors calculated for other macroscopic quantities discussed in the previous chapters, and the reason for this is tied to the specific features of the CompEl protocol. As mentioned in Section 5.1, ions are swapped with water molecules at regular intervals throughout the simulation as ions proceed to traverse the nanopore, in order to restore the point charge mismatch and allow the transmembrane potential difference to persist. However, these frequent swaps often cause the magnitude of the voltage to oscillate quite rapidly throughout the simulation, increasing the uncertainty of the instantaneous voltage in each window. This effect is usually more pronounced in smaller membrane patches, and in simulations with shorter timescales with fewer frames outputted. The voltage is calculated retrospectively by integrating twice over the charge density along the transmembrane pores' principal axis at each timestep (or number of timesteps). A smaller membrane patch means the charge density will be affected significantly by each ion-water swap event, and the fluctuations in voltage will be greater. The larger the membrane patch, the more negligible the effect of these events. The effect of shorter timescales and less frequent frame collections simply results in less configurations sampled, and therefore larger jumps in voltage as time goes on.

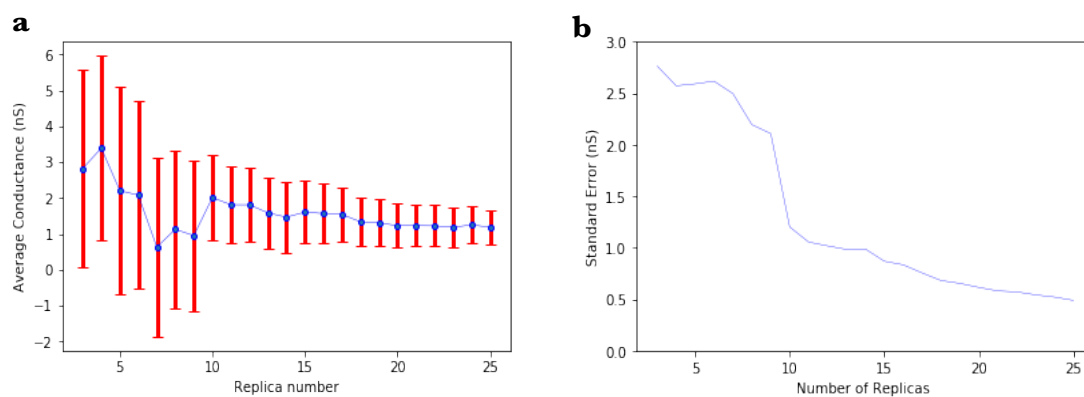


Figure 5.3:(a) Ensemble average plot illustrating the convergence of the average conductance (nS) as the ensemble size increases and **(b)** the decay of the standard error associated with the average conductance with increasing number of replicas (right). The ensemble averaged conductance was computed from a set of 100 x 20ns windows with an instantaneous voltage below 100 mV, converging at 1.19 ± 0.49 nS.

The conductive properties of the DNP were originally investigated experimentally, with standard painted lipid bilayer techniques for single-channel current recordings. The experimental value (1.59 ± 0.07 nS) was obtained from a conductance histogram constructed from 97 individual current recordings taken at + 20 mV. The computed I/V curve derived from the ensemble of CompEl simulations sampling within the voltage range 0-100 mV gave an average conductance G of 1.19 ± 0.49 nS. Taking the error into account, our computed value overlaps comfortably with the experimental value. The ion selectivity was calculated by taking the average ratio of the flux of each of the ion species (Na^+ and Cl^-) from the 100 x 20ns windows that were used to construct the I/V curves. Predictably, the negatively charged DNP exhibits a preference for cations due to its dense negative charge, with a ratio of cations to ions of around 3:1 observed, as illustrated by Figure 5.5. The ion selectivity has not yet been determined experimentally, however dye efflux assays with DNP-gated vesicles containing charged fluorophores have demonstrated that the DNP preferentially transports sulpho-rhodamine B (net charge of 1-) over carboxyfluorescein (net charge of 3-) with a ratio of around 10:1, indicative of a strong preference for less negatively charged cargo.[38] Overlaid snapshots from CompEl trajectories reveal that sodium ions have a propensity to diffuse through the DNP/membrane interface, due to the formation of toroidal lipid pores around the DNP (discussed in Chapter 4), while

chloride ions appear to only travel through the DNP lumen. The published experimental work mentions the emergence of sub-conductance states at higher voltages, generally in the 60 - 100 mV range.[3] Evidence of this phenomenon has also been observed in the CompEl simulations, illustrated in the I/V curves below. There is a pronounced reduction in the slope of the I/V curve with the inclusion of current measurements taken at higher voltages (above 50 mV), and the average conductance decreases as a result.

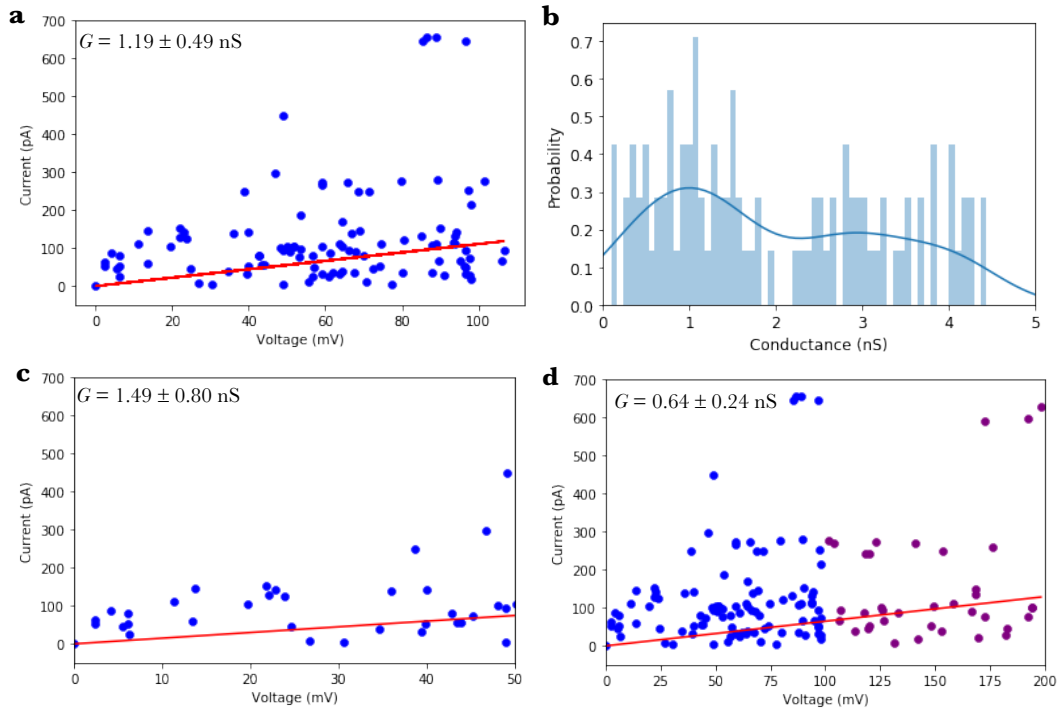


Figure 5.4: Series of plots illustrating the decline of the average conductance with the inclusion of measurements taken at high instantaneous voltages, due to the emergence of sub-conductance states. **(a)** I/V curve generated from 100 x 20ns slices with instantaneous voltages below 100 mV, yielding an average conductance G of 1.19 ± 0.49 nS. **(b)** Simulated conductance histogram constructed from the data from plot **a**. **(c)** Computed I/V curve generated from data taken from 40 x 20ns trajectory slices with instantaneous voltages under 50 mV (closest to the range used in experimental conditions), taken from 16 replicas. The slope of the red line gives an average conductance G of 1.49 ± 0.80 nS, which is closest in agreement with the experimental conductance, albeit with a larger standard error. **(d)** I/V curve generated from 136 x 20ns slices with voltages below 200 mV, with an average conductance G of 0.64 ± 0.24 nS. The marked decrease in average conductance is likely a result of the inclusion of sub-conductance states sampled at higher voltages.

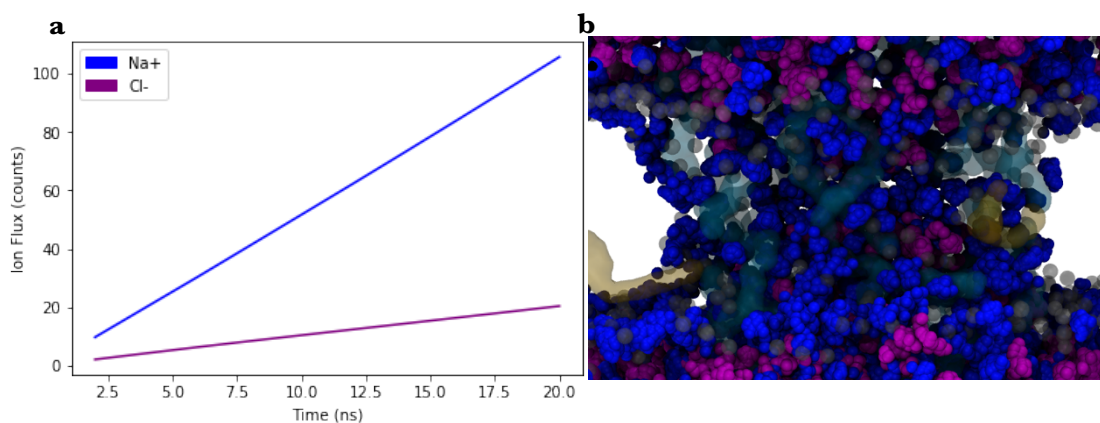


Figure 5.5(a): Ion flux calculated as a function of time, over the 100 windows taken from 25 replicas. The ratio of the gradient provides a good measure of the ion selectivity. The DNP exhibits a monovalent cation selectivity of 3.11 ± 0.23 at the ns timescale. **(b)** Overlaid frames from a CompEl trajectory illustrating the ion transport routes. The nanopore is shown in a transparent cyan surface representation, and the lipid headgroups of the toroidal lipid pore are shown as transparent grey spheres. The of sodium ions (blue) traversing the pore lumen outnumber the chlorine ions (purple), as shown in the ion flux plot, and it appears that sodium ions are transported along the DNP/toroid interface as well as through the DNP lumen.

While the original experimental study speculates that the diminished conductance at higher voltages is a result of a conformational changes within the DNP, no significant conformational changes were observed in the CompEl simulations. The most noteworthy observation that was made in these CompEl simulations was related to the motility of the water molecules. In several replicas featuring relatively high instantaneous voltages, the water molecules had a tendency to “jam” within the pore lumen – causing the density of the system to increase dramatically, and the lateral movements of all components of the system other than the lipid alkyl chains to dampen significantly. Similar behaviour has been reported by in simulations studies on graphene nanopores performed by Wilson *et al*[166], which reports water-compression gating that arises from the dielectrophoretic force of water moving through a narrow lumen under an applied voltage. The authors report a significant increase in the local density of water particles dwelling within the graphene pore lumen at voltages above 200 mV, and the magnitude of the pressure acting along the pore axis increases rapidly

as the voltage is ramped up from 500 mV to 1000 mV. This in turn caused a pore blockade that was severe enough to impede the translocation of *ds*DNA strands in constant-velocity steered MD simulations of DNA transport through graphene pores. This effect was observed in porous graphene membranes of varying thicknesses (from one atom to five atoms thick), and the diameter of the nanopore lumen was typically between 2.9 and 3.5 nm – which match the lumen diameter profile of the DNP very closely. To determine whether or not dielectrophoretic effect could be causing the emergence of the sub-conductance state at higher voltages, further analyses of local water densities and ion mobilities could be performed on the CompEl windows. As the windows with appropriately high voltages are of a relatively short duration, longer simulations under constant applied voltages may also be necessary, but this is beyond the scope of this thesis.

5.5 Conclusions

Here, I have demonstrated how an ensemble-based CG approach using the CompEl protocol for simulated electrophysiology can be used to accurately estimate the average conductance of a membrane-spanning DNP, validated by existing experimental literature values. The DNP displays selectivity towards sodium cations, which diffuse freely through the DNP lumen and the DNP/lipid interface. These CompEl simulations also captured the voltage-dependant suppression of conductance that has been observed in previous experimental studies [3], which could be the result of dielectrophoretic compression of water within the narrow channel lumen (evidenced by a decrease in solvent mobility and increase in the system density) but further work must be done to explain the mechanism that gives rise to this phenomenon. Though converged, the magnitude of the standard errors associated with the mean conductance values were somewhat higher than experimental the experimental errors, which was reported as the standard deviation of a sample of ~ 100 independent single channel current recordings. The relatively high errors reported here are tied to specific features of the CompEl protocol itself, as discussed previously in Section 5.4. These errors can be improved by running further simulations with a larger membrane patch, which lessens the influence of rapid water-ion exchanges on the transmembrane potential difference, causing less fluctuation and therefore less uncertainty in the instantaneous voltage.

Chapter 6

Constant-Velocity Steered Molecular Dynamics Simulations of Fluorophore Translocation through a Lipid Bilayer-Spanning DNA Nanopore

While the TEG-cholesterol anchored DNP studied in this thesis has many useful potential applications in biotechnology, their design and ongoing development is primarily geared towards drug delivery applications. As such, we must assess and understand their ability to mediate the transport of various types of molecular cargo and identify any size or charge selectivity the pores may exhibit before we can deploy them successfully for drug delivery. To date, the translocation of only a small number of small charged fluorophores through modified DNP's has been studied using fluorescence confocal microscopy, namely sulpho-rhodamine B (SRB) [3],[48], carboxyfluorescein (CF) [3] and ATTO 655 (a fluorescent dye manufactured by ATTO-TEC).

As discussed earlier in Chapter 1.4, fluorescent dye influx/efflux assays allow us to calculate the rate of fluorophore translocation through a ligand-gated DNP. Differences between the translocation rates of different fluorophores may indicate selectivity towards fluorophores with different charges and/or steric properties, and the existing literature postulates that the relatively high translocation rate of SRB compared to that of CF is due to charge selectivity of the DNP.[3], [48] SRB and CF are structurally very similar, but differ in size and net charge, with the former possessing two negatively charged groups and one positively charged group in solution, and the former possessing three negatively charged groups. In addition, the bulky sulphate and diethylamine moieties make SRB the larger of the two molecules, making the charge density of the CF molecule greater than that of SRB. It is proposed that density of negative charges distributed along the interior walls of the DNP repels the more negatively charged fluorophores from the pore lumen, impeding translocation. More recent (unpublished) dye efflux assays have successfully made use of ATTO 665, which has shown influx rates comparable to SRB.

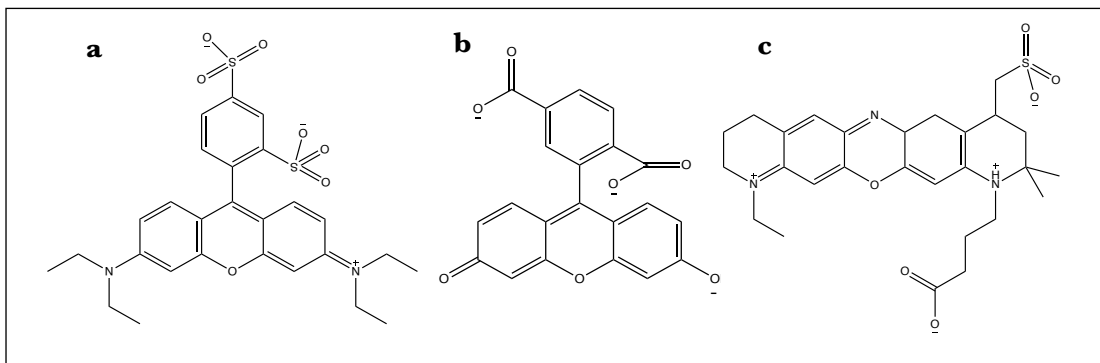


Figure 6.1: Skeletal representations of the three fluorophores. (a) SRB (b) CF and (c) ATTO 655. ATTO 655 has no net charge, while SRB and CF have net charges of -1 and -3 respectively.

The existing fluorophore efflux study[3] also proposed that the translocation of SRB through SUV-encapsulated DNP's proceeds mainly via the pore lumen, as the fluorescence signal of ligand-gated DNP's containing SRB increased substantially after the ligand gate was opened following the addition of a DNA “key” that hybridises to the lock strand that bundles the overhanging *ss*DNA loops at gate entrance. It is worth noting here that the design of the DNPs used in the efflux assays makes use of longer DNA strands, to introduce these overhanging loops, therefore these DNPs may have slightly different structural and functional properties to the DNPs studied in this thesis. The translocation of CF was generally quite poor and oddly, the rate of dye efflux through the pore in its open state (with the correct key added) was lower than the rate of efflux in the presence of the mismatched key, where the SUV-spanning nanopore is assumed to exist in its closed state (Fig 6.2c). This suggests the possibility of a secondary translocation pathway that proceeds outside of the DNP lumen, where the likelihood of one pathway occurring over another is dependent on the net charge and/or size of the translocating fluorophore. Additionally, the kinetic release trace (Fig. 6.2b) constructed for SRB showed that the background SRB emission at $t=0$ (immediately after the addition of the DNA key) was slightly higher than that of CF, and the low efflux rate of SRB through the closed pore (Fig. 6.2c) suggests that freely translocating fluorophores such as SRB can “leak” through the closed DNP, which may also be a result of a secondary translocation pathway through the DNP/toroid interface. Here, I discuss the putative basis of this behavior by analysing the results from a series of CG constant-velocity steered molecular dynamics (cv-SMD) simulations and subsequent umbrella sampling simulations, designed to model the single-molecule translocation mechanisms of SRB, CF and ATTO 655 through a bilayer spanning DNP.

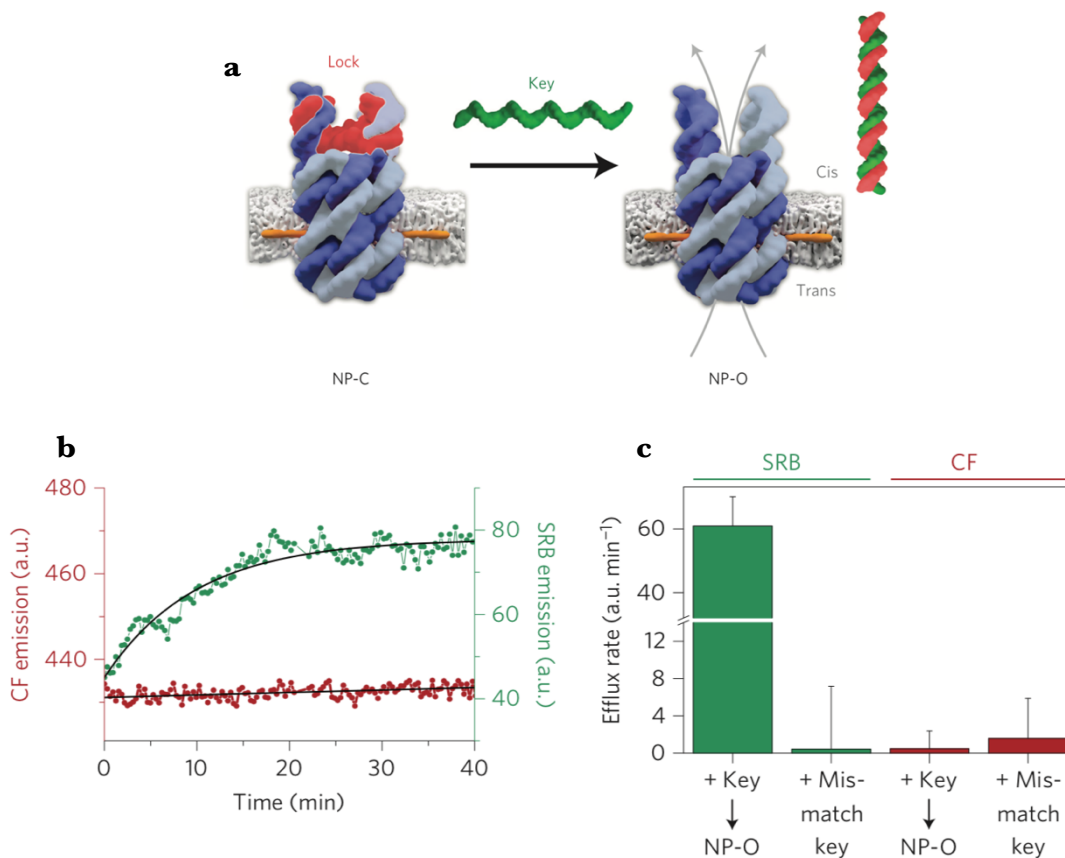


Figure 6.2. (a) Illustration of the ‘unlocking’ of the ligand-gated nanopore designed by Burns et al.[3]. The closed state of the nanopore is denoted NP-C, and the open state NP-O. (b) Kinetic release traces for CF (red) and SRB (green) yielded from UV-Vis absorbance spectroscopy experiments on SUV’s containing the fluorophores, where the gate-keeping nanopore is in the open state (NP-O) after the addition of the DNA key at $t=0$. (c) Efflux rates calculated for nanopore-gated SUV’s containing CF and SRB, demonstrating that the addition of the correct key strand (leading to formation of NP-O) leads to a strong increase in the efflux rate of SRB, compared to the NP-C state exposed to a mismatched key (which should not trigger formation of NP-O) suggesting that translocation of SRB proceeds predominantly through the pore lumen. In the case of CF, the presence of the mismatched key appears to increase the rate of efflux, suggesting that the NP-C actually favours translocation of this fluorophore. Figures adapted from Burns et al.[3]

6.1 Simulation Details and Analytical Methods

Constant-velocity steered MD (cv-SMD) is a simulation technique usually employed for the modeling of non-equilibrium, force-induced processes such as protein folding, receptor-substrate binding events or translocation pathways by applying a force to set of atoms (SMD atoms), while keeping another group of atoms fixed. By keeping one group fixed while pulling another, we force the system to adopt a series of configurations along the pre-defined vector of the proposed pathway, and this allows to sample the free energy landscape of the pathway more readily. Extracting reliable free-energy landscapes directly from standard equilibrium MD simulations is not usually feasible, as the likelihood sufficiently sampling the relevant configurations along the pathway without incurring huge computational expenses is very small.[97], [167] Here, I discuss the implementation of the pull code within GROMACS, and the methods used for subsequent calculation and analysis of force profiles and free energy profiles, to identify energy barriers along the translocation pathway for each fluorophore through a DNP.

6.1.1 *The Pull Code*

The pull code implemented within GROMACS allows the user to apply the pull types in a number of different ways to suit the intended purpose of the simulations. For the translocation simulations described in this chapter, the umbrella pulling scheme was employed – where the centre of mass of the pull group is harmonically restrained to a dummy atom, two which a force is applied in the direction of the translocation. Restraints are usually applied to groups that are in close proximity to the pull group i.e. the nanopore that is being traversed, as this facilitates pulling and ensure that interactions between the pull group and the surrounding biomolecular structures are properly sampled. In addition to the pulling vector, the user must specify the force constant of the spring, and the pull rate. A larger force constant coincides with a stiffer spring, which reduces the width of the sampling window at each coordinate, increasing the bias of the translocation pathway. Excessively fast pulling rates may lead to deformations in groups in close proximity of the pull group, so pull rates between 0.001 and 0.01 nm ps⁻¹ are typically used in biomolecular simulation studies.[168] I used a spring constant of 500 kJ/mol to maximise sampling of conformational space, thus allowing different translocation paths to be explored, and a pulling rate of 0.002 nm

ps⁻¹ over a duration of 5 ns, corresponding to a pulling distance of 10 nm from one pore opening to the other. The force experienced by the harmonic spring at each coordinate along the pulling axis is then outputted during the simulation. The potential of mean force (PMF) can then be extracted, either by direct analysis of the pulling trajectories (through the invocation of Jarzynski's equality, explained in Section 6.1.4), or by running further restrained equilibrium simulations of the system in a number of slightly overlapping adjacent configurations that occur along the pulling vector – a technique known as umbrella sampling (US).[63] My choice of pulling parameters made it possible to run an ensemble of pulling simulations for each fluorophore cheaply and rapidly, while generating enough intermediate system configurations to generate an appropriately sized ensemble of US simulations for each system.

6.1.2 Calculating the Potential of Mean Force

Consider a process for which a parameter λ changes incrementally with time between the start of the process λ_0 and the end of the process λ_t , where t is the time taken for the process to reach completion. According to the second law of thermodynamics, the average work W performed on a system undergoing a quasistatic process is equal to the difference between the free energies (F) corresponding to the initial and final states of the system (λ_0 and λ_t), as implied by the following expression:

$$\Delta F = F(\lambda_t) - F(\lambda_0) = \langle W \rangle \tag{6.1}$$

where $\langle W \rangle$ denotes an ensemble of measurements of W , and λ has a value of 0 at λ_0 and 1 at λ_t . Put simply, the potential of mean force is a measure of the change in the free energy of the system as a function of λ . It is obtained by taking the integral the work performed by the mean force acting on the system at each value of λ . For the case presented in this chapter, λ is represented by the translocation coordinate ξ along the translocation pathway, which runs from one side of the bilayer to the other. The translocation of a small molecule through a DNP can be considered a quasistatic process if it proceeds slowly enough for neighboring states at neighboring values of ξ to exist in equilibrium.[169]

6.1.3 Umbrella Sampling

The trajectories obtained from cv-SMD simulations of the translocation process serve as a starting point for a subsequent ensemble of umbrella sampling simulations. The user selects a number of frames within the cv-SMD trajectory to represent the incremental progression of the translocation coordinate ξ from the start of the translocation to the end point. The frames define the umbrella sampling “windows” (USWs), from which overlapping histograms of configurations are yielded during the US simulations. In order to keep the translocating molecule (pull group) centered within each umbrella sampling window, a harmonic restraint is applied to it, which biases the distribution of potential energies sampled within each window. This bias is removed during the post-processing phase, yielding an unbiased distribution of energies for each window from which the PMFs are calculated. PMFs obtained for each window are assembled into a continuous PMF function, which evolves as a function of ξ . [168] This post-processing is most often done through an implementation of the weighted histogram analysis method (WHAM).[63], [170]

To calculate the PMFs for the fluorophore translocation events simulated with cv-SMD, I divided each pulling trajectory into 52 frames, such that the distance between neighboring ξ windows was ~ 0.2 nm in each case. The harmonic umbrella biasing potential in each window was applied with a force constant of 500 kcal/mol/nm², to promote overlap between USWs while keeping the molecule centered in its sampling window, as this allows the WHAM algorithm to construct a smooth free energy profile from the histograms of configurations generated by US. As the translocation coordinate ξ is a function of the distance in the z-direction, the PMFs calculated here are one-dimensional, which assumes that fluctuations in the other two dimensions will not affect the free energy. The accuracy of this assumption cannot be commented on without a thorough investigation of collective variables using dimensionality reduction techniques such as principal component analysis. [171]

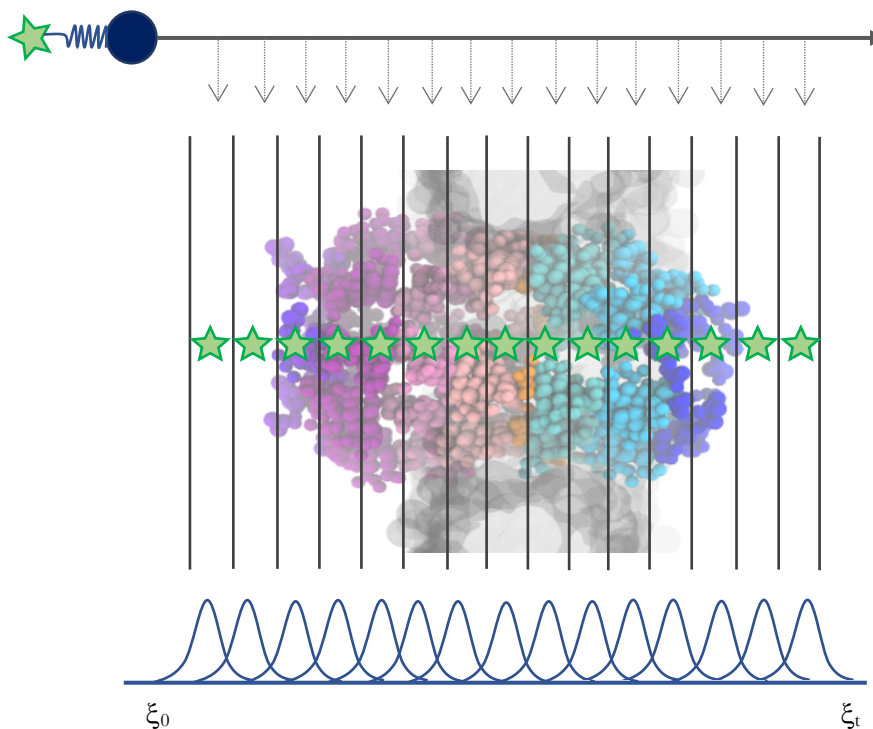


Figure 6.3: Setting up USWs from a pulling simulation. The image at the top represents the pulling simulation, where pulling force is applied to the dummy atom (blue), to which the fluorophore (green star) is harmonically restrained. Frames from the pulling trajectory (denoted by dashed arrows) are extracted and used to seed the separate umbrella sampling simulations (middle image), where the fluorophore’s centre of mass is restrained within its sampling window using an umbrella biasing potential. The simulations generate a series of overlapping (biased) histograms of configurations (bottom image), which are analysed and unbiased by WHAM, producing a smooth PMF curve describing the free energy surface of the translocation process.

6.1.4 Jarzynski’s Equality

A popular, faster and less computationally costly alternative to US simulations involves running an ensemble of cv-SMD simulations and invoking the Jarzynski equality (JE) [128]-[129], which relates the equilibrium free energy change with the work done by the system as it undergoes a non-equilibrium “fast-switching” process according to the following formula:

$$\langle e^{-W(\xi)/k_B T} \rangle = e^{\Delta G(\xi)/k_B T}$$

(6.2)

where W is the work done across an ensemble of pulling trajectories, k_B is the Boltzmann constant and ΔG is the Gibbs free energy difference between the two adjacent states in the pulling process. The left-hand side of the equation applies to the non-equilibrium scenario, where measurements of W are taken from an ensemble of non-equilibrium simulations, which are used to approximate the free energy ΔG (an equilibrium property, denoted on the right-hand side) associated with the non-equilibrium process. An ensemble of cv-SMD simulations is performed according to the methodology described in 6.1.1, and a free energy profile is obtained directly from these trajectories by calculating ΔG of the system at each translocation/pulling coordinate ξ according to the following relationship:

$$\Delta G = k_B T \ln \left[\frac{1}{N} \sum_i^N e^{W_i(\xi) / k_B T} \right]$$

(6.3)

where N is the total number of work samples taken from N replica simulations. Insufficient sampling along the coordinate ξ is a common concern when using JE to calculate free energy profiles, as well as the tendency for smaller values of W to dominate the exponential average disproportionately. To remedy this, the second-order cumulant expansion of the JE terms is used to calculate ΔG at each value of ξ to yield a less biased PMF – though loss of energy due to dissipated work is still quite difficult to account for. US yields more reliable PMFs and more accurate calculated ΔG values than direct calculation from cv-SMD trajectories with JE, simply because sampling is much more comprehensive and tunable with the former method, though it has been demonstrated that JE can reproduce US-derived PMFs if lower pulling rates and longer simulation durations are used – though this typically drives the cost of the simulation up to levels that are comparable to the cost of US simulations. [131], [134]-[135] That said, Martin *et al.* [175] described a rapid and cost-efficient ensemble-based protocol utilising JE in conjunction with shorter-duration AA cv-SMD simulations, that was capable of deriving experimentally validated PMF profiles for the translocation of a range of polynucleotides and single nucleotides through wild-type and engineered α HL nanopores. The discriminating features of the free energy profiles

associated with the translocation of poly(A) and poly(C) through the wild-type α HL nanopore elucidated in this study provided a qualitative explanation of the differences in the rate of translocation of these polynucleotides that were observed in experiment.

6.2 Model Building

As the MARTINI force-field does not feature standardised parameters for small molecule fluorophores, parameters for the SRB, CF and ATTO 655 molecules had to be derived from AA simulations. Two AA topologies were built for each fluorophore, one with general AMBER force-field (GAFF) parameters, for which partial charges were derived from QM data using the restrained electrostatic potential (RESP) procedure, and one with CHARMM general force field parameters derived using the CGenFF web server *without* prior QM calculation of partial charges. These two sets of force-field parameters were tested to determine the extent to which the choice of initial AA force field influenced the final CG MARTINI parameters. Once the CG MARTINI fluorophores were built, I assembled three fully coarse-grained models ready for cv-SMD simulations of the translocation pathways for SRB, CF and ATTO 655 through a membrane-spanning DNP.

6.2.1 Parameterisation of All-Atom Fluorophore Models

Fluorophores (commonly referred to as fluorescent dyes) are small-to-medium sized organic molecules that are capable of absorb specific wavelengths of visible light, and re-emitting light at longer wavelengths after experiencing a vibrational relaxation following the initial electronic excitation. Fluorescent molecules often feature highly conjugated fused rings, as the gap between the highest occupied molecular orbital and the lowest occupied molecule orbital in these molecules corresponds to wavelengths within the visible spectrum. [176] The availability of parameters for fluorophores within the biomolecular force fields is limited, due to the intrinsic difficulty of accurately modelling the high degree of polarisability within these molecules, which arises from rapid movement of delocalised electrons throughout the conjugated ring system. This difficulty stems from the fixed point charge approximation that is a feature of most standard AA MD force fields [71] [91],[140]so the rapid charge fluctuations that occur in fluorophores are difficult to account for.

Additionally, there is a shortage of published experimental data on common fluorophores that quantify the properties that are used for validation of force-field parameters, such as solvation free energies. Nevertheless, AA parameterisations within the CHARMM and AMBER force fields have been developed for the rhodamine, Alexa Fluor, Cy and ATTO families of fluorophores by several independent research groups, using a variety of parameterisation methodologies, though the reliability of these parameters is uncertain. Studies by Vaiana *et al* [178], [179] describe the automated frequency matching method (AFMM) they used to obtain CHARMM27 compatible parameters for Rhodamine 6G and an MR121 (a member of the ATTO dye family). The method consists of two stages: an initial quantum-chemical calculation of partial charges and normal modes at the restricted Hartree-Fock level of theory (RHF) and assignment of existing CHARMM27 bonded parameters and vdW constants by analogy, followed by an iterative refinement of these parameters until the molecular mechanics (MM) derived normal mode eigenvalue/eigenvector sets match the results from the quantum-chemical normal mode analysis. Furthermore, the parameters they derived allow them to calculate fluorescence lifetimes from MD simulations, which were in good agreement with experimental lifetimes, however these were calculated as a function of distance between the dye and the quencher, which is unrelated to normal modes.

A very similar method was used by Corry *et al*[180] to parameterise a series of Alexa Fluor dyes, only the quantum chemical charge-fitting procedure was based on density functional theory (DFT) rather than RHF, and the comparison of vibrational normal modes between QM and MD models was forgone as they were not relevant to the purposes of their study. The aim of their work was to uncover the orientational freedom and orientation factors of protein-tethered fluorophores, which are almost impossible to measure directly in an experimental setting but make a large contribution to uncertainties associated with the interpretation of FRET experiments. Their results were ultimately inconclusive, due to the scarcity of experimental orientational measurements for validation of their fluorophore parameters, and insufficient sampling in their MD simulations. A substantial effort towards establishing a set of reliable, modular and easy to implement AMBER-compatible parameters named “AMBER-DYES” for the Alexa Fluor, Cy and ATTO family of fluorophores was made by Graen *et al*[181] in 2014. Their strategy also made use of QM methods for calculation of

partial charges, and after this stage atom types and bonded parameters were assigned by analogy from GAFF. The charge-fitting procedure was validated by comparing the calculated partial charges to those of tryptophan – a bulky aromatic and “dye-like” amino acid for which thoroughly validated partial charges exist for all major biomolecular force fields. In addition, they characterised charge fluctuations for each fluorophore from principal component analysis of QM/MM trajectories and concluded that the majority of the charge variance exhibited throughout the ring systems could be accounted for with a small number of polarisable centres, though this feature has not been implemented within AMBER-DYES.

The study presented in this chapter uses comparatively unsophisticated biomolecular CG models. The MARTINI force-field used therein is non-polarisable (with the exception of the water molecules), and we use it to explore the dependence of the fluorophores’ net charge and on its translocation pathway through the membrane-spanning DNP. Hence, the capacity of the AA parameters to account for charge polarisation and vibrational normal modes of the fluorophores is likely be of little consequence in the derived CG parameters. The key common features of the parameterisation schemes detailed above are the derivation of partial charges using QM calculations, and the assignment of atom-types and bonded terms from general force fields (either AMBER or CHARMM), so these are the steps I took when parameterising the AA fluorophore models. In the AMBER and CHARMM parameterisation scheme, each residue or molecule is restrained to have a fixed integer charge, so that large biomolecular systems can be built in a modular fashion without needing to refit individual atomic charges as the system increases in complexity. The partial charges the atoms within the molecule/residue must accurately reflect the distribution of charge around the molecule, and they must sum to give the correct restrained net charge. For parameterisation within CHARMM, the CGenFF program assigns partial charges according to the bond-charge increment scheme described in Chapter 2.2.1.2. The AMBER force-field relies on quantum mechanical calculation of the electrostatic potential around the molecule by summing the contributions from the nuclei and the electronic wavefunction. The partial charges of the atoms are then optimised according to the restrained electrostatic potential (RESP) procedure[182], to achieve an optimal least-squares fit between the RESP model calculated electrostatic potential. Molecular models of the ATTO 665, CF and SRB structures were hand

built using PyMOL, with coordinates saved in the PDB format. The PDB files were converted to the Gaussian16 ASCII format required for the QM geometry optimization and ESP calculations. Geometry optimisation was performed with no symmetry, bond length, bond angle or dihedral restrictions in Gaussian16 [183] at the HF/6-31G* level of theory, in the gas phase, followed by ESP calculations. The optimised geometries and ESP data for each fluorophore were then fed into the Antechamber programme [53], which performed the fitting of RESP charges, atom typing and assignment of bonded parameters. The optimised geometries outputted by Gaussian were also converted to the Tripos Mol2 format required by the CGenFF web-server for assignment of CGenFF partial charges, atom types and bonded parameters by analogy.

Each AA fluorophore model was solvated in a 4 nm x 4 nm x 4 nm water box with a concentration of NaCl set at 1.0 M, using the *gmx solvate* and *gmx ionize* plugins.[63], [98], producing solvated fluorophore models consisting of ~ 9000 atoms in each case. The models were minimised (with a 2fs time-step) using the default conjugate-gradient algorithm, until the forces converged below the maximum threshold of 1000 kJ mol⁻¹ nm⁻¹. Each system was subjected to an initial 5ns of equilibration within the NVT ensemble, with the temperature set to 300K using the velocity-rescaling thermostat, followed by 10 ns of NPT equilibration with the Berendsen barostat controlling the pressure at 1.013 bar. Production simulations were run in the NPT ensemble for a duration of 50 ns, with temperature and pressure control maintained with the Nosé-Hoover thermostat and barostat. Electrostatic forces were calculated using the particle mesh Ewald (PME) algorithm, and a 1.2 nm shifted cutoff was used for VdW and short-range electrostatic interactions. A total of 25 replicas were required for each model to guarantee convergence of the radius of gyration (R_g), which was assessed to confirm that the CG models derived from the AA models replicated the gross conformation of the molecules in the AA models.

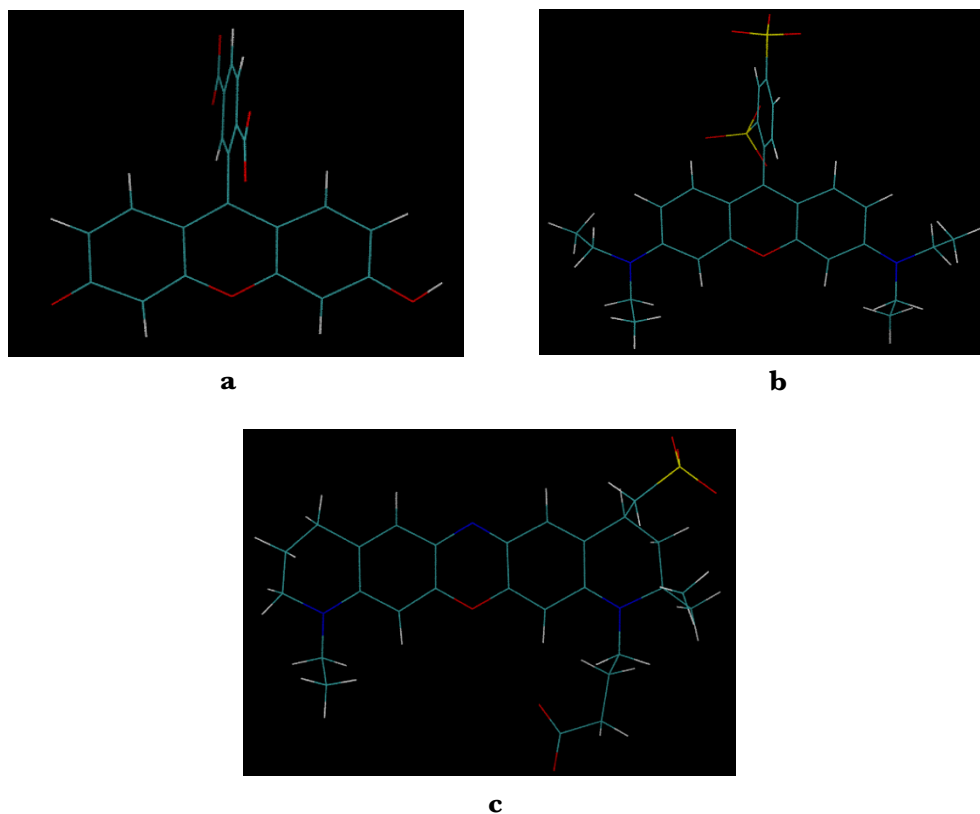


Figure 6.4: QM optimised geometries of the three fluorophores **(a)** CF **(b)** SRB and **(c)** ATTO 655.

6.2.2 Parameterisation of Coarse-Grained Fluorophore Models

The method used for the derivation of CG parameters for the CF, SRB and ATTO 655 molecules was identical to the method used to obtain the CG TEG-cholesterol parameters (using PyCGTool) described in Sections 2.2.2.3 and 3.1. Before bonded parameters could be extracted from the reference AA simulations, AA to CG mapping and bonding schemes were devised for the three fluorophores. The mappings were designed by eye, to preserve as much symmetry in the molecules as possible, while maintaining a faithful representation of the positions of charged and polar groups present within the structures. Tiny bead types (see Section 2.2.2.1) were used in the place of small bead types for the CG model of CF, to account for its' smaller size relative to SRB. Several mapping and bonding schemes were tried for each molecule, however only the mapping schemes shown below resulted in stable CG topologies, for which the distribution of gyration radii (R_g) in simulations matched the AA distribution. The corresponding bonding schemes are supplied in Appendices 5-7.

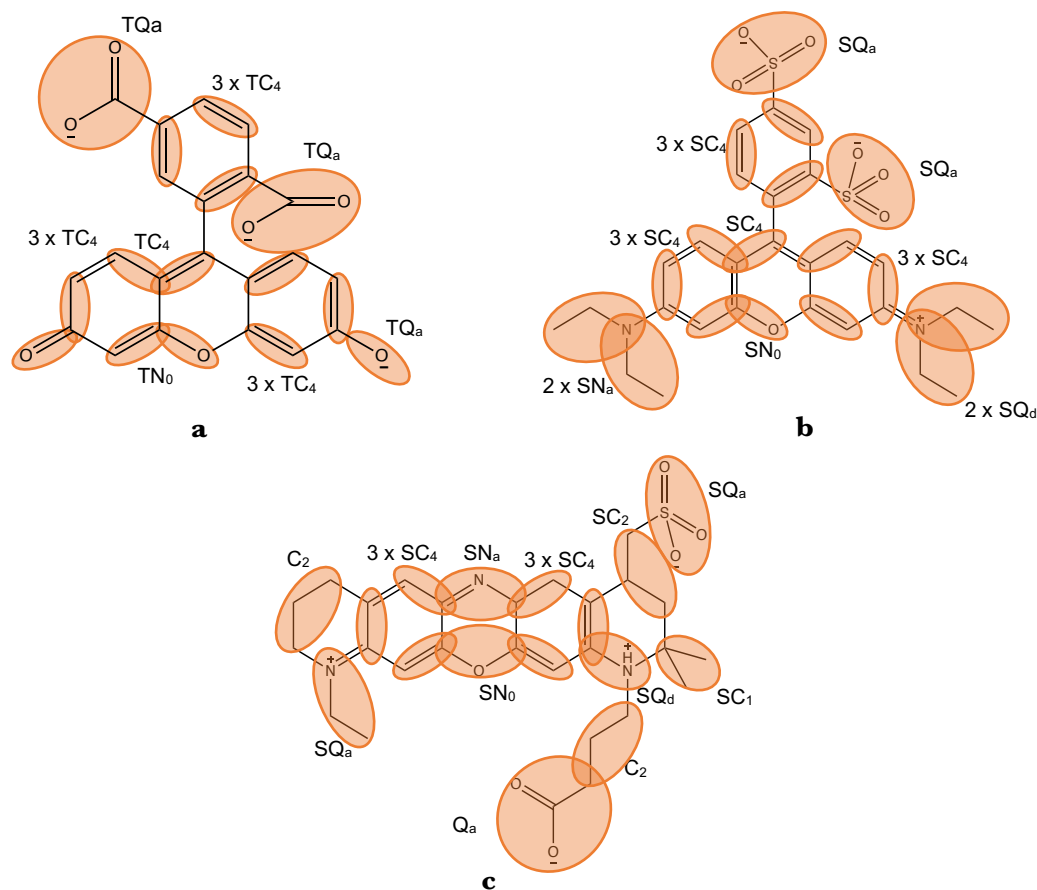


Figure 6.5: AA atom to MARTINI CG pseudo-atom mapping schemes used for parameterisation of the fluorophores with PyCGTool. **(a)** CF **(b)** SRB and **(c)** ATTO 655.

Two solvated CG models were built for each fluorophore using the *insane.py* script (one with CG parameters obtained from AA simulations with CHARMM parameters, the other obtained from AA simulations with AMBER parameters), each with a box size of 4 nm x 4 nm x 4 nm, filled with MARTINI polarisable water pseudo-molecules with a concentration of NaCl equivalent to 1.0 M. An ensemble of 25 production simulations lasting 50 ns each was performed for each model, and R_g data was extracted from the ensembles. For ATTO 655 and SRB, the AMBER-derived CG models had a larger average R_g than the CHARMM equivalent, whereas the average R_g of the two CG CF models were almost equivalent. It is worth mentioning that the CHARMM topology for CF generated by the CGenFF program had the lowest penalty score of the three, which may explain the higher degree of consistency in average R_g values between both CF models. This is illustrated in the Tukey box plot shown below – which shows that the overlap between the error bars of the median (indicated by notches) for the CF boxplots is strongest for CF.

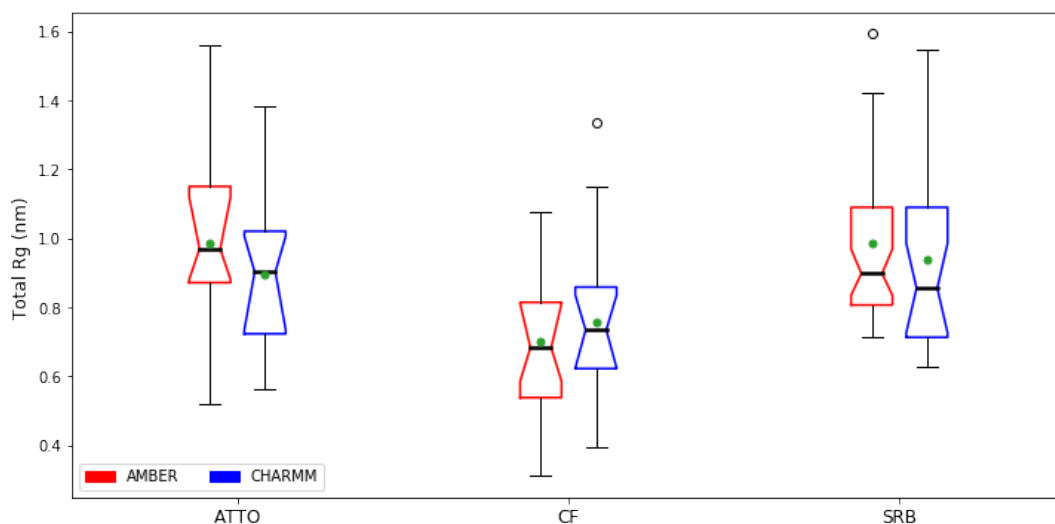


Figure 6.6: Tukey boxplots comparing the total R_g of the CHARMM and AMBER derived CG fluorophore models. The black lines represent the median R_g in each case, and the mean R_g is denoted by green circles. The height of the notches on the boxplot reflects the height of the error bars of the median, where the error bars represent the standard error of the median.

I chose to use the AMBER-derived CG topologies for the fluorophore translocation simulations, as the higher accuracy of the AMBER RESP-fitted partial charges did in fact have a noticeable influence on the R_g in the resulting CG models, as shown by the differences in the mean R_g between the CHARMM and AMBER derived models in Figure 6.6.

6.2.3 Assembly of Models for Pulling and Umbrella Sampling Simulations

The fully equilibrated CG 1.0 M NaCl DNP/membrane model (details in Chapter 4) served as the starting point for assembly of the cv-SMD simulation systems. The DNP and POPC bilayer coordinates were extracted from the final frame of a replica and converted to PDB format. Each fluorophore translocation model was assembled in the same way, with the equilibrated CG fluorophore positioned ~ 1.2 nm above the centre of the pore opening at the R1 terminus. The three fluorophore/DNP/bilayer models were solvated (with *insane.py*) in a 20 x 20 x 30 nm solvent box with a concentration of 1.0 M NaCl, using the polarisable MARTINI water model. The larger z-dimension was chosen to allow enough space along the pulling vector so that the fluorophore can be pulled continuously through its simulation box without interacting with DNPs in periodic boxes.

The simulation parameters (timesteps, integration algorithms, thermostats, barostats, electrostatics schemes and associated settings and cut-offs) used for the minimisation, equilibration and production simulations were largely the same as those listed in Chapters 3 and 4. However cv-SMD simulations require a slightly more specialised equilibration procedure to conserve the starting position of the pull group, and extra parameters must be specified for the pulling and umbrella sampling simulations (discussed in 6.1.1 and 6.1.3). After minimisation with the gradient descent algorithm, positional restraints with of a force constant of $1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ were placed on the fluorophore molecule, the DNP and the lipid headgroups for the first 5 ns of NVT equilibration. The restraints were kept as they were for the first 5 ns of the NPT equilibration, then reduced to $500 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ for the following 5 ns. The restraints on the lipid headgroups were gradually removed over the final 10 ns of NPT equilibration, but the positional restraints on the DNP were held in place for the pulling simulations.

Before pulling simulations were initiated, the positional restraints on the fluorophore molecules were replaced with harmonic restraints, with a spring constant of $500 \text{ kJ mol}^{-1} \text{ nm}^{-2}$. Pulling simulations were performed in the NPT ensemble, with the temperature held at 300 K and pressure at 1.013 bar with the Bussi velocity-rescaling [110] and Parrinello-Rahman barostat [112] combination, with temperature and pressure time coupling constants of 10 ps. The SMD atom was defined for each fluorophore; in each case this atom was a pseudo-atom located in one of the central aromatic rings. The SMD atom is the atom/pseudo-atom that is harmonically restrained to the dummy atom that represents the point in space that is shifted down the translocation coordinate. Each fluorophore was pulled at a rate of 0.005 nm ps^{-1} , over a duration of 5 ns, from one side of the box to the other. A total of 30 replicas (each with a different initial random velocity seed) was performed for each cv-SMD system, to explore the possibility of alternative transport pathways that may not proceed directly through the pore lumen (the “trans-lumen pathway”). Alternative transport pathways through the interface between the lipid bilayer and the outer surface of the DNP (the “interfacial pathway”) were observed for all three fluorophores, with the probability of transport occurring via the interfacial pathway varying considerably between the fluorophores (see Section 6.4). The emergence of two transport pathways in the cv-SMD ensembles necessitated two sets of umbrella

sampling simulations for each fluorophore, as direct comparisons of the PMFs yielded for the two pathways can give us a better idea of which route may be preferred by each fluorophore in an experimental setting. From each fluorophore cv-SMD ensemble, I selected one pulling trajectory in which the translocation occurs via trans-lumen pathway, and one trajectory featuring the interfacial pathway.

These selected trajectories were trimmed to remove frames in which the fluorophore was located was over 2 nm from the pore termini, and the trimmed trajectories were divided into 52 frames, resulting in 52 umbrella sampling simulation windows covering 12 nm of pulling distance. Each window was used to seed ten separate replicas (with different initial velocities) for each window, and each replica commenced with 5 ns of NPT equilibration before the umbrella sampling (production) simulation was initiated. All umbrella sampling simulations were run for 20 ns apiece. Each ensemble of US simulations used a harmonic umbrella biasing potential with a force constant of 500 kcal mol⁻¹ nm⁻² in each sampling window. Ten replicas for each USW were generally sufficient to produce a reproducible and differentiable averaged PMFs with reasonably low standard errors (between 0.93 kcal/mol and 1.98 kcal/mol), and the calculated values for the free energy change associated with the transport pathway ΔG largely followed a normal distribution, albeit with some spaces between histogram bins (see Appendix 8). Running more replicas for each USW would undoubtedly reduce the standard error associated with ΔG , though the additional computational expense incurred would provide little benefit here as we aim to make mainly qualitative comparisons between the translocation pathways for each fluorophore, and the characteristic PMFs derived from the existing dataset already provide a solid basis for this. More specific convergence criteria were difficult to establish in a timely manner as each of the ten PMFs constituting each averaged PMF were themselves derived from 52 US simulations, all with different starting configurations.

6.3 Computational Resources

A total of 4300 individual simulations were run for the investigation discussed in this chapter, and this workload was spread between two supercomputers: ARCHER[123] and Cartesius[124]. Each simulation was run on 120 CPU cores, as each ensemble of umbrella sampling simulations was submitted as an array job to make the workload as streamlined as possible. These simulations consumed 800,000 core hours in total. The computational resources used for this work were provided by the CompBioMed Centre of Excellence.[125]

6.4 Results and Discussion

As mentioned in Section 6.2.3, each cv-SMD fluorophore transport simulation proceeded via one of two pathways: the trans-lumen (TL) pathway or the interfacial (IF) pathway, shown in Figure 6.7. Since DNPs are designed for selective transport of small molecules by maintaining a structurally stable lumen of a fixed diameter while embedded in lipid bilayers, a trans-lumen pathway is desirable for their intended applications. The existing literature[3] posits that translocation of SRB proceeds via a TL pathway, but it provides no clear evidence of a TL pathway for CF (see Section 6.1). Performing a brief inspection of the trajectories provided me with an initial indication of which transport pathway might be preferred by each fluorophore. The likelihood of the TL and IF pathways expressed as percentages of the total number of replicas in the cv-SMD ensemble for each fluorophore are shown in Table 6.1. Both SRB and CF appear to adopt the TL pathway more frequently, though SRB has a higher propensity to proceed along this pathway than CF, as indicated by the higher percentage of cv-SMD replicas that feature it. In contrast, the translocation of ATTO 655 exhibited a strong preference for the IF pathway. Whether these preferences are reflective of reality or are merely the result of a bias imposed by the starting configuration, pulling rate and/or harmonic spring constant is not clear, as different combinations of these parameters were not tested. In any case, we can draw far more reliable conclusions from the PMFs.

	SRB	CF	ATTO
Trans-lumen	70%	62%	30%
Interfacial	30%	38%	70%

Table 6.1: Distribution of trans-lumen and interfacial transport events for each ensemble of cv-SMD replicas, represented as the percentage of replicas featuring the given transport pathway.

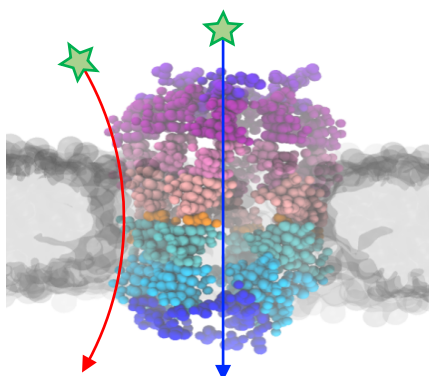


Figure 6.7: Simplified illustration of the idealised TL and IF pathways, represented by blue and red arrows respectively.

The averaged force profiles (Figure 6.8) of the IF and TL pathways for each fluorophore reveal a limited amount of information about the obstacles the fluorophore encounters by tracking the force experienced by the harmonic spring connecting the fluorophore COM to the dummy atom, which provides a measure of the force required to pull the fluorophore from the R8 terminus to the R1 terminus along the translocation vector. In all three of the TL force profiles, there are two distinctive force maxima that arise shortly after the fluorophore enters the R8 opening (at 1ns), and just before it exits through the R1 opening (~ 2.7 ns), suggesting that the constrictions at the termini impede translocation of small molecules through the DNP lumen to some extent. The IF force profiles all show a general upward drift in the force as the fluorophore traverses the boundary of the DNP and the lipid torus between 1ns and 3ns, with the peaks and troughs of distributed fairly randomly between the three fluorophores, suggesting there is no significant barrier to the IF translocation path imposed by the DNP/membrane interface.

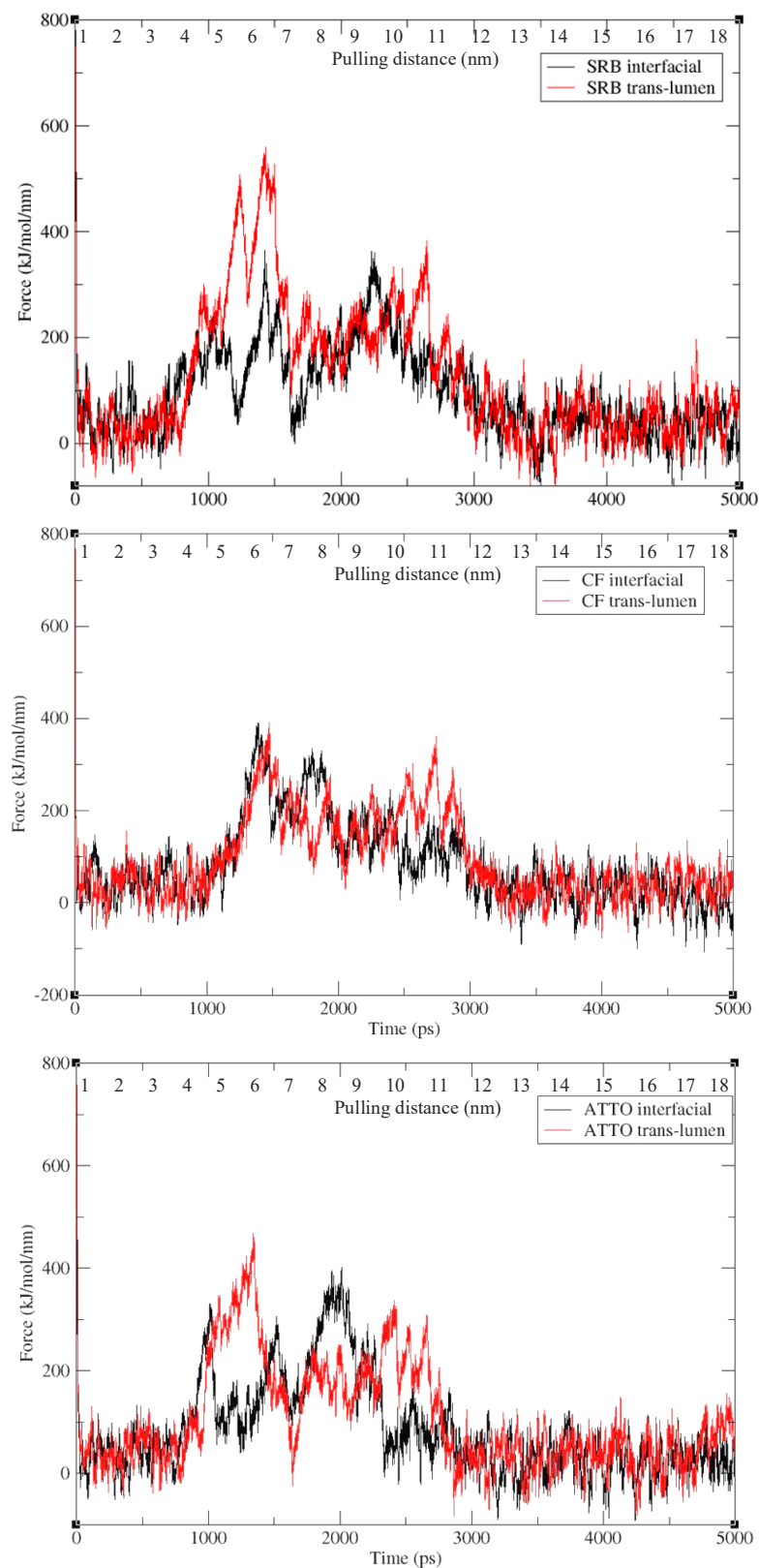


Figure 6.8: Averaged force profiles extracted from the ensembles of cv-SMD simulations of fluorophore translocation. The fluorophore enters the R8 pore entrance at 1 ns (at 4 nm pulling distance) and exits through the R1 terminus at around 3 ns (12 nm pulling distance).

The averaged force profiles (Figure 6.8) of the IF and TL pathways for each fluorophore reveals a limited amount of information about the obstacles the fluorophore encounters by tracking the force experienced by the harmonic spring connecting the fluorophore COM to the dummy atom, which provides a measure of the force required to pull the fluorophore from the R8 terminus to the R1 terminus along the translocation vector. In all three of the TL force profiles, there are two distinctive force maxima that arise shortly after the fluorophore enters the R8 opening (at 1 ns), and just before it exits through the R1 opening (~ 3 ns), suggesting that the constrictions at the termini impede translocation of small molecules through the DNP lumen to some extent. The IF force profiles all show a general upward drift in the force as the fluorophore traverses the boundary of the DNP and the lipid torus between 1 ns and 3 ns, with the peaks and troughs of distributed fairly randomly between the three fluorophores, suggesting there is no significant barrier to the IF translocation path imposed by the DNP/membrane interface. Direct comparison of the averaged PMFs associated obtained from the umbrella sampling simulations of the interfacial and trans-lumen pathways for each fluorophore provides a far more reliable account of the energetic obstacles and/or advantages associated with both pathways for each fluorophore. Evaluating this information alongside the calculated values for the overall free energy change (ΔG) allows us to predict which pathway is likely to be favoured. Before making a comparison across fluorophores, I will discuss the features of each individual PMF.

6.4.1 *Transport PMFs for SRB*

At first glance, we notice that the ensemble-averaged ΔG value associated with IF translocation pathway of SRB (Fig. 6.9b) is strongly negative (inclusive of errors), whereas the average ΔG calculated for the TL pathway fluctuates around zero within the bounds of the 95% confidence interval. Though there is little net energetic gain concomitant with the TL pathway, the fluctuations of PMF as a function of the translocation coordinate are rather modest, with upward spikes in free energy rarely exceeding 1 kcal/mol, meaning the SRB is unlikely to encounter any major energy barriers along the translocation path. An exception occurs between 3 nm and 4 nm on the x-axis, where there is sharp dip in the PMF. Analysis of the distribution of the histograms of force measurements taken in the ensemble of US simulations (shown in

Appendix 9) reveals a small gap between adjacent histograms in this region, indicating that more sampling is required here. A similar defect occurs in the IF PMF calculated for SRB (Fig. 6.9b), in the same region. Equal spacing between USWs does not always guarantee a smooth and defect-free PMF, so the ideal distribution of configurations must be established on a case-by-case basis through rigorous testing.[168]

The IF PMF features a dramatic fall in the free energy (~ 18 kcal/mol) that occurs shortly after the SRB molecule enters the gap between bilayer and the DNP, which then tails off after the molecule exits the interface (discounting the two small spikes between 3 and 4 nm caused by insufficient sampling). The reason behind this fall in free energy becomes clear when we consider the distribution of charges on the SRB molecule, and how they align with the charged sites of the surrounding DNP and POPC lipids. As shown in Figure 6.5b, the CG model of the SRB molecule has two negatively charged pseudo-atoms in close proximity to each other, representing the two sulphate groups of the 1,3-benzenedisulphonate moiety, which makes up one corner of the triangular molecular model. The repulsion between these two negative sulphate groups, as well as the repulsion between the entire benzenedisulphonate moiety and the negatively charged DNP backbone is attenuated by the presence of the positively charged choline headgroups of the zwitterionic POPC lipids. In addition, the solvated SRB molecule possesses a tertiary amine conjugated to a series of fused benzene rings, which forms a π -system that endows the nitrogen with a positive charge. This positive charge was also included within the CG model of SRB and is located in an adjacent corner of the model. Hence, the CG SRB molecule can easily orient its positively charged corner to face the DNP's negative backbone, and its densely negatively charged corner with the lipid headgroups to maximise favourable interactions, bringing the free energy down significantly as it traverses the interface.

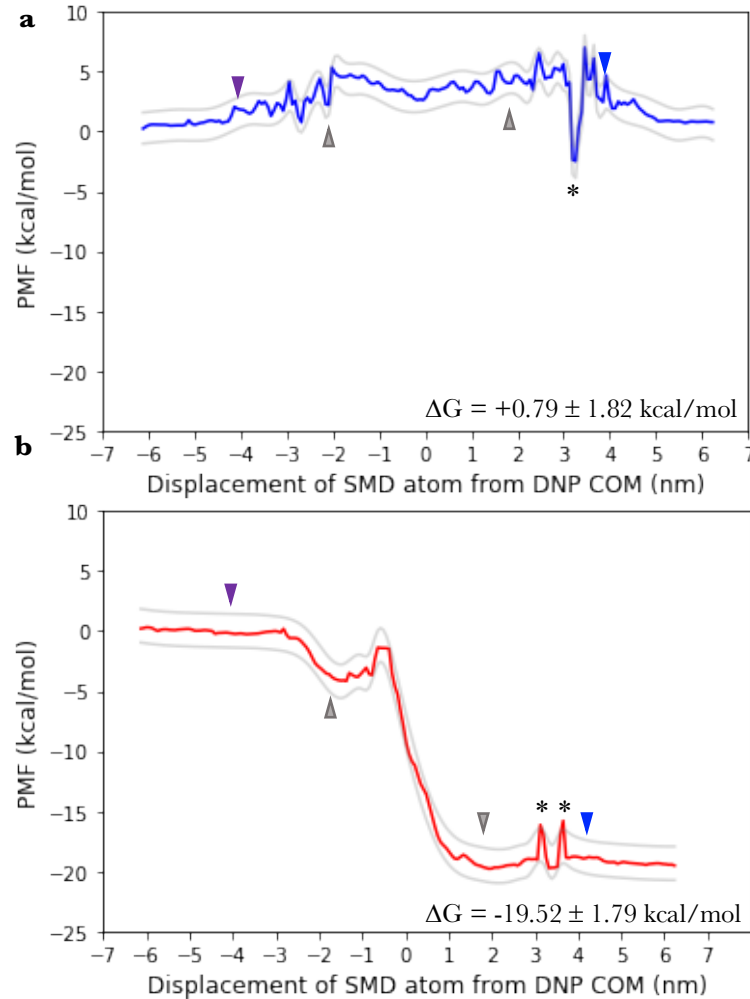


Figure 6.9: PMFs representing the 1D free energy landscapes associated with the transport of SRB via the TL pathway (**a**) and the IF pathway (**b**). The boundaries of the error bars are represented by the grey traces drawn above and below the PMF, where distance between the PMF curve and error traces are equal to the bootstrapped standard error of the ensemble PMF averaged over 1 nm blocks along the translocation coordinate. The distance along the translocation coordinate is defined as the displacement of the SMD atom from the DNP centre of mass, which corresponds to the symmetrised distance along the principal pore axis (the z axis) relative to the centre of the DNP. The error reported for the ΔG corresponds to the bootstrapped 95% confidence interval. The same error representation method is used for all following PMFs. The purple and blue arrows mark the approximate location of the outermost edges of the R8 and R1 regions of the DNP, and the grey arrows represent the surfaces of the lipid bilayer. The asterisks mark the loci of defects in the PMF, which arise from insufficient overlap between force histograms in the US simulations.

Once the SRB molecule escapes the interface, it appears to diffuse across the undulating zwitterionic surface of the bilayer (this was also seen for other fluorophores). In the simulations performed for USWs 37-41, where the fluorophore is initially restrained at specific points on the idealised IF translocation pathway between 4 nm and 5 nm displacement, the fluorophore drifts away from the position to which it was restrained (see Fig. 6.10d). This indicates that the repulsions between the DNP and the negative benzenedisulphonate moiety, and the attractions between the point charges on SRB and the zwitterionic lipid headgroups provide a strong enough driving force for the SRB molecule to overcome the harmonic restraints used in these US simulations and are absorbed to the bilayer surface. Hence, the IF pathway corresponds to a membrane adsorption event rather than a translocation event, and hence the calculated ΔG associated with these pathways represents the free energy of fluorophore adsorption. In the case of a single SRB molecule, the IF adsorption pathway evidently has the stronger energetic driving force associated with it due to the favourable electrostatic interaction, however the absence of significant energy barriers in the TL pathway and the associated range of ΔG values gravitating around zero indicated that the TL pathway is also feasible for SRB. Given that fluorescence experiments have already demonstrated a strong propensity for SRB to translocate rapidly through vesicle-bound DNPs in their open state, it is probable that SRB is very capable of translocating via the TL pathway. The IF adsorption pathway is of course more favourable, however commenting on the rate of IF adsorption relative to the rate of TL translocation is difficult here, as both free energy profiles are free of major energetic obstacles. Most importantly, we are considering the translocation of a single fluorophore through a single nanopore here, whereas kinetic efflux assays are established on the translocation of fluorophores contained within vesicles (with ~ 100 nm diameter) that are large enough to be punctured by multiple DNPs, down a concentration gradient over the course of minutes, meaning the fluorophore-fluorophore interactions and entropic effects that influence the kinetics of fluorophore efflux are ignored in these simulations, as are the effects of any aggregation of multiple DNPs along the surface of the vesicles. As mentioned in the introduction of this chapter, the kinetic efflux assays made use of a functionalised DNP with overhanging ssDNA strands on one of the termini, which may well interact with the surface of the bilayer when they are unbound and in their open state, potentially obstructing the IF pathway and making the TL pathway more accessible.

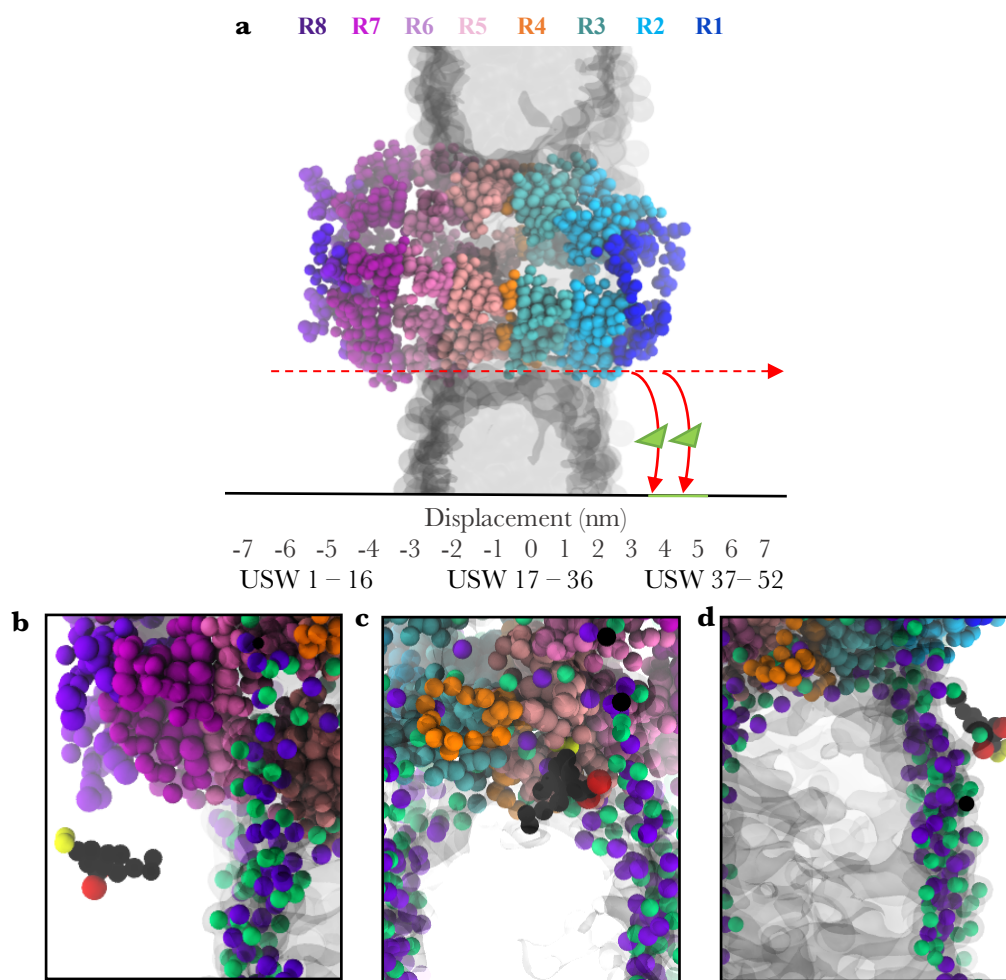


Figure 6.10: (a) Schematic illustrating the tendency for the fluorophore molecule (green triangle) to deviate from the idealised IF translocation pathway in US windows 36-41 (between 4 and 5 nm displacement) and associate with the POPC headgroups once it departs the DNP/membrane interface. The approximate distribution of the 52 USWs along the displacement values is shown below. (b) US simulation snapshot taken from a replica of USW 15 showing how the CG SRB molecule aligns its positively charged pseudo-atoms (shown in yellow) toward the DNP backbone, and its negatively charged pseudo-atoms (shown in red) away from the DNP. (c) US simulation snapshot extracted from USW 26, where the positive pseudo-atoms associate strongly with the DNP backbone, and the negative pseudo-atoms interact with the zwitterionic lipid headgroups. The positively charged choline groups are represented as green spheres, and the negative phosphate groups are represented as violet spheres. (d) Snapshot taken from USW 41, illustrating the SRB molecule drifting from the point in space to which it is harmonically restrained, to form stabilising interactions between its charged pseudo-atoms and the zwitterionic lipid headgroups. This drift in position typically occurred after ~ 5 ns of simulation time.

6.4.2 *Transport PMFs for CF*

The PMFs calculated for CF (shown in Figure 6.11) display very different characteristics to the PMFs calculated for SRB. The two fluorophore models appear structurally similar upon initial inspection (Fig. 6.3) with a similar triangular arrangement of aromatics rings, however their differences in net charge, charge distribution and charge density influence their interactions with the membrane-spanning DNP drastically, and this is reflected by the marked differences in the appearances of their PMFs and the associated ΔG values. The net charge of CF is three times larger than that of SRB, with three negative point charges distributed across two corners of the CG model, and no counterbalancing positive point charges anywhere in the molecule. CF is also smaller in size, and this difference in size was accounted for in the CG models through the use of T-prefix pseudo-atoms for the modeling of CF, as discussed in Section 6.2.2. This CG mapping reduced the volume of the CF model (approx. 0.33 nm^3) to half of the volume of SRB model (0.66 nm^3). These combined features endow the CF model with a higher negative charge density relative to SRB, and therefore it experiences stronger repulsive interactions with the membrane spanning DNP along either translocation pathway. The appearance of the PMFs for the TL and IF pathways of CF (Fig. 6.9) substantiates this theory well.

The TL pathway is more energetically favourable because it provides the path of least resistance; the lumen of the is wide enough ($2.21 \pm 0.01 \text{ nm}$ on average) to accommodate the passage of the small CF molecule while avoiding close contact with the negatively charged DNA backbone pseudo-atoms on the interior surface of the DNP. The size of the gap between the DNP and the lipid torus at the DNP/membrane interface fluctuates between 0.5 and 1.0 nm throughout the US simulations, meaning the CF molecule is forced to make extremely close contact with the negatively charged DNA backbone as it travels via the IF pathway. The spikes in free energy along the TL pathway coincide with constricted R8, R4 and R1 regions where the negative charges of the helices become concentrated within the lumen (see Fig. 4.11b), and repulsions between the CF and the DNP are strongest. As the CF molecule passes the constricted R8 opening, the free energy experiences a moderate increase before falling gradually as the CF molecule enters the less constricted part of the lumen, running from R7 to R5. As the CF approaches the R4 constriction (near the DNP centre of mass), the free energy rises sharply by $\sim 8.5 \text{ kcal/mol}$, presenting a significant energy barrier to

complete translocation through the lumen. Assuming this energy barrier is overcome, and the CF molecule is able to translocate past the R4 constriction, the CF molecule then encounters another smaller, but significant free energy barrier of ~ 4 kcal/mol as it approaches the R1 constriction.

Of all of the PMFs calculated for the six pathways studied across the three fluorophores, the IF pathway of CF is the most idiosyncratic, with the highest average standard error in free energy and a generally “noisy” appearance. The US force histograms generated from the ensemble of US simulations of this translocation pathway (see Appendix 9) are much narrower than the histograms generated for the other five translocation pathways, and consequently the overlap of these histograms is relatively poor throughout the entire translocation vector, causing sudden drops in the free energy between US windows. This can be rectified through the use of smaller spacings between US windows in future studies of fluorophore transport.

Like SRB, the CF molecule exhibited the same tendency to break its harmonic restraints and drift away from the DNP at displacements between 4 and 5 nm (shown in Figure 6.10a) in an attempt to reduce the magnitude of repulsive forces between itself and the DNP that arise in the IF pathway. Though they are numerous, the individual spikes in free energy rarely exceed 2.5 kcal/mol and are minor in comparison to the large energy barrier encountered by the CF along the TL pathway. We can also distinguish a very minor downward drift in the free energy profile once the CF passes the outer surface of the R4 region, that leads to a slightly negative average ΔG , however the relatively high error makes it difficult to ascertain whether or not the IF membrane adsorption pathway is a truly spontaneous process. Taking the relatively low energetic barriers and borderline ΔG value into account, we can assume that the IF pathway is marginally feasible and is possibly faster than its' TL equivalent. We can confidently conclude that both the IF and TL pathways are appreciably less accessible for CF than for SRB (and ATTO 655 as we will see shortly), due to their differences in charge density and distribution.

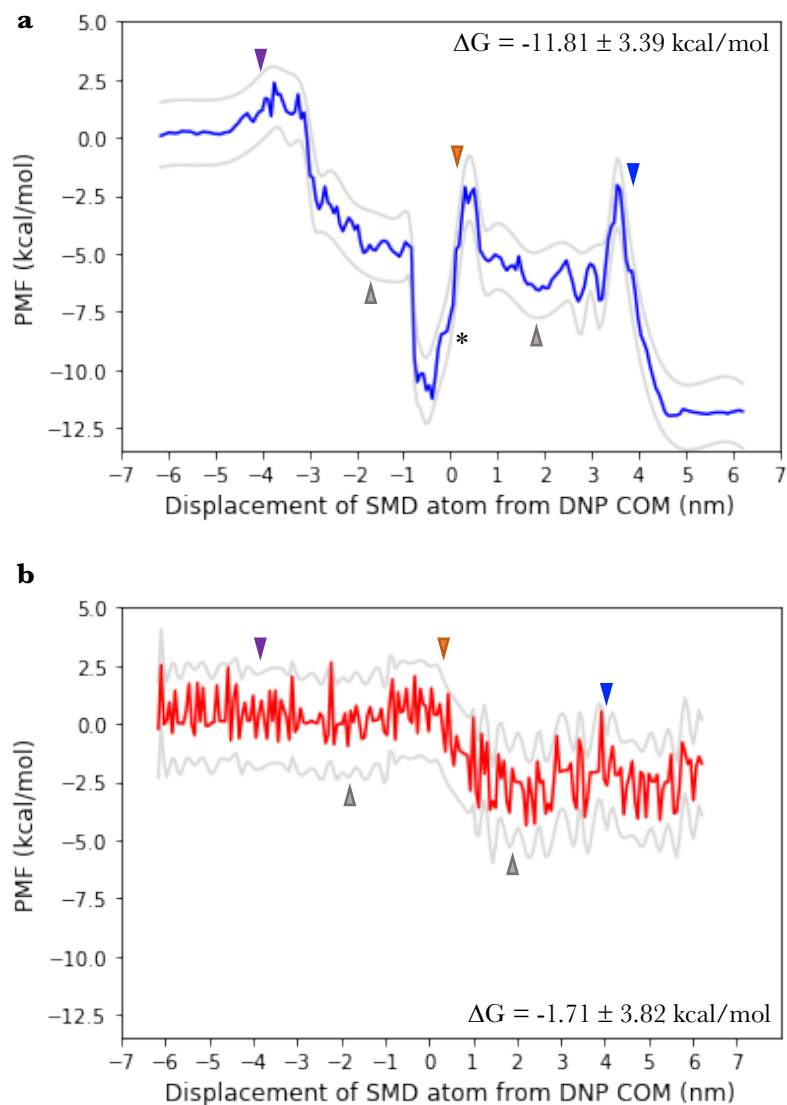


Figure 6.11: PMFs representing the 1D free energy landscapes associated with the transport CF via the TL translocation pathway (**a**) and the IF membrane adsorption pathway (**b**). The purple and blue arrows mark the approximate location of the outermost edges of the R8 and R1 regions of the DNP, and the grey arrows represent the surfaces of the lipid bilayer. The orange triangle in (**a**) indicates the position of the lumen constriction in the R4 region, where there CF molecule experiences strong repulsive interactions with the DNP midway through the translocation. The asterisk in **a** denotes a translocation coordinate (at around 0.5 nm displacement) at which insufficient overlap of force histograms may cause a minor defect in the PMF.

6.4.3 *Transport PMFs for ATTO 655*

ATTO 655 differs greatly from SRB and CF in that it has no net charge, and a relatively sparse distribution of positive and negative point charges across the length of its fused 5-ring system. The aromaticity of the ring system is not complete, and the piperidine rings on either end of the fused system are flexible enough to allow the ring system to curve slightly. In addition, the distribution of point charges is such that one end of the ~ 3 nm long fused ring system has a net charge of +1 while the other bears a net charge of -1, giving the molecule the characteristics of a flexible bar magnet. The effect that this unique feature of ATTO 665 has on the transport properties is evident in the PMFs. The calculated ΔG for both the TL and IF pathways are strongly negative within the bounds of the 95% confidence intervals, and therefore it is likely that both pathways are energetically feasible for this fluorophore. The magnitude of ΔG for the IF pathway is almost twice that of the TL pathway, however the free energy barriers that arise along the TL PMF are slightly lower in scale than the free energy barriers seen in the IF PMF. The first energetic barrier in the IF PMF (~ 2.5 kcal/mol) occurs shortly after the ATTO 655 molecule enters the lumen at the R8 opening and arises from steric clashes between the molecule and the interior of the DNP lumen.

As the ATTO 655 molecule enters the lumen, it has a tendency to align its positively charged groups directly adjacent to the negatively charged opposing DNA helices to increase favourable electrostatic interactions, however this orientation makes it difficult for it to clear the narrow opening, as shown in Figure 6.12a. The free energy then decreases gradually as it navigates the more spacious portion of the lumen, where it retains the same alignment without encountering steric clashes, before exiting through the constriction at R1, causing the free energy to rise again by ~ 2.5 kcal/mol as the negatively charged sulphate and carboxylate groups come into close contact with the DNA backbone pseudo-atoms (Figure 6.13c).

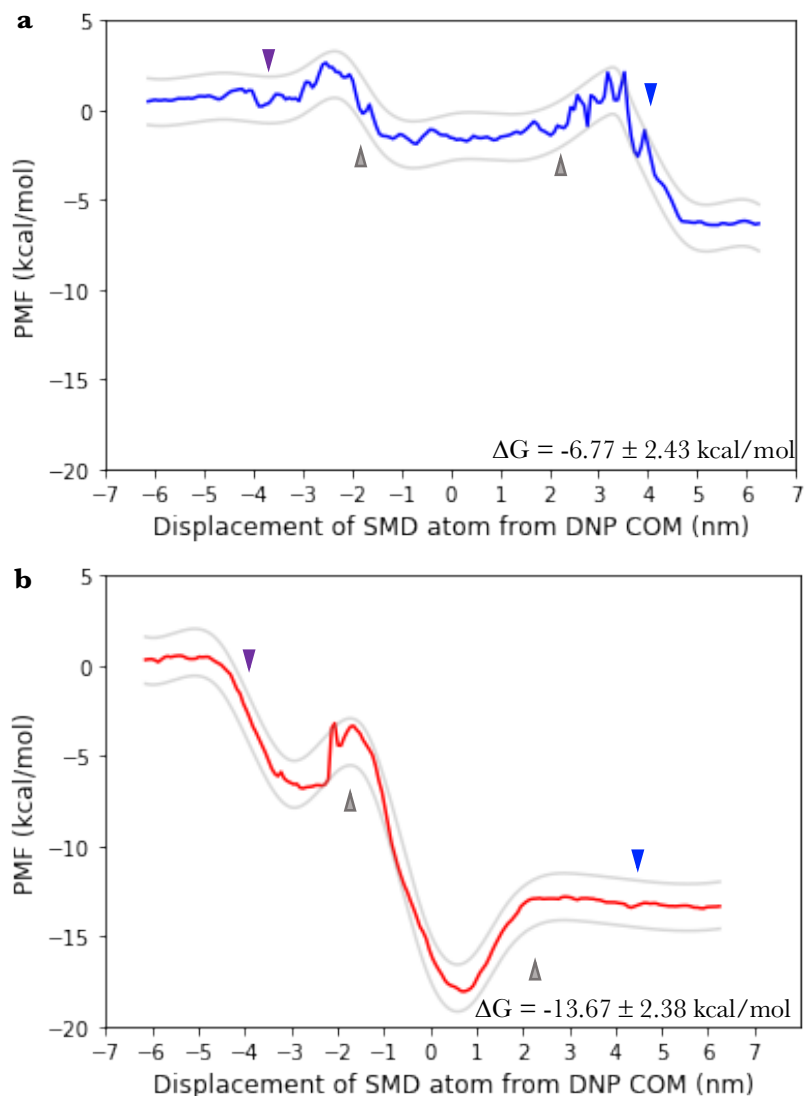


Figure 6.12: PMFs calculated for transport of ATTO 655 via the TL translocation pathway (**a**) and the IF membrane adsorption pathway (**b**). In the IF PMF, the first barrier is the result of the negatively charged end of the molecule being forced into close contact with the DNP at the interface. Once it enters the interface between the DNP and the zwitterionic lipid headgroups, it flexes to embed its hydrophobic midsection into the membrane and align its' charged ends with the headgroups, and these favourable interactions cause the free energy to fall. The departure of the ATTO 655 molecule from the interface at 2 nm displacement results in the loss of favourable hydrophobic interactions, and this coincides with the second free energy barrier seen here.

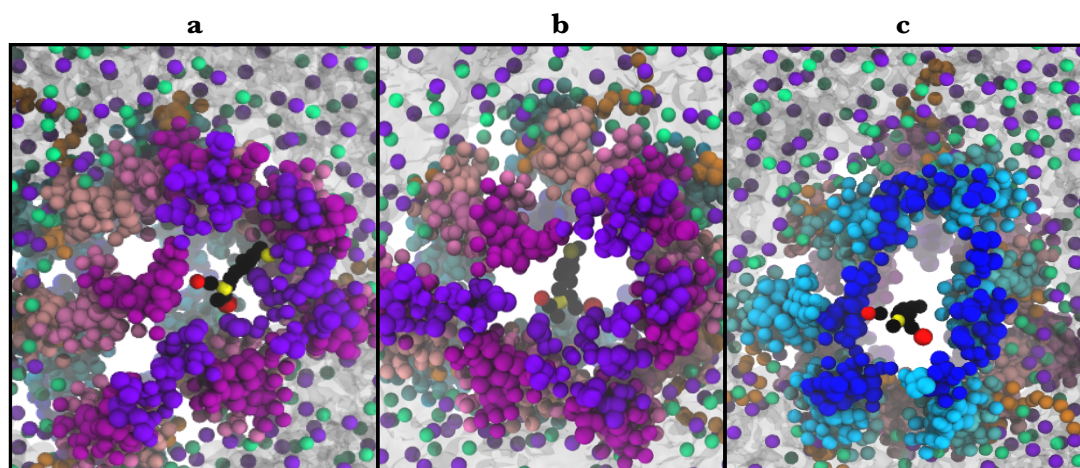


Figure 6.13: US simulation snapshots showing the orientation adopted by the ATTO 655 fluorophore as it clears the DNP lumen in the TL pathway. **(a)** As the fluorophore approaches the R8 entrance, it orients its positively charged groups (one protonated cyclic tertiary amine and one conjugated cyclic tertiary amine, indicated by yellow spheres) to face the negatively charged interior of the DNP lumen. The width of the fused ring system introduces steric clashes in this orientation, causing a small spike in the free energy profile (Fig. 6.13a). Negatively charged groups (sulphate and carboxylate) are represented as red spheres and are located adjacent to each other on one end of the fused ring system. **(b)** The fluorophore retains this orientation as it traverses the wider midsection of lumen freely. **(c)** The fluorophore tilts its negatively charged groups out of the constricted and densely charged R1 terminus as it exits the pore lumen, causing another small spike in the free energy due to repulsions.

The first of the two distinct free energy barriers associated with the IF pathway PMF (Fig. 6.12b) is comparable in scale to the free energy barriers seen in the TL pathway, which suggests that the probability of the molecule entering the DNP/membrane interface is more-or-less equal to the probability of it entering the lumen. In order to penetrate the interface, the negatively charged pseudo-atoms of the ATTO 655 molecule make close contact with the DNP backbone (Fig. 6.14a) just as they do in the TL pathway, and therefore we see similar initial free energy barriers for both pathways. After entering the interface, the PMF drops extremely rapidly by ~ 15 kcal/mol. This dramatic decrease in free energy is initiated by the flexion of the fluorophore molecule at either end of the fused ring system, which allows the hydrophobic, aromatic midsection of the molecule to embed itself into the hydrophobic interior of the lipid

torus, while the charged ends flex outwards to interact with the zwitterionic lipid headgroups (Fig. 6.14b). As the molecule exits the interface, the favourable hydrophobic interactions between the ATTO 655 midsection are lost and the negatively charged groups must again come into close contact with the DNP backbone, which gives rise to a slightly more severe secondary free energy barrier of ~ 4 kcal/mol. It is difficult to determine the extent to which this energy barrier might impede full cross-membrane transport of ATTO 655 along the IF pathway, but it is clear that the fluorophore is quite capable of entering the lumen and the interface with similar probabilities. This, along with the modest free energy barriers seen in the PMF, and the negative ΔG values calculated for both pathways implies that both are accessible, and their transport rates are comparable.

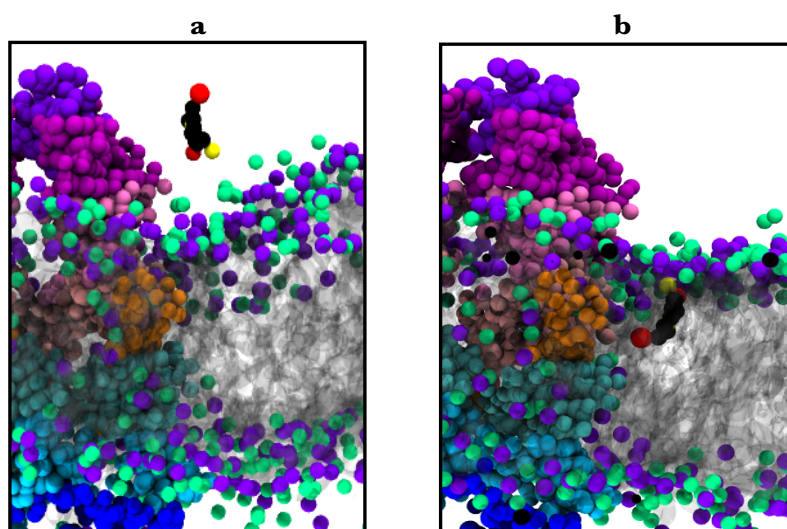


Figure 6.14: US simulation snapshots illustrating the important role of flexion of the ATTO 655 molecule as it enters the DNP/membrane interface. **(a)** The fluorophore curves its charged negatively charged groups (red) away from the DNP backbone as it approaches the interface. As it enters, the negative groups briefly make close contact with the DNP backbone. **(b)** Once the fluorophore enters the interface fully, it curves to embed its hydrophobic midsection (black) into the hydrophobic interior of the lipid torus, leaving the charged groups (red and yellow) to interact with the charged lipid headgroups (positive choline groups represented as green spheres, negative phosphate groups represented as violet spheres).

6.5 Conclusions

The qualitative comparisons of the PMF plots yielded from the six ensembles of US sampling simulations made here have allowed us to determine which fluorophores have a higher tendency to translocate across a bilayer-spanning DNP and relate these findings to experimental observations of charge-selective transport mediated by DNPs. I have ascertained that CF is significantly less likely to be transported across the bilayer via either of the two pathways (TL translocation or IF membrane adsorption) than SRB and ATTO 655 on account of its dense negative charge, meaning it is strongly repelled by the DNP and therefore unlikely to be transported through or around it. This aligns well with the observations of comparatively weak fluorescence in efflux assays performed for DNP-gated vesicles containing CF.[3] Furthermore, the large free energy barrier associated with more energetically feasible TL translocation suggests that any successful individual CF translocation events through the lumen are likely to proceed very slowly. The noisiness of the IF PMF makes it difficult to identify individual energetic barriers associated with interfacial membrane adsorption of CF, though the magnitude of the spikes in free energy that were observed were generally lower than those observed in the TL PMF. In addition, the small net decrease in the free energy after IF transport of CF indicates that it is marginally feasible. This provides a tentative explanation as to why the rate of CF efflux was greater when the DNP was in its closed state rather than its open state in the efflux assays [3] – the blockage of the slower but more energetically favourable TL pathway forces the CF to leak through the DNP/membrane interface via the faster but less favourable IF adsorption pathway.

Similar fluorescence efflux assays have revealed that both ATTO 655 and SRB have a strong tendency to translocate through the pore lumen, and the TL PMFs presented here have also demonstrated that TL translocation is likely to be both fast and feasible for these fluorophores when the DNP is in its open state. Interfacial adsorption also appears to be highly favourable for these two fluorophores, however this would be difficult to prove with existing fluorophore efflux assay data as it is not clear whether or not adsorption of these fluorophores to the outer surface of the vesicle would give rise to a fluorescence signal. Structural differences in the design of the DNP studied here and the design used in experiment may also diminish the propensity for IF adsorption. The gated DNPs used in fluorescence efflux assays made use longer DNA strands to introduce the gating functionality that was required for the efflux assays.

Finally, I was able to comment on the structural and electrostatic properties of ATTO 655 that enable its' facile transport through the DNP lumen, which gives us new insights into the type of molecular cargo that these DNPs are most suited to transporting across membranes.

Chapter 7

Summary and Future Outlook

In this thesis, I have studied the structure, dynamics, conductance and translocational properties of archetypal cholesterol-anchored DNA nanopores using an ensemble-based coarse-grained MD protocol, which enabled me to access microsecond timescales across an extensive range of simulation systems representing different experimental conditions, whilst maintaining tight control of errors throughout.

In bulk solution, the DNP adopts a bloated and anisotropic conformation, where repulsions between helices cause the DNP to deviate from the idealised barrel-like structure, with large spaces between adjacent helices and an undefined lumen. The ensembles of CG simulations of the DNP in 0.3 M and 1.0 M NaCl identified which helices are most susceptible to kinking and base-pair breakage, and back-mapping middle structures from clustered CG trajectories to their AA conformation allowed me to calculate accurate and precise dimensions and kink angles for the solvated DNP, which were validated against cryo-EM structures and fully atomistic simulations in 0.3 M NaCl. The nicks in the R4 region of the pore coincide with the centre of the helix kink angles, indicating that the presence of the nick sites contribute to the DNP's characteristic breathing motions. The poly(T) inter-helix crossovers at the R1 and R8 pore termini are highly strained and cause local constriction of the nanopore at the termini; the R8 terminus being the narrower of the two (by ~ 1 nm). Overall, the concentration of NaCl had little effect on the overall structure and dimensions of the solvated pore. However, this was not the case for the membrane-spanning DNP, where salt concentration had a profound effect on the lumen profile, dynamics and membrane-binding mode. In the lower salt condition ensemble, the DNP was extremely mobile within the bilayer across all replicas and was eventually ejected from the bilayer in the two thirds of the replicas to adopt the side-on membrane binding mode for the latter part of the affected replicas, while the transmembrane binding mode was retained in the remaining third. This was not observed in the ensemble of membrane/DNP simulations performed in 1.0 M NaCl, where the DNP remained in stable transmembrane orientation, with a smooth defined pore lumen, and relatively small inter-helix distance of ~ 2.4 nm. The expulsion of the DNP in 0.3 M NaCl, and the subsequent side-on membrane binding mode suggests that lower monovalent salt

concentrations do not provide sufficient electrostatic screening to the DNP/POPC bilayer system, and the inter-helix repulsion caused by compression of the DNP within surrounding lipid torus is too large for the system to withstand. Hence, higher salt concentrations appear to stabilise the transmembrane binding mode.

After characterising the structural features of the pore and their response to ionic strength, I demonstrated how an ensemble-based CG approach using the CompEl protocol for simulated electrophysiology could be used to accurately determine the average conductance of the membrane spanning DNP in 0.3 M NaCl. In addition to recreating the experimental conductance, the CompEl simulations also recreated the voltage-dependent suppression of conductance that was seen in previous experimental studies [3]. Finally, I discussed the free energy profiles associated with the DNP-mediated transport pathways of the SRB, CF and ATTO 655 fluorophores. The ensemble-based protocol utilised in this study identified two distinct transport pathways: the trans-lumen translocation pathway and the interfacial membrane-adsorption pathway through the DNP/membrane interface, which leads to binding of the fluorophore to the outer bilayer surface. In agreement with experimental data derived from kinetic efflux assays, the PMF plots show that DNP is very likely to facilitate transport (via either pathway) of SRB and ATTO 655 to a greater extent than CF. The density of negative point charges on the relatively small CF molecule gives rise to pronounced repulsive forces between the DNP and the approaching fluorophore molecule regardless of the pathway taken. While the trans-lumen pathway for CF is energetically favourable, it is likely to proceed slowly due to the presence of a large energy barrier at the translocation coordinate that coincides with the fluorophores' passage through the constriction at the R4 region of the lumen. The weakly negatively charged SRB and net charge neutral ATTO 655 experience a much lower degree of repulsion and are able to translocate freely via either pathway.

The open-ended and exploratory nature of this study allowed me to investigate a comprehensive range of the DNPs behavior and structural properties. That said, there are several aspects of this study that require further work and consideration, and protocols that require refinement to improve their computational efficiency and enhance the precision of the results they produce. I will briefly discuss these aspects and propose actions to address them.

7.1 Stiff Elastic Network vs Soft Elastic Network for MARTINI DNA

As explained in Chapters 2 and 3, the MARTINI force field makes use of an elastic network of restraints to reinforce base-pairing and maintain the persistence length of dsDNA strands, of which there are two varieties: the “stiff” elastic network and the “soft” elastic network. The stiff elastic network restrains both the DNA backbone and the nucleotide bases, such that rotation of the bases and base-flipping is hindered, whereas the soft elastic network only restrains the backbone, allowing for more realistic dynamics of the nucleotide bases. The main advantage of employing the stiff elastic network is efficiency; fluctuations in the structure of the MARTINI DNA are relatively minor between timesteps, and this allows for longer timesteps and ultimately faster simulations where longer timescales are more easily accessible.[84] As this study was aimed at gaining a comprehensive understanding of the DNPs various functionalities and responses to environmental conditions using an ensemble-based protocol, a lot of long-duration simulations were needed, and efficiency was a high priority. For this reason, the stiff elastic network was employed in all of the simulations conducted for this thesis. The consequences of this were seen in Chapter 3. While loss of resolution in the dimensions of the DNP was easily corrected by back-mapping clustered middle structures to their AA representation, there was an underestimation of the total RMSF in the CG models compared to the AA models that could not be corrected for so easily and the higher flexibility of the R1 terminus compared to the R8 terminus that was revealed in the AA simulations was missed in the CG simulations.

For future studies of this nature, both the stiff and soft elastic restraints should be tested, and validation of the results against the available experimental data should inform the choice of restraints going forward. The stiff elastic network may have also influenced the results obtained from the DNP/membrane simulations discussed in Chapter 4. With the rotations of nucleotide bases hindered, the repulsions between negatively charged backbones on adjacent helices within the DNP may have been somewhat exaggerated. We saw that in the 0.3 M NaCl ensemble, the repulsion between the helices caused the nanopore to exit the bilayer in a subset of the replicas. To confirm whether or not this behaviour is an artefact, further equilibrium simulations of the 0.3 M DNP/membrane system could be performed with the soft elastic restraints.

7.2 Energetics of Insertion and Expulsion of DNPs

Another more direct approach to establish the cause of the DNP expulsion events witnessed in Chapter 4 would be to perform cv-SMD simulations of the DNP inserting into and exiting the POPC bilayer, using similar methods to those described in Chapter 6, and in the 2017 paper by Maingi *et al.*[25] which investigated the insertion of a similar DNA nanopore into planar lipid bilayers. Free energy profiles could be yielded for the trans-membrane insertion (pulling in) and expulsion (pulling out) of the DNP at the two salt concentrations studied here (0.3 M NaCl and 1.0 M NaCl) using ensembles of cv-SMD simulations in conjunction with Jarzynski's equality [172], or with subsequent US simulations. If the energy barriers associated with the insertion PMFs exceed (in number or magnitude) the energy barriers involved in expulsion, then it is probable that the transmembrane orientation is energetically unfavorable in the given set of conditions. By comparing the free energy landscapes associated with pulling in at the two salt concentrations, we can determine the role that ionic strength plays to facilitate or obstruct trans-membrane insertion. Comparison of the pull-out PMFs would allow us to elucidate which conditions (if any) might increase the likelihood of expulsion from the membrane. It is also important to refine the protocol to ensure that long-range electrostatic interactions are described as realistically as possible, so a PME electrostatics scheme should be employed in future simulations of these systems.

7.3 CompEl and Applied Electric Field Simulations

In Chapter 5, I highlighted the utility and efficiency of the CompEl protocol for the accurate derivation of the current-voltage relationships for bilayer-spanning DNPs. The simulations used up a relatively small amount of computational resources (under 100,000 core hours) and produced reproducible results for the average conductance with converged standard errors, however the magnitude of the standard errors was larger than the errors reported in experiment due to finite size effects, and the fluctuations in the instantaneous voltage brought about by the water-ion swap events. The precision of these results could be improved by running further simulations with a larger membrane patch, which would reduce the influence of rapid water-ion

exchanges on the transmembrane potential difference, which in turn dampens the fluctuations in voltage. Less extreme fluctuations in voltage result in less uncertainty in the instantaneous voltage, and therefore less uncertainty in the average conductance.

Furthermore, the depression in the slope of the I-V plot that was caused by the inclusion of measurements taken at higher instantaneous voltage could be explored further by running longer duration equilibrium simulations of the membrane-spanning DNP under a consistent potential difference above 200 mV. This would allow us to analyse the equilibrium behaviour of individual system components under these voltage conditions, potentially shedding light on the cause of the sub-conductance states that have been seen at this voltage range in experiment and alluded to in the results discussed in Chapter 5. As mentioned previously, maintaining a constant transmembrane potential difference in a system with a fluctuating dielectric is not trivial. In both NAMD[100] and GROMACS[98], users have the option to apply a uniform external electric field along the membrane normal, along the principle vector of the transmembrane nanopore. Assuming the system is simulated for long enough, and enough replicas are performed, the influence exerted by fluctuations in the box vectors and dielectric distribution on the magnitude of the potential difference becomes negligible, and the average potential difference will reach a state of convergence.

7.4 Refinement of cv-SMD Protocol for Fluorophore Translocation

The umbrella sampling simulations discussed in Chapter 6 were aimed at elucidating the free energy profiles of two different translocation pathways: the trans-lumen (TL) and interfacial (IF) pathways. However, all three fluorophores displayed a tendency to drift away from the translocation coordinate in certain umbrella sampling windows in the IF pathway (corresponding to displacements of 4 nm to 5 nm from the DNP centre of mass), indicating that the spring constant of the positional restraints employed in these windows was insufficient to correctly model full interfacial translocation. Hence, the IF pathway studied here is not strictly a translocation pathway, as the free energy profiles do not return to zero once the fluorophore passes through the membrane via the DNP/membrane interface. Rather, it is a transport pathway that involves adsorption of the fluorophore to the bilayer surface, and hence the calculated free

energy change ΔG corresponds to a free energy of membrane adsorption. In order to compare the energetic feasibility of the IF and TL pathways, the umbrella sampling simulations for the IF pathways should be repeated with stronger positional restraints.

As discussed in Chapter 6, there is more than one tried and tested method used to obtain free energy profiles for non-equilibrium processes such as translocation from cv-SMD simulations. The ensemble-based US protocol used in this thesis yielded reproducible PMFs, however the standard deviation of the distributions of the free energy change (ΔG) values calculated for the CF pathways were relatively large in comparison to the equivalents for ATTO 655 and SRB, as were the associated standard errors. This can be attributed to the pulling parameters that were used for the cv-SMD simulations of CF translocation, which resulted in suboptimal spacing between USWs in the subsequent US simulation. The same pulling rate and harmonic spring constant spacing between USWs (~ 0.23 nm) was used in the pulling and US simulations run for all three fluorophores, and while these parameters returned generally good overlap between force histograms and smooth PMFs for simulations of ATTO 655 and SRB, the same cannot be said for the CF simulations.

Hence, it would be wise to test different combinations of pulling/umbrella sampling parameters for the CF translocation model, to improve the sampling and there reduce the uncertainty in the results. An alternative methodology for the derivation of free energy profiles, using ensembles of cv-SMD simulations in combination with Jarzynski equality could also be tested on this system. This methodology has the potential to produce high quality free energy profiles at a fraction of the computational cost of the more traditional US approach, provided the pulling parameters used in the cv-SMD have been thoroughly tested, and the ensemble of cv-SMD simulations is large enough to yield converged standard errors for the free energy.[175]

7.5 Final Remarks

The findings described in this thesis have shed some light on the relationship between the structural features of the cholesterol-anchored DNA nanopore and its' functional properties, enabling a deeper understanding of phenomena that have been witnessed in experimental studies of DNPs thus far. The pronounced bloating and flexibility of the midsection of the DNP around the nick sites in the bulk solution simulations

constitute an unexpected discovery, as the pore was initially assumed to adopt a rigid cylindrical structure with straight unkinked helices. The rigidity of the pore may be enhanced by adding additional crossovers between adjacent helices near the nick sites, or perhaps by switching from a six-helix bundle design to a five-helix bundle. The formation of a toroidal lipid pore around the POPC membrane-spanning DNP indicates that the POPC lipids undergo major rearrangements to decrease the hydrophobic mismatch between the DNP and the membrane, which explains why the insertion of the DNP into planar bilayers is often very slow. The compressive force of the toroidal lipid pore around the DNP gives rise to significant inter-helix repulsions, which may introduce a thermodynamic barrier to successful transmembrane insertion. We also expect that the use of higher ionic strength solutions (~ 1.0 M NaCl) in experimental settings is likely to decrease the inter-helix repulsion experienced by a membrane-spanning DNP, and this in turn may improve the frequency of successful insertions into planar bilayers. Finally, the results from the cv-SMD simulations confirm that the DNP does indeed discriminate between different kinds of molecular cargo according to their charge, and that it is highly suited to transporting slightly flexible molecules with zero net charge such as ATTO 655.

Bibliography

- [1] S. Howorka, “Building membrane nanopores,” *Nat. Nanotechnol.*, vol. 12, no. 7, pp. 619–630, 2017.
- [2] S. Howorka, “The Changing of the Guard,” *Science (80-.)*, vol. 352, no. 6288, p. 890, 2016.
- [3] J. R. Burns, A. Seifert, N. Fertig, and S. Howorka, “A biomimetic DNA-based channel for the ligand-controlled transport of charged molecular cargo across a biological membrane,” *Nat. Nanotechnol.*, vol. 11, no. 2, pp. 152–156, 2016.
- [4] O. Birkholz, J. R. Burns, C. P. Richter, S. Howorka, and J. Piehler, “Multi-functional DNA nanostructures that puncture and remodel lipid membranes into hybrid materials,” *Nat. Commun.*, vol. 9, no. 1521, 2018.
- [5] P. Chidchob and H. F. Sleiman, “Recent advances in DNA nanotechnology,” *Curr. Opin. Chem. Biol.*, vol. 46, pp. 63–70, 2018.
- [6] P. W. K. Rothmund, “Folding DNA to create nanoscale shapes and patterns,” vol. 440, no. March, pp. 297–302, 2006.
- [7] I. Gállego, B. Manning, J. D. Prades, M. Mir, J. Samitier, and R. Eritja, “DNA-Origami-Driven Lithography for Patterning on Gold Surfaces with Sub-10 nm Resolution,” *Adv. Mater.*, vol. 29, no. 11, pp. 1–7, 2017.
- [8] T. Lazaridis and A. Rahaman, “A thermodynamic approach to Alamethicin pore formation,” vol. 49, no. 18, pp. 1841–1850, 2009.
- [9] H. Bayley and P. S. Cremer, “Stochastic sensors inspired by biology,” *Nature*, vol. 413, no. 6852, pp. 226–230, 2001.
- [10] H. Bayley, O. Braha, and L. Q. Gu, “Stochastic sensing with protein pores,” *Adv. Mater.*, vol. 12, no. 2, pp. 139–142, 2000.
- [11] D. W. Deamer and M. Akeson, “Nanopores and nucleic acids: Prospects for ultrarapid sequencing,” *Trends Biotechnol.*, vol. 18, no. 4, pp. 147–151, 2000.
- [12] “Oxford Nanopore Technologies,” 2018. [Online]. Available: <https://nanoporetech.com>. [Accessed: 17-Sep-2018].
- [13] K. R. Lieberman, G. M. Cherf, M. J. Doody, F. Olasagasti, Y. Kolodji, and M. Akeson, “Processive replication of single DNA molecules in a nanopore catalyzed by phi29 DNA polymerase,” *J. Am. Chem. Soc.*, vol. 132, no. 50, pp. 17961–17972, 2010.
- [14] G. F. Schneider and C. Dekker, “DNA sequencing with nanopores,” *Nat. Biotechnol.*, vol. 30, no. 4, pp. 326–328, 2012.
- [15] S. Wang, Z. Shengyi, F. Haque, and P. Guo, “Engineering of Protein Nanopores for Sequencing, Chemical or Protein Sensing and Disease Diagnosis Shaoying,” vol. 51, pp. 80–89, 2019.
- [16] M. Soskine, A. Biesemans, B. Moeyaert, S. Cheley, H. Bayley, and G. Maglia, “An Engineered ClyA Nanopore Detects Folded Target Proteins by Selective External Association and Pore Entry,” pp. 1–6, 2012.
- [17] M. Soskine, A. Biesemans, M. De Maeyer, and G. Maglia, “Tuning the size and properties of ClyA nanopores assisted by directed evolution,” *J. Am. Chem. Soc.*, vol. 135, no. 36, pp. 13456–13463, 2013.
- [18] N. R. Zaccai *et al.*, “A de novo peptide hexamer with a mutable channel,” *Nat. Chem. Biol.*, vol. 7, no. 12, pp. 935–941, 2011.
- [19] M. Ayub and H. Bayley, “Engineered transmembrane pores,” *Curr. Opin. Chem. Biol.*, vol. 34, pp. 117–126, 2016.
- [20] C. Cao *et al.*, “Single-molecule sensing of peptides and nucleic acids by engineered aerolysin nanopores,” *Nat. Commun.*, vol. 10, no. 1, 2019.
- [21] T. Z. Butler, M. Pavlenok, I. M. Derrington, M. Niederweis, and J. H. Gundlach, “Single-molecule DNA detection with an engineered MspA protein nanopore,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 105, no. 52, pp. 20647–20652, 2008.
- [22] S. Hernández-Ainsa and U. F. Keyser, “DNA origami nanopores: developments, challenges and perspectives,” *Nanoscale*, vol. 6, no. 23, pp. 14121–14132, 2014.
- [23] P. S. Spinney, S. D. Collins, D. G. Howitt, and R. L. Smith, “Fabrication and

- characterization of a solid-state nanopore with self-aligned carbon nanoelectrodes for molecular detection,” *Nanotechnology*, vol. 23, no. 13, 2012.
- [24] J. R. Burns, E. Stulz, and S. Howorka, “Self-assembled DNA nanopores that span lipid bilayers,” *Nano Lett.*, vol. 13, no. 6, pp. 2351–2356, 2013.
- [25] V. Maingi, J. R. Burns, J. J. Uusitalo, S. Howorka, S. J. Marrink, and M. S. P. Sansom, “Stability and dynamics of membrane-spanning DNA nanopores,” *Nat. Commun.*, vol. 8, p. 14784, 2017.
- [26] K. Göpfrich *et al.*, “Ion Channels Made from a Single Membrane-Spanning DNA Duplex,” *Nano Lett.*, vol. 16, pp. 4665–4669, 2016.
- [27] J. R. Burns *et al.*, “Lipid-bilayer-spanning DNA nanopores with a bifunctional porphyrin anchor,” *Angew. Chemie - Int. Ed.*, vol. 52, no. 46, pp. 12069–12072, 2013.
- [28] V. Maingi, M. Lelimosin, S. Howorka, and M. S. P. Sansom, “Gating-like Motions and Wall Porosity in a DNA Nanopore Scaffold Revealed by Molecular Simulations,” *ACS Nano*, vol. 9, no. 11, pp. 11209–11217, 2015.
- [29] J. A. Killian, “Hydrophobic mismatch between proteins and lipids in membranes,” *Biochim. Biophys. Acta - Rev. Biomembr.*, vol. 1376, no. 3, pp. 401–415, 1998.
- [30] T. M. Weiss, P. C. A. Van der Wel, J. A. Killian, R. E. Koeppe, and H. W. Huang, “Hydrophobic mismatch between helices and lipid bilayers,” *Biophys. J.*, vol. 84, no. 1, pp. 379–385, 2003.
- [31] S. K. Kandasamy and R. G. Larson, “Molecular dynamics simulations of model trans-membrane peptides in lipid bilayers: A systematic investigation of hydrophobic mismatch,” *Biophys. J.*, vol. 90, no. 7, pp. 2326–2343, 2006.
- [32] M. Venturoli, B. Smit, and M. M. Sperotto, “Simulation studies of protein-induced bilayer deformations, and lipid-induced protein tilting, on a mesoscopic model for lipid bilayers with embedded proteins,” *Biophys. J.*, vol. 88, no. 3, pp. 1778–1798, 2005.
- [33] M. Jensen and O. G. Mouritsen, “Lipids do influence protein function - The hydrophobic matching hypothesis revisited,” *Biochim. Biophys. Acta - Biomembr.*, vol. 1666, no. 1–2, pp. 205–226, 2004.
- [34] H. J. Kaiser *et al.*, “Lateral sorting in model membranes by cholesterol-mediated hydrophobic matching,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 108, no. 40, pp. 16628–16633, 2011.
- [35] R. Briones, C. Aponte-Santamaría, and B. L. de Groot, “Localization and ordering of lipids around aquaporin-0: Protein and lipid mobility effects,” *Front. Physiol.*, vol. 8, no. MAR, 2017.
- [36] E. Zakharian, “Recording of Ion Channel Activity in Planar Lipid Bilayer Experiments,” in *Ion Channels: Methods and Protocols, Methods in Molecular Biology*, 2nd ed., vol. 998, N. Gamper, Ed. New York: Humana Press, 2013, pp. 109–118.
- [37] S. M. Bierbower and M. S. Shapiro, “Förster Resonance Energy Transfer-Based Imaging at the Cell Surface of Live Cells,” in *Ion Channels: Method and Protocols, Methods in Molecular Biology*, 2nd ed., vol. 998, N. Gamper, Ed. New York: Humana Press, 2013, pp. 209–216.
- [38] J. R. Burns, A. Seifert, N. Fertig, and S. Howorka, “A biomimetic DNA-based channel for the ligand-controlled transport of charged molecular cargo across a biological membrane,” *Nat. Nanotechnol.*, vol. 11, no. 2, pp. 152–156, 2016.
- [39] S. M. J. L. Van Den Wildenberg, B. Prevo, and E. J. G. Peterman, *Single Molecule Analysis*, vol. 783. New York: Humana Press, 2011.
- [40] N. Yilmaz and T. Kobayashi, “Assemblies of pore-forming toxins visualized by atomic force microscopy,” *Biochim. Biophys. Acta - Biomembr.*, vol. 1858, no. 3, pp. 500–511, 2016.
- [41] J. Milne, M. Borgnia, and A. Bartesaghi, “Cryo-electron microscopy: A primer for the non-microscopist,” *Febs J.*, vol. 280, no. 1, pp. 28–45, 2013.
- [42] W. Peng, M. de Souza Santos, Y. Li, D. R. Tomchick, and K. Orth, “High-resolution cryo-EM structures of the E. Coli hemolysin ClyA oligomers,” *PLoS One*, vol. 14, no. 5, pp. 1–17, 2019.

- [43] A. Seifert, K. Göpfrich, J. R. Burns, N. Fertig, U. F. Keyser, and S. Howorka, “Bilayer-spanning DNA nanopores with voltage-switching between open and closed state,” *ACS Nano*, vol. 9, no. 2, pp. 1117–1126, 2015.
- [44] V. L. Sukhorukov *et al.*, “Pore size of swelling-activated channels for organic osmolytes in Jurkat lymphocytes, probed by differential polymer exclusion,” *Biochim. Biophys. Acta - Biomembr.*, vol. 1788, no. 9, pp. 1841–1850, 2009.
- [45] N. A. W. Bell *et al.*, “DNA Origami Nanopores,” *Nano Lett.*, vol. 12, pp. 512–517, 2012.
- [46] M. Langecker *et al.*, “Synthetic lipid membrane channels formed by designed DNA nanostructures,” *Science (80-.)*, vol. 338, no. 6109, pp. 932–936, 2012.
- [47] J. R. Burns, N. Al-Juffali, S. M. Janes, and S. Howorka, “Membrane-spanning DNA nanopores with cytotoxic effect,” *Angew. Chemie - Int. Ed.*, vol. 53, no. 46, pp. 12466–12470, 2014.
- [48] P. M. Arnott and S. Howorka, “A Temperature-Gated Nanovalve Self-Assembled from DNA to Control Molecular Transport across Membranes,” *ACS Nano*, vol. 13, no. 3, pp. 3334–3340, 2019.
- [49] J. Yoo and A. Aksimentiev, “Molecular Dynamics of Membrane-Spanning DNA Channels: Conductance Mechanism, Electro-Osmotic Transport, and Mechanical Gating,” *J. Phys. Chem. Lett.*, vol. 6, no. 23, pp. 4680–4687, 2015.
- [50] K. Göpfrich *et al.*, “Large-Conductance Transmembrane Porin Made from DNA Origami,” *ACS Nano*, vol. 10, no. 9, pp. 8207–8214, 2016.
- [51] G. A. Venkatesan, G. J. Taylor, C. M. Basham, N. G. Brady, C. P. Collier, and S. A. Sarles, “Evaporation-induced monolayer compression improves droplet interface bilayer formation using unsaturated lipids,” *Biomicrofluidics*, vol. 12, no. 2, pp. 1–13, 2018.
- [52] T. Wang, D. Schiffels, S. Martinez Cuesta, D. Kuchnir Fygenon, and N. C. Seeman, “Design and characterization of 1D nanotubes and 2D periodic arrays self-assembled from DNA multi-helix bundles,” *J. Am. Chem. Soc.*, vol. 134, no. 3, pp. 1606–1616, 2012.
- [53] D. A. Case *et al.*, “AMBER 2016.” University of California, San Francisco, 2016.
- [54] H. Joshi and P. K. Maiti, “Structure and electrical properties of DNA nanotubes embedded in lipid bilayer membranes,” *Nucleic Acids Res.*, vol. 46, no. 5, pp. 2234–2242, 2018.
- [55] K. Göpfrich *et al.*, “Ion channels made from a single membrane-spanning DNA duplex,” *Nano Lett.*, vol. 16, no. 7, pp. 4665–4669, 2016.
- [56] S. Krishnan *et al.*, “Molecular transport through large-diameter DNA nanopores,” *Nat. Commun.*, vol. 7, p. 12787, 2016.
- [57] S. Naskar, M. Gosika, H. Joshi, and P. K. Maiti, “Tuning the Stability of DNA Nanotubes with Salt,” *J. Phys. Chem. C*, vol. 123, no. 14, pp. 9461–9470, 2019.
- [58] A. R. Leach, *Molecular Modelling, Principles and Applications*, 2nd ed. Harlow: Prentice Hall, 2001.
- [59] M. Allen, “Introduction to molecular dynamics simulation,” in *Computational Soft Matter: From Synthetic Polymers to Proteins*, vol. 23, no. 2, 2004, pp. 1–28.
- [60] S. Hug, “Classical Molecular Dynamics in a Nutshell,” in *Biomolecular Simulations, Methods and Protocols*, 1st ed., L. Monticelli, E. Salonen, and J. M. Walker, Eds. New York: Humana Press, 2013, pp. 127–163.
- [61] S. Piana, K. Lindorff-Larsen, R. M. Dirks, J. K. Salmon, R. O. Dror, and D. E. Shaw, “Evaluating the effects of cutoffs and treatment of long-range electrostatics in protein folding simulations,” *PLoS One*, vol. 7, no. 6, pp. 6–11, 2012.
- [62] P. P. Ewald, “Die Berechnung optischer und elektrostatischer Gitterpotentiale,” *Ann. Phys.*, vol. 369, no. 3, pp. 253–287, 1921.
- [63] M. T. C. Martins-Costa and M. F. Ruiz-López, “Gromacs 5.1 Manual,” *J. Comput. Chem.*, vol. 38, no. 10, pp. 659–668, 2017.
- [64] I. Fukuda and H. Nakamura, “Non-Ewald methods: Theory and applications to molecular systems,” *Biophys. Rev.*, vol. 4, no. 3, pp. 161–170, 2012.

- [65] L. Onsager, “Electric Moments of Molecules in Liquids,” *J. Am. Chem. Soc.*, vol. 58, no. 8, pp. 1486–1493, Aug. 1936.
- [66] M. Karttunen, J. Rottler, I. Vattulainen, and C. Sagui, “Chapter 2 Electrostatics in Biomolecular Simulations: Where Are We Now and Where Are We Heading?,” *Curr. Top. Membr.*, vol. 60, no. January 2018, pp. 49–89, 2008.
- [67] R. Galindo-Murillo *et al.*, “Assessing the Current State of Amber Force Field Modifications for DNA,” *J. Chem. Theory Comput.*, vol. 12, no. 8, pp. 4114–4127, 2016.
- [68] K. Hart, N. Foloppe, C. M. Baker, E. J. Denning, L. Nilsson, and A. D. MacKerell, “Optimization of the CHARMM additive force field for DNA: Improved treatment of the BI/BII conformational equilibrium,” *J. Chem. Theory Comput.*, vol. 8, no. 1, pp. 348–362, 2012.
- [69] S. Y. Reddy, F. Leclerc, and M. Karplus, “DNA polymorphism: A Comparison of Force Fields for Nucleic Acids,” vol. 84, no. 3, pp. 1421–1449, 2003.
- [70] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, “CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations,” vol. 4, no. 2, pp. 187–217, 1983.
- [71] B. R. Brooks *et al.*, “CHARMM: The Biomolecular Simulation Program,” 2009.
- [72] A. D. MacKerell *et al.*, “All-atom empirical potential for molecular modeling and dynamics studies of proteins,” *J. Phys. Chem. B*, vol. 102, no. 18, pp. 3586–3616, 1998.
- [73] K. Vanommeslaeghe *et al.*, “CHARMM General Force Field (CGenFF): A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields,” *J. Comput. Chem.*, vol. 31, no. 4, pp. 671–690, 2011.
- [74] K. Vanommeslaeghe, E. P. Raman, and A. D. MacKerell, “Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges,” *J. Chem. Inf. Model.*, vol. 52, no. 12, pp. 3155–3168, 2012.
- [75] K. Vanommeslaeghe and A. D. MacKerell, “Automation of the CHARMM general force field (CGenFF) I: Bond perception and atom typing,” *J. Chem. Inf. Model.*, vol. 52, no. 12, pp. 3144–3154, 2012.
- [76] C. Globisch, V. Krishnamani, M. Deserno, and C. Peter, “Optimization of an Elastic Network Augmented Coarse Grained Model to Study CCMV Capsid Deformation,” *PLoS One*, vol. 8, no. 4, 2013.
- [77] A. Debnath, S. Wiegand, H. Paulsen, K. Kremer, and C. Peter, “Derivation of coarse-grained simulation models of chlorophyll molecules in lipid bilayers for applications in light harvesting systems,” *Phys. Chem. Chem. Phys.*, vol. 17, no. 34, pp. 22054–22063, 2015.
- [78] T. E. Ouldridge, A. A. Louis, and J. P. K. Doye, “Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model,” *J. Chem. Phys.*, vol. 134, no. 8, 2011.
- [79] H. I. Ingólfsson *et al.*, “The power of coarse graining in biomolecular simulations,” *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, vol. 4, no. 3, pp. 225–248, 2014.
- [80] M. Pasi, R. Lavery, and N. Ceres, “PaLaCe: A coarse-grain protein model for studying mechanical properties,” *J. Chem. Theory Comput.*, vol. 9, no. 1, pp. 785–793, 2013.
- [81] J. Barnoud and L. Monticelli, “Coarse-Grained Force Fields for Molecular Simulations,” in *Molecular Modeling of Proteins*, A. Kukol, Ed. New York, NY: Springer New York, 2015, pp. 125–149.
- [82] S. J. Marrink, A. H. de Vries, and A. E. Mark, “Coarse Grained Model for Semiquantitative Lipid Simulations,” *J. Phys. Chem. B*, vol. 108, no. 2, pp. 750–760, 2004.
- [83] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. De Vries, “The MARTINI force field: Coarse grained model for biomolecular simulations,” *J. Phys. Chem. B*, vol. 111, no. 27, pp. 7812–7824, 2007.
- [84] J. J. Uusitalo, H. I. Ingólfsson, P. Akhshi, D. P. Tieleman, and S. J. Marrink, “Martini Coarse-Grained Force Field: Extension to DNA,” *J. Chem. Theory Comput.*, vol. 11, no. 8, pp. 3932–3945, 2015.

- [85] L. D. Schuler, X. Daura, and W. F. Van Gunsteren, “An Improved GROMOS96 Force Field for Aliphatic Hydrocarbons in the Condensed Phase,” *J. Comput. Chem.*, vol. 22, no. 11, pp. 1205–1218, 2001.
- [86] L. Degève, S. M. Vechi, and C. Q. Junior, “The hydration structure of the Na⁺ and K⁺ ions and the selectivity of their ionic channels,” *Biochim. Biophys. Acta - Bioenerg.*, vol. 1274, no. 3, pp. 149–156, 1996.
- [87] S. O. Yesylevskyy, L. V. Schäfer, D. Sengupta, and S. J. Marrink, “Polarizable Water Model for the Coarse-Grained MARTINI,” *PLoS Comput. Biol.*, vol. 6, no. 6, 2010.
- [88] M. A. González, “Force fields and molecular dynamics simulations,” *Collect. SFN*, vol. 12, pp. 169–200, 2011.
- [89] B. Hess, C. Holm, and N. Van Der Vegt, “Osmotic coefficients of atomistic NaCl (aq) force fields,” *J. Chem. Phys.*, vol. 124, no. 16, 2006.
- [90] N. A. Altwajry, M. Baron, D. W. Wright, P. V. Coveney, and A. Townsend-Nicholson, “An Ensemble-Based Protocol for the Computational Prediction of Helix-Helix Interactions in G Protein-Coupled Receptors using Coarse-Grained Molecular Dynamics,” *J. Chem. Theory Comput.*, vol. 13, pp. 2254–2270, 2017.
- [91] M. Xue, L. Cheng, I. Faustino, W. Guo, and S. J. Marrink, “Molecular Mechanism of Lipid Nanodisk Formation by Styrene-Maleic Acid Copolymers,” *Biophys. J.*, vol. 115, no. 3, pp. 494–502, 2018.
- [92] V. Corradi *et al.*, “Lipid-Protein Interactions Are Unique Fingerprints for Membrane Proteins,” *ACS Cent. Sci.*, vol. 4, no. 6, pp. 709–717, 2018.
- [93] S. Khalid, P. J. Bond, J. Holyoake, R. W. Hawtin, and M. S. P. Sansom, “DNA and lipid bilayers: Self-assembly and insertion,” *J. R. Soc. Interface*, vol. 5, no. SUPPL. 3, 2008.
- [94] D. H. de Jong *et al.*, “Improved Parameters for the Martini Coarse-Grained Protein Force Field,” *J. Chem. Theory Comput.*, vol. 9, no. 1, pp. 687–697, Jan. 2013.
- [95] J. A. Graham, J. W. Essex, and S. Khalid, “PyCGTOOL: Automated Generation of Coarse-Grained Molecular Dynamics Models from Atomistic Trajectories,” *J. Chem. Inf. Model.*, vol. 57, pp. 650–656, 2017.
- [96] T. A. Wassenaar, H. I. Ingólfsson, R. A. Böckmann, D. P. Tieleman, and S. J. Marrink, “Computational lipidomics with insane: A versatile tool for generating custom membranes for molecular simulations,” *J. Chem. Theory Comput.*, vol. 11, no. 5, pp. 2144–2155, 2015.
- [97] F. Jensen, *Introduction to Computational Chemistry*, 2nd ed. Chichester: John Wiley and Sons, 2007.
- [98] M. J. Abraham *et al.*, “Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers,” *SoftwareX*, vol. 1–2, pp. 19–25, 2015.
- [99] D. C. Rapaport, *The Art of Molecular Dynamics Simulation*, 2nd ed. Cambridge: Cambridge University Press, 2004.
- [100] J. C. Phillips *et al.*, “Scalable molecular dynamics with NAMD,” *J. Comput. Chem.*, vol. 26, no. 16, pp. 1781–1802, 2005.
- [101] J. M. Seddon and J. D. Gale, *Thermodynamics and Statistical Mechanics*, 1st ed. New York: Wiley-Interscience, 2002.
- [102] D. A. McQuarrie, *Statistical Mechanics*. Sausalito: University Science Books, 2000.
- [103] G. Lamoureux, J. D. Faraldo-Gómez, S. Krupin, and S. Y. Noskov, “Polarizable model of chloroform based on classical Drude oscillators,” *Chem. Phys. Lett.*, vol. 468, no. 4–6, pp. 270–274, 2009.
- [104] A. P. Bhati, S. Wan, D. W. Wright, and P. V. Coveney, “Rapid, accurate, precise, and reliable relative free energy prediction using ensemble based thermodynamic integration,” *J. Chem. Theory Comput.*, vol. 13, no. 1, pp. 210–222, 2017.
- [105] P. V. Coveney and S. Wan, “On the calculation of equilibrium thermodynamic properties from molecular dynamics,” *Phys. Chem. Chem. Phys.*, vol. 18, no. 44, pp. 30236–30240, 2016.
- [106] S. Wan *et al.*, “Rapid and Reliable Binding Affinity Prediction of Bromodomain

- Inhibitors: A Computational Study,” *J. Chem. Theory Comput.*, vol. 13, no. 2, pp. 784–795, 2017.
- [107] D. W. Wright, B. A. Hall, O. A. Kenway, S. Jha, and P. V Coveney, “Computing Clinically Relevant Binding Free Energies of HIV - 1 Protease Inhibitors,” 2014.
- [108] H. J. C. Berendsen, J. P. M. Postma, W. F. Van Gunsteren, A. Dinola, and J. R. Haak, “Molecular dynamics with coupling to an external bath,” *J. Chem. Phys.*, vol. 81, no. 8, pp. 3684–3690, 1984.
- [109] P. H. Hünenberger, “Thermostat algorithms for molecular dynamics simulations,” *Adv. Polym. Sci.*, vol. 173, pp. 105–147, 2005.
- [110] G. Bussi, D. Donadio, and M. Parrinello, “Canonical sampling through velocity rescaling,” *J. Chem. Phys.*, vol. 126, no. 1, 2007.
- [111] M. Fyta, “Atomistic methods,” *Computational Approaches in Physics*. Morgan & Claypool Publishers, pp. 3–14, 2016.
- [112] M. Parrinello and A. Rahman, “Polymorphic transitions in single crystals: A new molecular dynamics method,” *J. Appl. Phys.*, vol. 52, no. 12, pp. 7182–7190, 1981.
- [113] G. H. Lubke and I. Campbell, “Inference Based on the Best-Fitting Model can Contribute to the Replication Crisis: Assessing Model Selection Uncertainty Using a Bootstrap Approach Gitta,” *Struct. Equ. Model.*, vol. 23, no. 4, pp. 479–490, 2016.
- [114] G. H. Lubke, I. Campbell, D. McArtor, P. Miller, J. Lunningham, and S. M. van den Berg, “Assessing Model Selection Uncertainty Using a Bootstrap Approach: An update,” *Struct. Equ. Model.*, vol. 24, no. 2, pp. 230–245, 2017.
- [115] D. Wright, B. Hall, and O. Kenway, “Computing Clinically Relevant Binding Free Energies of HIV-1 Protease Inhibitors,” *J. Chem. Theory Comput.*, vol. 10, no. 3, pp. 1228–1241, 2014.
- [116] D. W. Wright, S. Wan, C. Meyer, G. Tresadern, P. V Coveney, and H. Van Vlijmen, “Application of ESMACS binding free energy protocols to diverse datasets : Bromodomain-containing protein 4,” *Sci. Rep.*, no. March, pp. 1–15, 2019.
- [117] D. W. Wright *et al.*, “Application of the ESMACS Binding Free Energy Protocol to a Multi-Binding Site Lactate Dehydrogenase A Ligand Dataset,” vol. 1900194, pp. 1–10, 2020.
- [118] S. Wan, G. Tresadern, L. Pérez-benito, H. Van Vlijmen, and P. V Coveney, “Accuracy and Precision of Alchemical Relative Free Energy Predictions with and without Replica-Exchange,” vol. 1900195, 2020.
- [119] “The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.” Schrödinger, LCC.
- [120] A. Mackerell, K. Vannomeslaeghe, V. Gazula, B. Chitre, R. Puthanthodiyil, and N. K. Polani, “CGenFF Interface.” [Online]. Available: <https://cgenff.umaryland.edu>. [Accessed: 20-Sep-2017].
- [121] W. Humphrey, A. Dalke, and K. Schulten, “VMD-Visual Molecular Dynamics,” *J. Mol. Graph.*, vol. 14, pp. 33–38, 1996.
- [122] N. Michaud-Agrawal, E. J. Degging, T. B. Woolf, and O. Beckstein, “Software News and Updates MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations NAVEEN,” *J. Comput. Chem.*, vol. 32, no. 10, pp. 174–182, 2012.
- [123] “ARCHER,” 2018. [Online]. Available: <http://www.archer.ac.uk>.
- [124] “Cartesius,” 2018. [Online]. Available: <https://userinfo.surfsara.nl/systems/cartesius>.
- [125] “ComBioMed.” [Online]. Available: <https://www.combiomed.eu>. [Accessed: 01-Dec-2018].
- [126] R. Owczarzy, B. G. Moreira, Y. You, M. A. Behlke, and J. A. Wälder, “Predicting stability of DNA duplexes in solutions containing magnesium and monovalent cations,” *Biochemistry*, vol. 47, no. 19, pp. 5336–5353, 2008.
- [127] A. E. Every and I. M. Russu, “Influence of Magnesium Ions on Spontaneous Opening of DNA Base Pairs,” *J. Phys. Chem. B*, vol. 112, no. 25, pp. 7689–7695, 2008.
- [128] K. Serec, S. Dolanski Babic, R. Podgornik, and S. Tomic, “Effect of magnesium ions

- on the structure of DNA thin films : an infrared spectroscopy study,” *Nucleic Acids Res.*, vol. 44, no. 17, pp. 8456–8464, 2016.
- [129] P. J. Hagerman, “Flexibility of DNA,” *Annu. Rev. Biophys. Chem.*, vol. 17, pp. 265–286, 1988.
- [130] M. Behe and G. Felsenfeld, “Effects of methylation on a synthetic polynucleotide : The B-Z transition in poly (dG-m5dC) poly (dG-m5dC),” *Proc. Natl. Acad. Sci.*, vol. 78, no. 3, pp. 1619–1623, 1981.
- [131] S. Diekmann, J. C. Wang, and W. Germany, “On the Sequence Determinants and Flexibility Kinetoplast DNA Fragment with Abnormal Electrophoretic Mobilities of the Gel,” 1985.
- [132] C. L. Lawson, H. M. Berman, and W. Chiu, “Evolving data standards for cryo-EM structures,” *Struct. Dyn.*, vol. 7, no. 1, 2020.
- [133] T. A. Wassenaar, K. Pluhackova, R. A. Bo, S. J. Marrink, and D. P. Tieleman, “Going Backward : A Flexible Geometric Approach to Reverse Transformation from Coarse Grained to Atomistic Models,” 2013.
- [134] J. Frank *et al.*, “SPIDER and WEB: Processing and Visualization of Images in 3D Electron Microscopy and Related Fields,” *J. Struct. Biol.*, vol. 116, no. 1, pp. 190–199, 1996.
- [135] W. Li and J. Frank, “Transfer RNA in the hybrid P/E state: Correlating molecular dynamics simulations with cryo-EM data,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 104, no. 42, pp. 16540–16545, 2007.
- [136] L. G. Trabuco, E. Villa, K. Mitra, and J. Frank, “Flexible Fitting of Atomic Structures into Electron Microscopy Maps Using Molecular Dynamics,” *Structure*, vol. 16, no. 5, pp. 1–20, 2008.
- [137] M. Javanainen, H. Martinez-Seara, and I. Vattulainen, “Umbrella Sampling Simulations of TM Domain Dimerization,” pp. 1–20, 2017.
- [138] J. B. Mills, E. Vacano, and P. J. Hagerman, “Flexibility of single-stranded DNA: Use of gapped duplex helices to determine the persistence lengths of Poly(dT) and Poly(dA),” *J. Mol. Biol.*, vol. 285, no. 1, pp. 245–257, 1999.
- [139] M. A. Kastenholz, T. U. Schwartz, and P. H. Hünenberger, “The transition between the B and Z conformations of DNA investigated by targeted molecular dynamics simulations with explicit solvation,” *Biophys. J.*, vol. 91, no. 8, pp. 2976–2990, 2006.
- [140] F. Kilchherr, C. Wachauf, B. Pelz, M. Rief, M. Zacharias, and H. Dietz, “Single-molecule dissection of stacking forces in DNA,” *Science (80-.)*, vol. 353, no. 6304, 2016.
- [141] Y. Yin and X. S. Zhao, “Kinetics and dynamics of DNA hybridization,” *Acc. Chem. Res.*, vol. 44, no. 11, pp. 1172–1181, 2011.
- [142] R. D. Hills and N. McGlinchey, “Model parameters for simulation of physiological lipids,” *J. Comput. Chem.*, vol. 37, no. 12, pp. 1112–1118, 2016.
- [143] F. Castro-Román, R. W. Benz, S. H. White, and D. J. Tobias, “Investigation of finite system-size effects in molecular dynamics simulations of lipid bilayers,” *J. Phys. Chem. B*, vol. 110, no. 47, pp. 24157–24164, 2006.
- [144] C. Monzel and K. Sengupta, “Measuring shape fluctuations in biological membranes,” *J. Phys. D. Appl. Phys.*, vol. 49, no. 24, 2016.
- [145] D. P. Tieleman and H. J. C. Berendsen, “Molecular dynamics simulations of a fully hydrated dipalmitoylphosphatidylcholine bilayer with different macroscopic boundary conditions and parameters,” *J. Chem. Phys.*, vol. 105, no. 11, pp. 4871–4880, 1996.
- [146] P. Larsson, L. C. Alskär, and C. A. S. Bergström, “Molecular Structuring and Phase Transition of Lipid-Based Formulations upon Water Dispersion: A Coarse-Grained Molecular Dynamics Simulation Approach,” *Mol. Pharm.*, vol. 14, no. 12, pp. 4145–4153, 2017.
- [147] L. Frey, S. Hiller, R. Riek, and S. Bibow, “Lipid- and Cholesterol-Mediated Time-Scale-Specific Modulation of the Outer Membrane Protein X Dynamics in Lipid Bilayers,” *J. Am. Chem. Soc.*, vol. 140, no. 45, pp. 15402–15411, 2018.

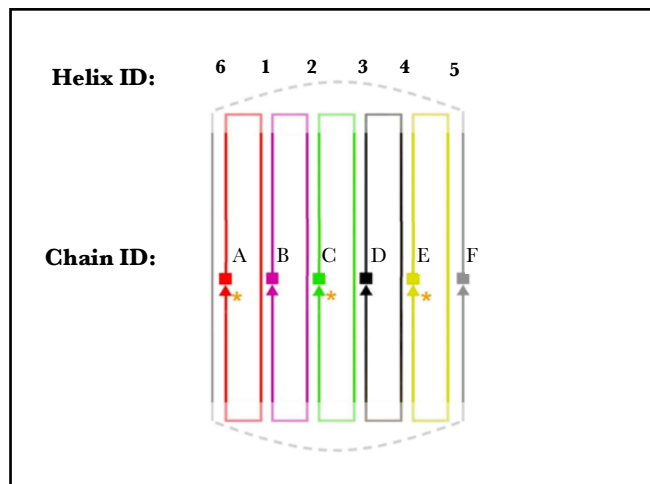
- [148] N. Kučerka, M. P. Nieh, and J. Katsaras, “Fluid phase lipid areas and bilayer thicknesses of commonly used phosphatidylcholines as a function of temperature,” *Biochim. Biophys. Acta - Biomembr.*, vol. 1808, no. 11, pp. 2761–2771, 2011.
- [149] M. Saeedimazine, A. Montanino, S. Kleiven, and A. Villa, “Role of lipid composition on the structural and mechanical features of axonal membranes: a molecular simulation study,” *Sci. Rep.*, vol. 9, no. 1, pp. 1–12, 2019.
- [150] H. Saito, T. Morishita, T. Mizukami, K. I. Nishiyama, K. Kawaguchi, and H. Nagao, “Molecular dynamics study of binary POPC bilayers: Molecular condensing effects on membrane structure and dynamics,” *J. Phys. Conf. Ser.*, vol. 1136, no. 1, 2018.
- [151] J. D. Perlmutter and J. N. Sachs, “Experimental verification of lipid bilayer structure through multi-scale modeling,” *Biochim. Biophys. Acta - Biomembr.*, vol. 1788, no. 10, pp. 2284–2290, 2009.
- [152] G. Pabst, A. Hodzic, J. Štrancar, S. Danner, M. Rappolt, and P. Laggner, “Rigidification of neutral lipid bilayers in the presence of salts,” *Biophys. J.*, vol. 93, no. 8, pp. 2688–2696, 2007.
- [153] A. Hac, T. Heimburg, H. Grubmüller, and R. A. Bockmann, “Effect of Sodium Chloride on a Lipid Bilayer,” *Biophys. J.*, vol. 85, no. September, pp. 1647–1655, 2003.
- [154] L. S. Vermeer, B. L. De Groot, V. Réat, A. Milon, and J. Czaplicki, “Acyl chain order parameter profiles in phospholipid bilayers: Computation from molecular dynamics simulations and comparison with 2H NMR experiments,” *Eur. Biophys. J.*, vol. 36, no. 8, pp. 919–931, 2007.
- [155] S. L. Grage, S. Afonin, S. Kara, G. Buth, and A. S. Ulrich, “Membrane thinning and thickening induced by membrane-active amphipathic peptides,” *Front. Cell Dev. Biol.*, vol. 4, no. JUN, 2016.
- [156] J. W. Klingelhoefer, T. Carpenter, and M. S. P. Sansom, “Peptide nanopores and lipid bilayers: Interactions by coarse-grained molecular-dynamics simulations,” *Biophys. J.*, vol. 96, no. 9, pp. 3519–3528, 2009.
- [157] V. M. Aguilera, C. Verdiá-Báguena, and A. Alcaraz, “Lipid charge regulation of non-specific biological ion channels,” *Phys. Chem. Chem. Phys.*, vol. 16, no. 9, pp. 3881–3893, 2014.
- [158] J. Seelig and W. Niederberger, “Deuterium-Labeled Lipids as Structural Probes in Liquid Crystalline Bilayers. A Deuterium Magnetic Resonance Study,” *J. Am. Chem. Soc.*, vol. 96, no. 7, pp. 2069–2072, 1974.
- [159] M. Bernetti, M. Masetti, W. Rocchia, and A. Cavalli, “Kinetics of Drug Binding and Residence Time,” *Annu. Rev. Phys. Chem.*, vol. 70, no. 1, pp. 143–171, 2019.
- [160] R. Casasnovas, V. Limongelli, P. Tiwary, P. Carloni, and M. Parrinello, “Unbinding Kinetics of a p38 MAP Kinase Type II Inhibitor from Metadynamics Simulations,” *J. Am. Chem. Soc.*, vol. 139, no. 13, pp. 4780–4788, 2017.
- [161] M. I. Angelova *et al.*, “pH sensing by lipids in membranes: The fundamentals of pH-driven migration, polarization and deformations of lipid bilayer assemblies,” *Biochim. Biophys. Acta - Biomembr.*, vol. 1860, no. 10, pp. 2042–2063, 2018.
- [162] C. Kutzner, H. Grubmüller, B. L. De Groot, and U. Zachariae, “Computational Electrophysiology: The Molecular Dynamics of Ion Channel Permeation and Selectivity in Atomistic Detail,” *Biophys. J.*, vol. 101, pp. 809–817, 2011.
- [163] C. Kutzner, D. A. Köpfer, J. Machtens, B. L. De Groot, C. Song, and U. Zachariae, “Insights into the function of ion channels by computational electrophysiology simulations ☆,” *Biochim. Biophys. Acta*, vol. 1858, pp. 1741–1752, 2016.
- [164] J. Gumbart, F. Khalili-Araghi, M. Sotomayor, and B. Roux, “Constant electric field simulations of the membrane potential illustrated with simple systems,” *Biochim. Biophys. Acta - Biomembr.*, vol. 1818, no. 2, pp. 294–302, 2012.
- [165] J. Wang, L. Zhang, J. Xue, and G. Hu, “Ion diffusion coefficient measurements in nanochannels at various concentrations,” *Biomicrofluidics*, vol. 8, no. 2, p. 1, 2014.
- [166] J. Wilson and A. Aksimentiev, “Water-compression gating of nanopore transport,”

- Phys. Rev. Lett.*, vol. 120, no. 26, pp. 139–148, 2018.
- [167] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed. San Diego: Academic Press, 2001.
- [168] J. A. Lemkul and D. R. Bevan, “Assessing the stability of Alzheimer’s amyloid protofibrils using molecular dynamics,” *J. Phys. Chem. B*, vol. 114, no. 4, pp. 1652–1660, 2010.
- [169] S. Park, F. Khalili-araghi, E. Tajkhorshid, K. Schulten, S. Park, and F. Khalili-araghi, “Free energy calculation from steered molecular dynamics simulations using Jarzynski’s equality,” vol. 3559, no. 2003, pp. 1–10, 2014.
- [170] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, “THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method,” *J. Comput. Chem.*, vol. 13, no. 8, pp. 1011–1021, 1992.
- [171] G. Fiorin, M. L. Klein, and J. Hénin, “Using collective variables to drive molecular dynamics simulations,” *Mol. Phys.*, vol. 111, no. 22–23, pp. 3345–3362, Dec. 2013.
- [172] C. Jarzynski, “Nonequilibrium equality for free energy differences,” *Phys. Rev. Lett.*, vol. 78, no. 14, pp. 2690–2693, 1997.
- [173] T. Batuğ, P. C. Chen, S. M. Patra, and S. Kuyucak, “Potential of mean force calculations of ligand binding to ion channels from Jarzynski’s equality and umbrella sampling,” *J. Chem. Phys.*, vol. 128, no. 15, 2008.
- [174] D. P. Tran, K. Takemura, K. Kuwata, and A. Kitao, “Protein-Ligand Dissociation Simulated by Parallel Cascade Selection Molecular Dynamics,” *J. Chem. Theory Comput.*, vol. 14, no. 1, pp. 404–417, 2018.
- [175] H. S. C. Martin, S. Jha, S. Howorka, and P. V. Coveney, “Determination of free energy profiles for the translocation of polynucleotides through α -hemolysin nanopores using non-equilibrium molecular dynamics simulations,” *J. Chem. Theory Comput.*, vol. 5, no. 8, pp. 2135–2148, 2009.
- [176] J. Clayden, N. Greeves, S. Warren, and P. Wothers, *Organic Chemistry*, 2nd ed. Oxford: Oxford University Press, 2012.
- [177] A. D. Mackerell, “Empirical Force Fields for Biological Macromolecules :,” 2004.
- [178] A. C. Vaiana, A. Schulz, J. Wolfrum, M. Sauer, and J. C. Smith, “Molecular mechanics force field parameterization of the fluorescent probe rhodamine 6G using automated frequency matching,” *J. Comput. Chem.*, vol. 24, no. 5, pp. 632–639, 2003.
- [179] A. C. Vaiana, H. Neuweiler, A. Schulz, J. Wolfrum, M. Sauer, and J. C. Smith, “Fluorescence Quenching of Dyes by Tryptophan: Interactions at Atomic Detail from Combination of Experiment and Computer Simulation,” *J. Am. Chem. Soc.*, vol. 125, no. 47, pp. 14564–14572, 2003.
- [180] B. Corry and D. Jayatilaka, “Simulation of structure, orientation, and energy transfer between AlexaFluor molecules attached to MscL,” *Biophys. J.*, vol. 95, no. 6, pp. 2711–2721, 2008.
- [181] T. Graen, M. Hoefling, and H. Grubmüller, “AMBER-DYES: Characterization of charge fluctuations and force field parameterization of fluorescent dyes for molecular dynamics simulations,” *J. Chem. Theory Comput.*, vol. 10, no. 12, pp. 5505–5512, 2014.
- [182] C. I. Bayly, P. Cieplak, W. D. Cornell, and P. A. Kollman, “A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model,” *J. Phys. Chem.*, vol. 97, no. 40, pp. 10269–10280, 1993.
- [183] M. J. Frisch *et al.*, “Gaussian 16.” Gaussian, Inc., Wallingford, CT, 2016.

Appendices

Appendix 1: Specific nucleotide sequences of each chain making up the DNP, and CaDNAno schematic illustrating how the oligonucleotides chains arrange themselves to form the six-helix bundle.

Chain ID	Sequence 5' → 3'
A	AGCGAACGTGGATTTTGTCCGACATCGGCAAGCTCCCTTTTTCGACTATT*
B	CCGATGTCGGACTTTTACACGATCTTCGCCTGCTGGGTTTTGGGAGCTTG
C	CGAAGATCGTGTTTTTCCACAGTTGATTGCCCTTCACTTTTCCCAGCAGG*
D	AATCAACTGTGGTTTTTCTCACTGGTGATTAGAATGCTTTTGTGAAGGGC
E	TCACCAGTGAGATTTTGTTCGTACCAGGTGCATGGATTTTGCATTCTAA*
F	CCTGGTACGACATTTTTCCACGTTCGCTAATAGTCGATTTTATCCATGCA



Appendix 2: CHARMM36 residue topology file entry and parameter file for the TEG-cholesterol 3' DNA backbone modification, generated by the CGenFF web-server.

```
PRES TCL    -1.000

ATOM C3' CN7      0.01
ATOM H3' HN7      0.09
!ATOM O3' OG303  -0.577
ATOM P01 PG1      1.517
ATOM O01 OG2P1   -0.791
ATOM O9G OG2P1   -0.791
ATOM O3' ON2      -0.57
ATOM O7G OG303   -0.565
ATOM C39G CG321  -0.080
ATOM H68G HGA2    0.090
ATOM H67G HGA2    0.090
ATOM C38G CG321  -0.011
ATOM H65G HGA2    0.090
ATOM H66G HGA2    0.090
ATOM O6G OG301   -0.338
ATOM C37G CG321  -0.011
ATOM H63G HGA2    0.090
ATOM H64G HGA2    0.090
ATOM C36G CG321  -0.011
ATOM H62G HGA2    0.090
ATOM H61G HGA2    0.090
ATOM O5G OG301   -0.338
ATOM C35G CG321  -0.011
ATOM H60G HGA2    0.090
ATOM H59G HGA2    0.090
ATOM C34G CG321  -0.011
ATOM H57G HGA2    0.090
ATOM H58G HGA2    0.090
ATOM O4G OG301   -0.338
ATOM C33G CG321  -0.011
ATOM H55G HGA2    0.090
ATOM H56G HGA2    0.090
ATOM C32G CG321  -0.011
ATOM H54G HGA2    0.090
ATOM H53G HGA2    0.090
ATOM O3G OG301   -0.338
ATOM C31G CG321  -0.026
ATOM H52G HGA2    0.090
ATOM H51G HGA2    0.090
ATOM C30G CG321  -0.173
ATOM H50G HGA2    0.090
ATOM H49G HGA2    0.090
ATOM C29G CG321  0.083
```

ATOM H36G HGA2	0.090
ATOM H48G HGA2	0.090
ATOM N0G NG2S1	-0.415
ATOM H35G HGP1	0.329
ATOM C28G CG2O6	0.207
ATOM O2G OG2D1	-0.390
ATOM O0G OG302	-0.293
ATOM C24G CG311	0.224
ATOM H47G HGA1	0.090
ATOM C25G CG321	-0.193
ATOM C26G CG321	-0.174
ATOM C27G CG331	-0.272
ATOM C23G CG321	-0.185
ATOM C22G CG301	-0.002
ATOM C21G CG2D1	-0.001
ATOM C20G CG2D1	-0.150
ATOM C19G CG321	-0.180
ATOM C18G CG331	-0.271
ATOM C17G CG321	-0.182
ATOM C16G CG321	-0.180
ATOM C15G CG311	-0.089
ATOM C14G CG311	-0.088
ATOM C13G CG3RC1	-0.001
ATOM C12G CG3RC1	-0.098
ATOM C11G CG3C52	-0.182
ATOM C10G CG3C52	-0.179
ATOM C9G CG3C51	-0.091
ATOM C8G CG331	-0.272
ATOM C7G CG311	-0.094
ATOM C6G CG321	-0.172
ATOM C5G CG321	-0.175
ATOM C4G CG331	-0.271
ATOM C3G CG331	-0.271
ATOM C2G CG321	-0.171
ATOM C0G CG311	-0.083
ATOM HG HGA3	0.090
ATOM H2G HGA3	0.090
ATOM H3G HGA3	0.090
ATOM H4G HGA3	0.090
ATOM H5G HGA3	0.090
ATOM H6G HGA3	0.090
ATOM H7G HGA1	0.090
ATOM H8G HGA2	0.090
ATOM H9G HGA2	0.090
ATOM H10G HGA2	0.090
ATOM H11G HGA2	0.090
ATOM H12G HGA3	0.090
ATOM H13G HGA3	0.090
ATOM H14G HGA3	0.090
ATOM H15G HGA1	0.090
ATOM H16G HGA2	0.090

ATOM H17G HGA2	0.090
ATOM H18G HGA2	0.090
ATOM H19G HGA2	0.090
ATOM H20G HGA1	0.090
ATOM H21G HGA2	0.090
ATOM H22G HGA2	0.090
ATOM H23G HGA3	0.090
ATOM H24G HGA3	0.090
ATOM H25G HGA3	0.090
ATOM H26G HGA1	0.090
ATOM H27G HGA2	0.090
ATOM H28G HGA2	0.090
ATOM H29G HGA2	0.090
ATOM H30G HGA2	0.090
ATOM H31G HGA1	0.090
ATOM H32G HGA2	0.090
ATOM H33G HGA2	0.090
ATOM H34G HGA4	0.150
ATOM H37G HGA3	0.090
ATOM H38G HGA3	0.090
ATOM H39G HGA3	0.090
ATOM H40G HGA1	0.090
ATOM H41G HGA2	0.090
ATOM H42G HGA2	0.090
ATOM H43G HGA2	0.090
ATOM H44G HGA2	0.090
ATOM H45G HGA2	0.090
ATOM H46G HGA2	0.090

BOND O3' P01

BOND P01 O01 P01 O9G O7G P01 O7G C39G

BOND C39G H67G C39G H68G C38G C39G H65G C38G H66G C38G

BOND C38G O6G C37G O6G H64G C37G H63G C37G C37G C36G

BOND H56G C33G O5G C36G O5G C35G H57G C34G C34G C35G

BOND C34G H58G

BOND C34G O4G

BOND H55G C33G

BOND C36G H61G

BOND C36G H62G

BOND H50G C30G

BOND C33G O4G

BOND C33G C32G

BOND C35G H59G

BOND C35G H60G

BOND H49G C30G

BOND H45G C25G

BOND O3G C32G

BOND O3G C31G

BOND C30G C31G

BOND C30G C29G

BOND C32G H53G

BOND C32G H54G
BOND H48G C29G
BOND C25G H46G
BOND C25G C26G
BOND C25G C24G
BOND C31G H51G
BOND C31G H52G
BOND H43G C26G
BOND H44G C26G
BOND N0G C29G
BOND N0G C28G
BOND N0G H35G
BOND C29G H36G
BOND C28G O2G
BOND C28G O0G
BOND O0G C24G
BOND H47G C24G
BOND C26G C22G
BOND C24G C23G
BOND H30G C16G
BOND H40G C15G
BOND H38G C27G
BOND C22G C15G
BOND C22G C27G
BOND C22G C21G
BOND H29G C16G
BOND C23G H42G
BOND C23G C21G
BOND C23G H41G
BOND C16G C15G
BOND C16G C17G
BOND C15G C14G
BOND C27G H39G
BOND C27G H37G
BOND C21G C20G
BOND H32G C19G
BOND C20G C19G
BOND C20G H34G
BOND H28G C17G
BOND C17G H27G
BOND C17G C13G
BOND C14G H31G
BOND C14G C19G
BOND C14G C12G
BOND C19G H33G
BOND H3G C4G
BOND HG C4G
BOND H23G C18G
BOND C4G H2G
BOND C4G C0G
BOND H9G C2G

BOND H26G C12G
 BOND H24G C18G
 BOND C13G C12G
 BOND C13G C18G
 BOND C13G C9G
 BOND C12G C11G
 BOND C18G H25G
 BOND H5G C3G
 BOND H8G C2G
 BOND H17G C6G
 BOND C2G C0G
 BOND C2G C5G
 BOND C0G C3G
 BOND C0G H7G
 BOND H6G C3G
 BOND C3G H4G
 BOND H16G C6G
 BOND C6G C5G
 BOND C6G C7G
 BOND H20G C9G
 BOND C9G C7G
 BOND C9G C10G
 BOND C11G H21G
 BOND C11G H22G
 BOND C11G C10G
 BOND C5G H10G
 BOND C5G H11G
 BOND C7G H15G
 BOND C7G C8G
 BOND C10G H19G
 BOND C10G H18G
 BOND C8G H14G
 BOND C8G H13G
 BOND C8G H12G
 IMPR C28G N0G O2G O0G
 end

* Parameters generated by analogy by
 * CHARMM General Force Field (CGenFF) program version 1.0.0

! Penalties lower than 10 indicate the analogy is fair; penalties between 10
 ! and 50 mean some basic validation is recommended; penalties higher than
 ! 50 indicate poor analogy and mandate extensive validation/optimization.

BONDS

CN7 OG303 310.0 1.433 ! C3' to O3' (cholesterol mod)

ANGLES

CG2O6 OG302 CG311 40.00 111.00 ! /Users , from CG2O6 OG302 CG321,
 penalty= 0.6
 CN7 ON2 PG1 20.00 120.00 35.00 2.33000 ! Link

OG2P1	PG1	ON2	98.90	107.50 ! Link
ON2	PG1	OG303	80.00	104.30 ! Link
CN7	CN7	OG303	115.00	109.70 ! Link
CN8	CN7	OG303	115.00	109.70 ! Link
HN7	CN7	OG303	60.0	109.50

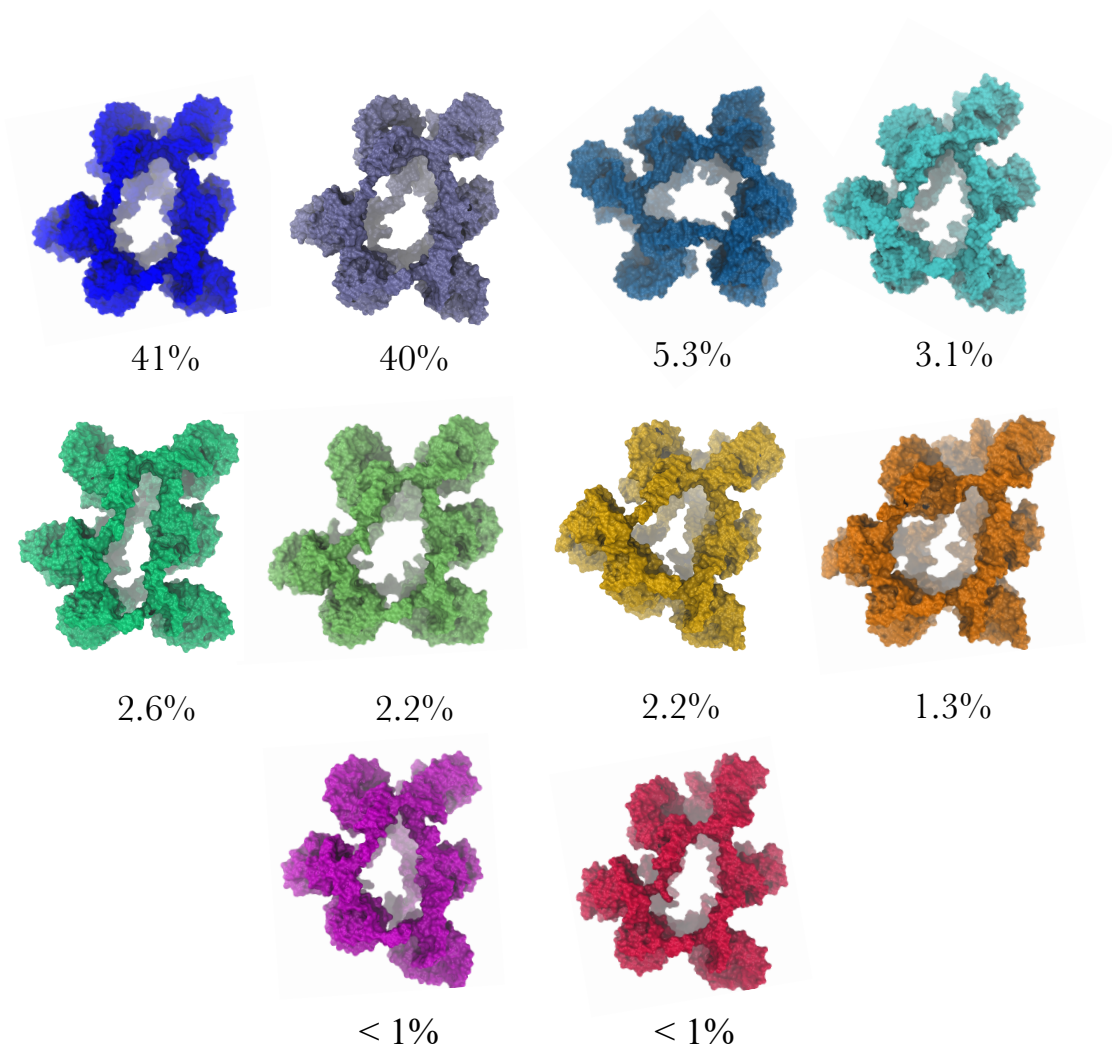
DIHEDRALS

NG2S1	CG2O6	OG302	CG311	0.1500	1	180.00 ! /Users , from NG2S1 CG2O6 OG302 CG321, penalty= 0.6
NG2S1	CG2O6	OG302	CG311	2.2000	2	180.00 ! /Users , from NG2S1 CG2O6 OG302 CG321, penalty= 0.6
NG2S1	CG2O6	OG302	CG311	0.1000	3	180.00 ! /Users , from NG2S1 CG2O6 OG302 CG321, penalty= 0.6
OG2D1	CG2O6	OG302	CG311	0.1500	1	0.00 ! /Users , from OG2D1 CG2O6 OG302 CG321, penalty= 0.6
OG2D1	CG2O6	OG302	CG311	2.2000	2	180.00 ! /Users , from OG2D1 CG2O6 OG302 CG321, penalty= 0.6
OG2D1	CG2O6	OG302	CG311	0.1000	3	0.00 ! /Users , from OG2D1 CG2O6 OG302 CG321, penalty= 0.6
OG302	CG311	CG321	CG2D1	0.2000	3	0.00 ! /Users , from OG311 CG311 CG321 CG2D1, penalty= 1.5
OG302	CG311	CG321	CG321	0.1950	3	0.00 ! /Users , from CG321 CG321 CG321 OG302, penalty= 4
CG321	CG311	OG302	CG2O6	0.7000	1	180.00 ! /Users , from CG321 CG311 OG302 CG2O2, penalty= 4
HGA1	CG311	OG302	CG2O6	0.0000	3	0.00 ! /Users , from HGA1 CG311 OG302 CG2O2, penalty= 4
OG301	CG321	CG321	OG303	0.1950	3	0.00 ! /Users , from OG303 CG321 CG321 OG303, penalty= 1.5
CG321	CG321	NG2S1	CG2O6	0.3500	1	180.00 ! /Users , from CG331 CG321 NG2S1 CG2O6, penalty= 0.9
CG321	CG321	NG2S1	CG2O6	0.7500	2	0.00 ! /Users , from CG331 CG321 NG2S1 CG2O6, penalty= 0.9
CG321	CG321	NG2S1	CG2O6	0.1500	4	0.00 ! /Users , from CG331 CG321 NG2S1 CG2O6, penalty= 0.9
HN7	CN7	ON2	PG1	0.0000	3	0.00 ! Link
CN7	CN7	ON2	PG1	2.5000	1	180.00 ! Link
CN8	CN7	ON2	PG1	2.5000	1	180.00 ! Link
ON2	PG1	OG303	CG321	1.2000	1	180.00 ! Link
ON2	PG1	OG303	CG321	0.1000	2	180.00 ! Link
ON2	PG1	OG303	CG321	0.1000	3	180.00 ! Link
CN7	ON2	PG1	OG303	1.2000	1	180.00 ! Link
CN7	ON2	PG1	OG303	0.1000	2	180.00 ! Link
CN7	ON2	PG1	OG303	0.1000	3	180.00 ! Link
CN7	ON2	PG1	OG2P1	0.1000	3	0.00 ! Link
OG303	CN7	CN7	CN8B	0.2000	4	0.00 ! Link
HN7	CN7	CN7	OG303	0.1950	3	0.00 ! Link
OG303	CN7	CN7	ON6	0.2000	3	0.00 ! Link

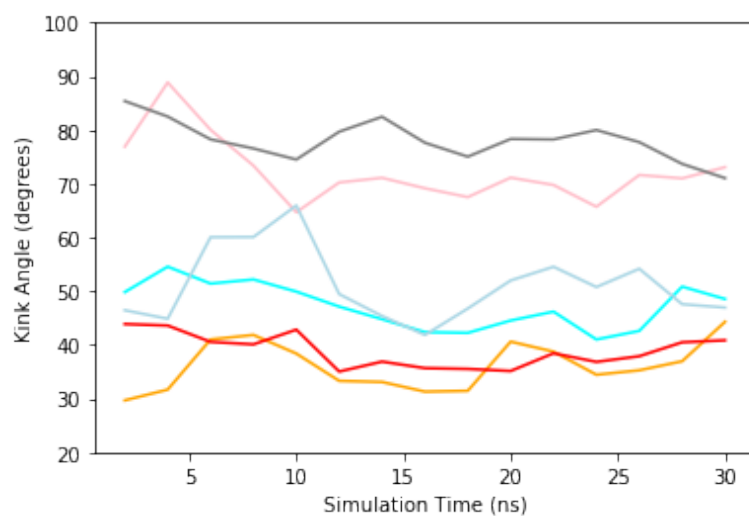
IMPROPERS

END

Appendix 3: Ten of the twelve most highly populated middle structures yielded after performing cluster analysis on the ensemble of CG trajectories using the GROMOS algorithm, using an RMSD cut-off of 0.45 Å. The relative stability of each clustered conformation is reflected by the percentage population of each cluster.



Appendix 4: Fluctuations in the average helix kink angle as a function of simulation time for the shorter duration AA simulations of the solvated DNP in 0.3 NaCl.



Appendix 5: PyCGTool bonding scheme for the MARTINI CG model of SRB.

```
[ SRB ]
SU1 AR1
AR1 AR2
AR2 AR3
AR1 AR3
AR2 SU2
AR2 AR4
AR4 AR9
AR4 AR5
AR4 AO1
AO1 AR7
AR7 AR6
AR5 AR6
AR6 NN1
AR6 NN2
NN1 NN2
AO1 AR8
AR8 AR9
AR9 AR10
AR8 AR10
AR10 PN1
AR10 PN2
PN1 PN2
```

Appendix 6: PyCGTool bonding scheme for the MARTINI CG model of CF.

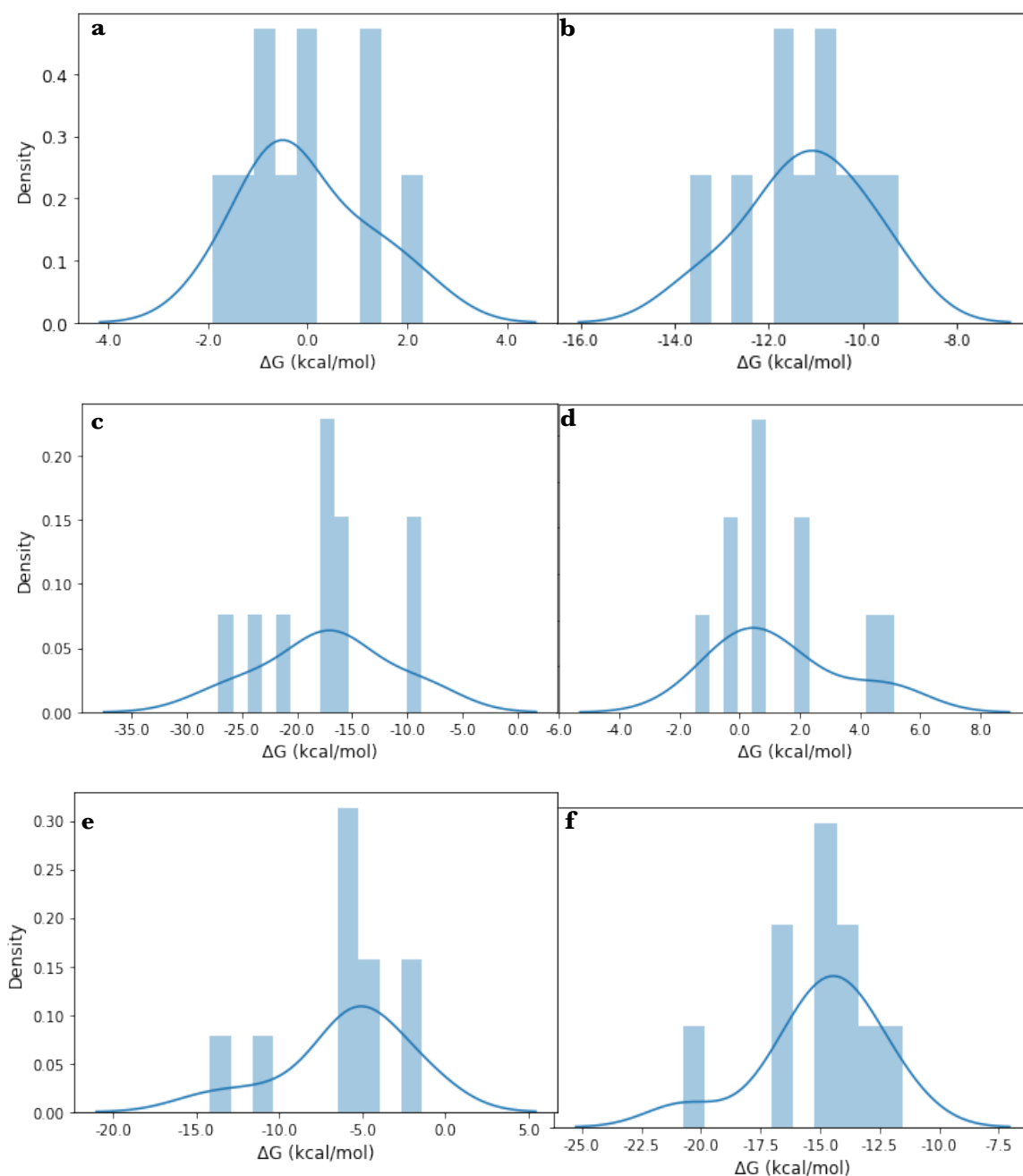
[CF]

CA1 AR2
AR2 AR1
AR2 AR0
AR0 AR1
AR2 AR1
AR1 AR3
AR1 CA2
AR3 AR5
AR3 AO1
AR3 AR8
AR5 AR4
AR5 AR6
AR4 AR6
AR6 CO1
AR4 AO1
AO1 AR7
AR8 AR7
AR8 AR9
AR9 AR7
AR9 CO2

Appendix 7: PyCGTool bonding scheme for the MARTINI CG model of CF.

[ATTO]
CA1 CH1
CH1 AM1
AM1 R1
R1 R2
R2 SU1
R2 AR1
AM1 AR1
AM1 AR2
AO1 AR2
AR1 AN1
AN1 AR2
AO1 AR3
AN1 AR4
AR4 AR3
AR3 AR5
AR5 AR4
AR5 C2
AR5 AM2
AM2 C2

Appendix 8: Histograms and overlaid estimated probability density functions representing the distribution of the free energies of translocation (ΔG_{trans}) for the six translocation pathways studied with cv-SMD and US simulations. Probability density functions were calculated using the kernel density estimator, and generally resemble normal distributions (albeit with some degree of skewness, which can be corrected by running additional US ensembles). **(a)** Distribution of ΔG_{trans} for the trans-lumen translocation of SRB. **(b)** Distribution of ΔG_{trans} for the interfacial translocation of SRB. **(c)** Distribution of ΔG_{trans} for the trans-lumen translocation of CF. **(d)** Distribution of ΔG_{trans} for the interfacial translocation of CF. **(e)** Distribution of ΔG_{trans} for the trans-lumen translocation of ATTO 655. **(f)** Distribution of ΔG_{trans} for the interfacial of ATTO 655.



Appendix 9: Distributions of the force histograms along the fluorophore translocation coordinate yielded from WHAM analyses of the umbrella sampling simulation data described in Chapter 6.

