Imperial College London

Imperial College London

Department of Life Sciences

STRUCTURAL AND MECHANISTIC INSIGHTS INTO PORE FORMATION BY PROTEINS OF THE MEMBRANE ATTACK COMPLEX/PERFORIN SUPERFAMILY

Tomas B. Voisin

February 2023

A thesis submitted to Imperial College London for the degree of Doctor of Philosophy

Copyright declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC). Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Declaration of originality

I declare that the work presented in this thesis is my own, conducted under the supervision of Dr. Doryen Bubeck at Imperial College London. Where work performed and/or published by others is discussed, it is appropriately stated and referenced. Where research was done in collaboration with others, all contributions are clearly acknowledged.

Acknowledgements

I would first like to thank my supervisor, Dr. Doryen Bubeck. Your guidance, mentorship and support have been a driving force in helping me complete this project. Thank you for your invaluable advice; you have given me the confidence I needed to tackle new scientific challenges!

I am grateful to Prof. Richard Smith for the generous gift of $CD59^{cyto}$. I would like to thank Dr. Edward Parsons and Prof. Bart Hoogenboom for recording and providing atomic force microscopy data, and Philip Haynes for his help with data analysis. I would like to express my gratitude to Prof. Phill Stansfeld for helping me get started with molecular dynamics simulations. I am also grateful to Dr. Morgan Beeby and Dr. Ernesto Cota for their advice throughout my PhD. Thanks to Paul Simpson for training me on the microscopes, helping me whenever I could not align them, and not getting angry when I lost a clip ring in the column. Thank you also to Dr. Suhail Islam for his outstanding work keeping the computers running and setting up everything I needed for my molecular dynamics projects.

I cannot thank the Bubeck lab members —past and present— enough for their support during this PhD. I would like to thank Nita, Anaïs and Jasmine for being among the first to welcome me into the lab. Shannon, Tarhan, Tanvi, Scott, Yibo, Vicky, Harry and Xueqing, you have all helped me, supported me and made me laugh so much. Thanks for putting up with the mess I made every time I went into the lab! And last but, by no means least, I would like to thank Emma Couves, who has been more than a labmate, but a true friend for over seven years. More broadly, I would like to thank all members of the fifth floor of the Sir Ernst Chain building for keeping me entertained and (relatively) sane during these years. Tina, our

thesis writing club may not have been the most successful, but at least we got much better at crosswords! Eli, Nanki, Trishant—thank you for being there when I needed to drink beer and watch bad TV instead of doing work. Iñigo, Anna and Nayim—thank you for making my first year so much fun!

I reserve my biggest thanks to my friends and family. Ed, thank you for everything over the last seven years—from splitting the cost of laundry to being an amazing friend. I would like to thank Ivet and Eliza for being great flatmates and not kicking me out despite the chaos I sometimes brought. Fio, thank you for always making me laugh and bringing snacks to DnD. Laura, thank you for coming to visit me in London all those times. Bradley, thank you so much for your love and support over the last three years—I don't know how I would have done it without you!

Finally, I would like to thank my parents, Hugo and Stéphanie, as well as my sisters, Célia and Marie, who always believed in me—un grand merci !

Collaboration acknowledgements

- Dr. Emma Couves (Bubeck Lab, Imperial College London) removed chain breaks from the CD59-C5b8 structure for molecular dynamics simulations in Chapter 2.

- Dr. Nita Shah (Bubeck Lab, Imperial College London) generated the following constructs in Chapter 3: ILY^{WT}, ILY^{EP}, ILY^{HB1} and ILY^{HB2}.

- Dr. Edward Parsons (Hoogenboom lab, University College London) recorded the AFM data in Section 3.3.1.

- The laboratory of Prof. Richard Smith (King's College London) generated the $CD59^{cyto}$ construct used in Chapters 3 and 4.

- Dr. Scott Gardner and Dr. Yibo Jin ran the 200-kV and 300-kV data collection sessions in Chapter 4.

Abstract

Members of the membrane attack complex/perforin (MACPF) superfamily of pore-forming proteins are characterised by a common three-dimensional fold able to puncture lipid membranes. They are found in bacterial and eukaryotes, and include immune effectors, toxins and pathogenic virulence factors. Their conserved pore-forming domain follows the same mechanism whereby two bundles of α-helices unfurl into membrane-spanning β-hairpins. This thesis provides insights into the effects of MACPF proteins on biological membranes.

Coarse-grain molecular dynamics simulations of the membrane attack complex (MAC) bound to its inhibitor CD59 reveal protein-lipid interactions and local changes in membrane thickness. These may serve as signals to recruit CD59 or the molecular machinery for MAC clearance. Some bacterial MACPF proteins called cholesterol-dependent cytolysins (CDCs) hijack CD59 on human cells as part of their pore formation pathway. Atomistic simulations of CD59 in a lipid bilayer show that it samples various orientations relative to the membrane, dictating whether its binding site is available for engaging MAC or CDCs, and thus for inhibiting or promoting pore formation.

CDCs assemble on cholesterol-rich lipid membranes and undergo sequential conformational changes to puncture bilayers. Site-directed mutagenesis of two CDCs reveals that an amphipathic helix in the pore-forming helical bundles is responsible for tuning the lytic activity of these proteins. Understanding the molecular basis for the function of this helix will require the high-resolution structure of a CDC late prepore intermediate. The first steps towards solving this structure by cryo-electron microscopy are presented in this thesis.

Contents

Copyright declaration **2**

Declaration of originality **3**

Acknowledgements **4**

Collaboration acknowledgements **6**

Abstract **7**

1 Introduction **13**
1.1 Pore-forming proteins . . . 14
1.1.1 α-pore forming proteins . . . 14
1.1.2 β-pore forming proteins . . . 16
1.1.3 Structural biology of the MACPF superfamily . . . 17
1.2 Cryo-electron microscopy . . . 21
1.2.1 Interactions between electrons and matter . . . 21
1.2.2 Image formation in the TEM . . . 24
1.2.3 Recent advances in cryoEM . . . 28
1.3 Molecular Dynamics Simulations . . . 31
1.3.1 Forcefields describe the forces applied on each atom . . . 32
1.3.2 Integrators solve Newton's equations of motion . . . 36
1.3.3 Typical steps in molecular dynamics . . . 37
1.3.4 Coarse-grain MD and the Martini forcefield . . . 39

2 Molecular Dynamics simulations of the inhibited membrane attack complex **44**
2.1 Introduction . . . 45
2.1.1 The Complement cascade . . . 45
2.1.2 Inhibition of MAC by CD59 . . . 50
2.1.3 MD simulations as a tool to study C5b8-CD59 . . . 55
2.1.4 Aims . . . 57
2.2 Computational Approach . . . 58
2.2.1 Atomistic simulations of GPI-anchored CD59 . . . 58
2.2.2 Parametrisation of the CD59-GPI EtNP linker . . . 59
2.2.3 Coarse-grain simulations of C5b8-CD59 . . . 61
2.3 Results . . . 62
2.3.1 CD59 samples orientations allowing it to bind C8βand ILY . . . 62
2.3.2 Interactions between CD59 and the membrane . . . 65
2.3.3 Orientation sampling of GPI-anchored CD59 . . . 65
2.3.4 Parametrising the CD59-GPI EtNP linker . . . 67
2.3.5 Building C5b8-CD59 and placing it in the membrane . . . 69
2.3.6 Lipid binding . . . 71
2.3.7 Membrane thinning . . . 74
2.4 Discussion . . . 75

3 Functional characterisation of the horizontal helix of cholesterol-dependent cytolysins **83**
3.1 Introduction . . . 84
3.1.1 Pore formation by cholesterol-dependent cytolysins . . . 84
3.1.2 The h-helix and disulfide-locked mutants . . . 88
3.1.3 Aims . . . 91
3.2 Materials and Methods . . . 91
3.2.1 Atomic force microscopy data analysis . . . 91
3.2.2 PLY and ILY cloning and mutagenesis . . . 92
3.2.3 ILY and PLY variants expression and purification . . . 93
3.2.4 Lipid monolayer negative stain electron microscopy . . . 94
3.2.5 Liposome lysis assays . . . 94
3.2.6 Coarse-grain MD simulations of a hybrid PLY conformation . . . 95
3.3 Results . . . 97
3.3.1 The HB-locked ILY mutants form collapsed assemblies . . . 97
3.3.2 Recombinantly-expressed PLY is active . . . 99
3.3.3 The h-helix tunes the lytic activity of CDCs . . . 100
3.3.4 The PLY^{hybrid} model does not explain the function of the h-helix . . . 103
3.4 Discussion . . . 105

4 Towards the structure of the intermedilysin late prepore by cryoEM **110**
4.1 Introduction . . . 111
4.1.1 The intermedilysin early and late prepores . . . 111
4.1.2 Sample preparation for cryoEM . . . 114
4.1.3 Data collection and image processing . . . 118
4.1.4 Aims . . . 120
4.2 Materials and Methods . . . 120
4.2.1 Protein expression and purification . . . 120
4.2.2 Nanodisc preparation and ILY complex assembly . . . 121
4.2.3 Sucrose density centrifugation and sucrose removal . . . 122
4.2.4 Negative stain EM imaging and image processing . . . 123
4.2.5 CryoEM grid preparation and imaging on a 120 kV instrument . . . 123
4.2.6 200 kV and 300 kV cryoEM data collections . . . 124
4.2.7 CryoEM image processing . . . 124
4.3 Results . . . 127
4.3.1 Purifying ILY^{HB1}-*nocharge*, MSP2N2 and nanodiscs . . . 127
4.3.2 Assembling the complex and sucrose gradient . . . 127
4.3.3 Sucrose removal and cryoEM grid optimisation . . . 133
4.3.4 Glacios data collections . . . 137
4.3.5 Double-crosslinking and eBIC data collection . . . 145
4.4 Discussion . . . 152

5 Conclusions and future directions **160**
5.1 Thesis summary . . . 160
5.2 Future directions . . . 162

References **164**

List of Figures

1.1 Examples of α-PFPs . . . 15
1.2 Pore formation by haemolysin . . . 17
1.3 The structurally conserved MACPF/CDC fold . . . 18
1.4 Pore formation by MPEG1/Perforin-2 . . . 20
1.5 Interactions between the electron beam and the sample . . . 23
1.6 Image formation in the electron microscope . . . 24
1.7 The contrast transfer function (CTF) . . . 27
1.8 Principles of 3D reconstruction . . . 28
1.9 *Ab initio* reconstruction by stochastic gradient descent (SGD) . . . 31
1.10 Bonded parameters in a typical forcefield . . . 33
1.11 Non-bonded parameters in a typical forcefield . . . 35
1.12 The leap-frog algorithm . . . 37
1.13 Multiscale modelling . . . 39
1.14 The Martini approach to CGMD . . . 42

2.1 The three complement activation pathways converge to form the membrane attack complex . . . 46
2.2 The terminal complement pathway . . . 49
2.3 Structure of the membrane attack complex (MAC) . . . 50
2.4 Structure of CD59 solved by solution NMR . . . 51
2.5 Structure of the CD59 GPI anchor . . . 52
2.6 Structural basis of MAC inhibition by CD59 . . . 53
2.7 CD59 adopts a different orientation when in complex with ILY and C8α . . . 53
2.8 Effects of inhibited MAC on lipid bilayers . . . 55
2.9 GPI anchors exchange between extended, *lollipop* conformations and *flop-down* conformations . . . 57
2.10 Parametrisation of the EtNP linker . . . 60
2.11 GPI-anchored CD59 samples the ILY-bound and C8α-bound orientations . . . 63
2.12 CD59 flexibility in atomistic MD simulations . . . 64
2.13 CD59 residues involved in hydrogen bonds with DOPC . . . 66
2.14 Orientations of CD59 relative to the membrane . . . 68
2.15 Comparison of atomistic and coarse-grain simulations of the EtNP linker . . . 70
2.16 Placing the C5b8-CD59 complex in the membrane . . . 72
2.17 PyLipID highlights potential lipid binding site . . . 73
2.18 C5b8-CD59 induces a local thinning of the membrane . . . 74
2.19 Schematic summary of CD59 interactions . . . 82

3.1 Domain composition of the archetypal CDC pneumolysin (PLY) . . . 85
3.2 Structure of the CDC domain 4 (D4) . . . 85
3.3 ILY forms an oligomeric early prepore . . . 87
3.4 Structure of the PLY pore . . . 89
3.5 The h-helix in the ILY early prepore . . . 90
3.6 The HB-locked mutants of ILY . . . 98
3.7 Purification of recombinant histidine-tagged pneumolysin (PLY) . . . 99
3.8 Recombinant PLY is active . . . 100

3.9 Mutating the h-helix has no effect on the lytic activity of ILY^{EP} 101
3.10 The role of the h-helix is conserved 102
3.11 Coarse-grain MD simulations of PLY^{hybrid} 104
3.12 Amphipathic helices are found in various membrane-lysing proteins 107
3.13 Overview of the CDC pore formation pathway 109

4.1 Comparison of native GPI-anchored CD59 and $CD59^{cyto}$ 114
4.2 Lipid nanodiscs as a model membrane system 116
4.3 CryoEM grid preparation 117
4.4 ILY^{HB1}-*nocharge* purification 128
4.5 Generation of DOPC-cholesterol nanodiscs 129
4.6 Negative stain EM imaging of the ILY^{HB1}-*nocharge* complex on lipid nanodiscs 130
4.7 Complex purification by sucrose density gradient centrifugation 131
4.8 Negative stain reconstruction of ILY^{HB1}-*nocharge* on lipid nanodiscs 132
4.9 Negative stain EM imaging to check sample stability 133
4.10 GraFix crosslinking of the sample 135
4.11 Screening of cryoEM grid types 136
4.12 Sucrose removal by dilution and concentration yields a sample suitable for structure determination 137
4.13 2D classification of particles from the Gold Lacey dataset 139
4.14 Initial processing to deal with conformational heterogeneity of the Gold Lacey dataset 140
4.15 3D reconstruction from the Gold Lacey dataset (binned by 4) 142
4.16 Image processing for the GrOx Quantifoil datasets 144
4.17 Sample after double-crosslinking with glutaraldehyde and BS^3 146
4.18 Overall processing workflow for the eBIC datasets 147
4.19 Using Topaz to recover side views from the untilted data 148
4.20 Initial reconstructions from the eBIC datasets 149
4.21 Final reconstruction of the early prepore 150
4.22 Final reconstruction of the potential new ILY conformation 151
4.23 Fitting the early prepore structure into the cryoEM reconstructions 153

List of Tables

2.1 Coarse-grain parameters of the EtNP linker 69

3.1 List of primers used for site-directed mutagenesis 93

4.1 Data collection details for the 200 kV Glacios datasets 138
4.2 Data collection details for the 300 kV eBIC datasets 146

1 || Introduction

Chapter summary

Pore-forming proteins (PFPs) are a structurally diverse class of proteins which includes cell-killing toxins, immune effectors and regulated cell death activators. They are broadly divided into two groups according to the secondary structure of the pores they form: α-PFPs and β-PFPs. Despite their structural diversity, PFPs follow a similar mechanism of action whereby soluble monomers bind a target membrane, where they oligomerise and undergo drastic conformational changes to rupture the lipid bilayer. This Chapter reviews some of the key mechanistic and structural features of PFPs, with an emphasis on the membrane attack complex/perforin superfamily of β-PFPs.

Our understanding of PFPs is a result of developments in techniques such as cryo-electron microscopy and molecular dynamics simulations. This Chapter discusses the fundamental concepts of both techniques, providing a theoretical framework for the work presented in the following Chapters.

1.1 Pore-forming proteins

The emergence of biological membranes was a defining event in the evolution of life. All modern cells are surrounded by a membrane separating the cytosol from the extracellular environment, which maintains the conditions necessary for the biochemical reactions of life to take place. Because of their ubiquity and their importance, cell membranes have become the targets of many molecular weapons in intercellular conflicts. Pore-forming proteins (PFPs) are found in all kingdoms of life [1, 2]. They have evolved a variety of ways to pierce lipid membranes, but they follow the same general mechanism. They are secreted as soluble monomers, which oligomerise on a target membrane and undergo conformational changes to insert themselves into the bilayer and permeabilise it. PFPs are broadly separated into two classes, α-PFPs and β-PFPs, based on the secondary structure of the pores they form. Amongst the diversity of PFPs are major virulence factors of many pathogenic bacteria, but also eukaryotic defence factors and endogenous effectors of regulated cell death [3, 4, 5].

1.1.1 α-pore forming proteins

α-PFPs are characterised by the α-helical structure of the pores they form to permeabilise membranes. One of the most extensively studied α-PFPs is cytolysin A (ClyA) produced by some strains of *Escherichia coli*, *Salmonella enterica* and *Shigella flexneri*, the prototypical member of the ClyA family of PFPs [2, 6]. This family also includes non-haemolytic tripartite enterotoxin and the B component of haemolysin BL enterotoxin produced by *Bacillus cereus* [7, 8]. ClyA undergoes dramatic conformational changes to transition from soluble monomer to membrane-inserted dodecamer (Figure 1.1A) [9, 10]. Its N-terminal helix swings towards the membrane, moving the N-terminus by more than 140 Å. The head region of the monomer, made up of a β-tongue flanked by two α-helices, becomes almost

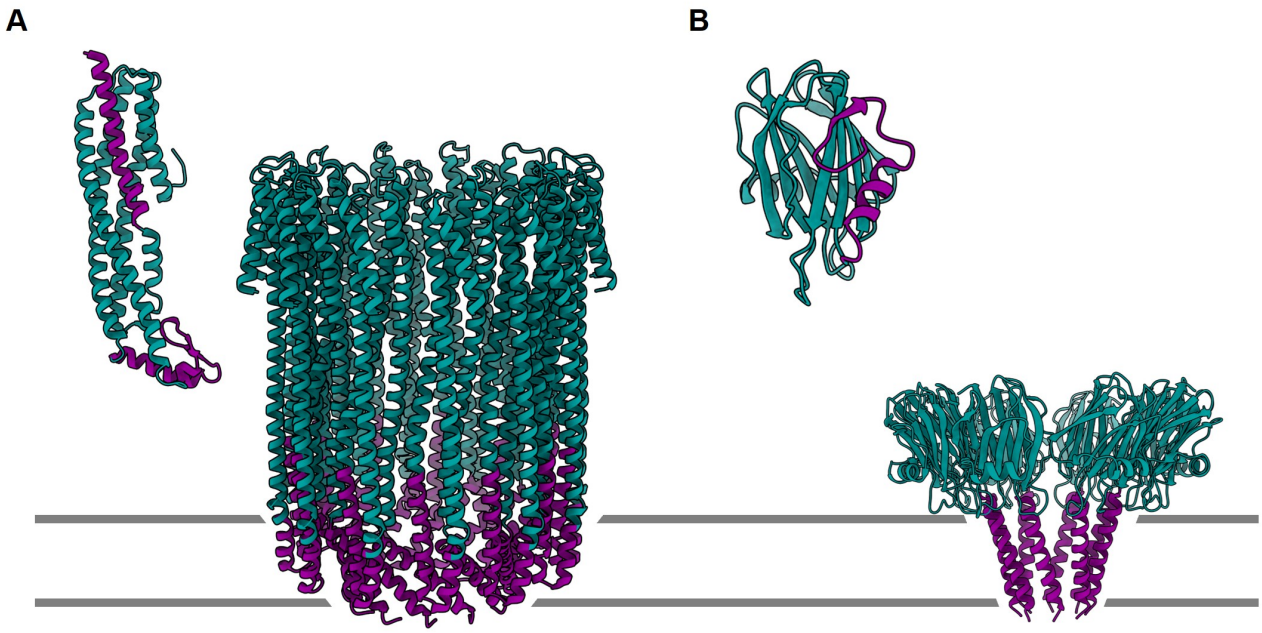


Figure 1.1 | Examples of α-PFPs. **A.** Structures of soluble ClyA (left, PDB 1QOY) and the dodecameric ClyA pore (right, PDB 2WCD). **B.** Structures of the actinoporin FraC in its soluble conformation (left, PDB 4TSL) and as an octameric pore (right, PDB 4TSY). The structural elements undergoing conformational changes to form the pores are highlighted in purple.

entirely α-helical to penetrate the membrane. However, the sequential order of conformational changes in ClyA pore formation remains a debated topic [6]. Several models have been proposed, including one in which the β-tongue transitions upon membrane binding of a monomer, followed by oligomerisation into a prepore and eventually the movement of the N-terminal helix, which opens the pore. Another hypothesis is that the N-terminal helix of a membrane-bound monomer moves before the addition of other ClyA copies, and the pore grows as subunits are assembled.

Other α-PFPs include the colicin family, a group of plasmid-encoded toxins secreted by *E. coli* to kill other *E. coli* strains competing for the same environment [2, 11]. Colicins are structurally diverse and perform a variety of functions, including blocking peptidoglycan synthesis, inhibiting protein synthesis by degrading tRNA or rRNA molecules, DNA degradation and pore formation. The colicin pore-forming fold has also been found in other proteins, e.g. the diphteria toxin [12] and the insecticidal Cry toxins from *Bacillus thuringiensis* [13].

The third family of α-PFPs is that of actinoporins, proteins produced by sea anemones to digest their prey and protect themselves from predators [14]. They are made up of a β-sandwich flanked by two α-helices, one of which swings away from the rest of the protein to penetrate the bilayer and form the pore (Figure 1.1B) [15]. This conformational change is followed by oligomerisation, according to a growing pore mechanism [16].

1.1.2 β-pore forming proteins

Unlike α-PFPs, β-PFPs permeabilise membranes *via* a β-barrel pore. Haemolysin from *Staphiloccocus aureus* is the prototypical member of a protein family that includes the necrotic enteritis toxin B [17] and δ-toxins from *Clostridium perfringens* [18, 19], as well as the *Vibrio cholerae* cytolysin [20]. The soluble, monomeric *S. aureus* haemolysin adopts a heavily β-stranded structure, including a three-stranded sheet termed the *pre-stem domain* (Figure 1.2A). Proteins from the haemolysin family assemble into heptameric or octameric prepores where the pre-stem domain moves away from the core of the protein and refolds into a β-hairpin to form a β-barrel with the other subunits (Figure 1.2B). This β-barrel has a hydrophilic lumen and a hydrophobic outer face, facilitating its spontaneous insertion in the membrane to form the final pore (Figure 1.2C) [21, 22].

The aerolysin family of β-PFPs includes members found in bacteria, archaea, eukaryotes and some viruses [23], such as the prototypical aerolysin from *Aeromonas* spp., monalysin from *Pseudomonas entomophila* [24], enterolobin from the seeds of the South American tree *Enterolobium contortisiliquum* [25], biomphalysin from the snail *Biomphalaria glabrata* [26] and βγ-CAT from the frog *Bombina maxima* [27]. A combination of X-ray crystallography, low-resolution cryo-electron microscopy (cryoEM), molecular dynamics and computational modelling suggested a mechanism whereby aerolysin forms a heptameric prepore on the plasma membrane before large rearrangements flatten the extramembranous region, as the

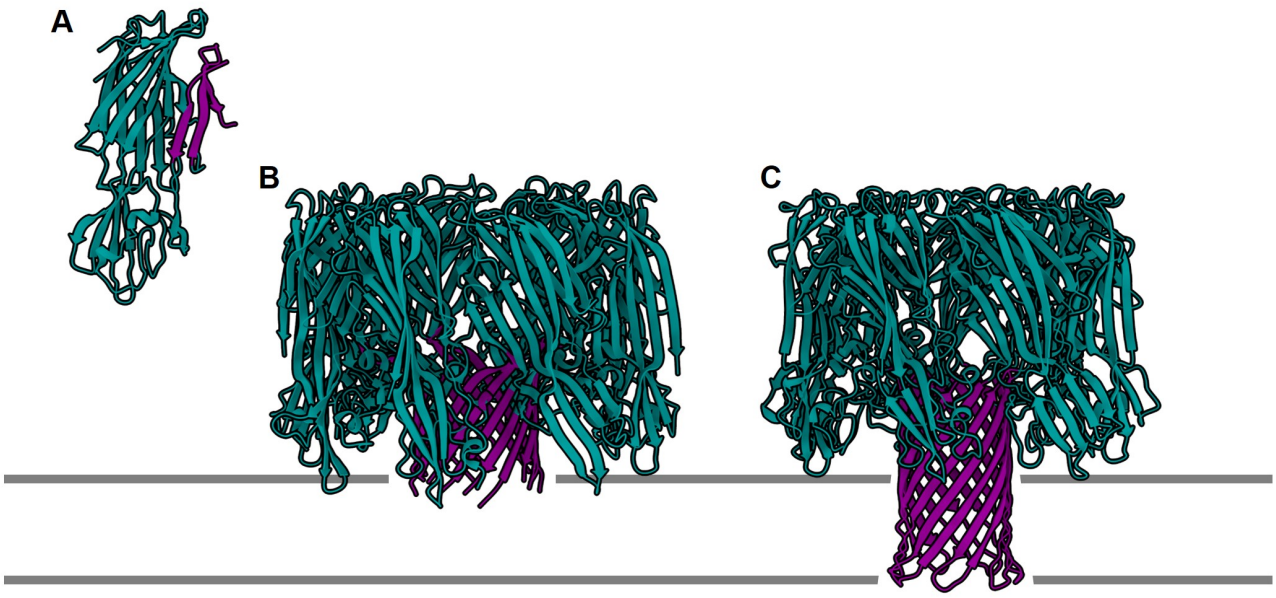


Figure 1.2 | Pore formation by haemolysin. **A.** Monomeric α-haemolysin from *S. aureus* shows the pre-stem (purple) close to the core of the protein (PDB 4IDJ). **B.** The prepore structure of γ-haemolysin reveals the transition of the pre-stem to a β-hairpin (purple) which associates with the β-hairpins from other monomers to form a β-barrel (PDB 4P1Y). **C.** In its pore conformation, the β-barrel (purple) of α-haemolysin spans the lipid bilayer (PDB 7AHL).

pre-stem loop of each monomer transitions to a β-hairpin to form the pore [28].

The third class of β-pore forming proteins is the membrane attack complex/perforin (MACPF) superfamily [2, 29]. It comprises many diverse members, including eukaryotic proteins from the complement system which form the membrane attack complex (MAC) [30], the immune proteins perforin-1 and perforin-2 [31], the bacterial cholesterol-dependent cytolysins (CDCs) [32], the fungal toxin pleurotolysin [33] and apicomplexan proteins involved in parasite egress [34]. While there is little sequence conservation between some members of this superfamily, they all share the same 3D fold responsible for pore formation: the MACPF/CDC domain.

1.1.3 Structural biology of the MACPF superfamily

The MACPF/CDC domain is defined by a twisted β-sheet sandwiched by two α-helical clusters (Figure 1.3) [29, 30, 35, 36]. In its soluble form, the lower cluster of helices is made up of two bundles which, during pore formation, unfurl into β-hairpins (Figure 1.3D). Each

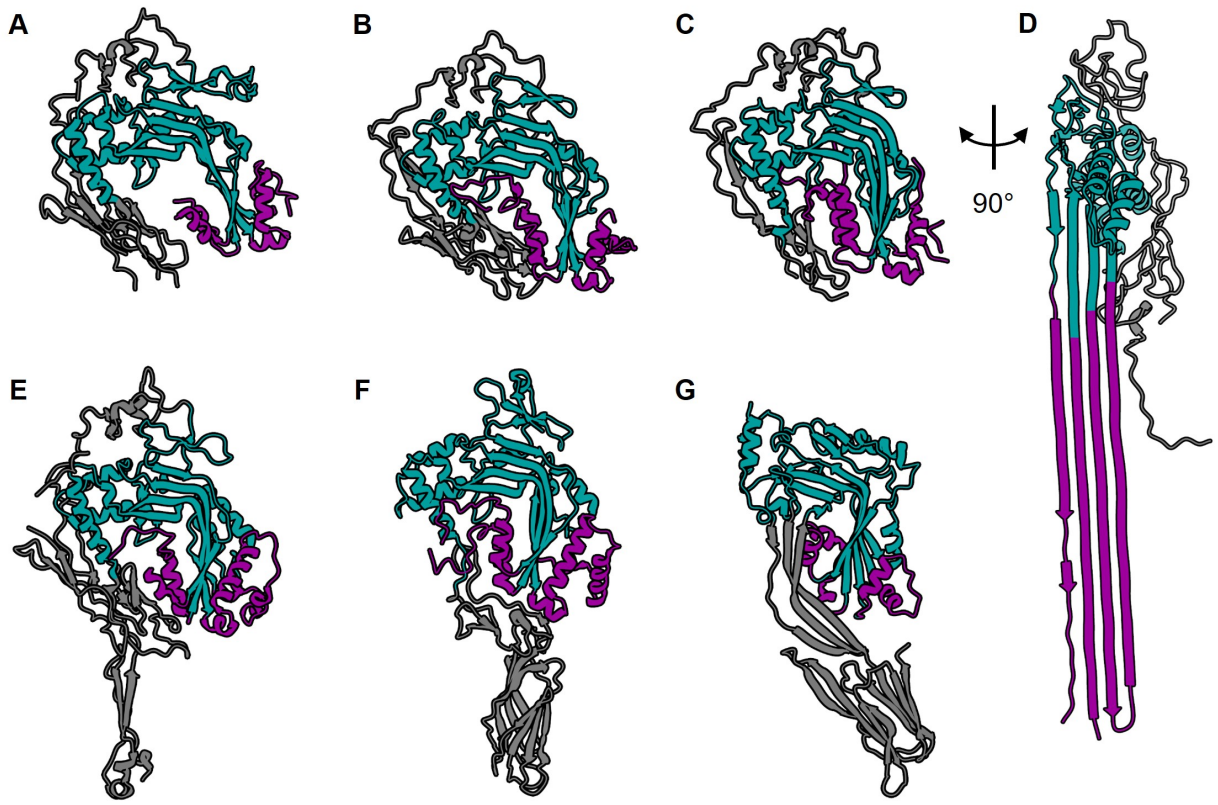


Figure 1.3 | The structurally conserved MACPF/CDC fold. **A.** Soluble form of the complement component C8α (PDB 3OJY). **B.** Soluble form of the complement component C8β (PDB 3OJY). **C.** Soluble form of the complement component C9 (PDB 6CXO). **D.** Membrane-inserted conformation of the complement component C9 (PDB 6H03). **E.** Soluble form of the complement component C6 (PDB 3T5O). Only the MACPF/CDC, low-density lipoprotein receptor class A, TS2 and TS3 domains are shown for clarity. **F**. Soluble form of perforin-1 (PDB 3NSJ). **G.** Soluble form of the bacterial toxin pneumolysin (PDB 5CR6). In each panel, the MACPF/CDC is coloured (green and purple), the residues that unfurl to form the pore are in purple, and all other domains are in grey.

protein from the MACPF superfamily contributes two β-hairpins to the β-barrel pore it forms as an oligomer. While the MACP/CDC domain is structurally conserved, differences in other domains between MACPF proteins lead to diverse mechanisms of pore formation.

The membrane attack complex (MAC) is a giant immune pore that forms on the plasma membrane of pathogens and damaged cells [37]. It is the end result of the complement pathway, a system of membrane-bound and blood-circulating proteins which bridges the innate and adaptive immune responses. MAC is a hetero-oligomeric complex, which assembles as a growing pore: each subunit unfurls its MACPF/CDC helical bundles before the addition of the next protein [38]. The structure and assembly mechanism of MAC are reviewed in detail

in Chapter 2.

Perforin-1 is an immune PFP expressed by cytotoxic T lymphocytes and delivered to infected or cancer cells *via* secretory vesicles sometimes called *lytic granules* [39, 40, 41]. Unlike MAC, perforin-1 assembles into a homo-oligomer and forms a prepore complex prior to membrane rupture [42, 43]. Perforin-1 prepores often only include fewer than ten monomers before the MACPF/CDC helical bundles unfurl to form arc-shaped pores. These assemblies then recruit other small prepores, which similarly transition to the pore conformation until full rings are formed.

MAC proteins and perforin-1 are thought to have evolved from perforin-2, also known as macrophage-expressed gene 1 (MPEG1), after gene duplication events [31]. Unlike other MACPF proteins, MPEG1 is a single-pass transmembrane protein found in the lumen of phagosomes [44]. Soluble MPEG1 monomers are believed to be released by proteolytic cleavage of the transmembrane domain when bacteria are engulfed by the phagosome. CryoEM studies have revealed that MPEG1 forms a prepore, where the MACPF/CDC domain is pointing away from the membrane-binding domain, i.e. away from the bilayer to rupture (Figure 1.4) [45, 46, 47]. Phagosome acidification triggers a $\sim$180° re-orientation of the MACPF/CDC domain as the helical bundles unfurl into pore-forming β-hairpins.

The fungal toxin pleurotolysin (Ply) differs from most other members of the MACPF superfamily as it is a two-part toxin, made up of a membrane-binding polypeptide chain (PlyA) and another subunit containing the MACPF/CDC domain (PlyB) [48, 49]. CryoEM structures of Ply assembled on liposomes revealed that each PlyB monomer interacts with a PlyA dimer to form a prepore, which then transitions into a pore as the MACPF/CDC helical bundles unfurl [33].

Among the best structurally characterised members of the MACPF superfamily are the bacterial cholesterol-dependent cytolysins (CDCs) [32]. These toxins are secreted as monomers

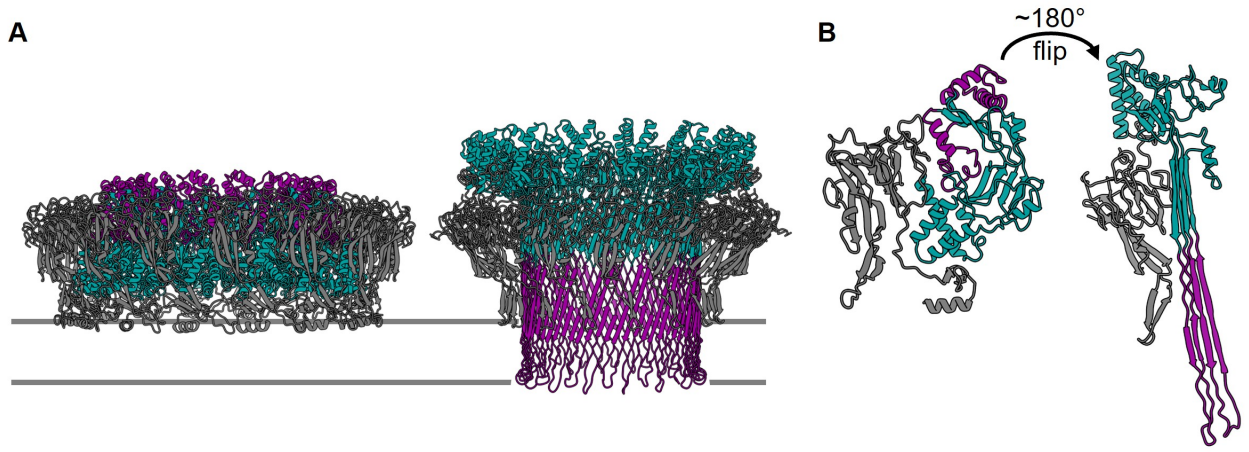


Figure 1.4 | Pore formation by MPEG1/Perforin-2. **A.** MPEG1 prepore (left, PDB 6SB3) and pore (right, PDB 6SB5) solved by cryoEM. **B.** The transition from the prepore (left) to the pore (right) conformation reveals that the MACPF/CDC domain (coloured) undergoes a $\sim$180° flip to reposition its helical bundles (purple) towards the membrane, which unfurl into β-hairpins.

which bind cholesterol-rich membranes, where they assemble to form a homo-oligomeric *early prepore* [50]. Unlike perforin-1 and pleurotolysin, the helical bundles of CDCs are too short to form transmembrane hairpins from their position in the early prepore. Rather, a conformational change in the rest of the protein leads to the collapse of the assembly into a *late prepore* conformation, where the MACPF/CDC domain is brought close to the membrane. This enables the helical bundles to unfurl into membrane-spanning β-hairpins, forming the final pore [51]. The pore-forming mechanism of CDCs is reviewed in detail in Chapter 3.

Extensive sequence analysis of CDCs recently identified a conserved motif, also found in nearly 300 uncharacterised proteins, which have been termed CDC-like proteins (CDCL) [52]. X-ray crystallography of a CDCL from the bacterium *Elizabethkingia anophelis* revealed that it contains the MACPF/CDC domain and a novel C-terminal domain, presumably responsible for membrane binding. The study also identified the existence of shorter CDCL proteins (CDCL^{S}), which lack the C-terminal domain. While the *E. anophelis* CDCL and CDCL^{S} did not lyse liposomes alone, the combination of both proteins resulted in the formation of lytic pores. However, the individual contributions of CDCL and CDCL^{S} to pore formation and the mechanism of action of the assembly remain unknown.

The MACPF superfamily of PFPs comprises many diverse members, which carry out important functions for immunity and virulence. While their pore formation pathways differ, they are all characterised by the structurally conserved MACPF/CDC domain. This makes them attractive systems to study the structural and mechanistic principles underlying membrane perforation by β-pore forming proteins.

1.2 Cryo-electron microscopy

Structural biology has been instrumental in providing a molecular understanding of a wide variety of biological processes essential to life. Solving the three-dimensional (3D) structure of biological macromolecules, in particular proteins, is often necessary to derive a mechanistic explanation of their functions. In recent years, cryo-electron microscopy (cryoEM) has opened many new alleys for structural biologists to study proteins without the constraints imposed by the need to crystallise them. It has been a powerful tool for elucidating the structure of membrane proteins, large assemblies and other samples that do not crystallise easily. It also has the advantage of keeping proteins in a more native environment, including the cellular context in which they perform their functions.

1.2.1 Interactions between electrons and matter

Why use electrons?

Electron microscopy, like visible-light microscopy, aims to magnify objects to resolve details that are not visible to the human eye. Using electrons over photons has several advantages, which make electron microscopy suitable for structure determination of biological molecules. In a perfect microscope, the smallest distance between two distinguishable objects (called the *resolution*) is proportional to the wavelength of the radiation used [53]. In the case

of an electron, the wavelength is inversely proportional to the square root of the voltage accelerating it. For visible light in the range of 400 nm, the maximum resolution will be close to 200 nm, whereas an electron accelerated by a 200 kV voltage will have a wavelength of 2.5 pm, giving a theoretical resolution close to 1 Å.

For decades, X-ray crystallography was the most widely used structural biology technique. It relies on shining X-ray radiation at a crystalline sample and recording the resulting diffraction pattern to determine the structure of the crystallised molecule. The requirement for crystals arises from the fact that X-rays are scattered only by the electron cloud of the sample, resulting in relatively weak scattering. The regular arrangement of molecules in a crystal is necessary to increase the signal so it can be detected for structure determination. Electrons, on the other hand, are scattered by both electrons and nuclei in the sample, due to their negative charge. As a result, electron scattering is much stronger than X-ray scattering, eliminating the need for crystals. As many proteins, in particular large membrane protein complexes, are notoriously difficult to crystallise, using electrons for structure determination widely increases the power of structural biology.

Electron scattering

A key characteristic of an electron beam is its coherence, which is described by two parameters. Spatial coherence is a result of the size of the electron source: perfect spatial coherence would be achieved if all electrons emanated from the exact same point at the source. Temporal coherence describes electrons with the same wavelength, and thus electrons that have a defined phase relationship.

Image formation in the transmission electron microscope (TEM) arises from samples scattering electrons from the incident beam (Figure 1.5). When the incident beam encounters a thin specimen, some electrons are transmitted without being deviated: they form the

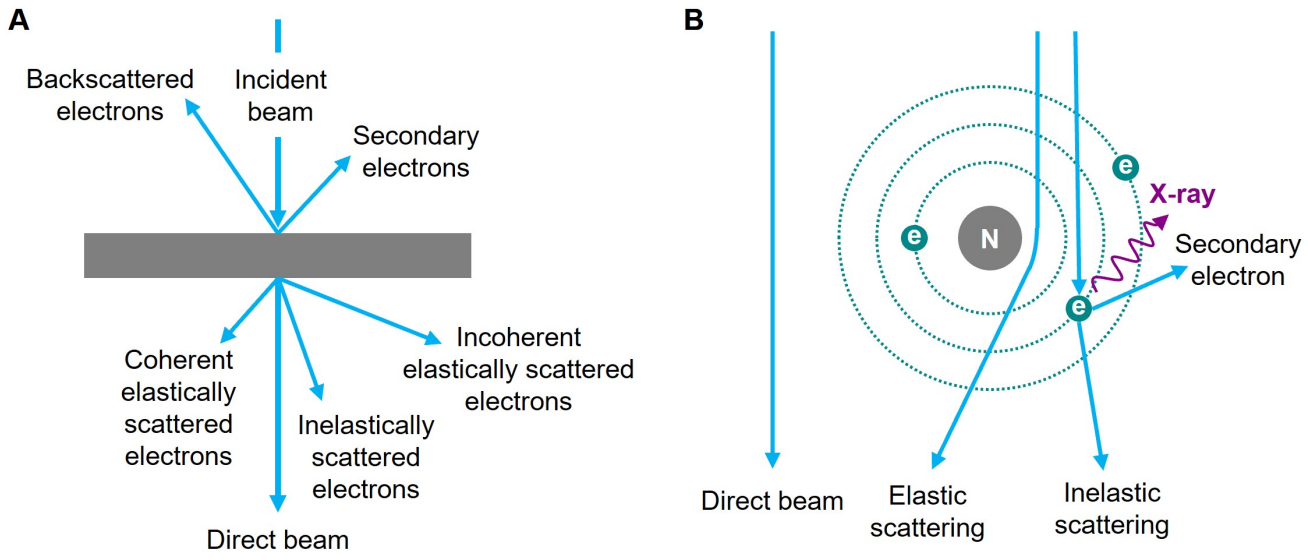


Figure 1.5 | Interactions between the electron beam and the sample [53]. **A.** Electrons from the incident beam can be transmitted without deflection (direct beam), backscattered, inelastically scattered forward (sometimes resulting in secondary electrons) or elastically scattered forward (giving either a coherent or incoherent beam). **B.** The direct beam is the result of electrons that do not interact with the sample. Electrons deflected by the sample without energy loss are elastically scattered. Collisions with electrons in the sample lead to inelastic scattering (energy loss), and can result in emission of secondary electrons and X-ray radiation.

direct beam. Some electrons interact with the sample and are scattered, either forward or backward (*backscattering*), depending on the scattering angle. Among forward-scattered electrons, most are inelastically scattered, i.e. they transfer energy to the specimen. This can result in secondary electron scattering, X-ray emission or bond breaking and sample damage. Inelastic scattering also changes the wavelength of the electron as it loses energy, leading to incorrect focusing by the lens systems of the microscope and causing noise in the image. A minority of electrons will be elastically scattered: they are deviated by the sample but do not lose energy. Coherence between elastically scattered electrons is dependent on the scattering angle. Those scattered at a small angle (up to $\sim 10^{\circ}$) tend to be coherent, whereas at higher angles, elastic scattering is often less coherent. Coherent, elastically scattered electrons are the only electrons that contribute to signal in transmission electron microscopy.

1.2.2 Image formation in the TEM

Anatomy of an electron microscope

In a TEM, a beam of electrons is produced by an electron source and is then focused by a system of lenses (the condenser system) before being scattered by the sample. The resulting scattered electrons are focused again by the objective lens and the image is magnified by the projector lens system before being recorded by the detector (Figure 1.6A) [53].

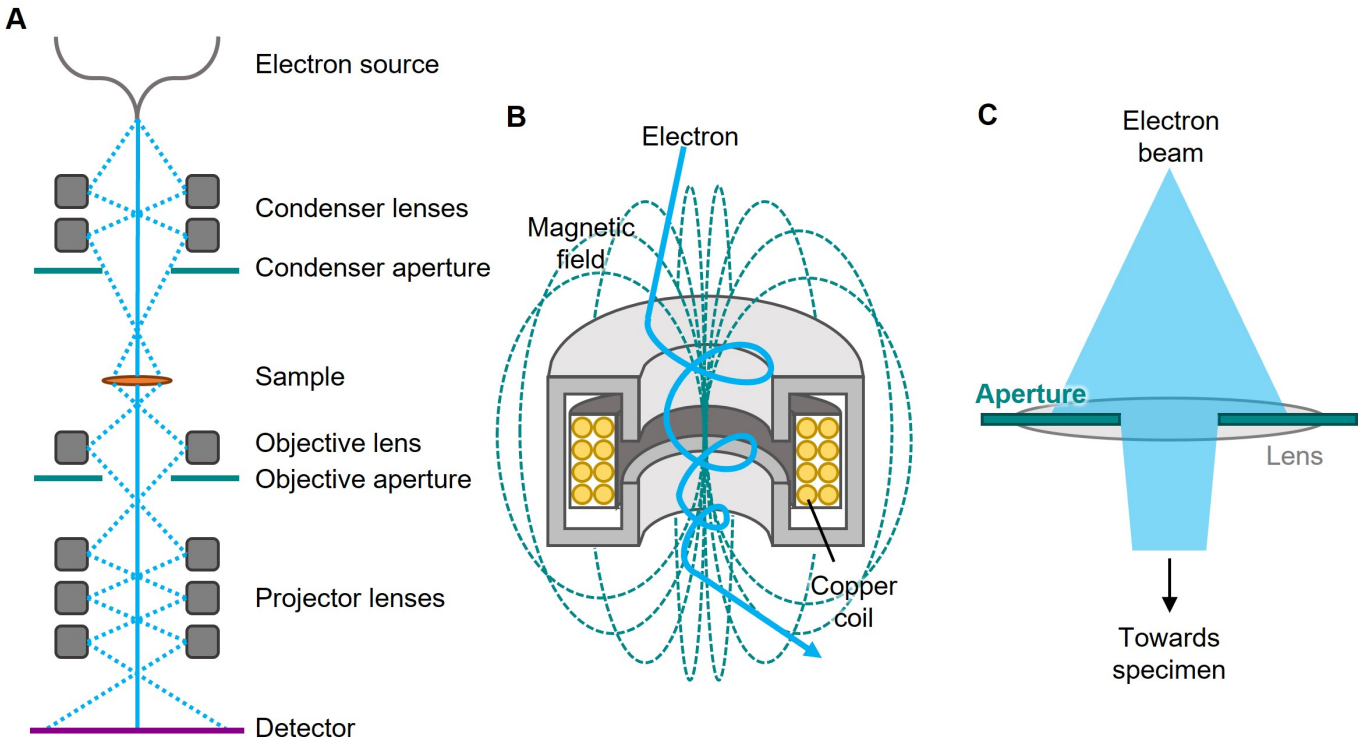


Figure 1.6 | Image formation in the electron microscope. **A.** Anatomy of a transmission electron microscope. A beam is emitted by an electron source and focused by the condenser system. The electrons then reach the sample, where they are scattered. They are focused again by the objective system and the image is magnified by the projector lenses before being recorded by the detector. **B.** Schematic of an electron lens. Current passing through a copper coil creates a magnetic field which accelerates the electrons and magnifies the image. **C.** An aperture restricts the portion of the beam passing through the optical system, resulting in a more parallel beam.

There are two kinds of electron sources: thermionic sources and field-emission guns. A thermionic source works by heating a material (either a tungsten filament or, more commonly, a crystal of lanthanum hexaboride) until the electrons within it acquire enough energy

to overcome the work function of the source, which is the barrier preventing the electrons from being ejected. The electrons exiting the source are first focused, then accelerated by an anode plate. In a field-emission gun, on the other hand, a voltage is applied to a very thin, needle-like tip to induce the release of electrons. The tip works as a cathode, and a nearby anode provides the extraction voltage needed to draw electrons from the tip. A second anode then accelerates the electrons to the appropriate voltage. An important characteristic of electron sources is the coherence of the resulting beam. Temporal coherence is key in a TEM as electrons with different wavelengths will be focused on different planes by the same lens. Electrons also need to be as spatially coherent as possible, i.e. they must be ejected at the same point (or as close as possible) in space. This is because electrons coming from different positions will form an image of the sample at different positions on the detector, resulting in blurring.

The electron microscope contains three main lens systems: the condenser system, the objective lens, and the projector system. Electron lenses contain a copper coil, through which a current is passed. This creates a magnetic field which alters the trajectory of electrons (Figure 1.6B). When an electron encounters a magnetic field, it experiences a force (the Lorentz force), which depends on the velocity of the electron and the strength of the magnetic field. The Lorentz force vector is normal to those of the velocity and the magnetic field, causing the electron to spiral as it travels through the lens. The beam is magnified and the image is rotated as a result. The condenser system and the objective lens also include apertures which make the beam more parallel by only letting a fraction of the beam pass through (Figure 1.6C).

Electron lenses suffer from imperfections, which lead to various kinds of optical defects. Electrons scattered at different spatial frequencies are focused at different points: the further an electron is from the centre of the lens, the closer it is focused, a defect called *spherical*

aberration. Another kind of aberration, termed *chromatic aberration*, refers to the fact that electrons with different wavelengths are focused at different points by the same lens. These aberrations have important effects on contrast, as explained below.

Phase contrast and the contrast transfer function

Scattering of the incident electron beam by the specimen results in contrast, i.e. differences in intensity in adjacent areas of the image. Phase contrast arises from differences in the phase shift of electrons scattered at different spatial frequencies [54]. As the waves recombine on the image plane, they all contribute to an interference pattern due to differences in their phase shifts.

The contrast, and thus the image of the specimen, is affected by the lens aberrations mentioned above. The contrast transfer function (CTF, Equation 1.1) describes how the image is modified as a function of spatial frequency:

$$\mathrm{CTF} = -2\sin\left[\pi\left(\Delta z\lambda k^2 - \frac{C_s\lambda^3 k^4}{2}\right)\right] \tag{1.1}$$

where Δz is the defocus, λ is the wavelength of the electrons, C_s is the spherical aberration coefficient and k is the spatial frequency.

The CTF oscillates and reaches zero several times, leading to total loss of information for some spatial frequencies. In the Fourier transform of the image, this corresponds to dark gaps between bright Thon rings (Figure 1.7). Changes in defocus affect the CTF by changing the spatial frequencies at which signal is lost, so cryoEM data are usually collected at a range of defocus values in order to record information at all relevant spatial frequencies. The CTF also suffers from dampening, arising from both spatial and temporal incoherence. When electrons come from different directions (spatial incoherence), the image is an average of several images, each slightly shifted in space relative to the others. Low spatial frequency

information suffers little from this, but high spatial frequency information is averaged out and lost. Electrons with different energies (temporal incoherence) are focused on different planes due to chromatic aberration in lenses, and therefore hit the detector with different defoci. As the positions of the peaks and zeroes of the CTF depend on defocus, the average CTF shows a progressive decrease in signal at higher spatial frequencies.

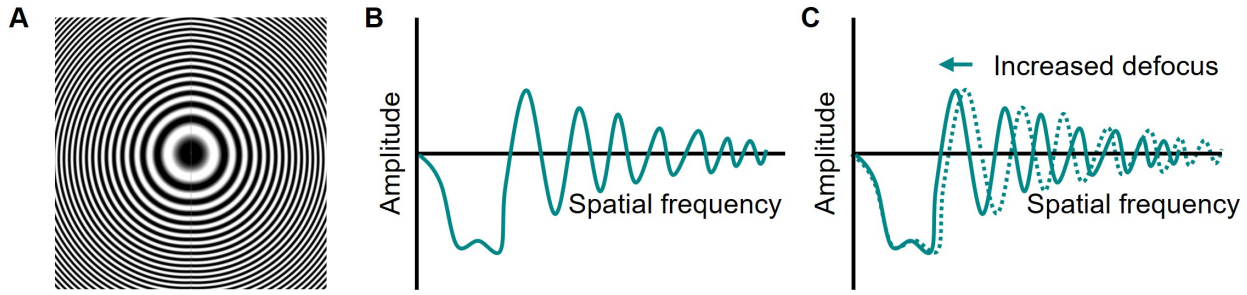


Figure 1.7 | The contrast transfer function (CTF). **A.** In the Fourier transform of a micrograph, Thon rings correspond to alternating peaks and 'zeroes' of the CTF. **B.** Schematic representation of the CTF. **C.** Changing the defocus value affects the CTF: by collecting further away from focus, the first 'zero' of the CTF is shifted to a lower spatial frequency.

The projection theorem

The detector of a TEM records 2D images, called *projections*, of 3D objects. Projections are related to the original sample by the projection-slice theorem, which states that the Fourier transform of each projection is a central slice through the 3D Fourier transform of the object (Figure 1.8A) [55]. In single particle analysis cryoEM, particles ideally adopt random orientations so that all slices from the 3D Fourier transform are present in the dataset.

Generating a 3D reconstruction from 2D projections requires calculating the angles between them. The 'common line' approach relies on the fact that each pair of projections shares a line of identical pixels, corresponding to their intersection in the 3D Fourier transform of the object (Figure 1.8B). Finding this common line between projections helps determine their relative orientations [55]. These angular orientations are then refined to improve the resolution of the reconstruction.

Accurate determination of particle orientations is achieved with projection mapping [54,

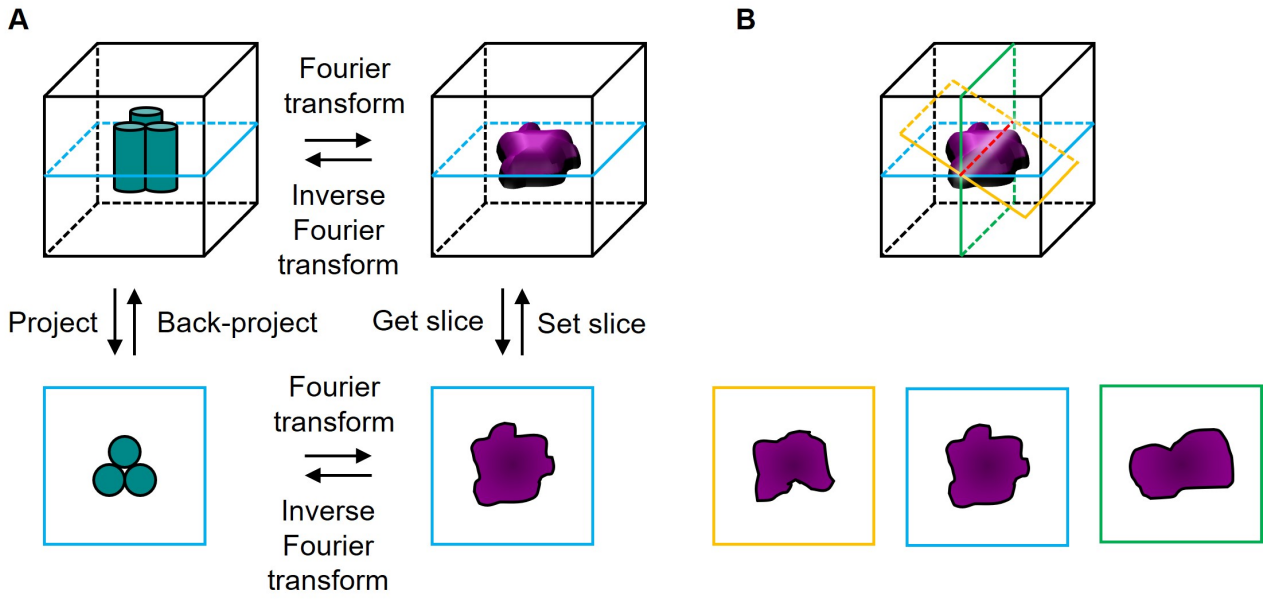


Figure 1.8 | Principles of 3D reconstruction. **A.** The projection theorem: the Fourier transform of a 2D projection is a slice through the 3D transform of the original object. Real space objects are in cyan and reciprocal space objects are in purple. **B.** Fourier transforms of 2D projections (bottom) share a common line (red) corresponding to their intersection in the 3D Fourier transform (top).

55]. A reference structure is used to generate projections in all directions. Each experimental projection is then aligned to these computationally-generated projections, and an algorithm maximises the correlation between them, in order to determine the orientation of the experimental projection with respect to the reference 3D structure. These projections are then used to calculate a new 3D structure and this process is repeated iteratively to improve angular assignments. The low signal-to-noise ratio in single-particle cryoEM requires averaging over tens of thousands of particles to extract signal.

1.2.3 Recent advances in cryoEM

Hardware developments

The invention of direct electron detectors (DEDs) was one of the most significant technological advances in cryoEM in recent years, leading to what is often called the *resolution revolution* [56, 57]. Previously, a typical TEM detector consisted in a scintillator, which con-

verted electrons to light signal, coupled to a CCD camera. The vastly improved quantum efficiency of DEDs allows the use of much lower electron doses, reducing sample damage. Moreover, DEDs collect movies made of multiple frames, rather than static images. As a result, particle motions on the grid can be corrected computationally, reducing blurring of the signal. Finally, the speed of DEDs, combined with developments in data acquisition (e.g. aberration-free image shift, whereby the beam is tilted from grid hole to grid hole instead of moving the stage each time) have dramatically increased the speed of collection and therefore the amount of data that can be generated in a single session.

The quality of the data collected on TEM has been improved by the introduction of energy filters [58]. Most scattering in the microscope is inelastic, and inelastically scattered electrons contribute noise to the image as they are not temporally coherent with elastically scattered electrons. Energy filters contain magnetic prisms which curve the path of electrons. Electrons with different energies are curved at different angles, and a slit is positioned to let through only electrons that have not lost energy through inelastic scattering. This leads to improvements in the signal-to-noise ratio of cryoEM data, facilitating high-resolution structure determination.

Software developments

Advances in image processing software have also played a pivotal role in the resolution revolution. The first improvement was the introduction of Bayesian statistics and maximum likelihood to increase the accuracy of particle alignments. This approach aims to maximise a function describing the probability of the model being correct in light of the data and prior information (corresponding to the smoothness of the molecule) [59]. As a result, particles are assigned probability-weighted orientations, which are refined with each iteration, increasing the accuracy of angular assignments and alignments. This approach also leads to better

classification results, both in two and three dimensions.

Ab initio 3D reconstruction is a computationally difficult task. Early algorithms adopted an iterative refinement approach which aimed to improve an initial starting model, and were prone to 'getting stuck' in local probability maxima (Figure 1.9). The cryoEM software Cryosparc [60] first introduced stochastic gradient descent (SGD) to *ab initio* reconstruction. In this approach, an arbitrary random initialisation (a randomly generated 3D model) is used with a random subset of particles to calculate a Bayesian likelihood function (called the *objective function*) similar to the one described above. The gradient (i.e. the derivative) of that function is approximated with respect to the parameters (i.e. orientations and translations of the particles) for the subset of particles. This gradient points in a direction in the space of 3D structures corresponding to an increase in the objective function. By iteratively approximating the gradient for random subsets of particles and updating the 3D structure accordingly, the algorithm samples the space of 3D structures until it reaches the correct reconstruction. Using only subsets of particles for each iteration is computationally inexpensive and results in large changes between each step. Those large, noisy changes are thought to be the reason why SGD algorithms are insensitive to local maxima.

New computational developments in cryoEM include the use of machine learning at various steps in the structure determination pipeline [61]. For instance, TOPAZ [62] and crYOLO [63] both use neural networks to automate particle picking. CryoDRGN is a machine-learning program that is able to uncover both discrete and continuous heterogeneity in datasets, i.e. heterogeneity due to different conformations being present and heterogeneity arising from flexibility of the macromolecule [64].

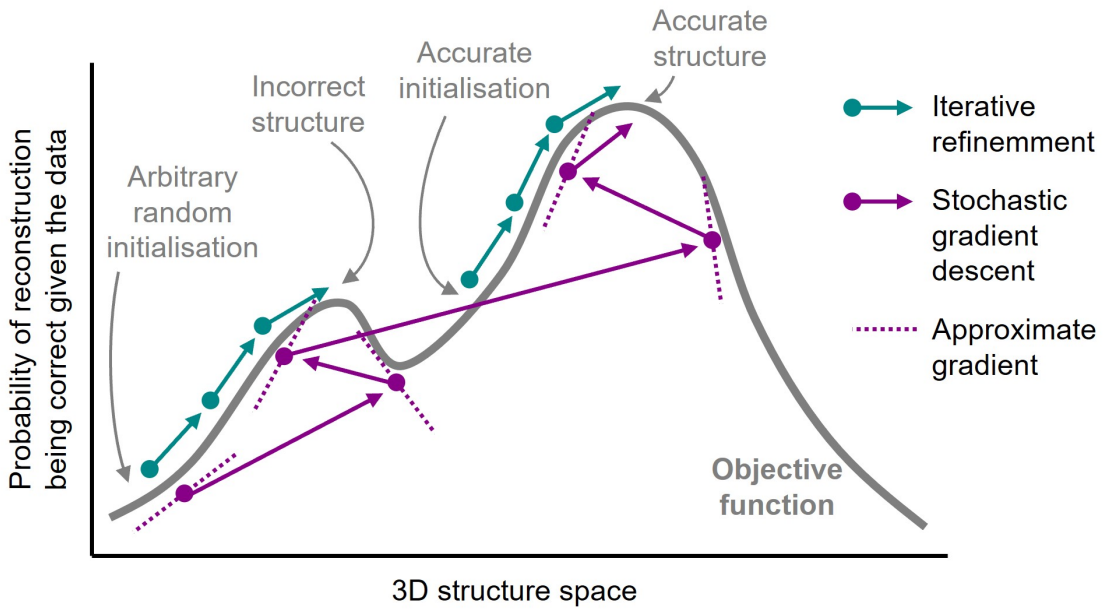


Figure 1.9 | *Ab initio* reconstruction by stochastic gradient descent (SGD). Iterative refinement algorithms aim to maximise an objective function (grey) by making small changes (turquoise arrows) in the parameters (i.e. particle orientations and translations) and are prone to 'getting stuck' in local maxima, making them very sensitive to the initial model used. SGD algorithms iteratively approximate the gradient of the objective function for random subsets of particles and use its direction to sample the 3D structure space much faster (purple arrows). The large changes made between iterations are believed to make SGD algorithms insensitive to the initialisation model and local maxima.

1.3 Molecular Dynamics Simulations

Experimental structural biology techniques such as X-ray crystallography and cryoEM give rise to static structures of proteins or protein complexes. While there have been efforts to resolve conformational ensembles of proteins, each structure remains a snapshot of a given conformation adopted by one or more macromolecules. However, protein dynamics are often key to their functions and can be lost in models derived from cryoEM or crystallography. Molecular dynamics (MD) simulations are a powerful tool to study the movements of individual atoms making up proteins and therefore resolve their overall dynamics.

1.3.1 Forcefields describe the forces applied on each atom

Classical MD simulations are a computational technique based on Newtonian mechanics. Conceptually, MD simulations work by calculating all the forces applied onto each atom and then solving Newton's equations of motion to compute atomic movements resulting from the forces. MD simulations rely on forcefields, which are series of equations describing all the forces affecting an individual atom. A variety of forcefields have been created, for general purposes or specific applications. The most commonly used ones in atomistic biological MD simulations are the Amber [65, 66], CHARMM [67], GROMOS [68] and OPLS/AA [69, 70] forcefield families. The general form of a forcefield typically used for biomolecular simulations is the potential energy function $V(\vec{r}^N)$ in Equation 1.2:

$$V(\vec{r}^N) = V_{\text{bonded}}(\vec{r}^N) + V_{\text{non-bonded}}(\vec{r}^N) \tag{1.2}$$

where $\vec{r}^N = (\vec{r}_1, \vec{r}_2, ..., \vec{r}_N)$ is the complete set of $3N$ atomic coordinates. $V_{\text{bonded}}(\vec{r}^N)$ and $V_{\text{non-bonded}}(\vec{r}^N)$ are potential energy functions corresponding to bonded (e.g. bond lengths, angles, dihedrals) and non-bonded parameters (e.g. Lennard-Jones potential, Coulombic potential), respectively.

Bonded parameters

Bonded parameters can be modelled in a variety of ways, but bond lengths and simple angles (between three atoms) are usually represented as harmonic oscillators (Figure 1.10). This means that, as they deviate from the empirically defined optimal value (e.g. as a bond is stretched or compressed), the potential energy increases, thus imposing an energy penalty on the system. Different forcefields sometimes use different ways of modelling dihedral (torsion) angles, but a common approach is the use of a simple periodic cosine function.

Finally, an improper dihedral is the angle formed by three atoms bonded to a fourth, central atom. Improper dihedrals are used to impose planarity in a ring and prevent plane flipping, and the simplest equation to describe them is a harmonic oscillator. This means that a typical potential energy function for bonded parameters may take the form of Equation 1.3:

$$V_{\text{bonded}} = \sum_{\text{bonds}} \frac{1}{2} K_l (l - l_0)^2 + \sum_{\text{angles}} \frac{1}{2} K_\theta (\theta - \theta_0)^2 + \sum_{\text{dihedrals}} K_\phi [1 + \cos(n\phi - \delta_n)] + \sum_{\text{improper}} \frac{1}{2} K_\omega (\omega - \omega_0)^2 \tag{1.3}$$

where K_l, K_θ, K_ϕ and K_ω are the bond length, angle, proper dihedral and improper dihedral force constants, respectively; l, θ, ϕ and ω are the bond length, bond angle, proper dihedral angle and improper dihedral angle values, respectively; the subscript zero denotes the equilibrium value for each term; n is the multiplicity (i.e. the number of minima as the bond is rotated through 360°) and δ_n is the phase for each n.

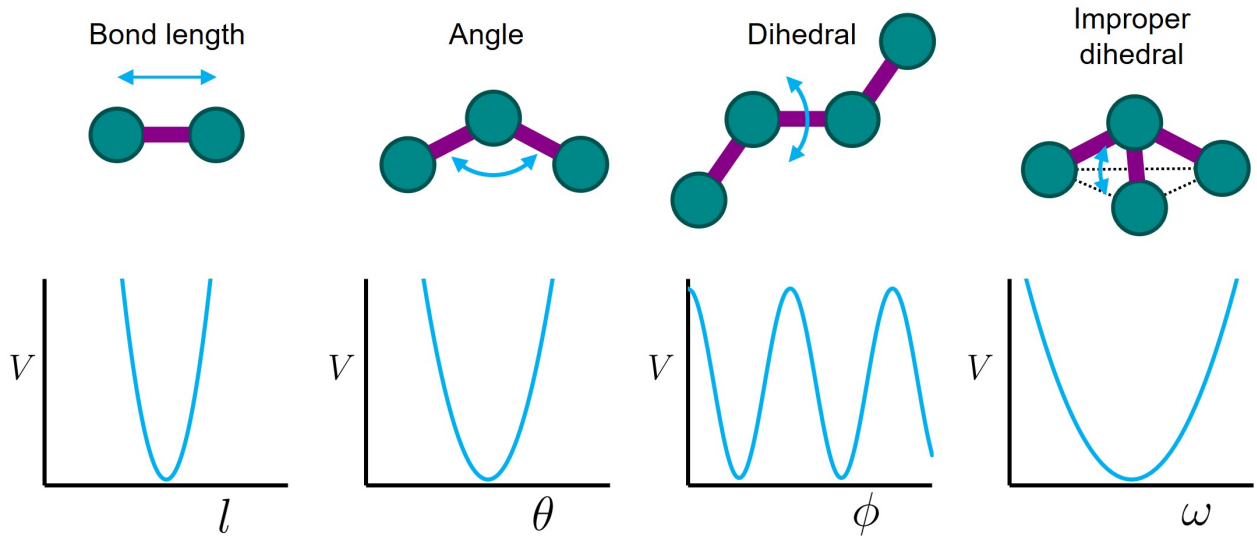


Figure 1.10 | Bonded parameters in a typical forcefield. The top panels show the physical parameters (blue arrows) while the bottom panels are example energy potential functions. Bond lengths, angles and improper dihedrals are often modelled as harmonic oscillators, while proper dihedrals are represented as periodic cosine functions in this example. Green beads are atoms and purple sticks are the bonds between them.

Non-bonded parameters

Inter-atomic interactions also occur between non-bonded atoms: those are taken into account in the non-bonded potential energy function $V_{\mathrm{non-bonded}}$. Contributions to $V_{\mathrm{non-bonded}}$ can arise from externally applied potential fields, pairwise interactions between atoms and higher-order interactions (i.e. between more than two atoms). However, in the way most biomolecular simulations are set up, no external field is applied and higher-order interactions are ignored.

Van der Waals interactions are a combination of non-bonded attractive and repulsive forces between atoms. The energy resulting from this type of interaction is dependent on the distance between the two atoms being considered, and is often modelled using the Lennard-Jones potential. Electrostatic interactions between charged particles are usually accounted for by the Coulombic law, which is also distance-dependent (Figure 1.11). Together, the Lennard-Jones and Coulombic potentials make up the non-bonded potential energy function between atoms i and j in Equation 1.4:

$$V_{\mathrm{non-bonded}} = \sum_{\mathrm{Lennard-Jones}} \epsilon_{ij} \left[\left(\frac{R_{\mathrm{min},ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{\mathrm{min},ij}}{r_{ij}} \right)^{6} \right] + \sum_{\mathrm{Coulombic}} \frac{q_i q_j}{4\pi D r_{ij}} \tag{1.4}$$

where ϵ_{ij} is the well depth of the Lennard-Jones potential and $R_{\mathrm{min},ij}$ is the radius; q_i and q_j are the charge of atoms i and j, respectively; D is the dielectric constant; and $r_{ij} = \|\vec{r}_i - \vec{r}_j\|$ is the distance between atoms i and j.

Differences between forcefields

The potential energy functions in Equations 1.3 and 1.4 are typical examples used in biomolecular simulations, but different forcefields model various potentials differently. For instance,

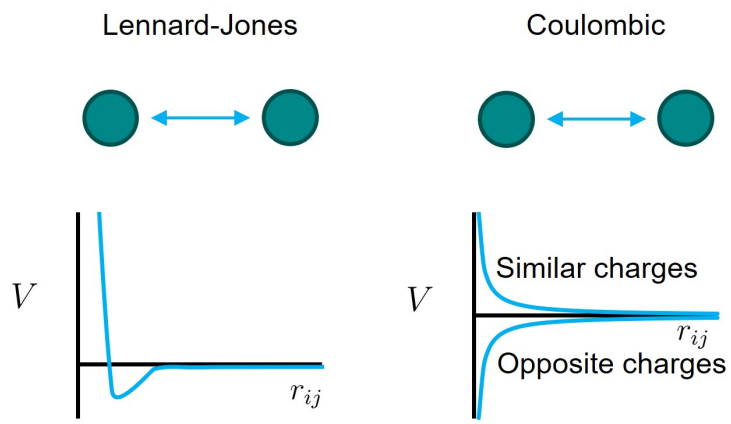


Figure 1.11 | Non-bonded parameters in a typical forcefield. Van der Waals interactions are often modelled using the Lennard-Jones potential (left) while electrostatic interactions are calculated using Coulomb's law (right). In both cases, the potential energy function V is dependent on the distance r_{ij} between the atoms. Coulombic interactions can be attractive or repulsive, depending on whether the charges are of opposite or similar sign, respectively. Green beads are atoms and blue arrows indicate the interactions between them.

the CHARMM forcefield adds a harmonic oscillator function between the two terminal atoms of a three-atom angle to improve accuracy [71].

While the general form of the potential energy function is common to most forcefields, the parameters they use can vary, leading to differences in the behaviour of biological macromolecules in simulations [72, 73, 74]. Parameters are often derived from a mixture of quantum chemistry calculations and experimental data of small fragments, and are usually defined for each atom type. Forcefields are continuously being improved and upgraded to model biological macromolecules more accurately; and some forcefields are better suited to specific types of molecules or application. For instance, a comparison between CHARMM27, CHARMM36 and two forcefields from the AMBER family revealed large differences in their accuracy for simulating DNA [75]. The CHARMM36 forcefield failed to preserve the structure of DNA over the course of the simulations, while CHARMM27 performed the best out of the four forcefields tested. A comparison of CHARMM36, GROMOS 54a7/GROMOS 56a6$_{\text{CARBO/CARBO_R}}$ and the GLYCAM06 forcefield from the AMBER family showed that all three perform similarly to describe protein-carbohydrate interactions

but display significant differences in the pattern of contacts and in unbinding free energies [76]. Comparing simulations of folded and intrinsically disordered proteins (IDP) with nuclear magnetic resonance data highlighted that CHARMM22* performs better than CHARMM36m and four AMBER forcefields (including three IDP-specific forcefields), but all six forcefields tested showed similar results for folded proteins [74]. Together, these results show that the choice of forcefield is an important part of preparing a simulation.

1.3.2 Integrators solve Newton's equations of motion

Newton's second law of motion states that, for a constant mass m_i, the forces $\vec{F}$ applied on a particle i are equal to the product of its mass and its acceleration $\vec{a}_i$. The sum of the forces applied to that particle can be calculated from the coordinates of the system using a forcefield; it is the derivative of the potential energy function with respect to the change in the particle's position. This means that there is a relationship between the potential energy function V and the acceleration of a particle, as shown in Equation 1.5. The acceleration of a particle is the derivative of its velocity, itself the derivative of its position with respect to time.

$$\vec{F_i} = m_i \vec{a}_i = -\frac{d}{d\vec{r_i}} V \tag{1.5}$$

MD integrators are algorithms that solve Newton's equations of motion. In general, integrators need to conserve energy and momentum, and must be reversible: when going back in time, the system should return to its original state [77]. Integrators should also be computationally efficient, be compatible with a long time step Δt for integration, and only calculate forces once per time step. The time step is related to the fastest motion of the molecule. For proteins, the fastest motion is the stretching vibrations of bonds between hydrogen atoms and heavy atoms, $\tau \approx 10$ fs. To avoid errors, the time step should be at least around 20

times smaller than τ, so the maximum time step is $\Delta t \approx 0.5$ fs [78].

In the GROMACS software [79], the standard MD integrator is the leap-frog algorithm. It uses positions $\vec{r}$ at time t, velocities $\vec{v}$ at time $t - \frac{1}{2}\Delta t$, and forces $\vec{F}(t)$ calculated from the positions at time t. It updates velocities and positions using Equations 1.6 (which includes masses m) and 1.7, respectively.

$$\vec{v}\left(t + \frac{1}{2}\Delta t\right) = \vec{v}\left(t - \frac{1}{2}\Delta t\right) + \frac{\Delta t}{m}\vec{F}(t) \tag{1.6}$$

$$\vec{r}\,(t + \Delta t) = \vec{r}\,(t) + \Delta t \vec{v}\left(t - \frac{1}{2}\Delta t\right) \tag{1.7}$$

In the leap-frog algorithm, velocities are staggered with respect to positions. They both then leap like frogs over each other's back, as schematised in Figure 1.12. This means that, at $t = t_0$, the algorithm needs velocities from $t = t_0 - \frac{1}{2}\Delta t$. If they are not available (e.g. from a previous run of that system), they are generated with a Maxwell-Boltzmann distribution at a given temperature. Other integrators include Verlet, velocity Verlet and Langevin algorithms.

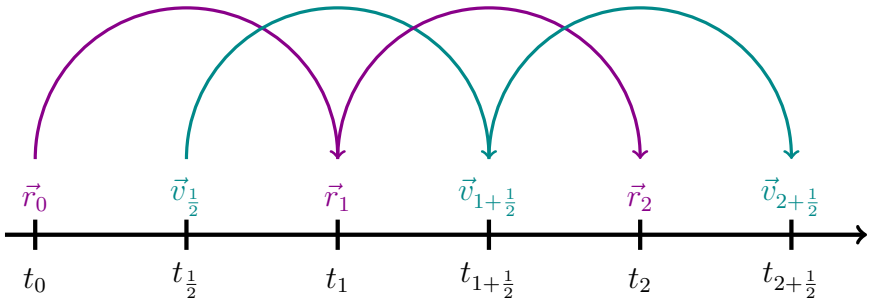


Figure 1.12 | The leap-frog algorithm calculates velocities $\vec{v}$ and positions $\vec{r}$ in a staggered way: the are updated like frogs leaping over each other.

1.3.3 Typical steps in molecular dynamics

A typical MD workflow comprises four main steps: (1) system preparation, (2) energy minimisation, (3) equilibration and (4) production [80]. System preparation is completely depen-

dent on the biological question to answer and the approach often needs to be specific to the molecules being simulated. Generally, the starting configuration needs to be built, including the protein of interest, usually solvated in water, sometimes in the presence of ions. If necessary, multiple copies of the molecule of interest can be placed in the simulation box, ligands can be added, a membrane can be included, etc. The resulting configuration needs to be described in terms of forcefield parameters, provided that an appropriate forcefield exists.

Energy minimisation, or relaxation, aims to find a local energy minimum of the starting structure. This ensures that, at the start of the simulation, there are no forces so great that atoms would have to move an unreasonably large distance. In GROMACS, the standard minimisation algorithm is a steepest descent one in which atoms are moved by a distance up to a set maximum, and forces are then calculated. If the potential energy of the system is decreased by the atomic movements, the new configuration is accepted; if not, the system reverts to its previous state and a different displacement is trialled. This is done a number of times set by the user, or until the maximum of the absolute values of the force components is below a target value.

The minimised system then needs to be brought to equilibrium (after generating starting velocities as mentioned in Section 1.3.2). This process is done by running one or more short simulations, sometimes with a small time step compared to the final production run. The aim of equilibration is to bring the system to the desired conditions (e.g. temperature, pressure) in which the run will be performed. Equilibration can be done in various thermodynamic ensembles: NVE (constant number of particles, volume and energy), NVT (constant number of particles, volume and temperature) or NPT (constant number of particles, pressure and temperature).

The production run mainly differs from the equilibration in that the system is considered to be at equilibrium and so the simulation can be used to extract properties of that system.

The early part of a production run is sometimes excluded from the analysis as additional equilibration. The length of the production run is usually determined based on the sampling required (longer simulations usually offer more sampling) and the computational resources available.

1.3.4 Coarse-grain MD and the Martini forcefield

Biological processes happen over a wide range of temporal and spatial scales, from electron transfer to movement at the organism level. Different computational approaches can be used to model biological processes at different levels: this is called *multiscale modelling* (Figure 1.13).

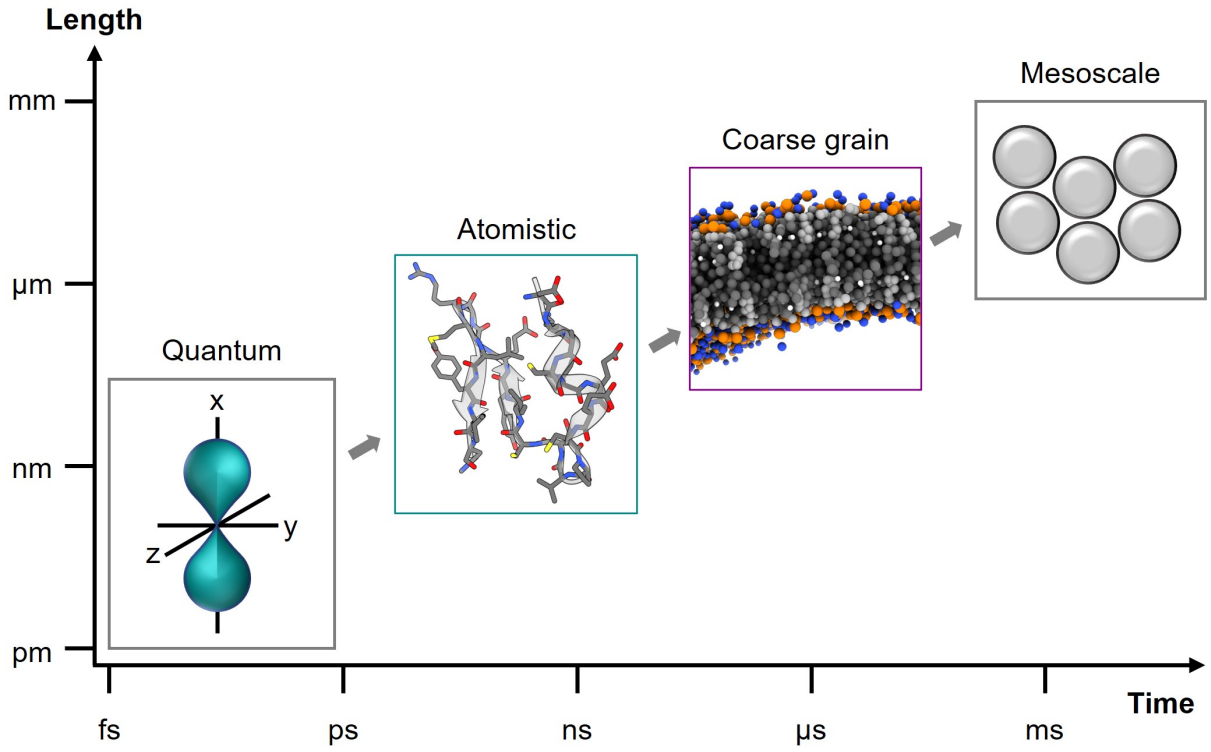


Figure 1.13 | Multiscale modelling comprises several modelling techniques that describe matter at specific scales in time and space.

The most fundamental description of matter comes from quantum mechanics, where a single electron is the smallest unit. Larger systems can be modelled by atomistic MD simulations, where the smallest unit is the atom, and whose principles are detailed above. Atomistic

MD simulations can be used to study the flexibility of proteins, ligand binding and unbinding, conformational changes, allostery, protein folding and a variety of processes at the protein level [81]. Coarse-grain molecular dynamics (CGMD) allows the study of larger systems over longer timescales when compared to atomistic MD. In CGMD, atoms are grouped together into beads or pseudo-atoms, and the motions of those beads are calculated using adapted forcefields. While atomistic detail is lost, the beads are parametrised to match the physico-chemical properties of their constituent atoms, making CGMD a powerful tool to study membrane or micelle self-assembly, pore formation, membrane deformation or protein-lipid interactions [82]. Mesoscale processes can be modelled using methods like dissipative particle dynamics, where each particle represents a whole molecule or a fluid region. Chemical details are lost and only physical properties like hydrodynamics and thermal fluctuations are modelled, making this approach well suited for complex fluids and polymers [83]. Some modelling approaches may combine elements from multiple scales, for instance quantum mechanics/molecular mechanics (QM/MM) simulations include a small portion of the system treated with a quantum mechanical method within an otherwise atomistic box [84]. The detail of atomistic MD can also be combined with the speed of CGMD, resulting in hybrid systems where an all-atom molecule is placed in a coarse-grain environment [85].

The Martini model

The most widely used model in biomolecular CGMD is the Martini model, which relies on the idea that small chemical fragments within molecules can be represented by specific beads, regardless of the rest of the molecule [82]. It was built using a combination of top-down and bottom-up parametrisation approaches: the non-bonded interactions are modelled as Lennard-Jones potentials based on experimental thermodynamic data (top-down) while the bonded parameters are derived from atomistic simulations (bottom-up). The original

forcefield developed in 2002 contained only four bead types (apolar, intermediate polar, polar and charged) and was restricted to lipids. It was used to simulate the spontaneous assembly of lipid vesicles [86], as well as the process by which they fuse with each other [87].

Martini 2 was introduced in 2007 to capture the chemical diversity of lipids more accurately [88]. The new forcefield included 18 bead types and successfully reproduced the separation of lipid-ordered and lipid-disordered phases and the formation of lipid rafts in model membranes [89]. The Martini 2 forcefield uses four-to-one mapping: on average, four heavy (i.e. non hydrogen) atoms are grouped together into a single bead (Figure 1.14A). In addition to the 18 bead types, it includes three bead sizes, resulting in 54 possible beads overall. The forcefield takes on a similar form to that presented in Section 1.3.1 with parameters adapted for the bead types used. The use of pseudo-atom beads results in smoother energy potentials (Figure 1.14B) which, along with the reduced number of particles when compared to an equivalent atomistic system, leads to a speed-up in computation. Martini 2 was improved in 2008, then again in 2013, to include proteins [90, 91]. As hydrogen atoms are not represented in the beads, bonded parameters had to be added specifically to maintain secondary structure elements. To stabilise tertiary structure, an approach using an elastic network of additional harmonic bonds to maintain the overall fold (termed ElNeDyn, or elastic network in dynamics) was also introduced [92]. Martini 2 was also extended to other biomolecules, including carbohydrates in 2009 [93], DNA in 2015 [94] and RNA in 2017 [95]. One of the limitations of Martini 2 was the way it originally modelled water: it was an uncharged bead corresponding to four water molecules. This did not capture the electrostatics of water, so a model of polarisable water was introduced in 2010, where each water molecule is represented by a neutral bead bonded to a positively-charged and a negatively-charged bead, creating a dipole moment [96]. This model was better able to simulate ion transport across lipid membranes and membrane electroporation. Overall, the Martini 2 model was used

in hundreds of studies to simulate various biological systems and processes, including a eukaryotic plasma membrane with over 60 different lipids [97], lipid scrambling [98], drug delivery by lipid nanodiscs [99], complex formation between a G protein-coupled receptor and β-arrestin [100] or the human nuclear pore complex [101].

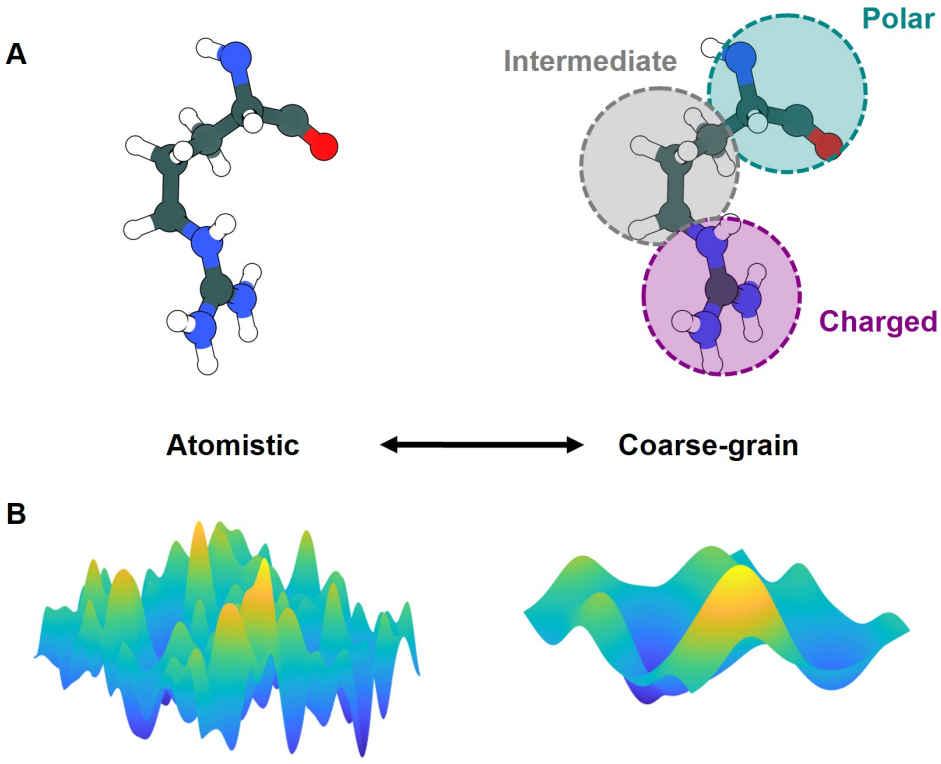


Figure 1.14 | The Martini approach to CGMD. **A.** An arginine residue in a protein is represented by three beads: one polar bead for the backbone (green), a bead of intermediate hydrophobicity for the aliphatic part of the side chain (grey) and a charged bead for the guanidinium group (purple). **B.** The energy potential landscape of a coarse-grain system in the Martini forcefield (right) is smoother than the corresponding atomistic landscape (left), leading to computational speed-up.

Two key limitations of Martini 2 are its tendency to overestimate intermolecular interactions (sometimes called the *stickiness problem*) and the limited chemical space available, especially for small-molecule ligands [82]. For proteins specifically, Martini 2 often fails to reproduce protein flexibility and to model protein cavities accurately. To solve these problems, a new version of the forcefield, Martini 3, was recently released [102]. It includes a total of 843 beads with a more specific interaction matrix and improved parameters for many beads. The representation of proteins was also improved by a full re-parametrisation

of side chains and changes to backbone bead types, which no longer depend on secondary structure. As an alternative to the ElNeDyn approach, Martini 3 can be combined with a network of Lennard-Jones interactions based on the contact map of the native structure, which improves the way protein flexibility is modelled [103, 104]. Since its release, Martini 3 has been tested for protein dimerisation, revealing that it has solved the stickiness problem of Martini 2 and describes well the free energy of association of transmembrane proteins, although it underestimates it for soluble proteins [105]. It was also shown to perform well when describing the interactions of peripheral membrane proteins with lipid bilayers [106] and binding of small molecules to protein pockets [107]. It is still in active development, and was recently extended to carbohydrates [108, 109]. However, despite the theoretical increase in chemical space available in Martini 3, Martini 2 still has a much larger number of published and validated parameters, and so remains a valuable tool in CGMD.

2 || Molecular Dynamics simulations of the inhibited membrane attack complex

Chapter summary

The membrane attack complex (MAC) is a giant immune pore, which forms on pathogens and damaged cells as a result of complement activation. The last-resort inhibitor of MAC is CD59, a small glycosylphosphatidylinositol (GPI)-anchored protein ubiquitously found in the cell membrane of human cells. Some cholesterol-dependent cytolysins (CDCs) from pathogenic bacteria hijack CD59 to aid pore formation and lyse cells. Recent cryoEM work has revealed that CD59 adopts drastically different orientations relative to the membrane when inhibiting MAC and when promoting CDC pore formation. I used atomistic molecular dynamics (MD) simulations of GPI-anchored CD59 to characterise its flexibility, showing that it adopts a wide range of orientations. CD59 preferentially samples specific poses, including those compatible with MAC and CDC binding.

MAC deposition on lipid bilayers is accompanied with changes in the physico-chemical properties of membranes. The cryoEM maps of inhibited MAC on lipid nanodiscs suggested interactions between complement proteins and lipids. I used coarse-grain MD simulations of the CD59-C5b8 complex in a lipid bilayer to characterise the effects of inhibited MAC on its local lipid environment. The simulations showed membrane thinning and lipid binding, which may act as signals for CD59 recruitment or inhibited MAC clearance.

Related publication: Emma C. Couves, Scott Gardner, Tomas B. Voisin, Phillip J. Stansfeld, Edward W, Tate, Doryen Bubeck. In press. Structural basis of membrane attack complex inhibition by CD59. *Nature Communications*.

2.1 Introduction

2.1.1 The Complement cascade

Complement is a system of over 40 membrane-bound and blood-circulating proteins. It is the immune system's first line of defence against many pathogens and is involved in tissue homeostasis by recognising damaged or abnormal cells [37]. It bridges the gap between innate and adaptive immunity, and regulates the adaptive response [110]. Complement can be activated by three pathways, which all trigger a cascade of proteolytic activation events, ultimately converging in the formation of the membrane attack complex (MAC) (Figure 2.1).

Complement activation occurs on target membranes

In both the classical and the lectin pathways, complement activation starts with the recognition of a molecular 'danger' signal on the surface of a cell membrane. In the classical pathway, antibodies bind antigens on pathogens or damaged cells and oligomerise to form a platform which recruits complement component C1 (Figure 2.1A) [111, 112]. C1 circulates in serum as a complex of C1q and the serine proteases C1r and C1s, which are activated upon C1q binding to the membrane-associated antibody complex [113, 114]. Cryo-electron tomography (cryoET) data show that C1 forms localised clusters on liposomes, which may allow activation of C1r and C1s by the serine proteases from neighbouring complexes [112]. In addition to antibodies, C1q directly recognises a variety of membrane proteins and lipids [115, 116, 117]. These include negatively-charged lipids in liposomes [118], lipid A in the lipopolysaccharide of *Escherichia coli* [119] and phosphatidylserine on the surface of apoptotic cells [120]. While the membrane and lipid environment clearly play a role in complement activation, the nature of the protein-lipid interactions and their effects on complement function remain unresolved.

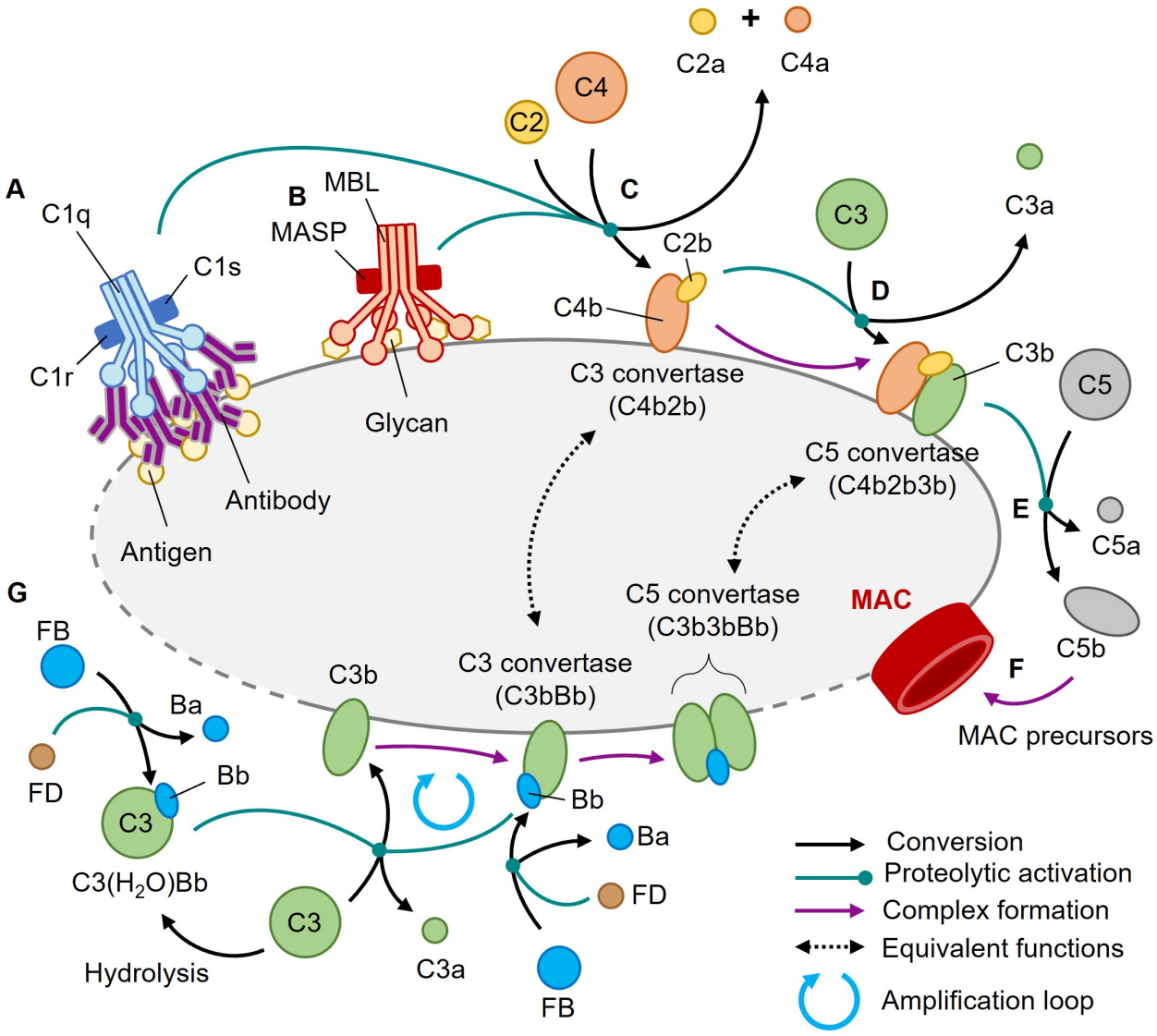


Figure 2.1 | The three complement activation pathways converge to form the membrane attack complex. **A.** In the classical pathway, C1q recognises antibodies bound to antigens. **B.** In the lectin pathway, the mannose-binding lectin (MPL) recognises pathogen-associated glycans. **C.** In both cases, serine proteases (C1s, MASP) proteolytically activate C4 and C2 to form the C4b2b C3 convertase. **D.** The resulting protein complex activates C3 to form the C4b2b3b C5 convertase. **E.** The C5 convertase cleaves C5 into C5a and C5b. **F.** C5b assembles with other complement proteins to form the membrane attack complex (MAC). **G.** In the alternative pathway, spontaneous hydrolysis of C3 leads to the formation of alternative C3 and C5 convertases which perform the same functions as those generated by the other two pathways.

In the lectin pathway, the mannose-binding lectin (MBL) binds pathogen-associated sugars [121] or proteins [122], but was also proposed to recognise the apoptotic signals calreticulin and CD91 [123] and even some phospholipids [124]. Similarly to C1q, MBL is in complex with serine proteases (the MBL-associated proteases, MASPs), which are activated

by the recognition of a 'danger' signal (Figure 2.1B) [125, 126]. Both the classical and lectin pathways trigger the same proteolytic cascade whereby blood-circulating C4 is cleaved into fragments C4a and C4b (Figure 2.1C) [127]. C4b covalently attaches to the membrane, where it binds C2. The serine proteases then cleave C2 to release C2a while C2b remains associated to C4b to form the C4b2b C3 convertase [128]. The initial cleavage of C4 reveals a thioester, which reacts with the target membrane to anchor C4b covalently. That thioester is rapidly hydrolysed in the absence of substrate [129] but C4b release from its activating MASP is relatively slow, which favours the formation of the convertase near the site of complement activation [130]. This suggests that the complement cascade is propagated near the initial site of activation, but the physiological implications on downstream complement proteins are unknown.

The alternative pathway is constantly active *via* a process called *tick-over*. A small proportion of serum C3 spontaneously hydrolyses to form C3(H_2O) in solution, which binds to Factor B (Figure 2.1G). Factor B is then cleaved by serine protease Factor D, and Fragment Ba is released as the soluble C3(H_2O)Bb complex is formed. C3(H_2O)Bb acts as a C3 convertase in solution: it cleaves serum C3 into C3a and C3b, allowing C3b deposition on a nearby cell surface. Fragment Bb, released from Factor B by Factor D-mediated hydrolysis, associates with the membrane-bound C3b to form the C3bBb C3 convertase. The convertase is part of an amplification loop: it proteolytically activates more serum C3 to form membrane-attached C3b, that can then become part of new convertase complexes. Addition of another C3b molecule to an already-formed convertase changes the specificity of the complex and results in the formation of the C3b3bBb C5 convertase (also called the *alternative* C5 convertase), which cleaves C5 into C5a and C5b. However, the alternative C3 convertase itself also exhibits some C5 convertase activity, and comparing its kinetics in solution and on a surface showed that the surface-associated convertase has a much higher

affinity for C5 [131]. The orientation of C3b also influences the convertase activity [132], indicating that the membrane on which complement is activated functions as a physical scaffold allowing complement proteins to orient themselves efficiently. The local protein environment of the convertase within the membrane also influences its activity: C3b needs to be densely packed on the target surface for efficient C5 convertase activity, but not C3 convertase activity [132, 133]. These results demonstrate that assembly on a target surface is important for activity of the alternative C5 convertase, but disentangling the precise contributions of protein-protein and protein-lipid interactions has been challenging.

Similarly to the alternative C3 convertase, addition of C3b to the classical C4b2b C3 convertase leads to the formation of the C4b2b3b C5 convertase (Figure 2.1D). The classical and alternative C5 convertases both cleave C5 into two products: the small anaphylatoxin C5a and the larger fragment C5b [134]. C5b is the first component of the membrane attack complex (MAC) but it has a short half-life as a soluble monomer. As a result, MAC forms near the convertase, and so the location of complement activation dictates the environment in which the lytic pore will be assembled. However, the molecular details of the interactions between membrane lipids and complement proteins, as well as their impact on MAC activity, have not been characterised.

The terminal complement pathway

All three activation pathways converge at the formation of the C5 convertase, which activates the terminal complement pathway by cleaving C5 (Figure 2.2). C6 rapidly binds and stabilises labile C5b and unfurls helices in its MACPF/CDC domain into β-hairpins that bind the membrane near the convertase [135]. C7 then binds the complex and undergoes a similar conformational change to form C5b7 [136]. C8 is a heterotrimeric complex made up of the C8α, C8β and C8γ polypeptide chains [30, 137]. C8β binds to C7 and unfurls its

MACPF/CDC helices into β-hairpins, as does C8α [38, 138, 139]. C8γ is, along with C5b, one of only two MAC proteins lacking the MACPF/CDC domain and, as such, does not interact with the membrane directly. A first C9 molecule associates with C5b8 to form the $C5b9_1$ complex. Further oligomerisation of C9 subunits then results in the full membrane attack complex (MAC) pore [140], with a diameter of 110 Å [38].

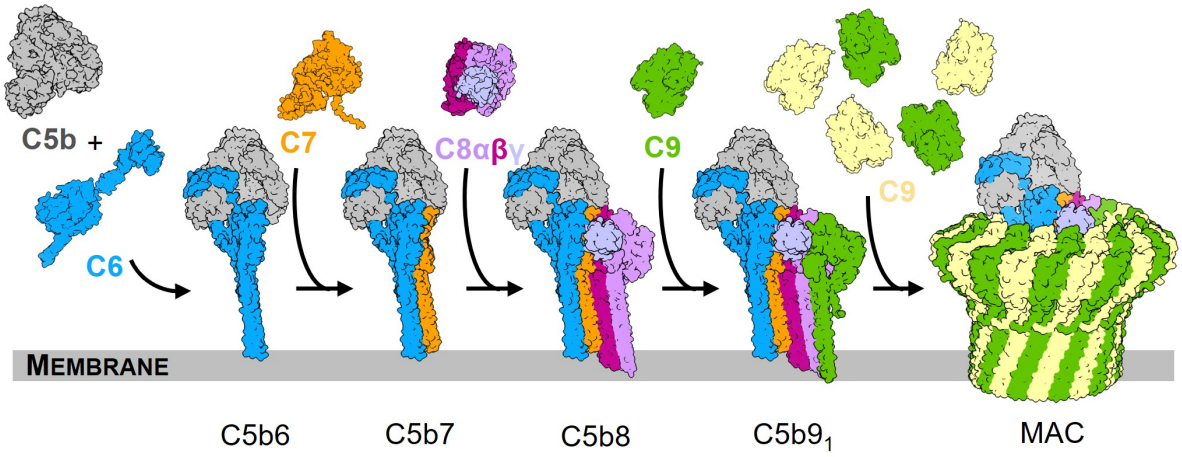


Figure 2.2 | The terminal complement pathway is the sequential assembly of complement proteins to form the membrane attack complex (MAC). Labile C5b is captured by C6, which stabilises it and anchors the complex to the membrane by unfurling its MACPF/CDC domain helices. C7 undergoes a similar conformational change to bind C5b6 and form the C5b7 complex. Heterotrimeric C8 then binds and unfurls its MACPF helices to form the C5b8 complex. Several copies of C9 then oligomerise to form the complete pore. All assemblies are derived from PDB 6H03. Soluble structures of proteins are derived from PDB 3T5O (C6), 3OJY (C8) and 6CXO (C9). The soluble C7 structure is a prediction from the AlphaFold Protein Structure Database (ID P10643) [141, 142]. This original figure was also used for publication in [143] under a CCBY 4.0 license.

MAC inserts into the membrane *via* the tips of its β-hairpins, which form a hydrophobic belt on the outer face of the pore, evidenced by the presence of a detergent belt in cryoEM structures (Figure 2.3A) [38]. MAC adopts a 'split-washer' shape and, as a result, the hydrophobic belt is not planar, unlike that of the CDC pneumolysin, which shares a common pore-forming fold [51]. Another immune pore-forming protein, Perforin-2, also shows a twisted conformation and a break in planarity [47], but the impact of these non-planar hydrophobic belts on the membrane is unknown. MAC can adopt two conformations: an open conformation, characterised by a ∼30 Å chasm between the edge of the final C9 and that of

C6, and a closed conformation where the β-barrel is sealed (Figure 2.3B). The physiological relevance of the two conformations is unclear.

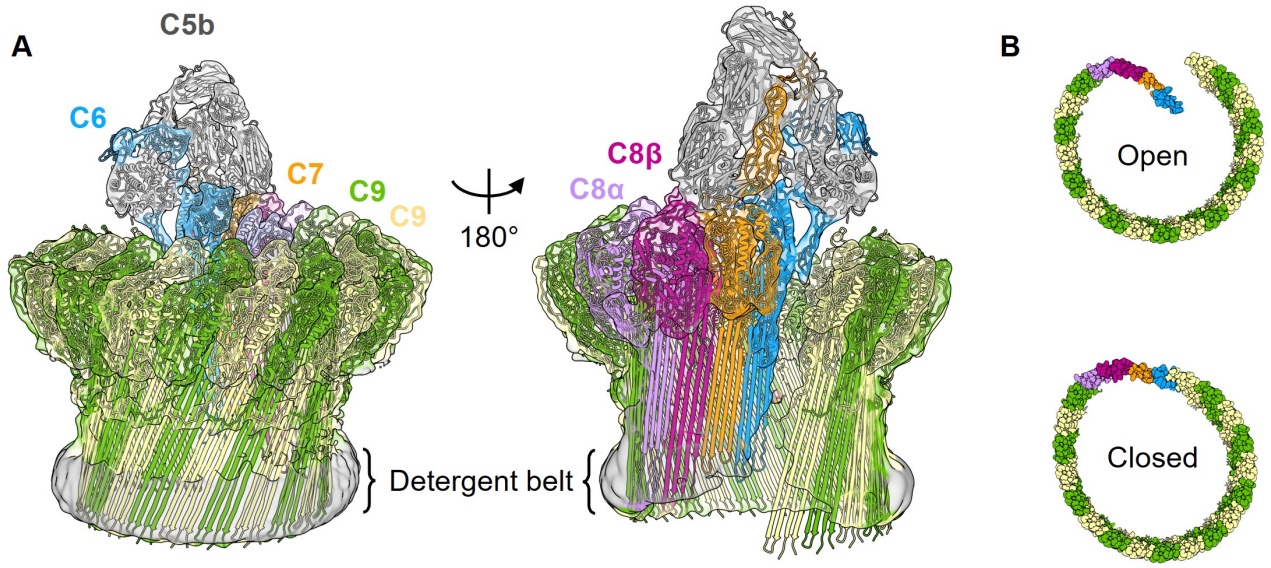


Figure 2.3 | Structure of the membrane attack complex (MAC). **A.** Open conformation of MAC, with the cryoEM map (EMDB 0106) overlayed with the model (PDB 6H03). **B.** Cross-sections of the open (top, PDB 6H03) and closed (bottom, PDB 6H04) conformations of MAC viewed from the bottom of the pore.

2.1.2 Inhibition of MAC by CD59

CD59 is a GPI-anchored protein

Complement activation is heavily regulated to avoid causing damage to healthy self-cells, and the last-resort inhibitor of MAC is a small protein called CD59, or *protectin* [144]. CD59 is an 11-kDa globular protein attached to the outer leaflet of the plasma membrane *via* a glycosylphosphatidylinositol (GPI) anchor [145]. It is made up of a central, three-stranded β-sheet faced by an α-helix, and its fold is maintained by five disulfide bonds (Figure 2.4). Structures of CD59 from an NMR ensemble suggest a high level of flexibility in the C-terminus [146]. CD59 contains an N-linked glycosylation site at Asn-18, and other potential sites for O-linked glycosylation [145], although the role of the glycans is unclear.

The C-terminal residue of CD59 (Asn-77) is covalently linked to the GPI anchor [145].

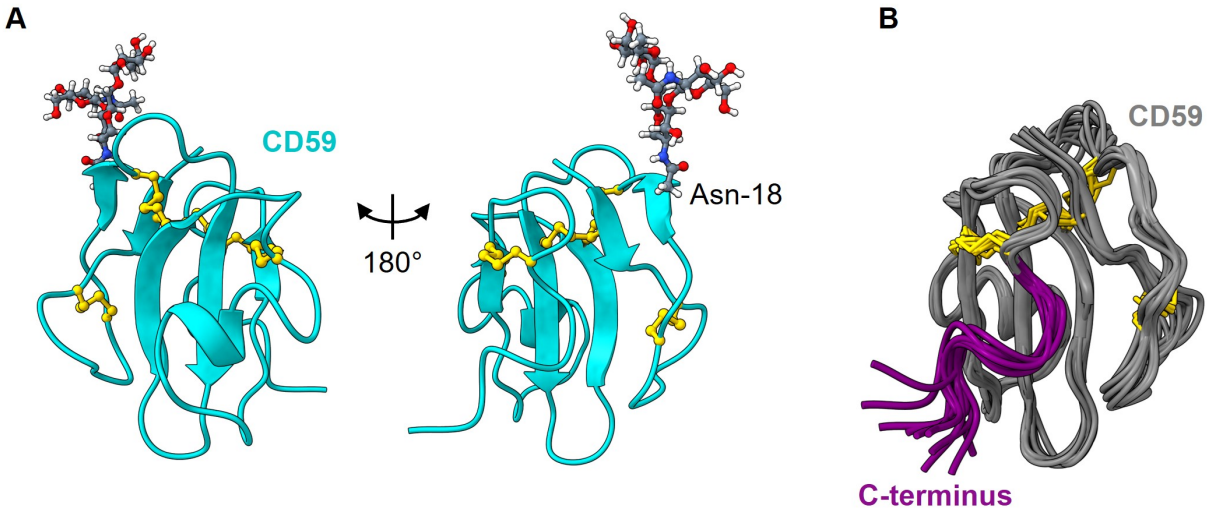


Figure 2.4 | Structure of CD59 solved by solution NMR (PDB 1CDR). **A.** CD59 (cyan) is a globular protein containing five disulfide bonds (yellow) with a glycosylation site at Asn-18. **B.** The NMR ensemble reveals a high degree of flexibility in the C-terminus (purple).

The backbone carboxylic carbon forms an amide bond with the nitrogen of a phosphoethanolamine (EtNP) linker connecting the protein to the sugar moieties of the anchor (Figure 2.5). The GPI core is made up three mannose (Man) residues, a glucosamine (GlcN) and a myoinositol (Ino), in the sequence Man-α(1→2)Man-α(1→6)Man-α(1→4)GlcN-α(1→6)Ino. The second mannose residue is decorated with another EtNP moiety and, in some instances, an N-acetylgalactosamine (GalNAc) residue. The myoinositol residue is connected to two lipid tails *via* a phosphoglycerol linker at carbon C4 and to a third tail at carbon C3.

CryoEM structures of CD59-inhibited MAC

Early binding assays involving CD59 and various MAC-derived peptides suggested that it interacted with both C8α and C9 *via* the same binding site [147, 148]. To characterise its precise mechanism of action, Dr. Emma Couves and Dr. Scott Gardner in the Bubeck lab solved multiple structures of CD59-inhibited MAC precursors [143]. Their cryoEM maps of inhibited complexes with different numbers of C9 molecules (C5b8, $C5b9_1$, $C5b9_2$ and $C5b9_3$) show that CD59 engages the leading edge of C8α with its final β-strand, positioning itself to deflect the hairpins of C9 away from the membrane (Figure 2.6). It also stabilises an

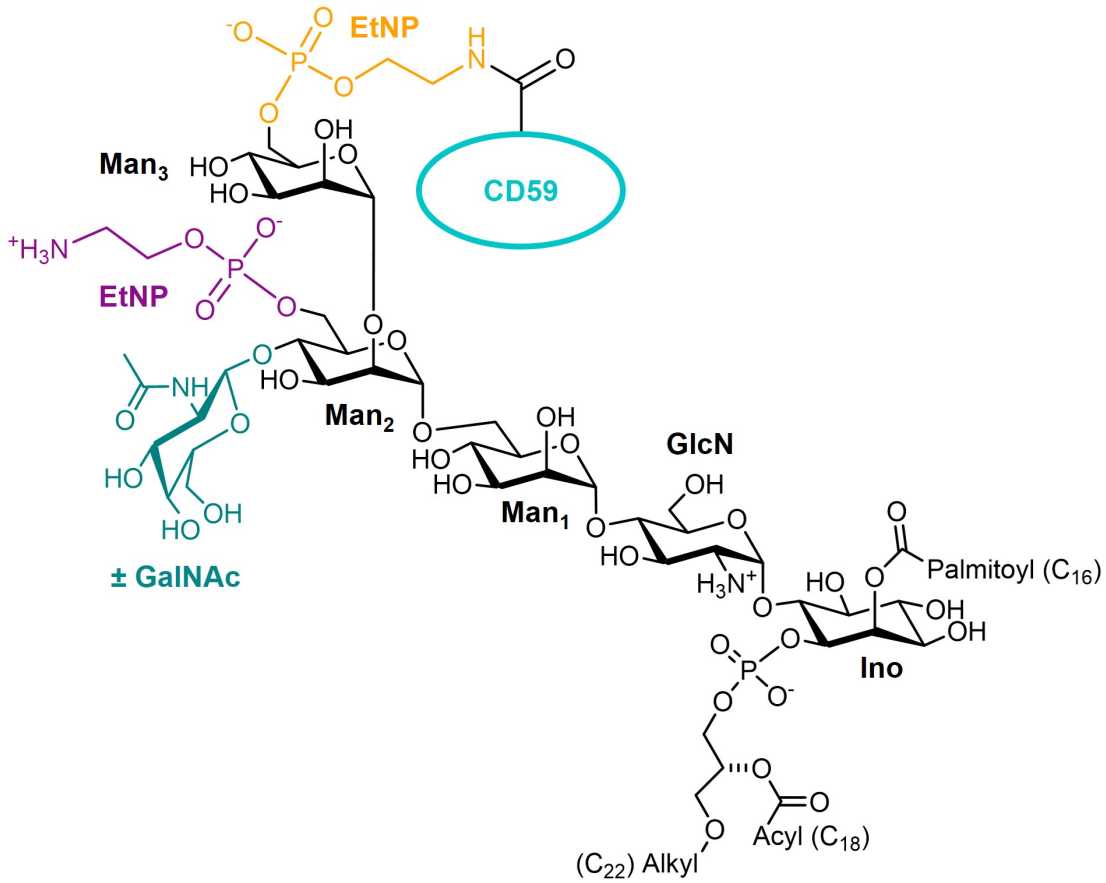


Figure 2.5 | Structure of the CD59 GPI anchor. A phosphoethanolamine (EtNP) linker (yellow) connects the protein to the anchor, while another (purple) decorates the second mannose residue. An N-acetylgalactosamine residue (turquoise) sometimes decorates the core of the anchor. Man: mannose; GlcN: glucosamine; Ino: myoinositol; GalNAc: N-acetylgalactosamine.

intermediate conformation in the final C9, preventing further oligomerisation.

CD59 binds the bacterial cholesterol-dependent cytolysins (CDCs) intermedilysin (ILY) and vaginolysin (VLY) *via* the same β-strand. While it engages C8α through similar residues, it adopts a very different binding pose for MAC inhibition and is rotated over 100° relative to the membrane (Figure 2.7) [143]. The crystal structure of CD59 fitted in the map suggests that its C-terminus, connected to the GPI anchor, is pointing away from the membrane, although the NMR structure in Figure 2.4 indicates flexibility. However, the propensity of CD59 to adopt these two orientations in the absence of binding partners is unknown.

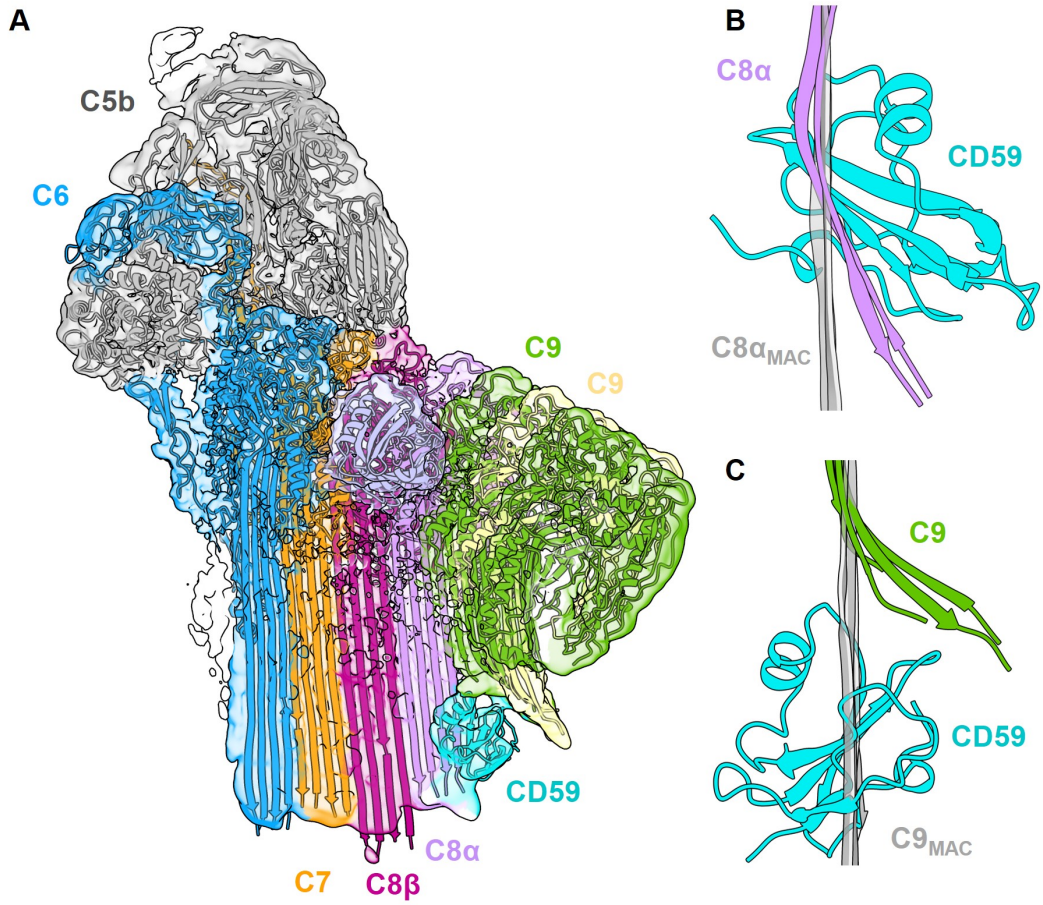


Figure 2.6 | Structural basis of MAC inhibition by CD59. **A.** Map-model overlay of the C5b9$_3$-CD59 structure solved by cryoEM (PDB 8B0G). **B.** CD59 binds to C8α and deviates its trajectory (purple, PDB 8H0F) compared to its conformation in MAC (grey, PDB 6H03). **C.** CD59 deflects the hairpins of C9 away from the membrane (green, PDB 8B0G), preventing pore formation.

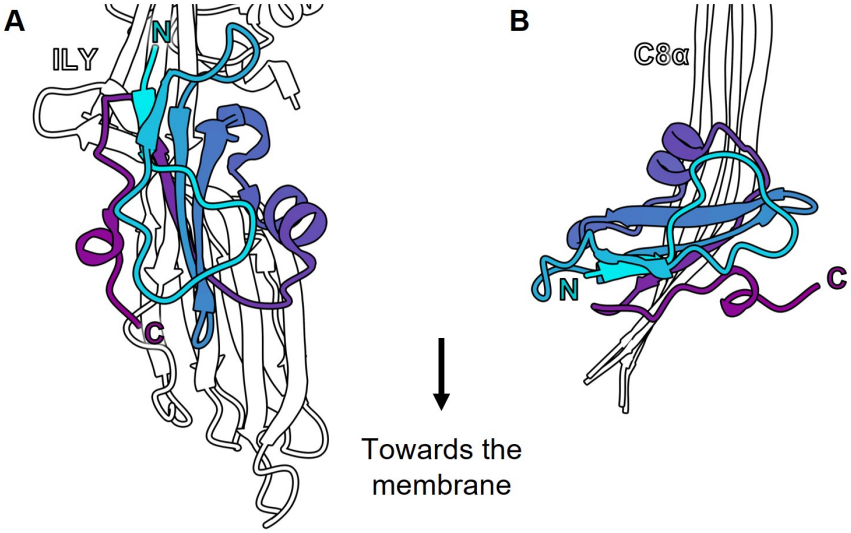


Figure 2.7 | CD59 adopts a different orientation when in complex with ILY (**A**) and C8α (**B**). CD59 is coloured from N- (cyan) to C-terminus (purple).

Effects of MAC precursors on the membrane

Various biophysical techniques have contributed to our understanding of the effects MAC precursors have on lipid membranes. Flicker spectroscopy of giant unilamellar vesicles revealed that C5b6 binding to the membrane decreases its bending modulus, i.e. it lowers the energy required for bending [38]. The addition of C7 does not affect the bending modulus but it was previously shown to tether the MAC precursor irreversibly [149]. C8 binding, on the other hand, increases the bending modulus of the membrane, making it more rigid. The C5b8 complex also forms a non-selective ion channel, unless it is captured by CD59, which then deflects the C8α hairpins (Figure 2.6) [150]. Finally, as C9 oligomerises and the pore widens, the membrane rigidifies further [38].

The cryoEM maps of CD59-inhibited MAC precursors generated by Dr. Emma Couves revealed new possible effects of MAC proteins on the membrane (Figure 2.8). She used the machine-learning algorithm CryoDRGN to resolve transient states in heterogeneous datasets [64]. The resulting reconstructions revealed density unaccounted for by the protein subunits (Figure 2.8A). Since the extra density is near the tips of the MACPF/CDC domains, it may represent ordered lipids. However, CryoDRGN-generated maps do not undergo gold standard refinements, so hypotheses derived from them should be tested with other techniques. Dr. Couves also applied a positive B-factor on the C5b9$_3$-CD59 map to visualise the shape of the nanodisc. The density shows a reduction in membrane thickness below the tip of C8β (Figure 2.8B), suggesting that MAC proteins affect the local properties of the membrane before pore formation. However, it is unclear whether this is an artefact of the spatial constraints imposed on the lipids by the nanodisc, or whether it is a physiological phenomenon.

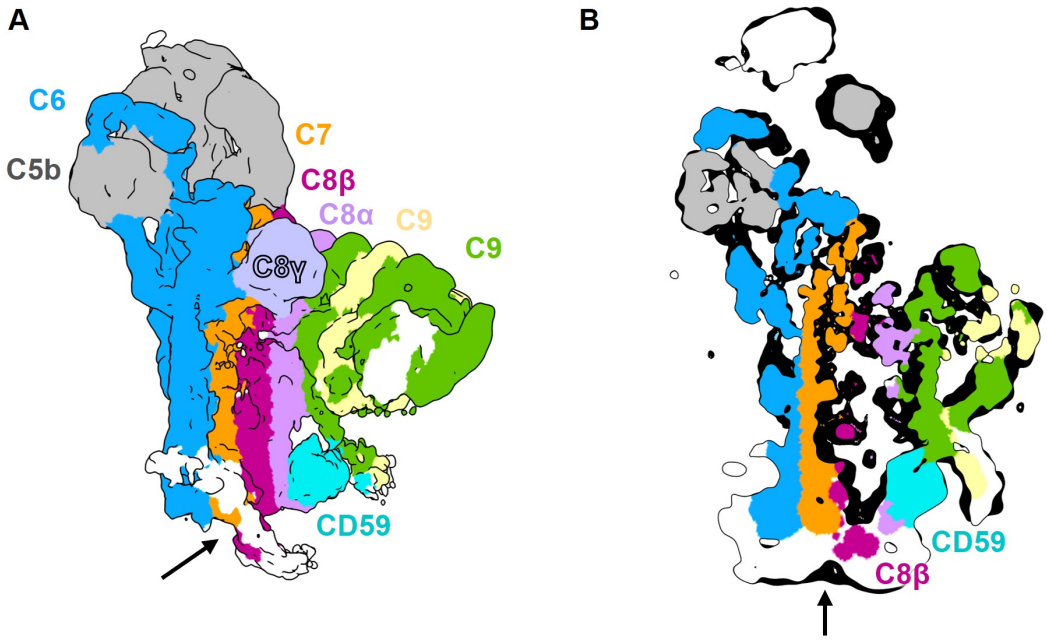


Figure 2.8 | Effects of inhibited MAC on lipid bilayers. **A.** CryoDRGN map coloured by subunit. The extra density (white) indicated by the arrow may correspond to ordered lipids. **B.** Cross-section of the $C5b9_3$-CD59 map (coloured by subunit) after applying a positive B-factor to enhance for the flexible regions. The white density at the bottom corresponds to the nanodisc. The arrow points to local thinning of the membrane.

2.1.3 MD simulations as a tool to study C5b8-CD59

MD simulations of pore-forming proteins

Molecular Dynamics (MD) simulations are a powerful tool for studying pore-forming proteins in the context of their target membranes [151]. For instance, simulations of the α-pore forming protein Cytolysin A combined with single-molecule imaging showed the importance of cholesterol binding for pore formation [152]. In the case of MACPF/CDC proteins specifically, atomistic MD simulations of the bacterial CDCs pneumolysin (PLY) and listeriolysin O (LLO) have been used to characterise protein-lipid interactions, as well as the molecular detail of pore formation [153, 154].

A major limitation to the use of MD simulations for studying pore-forming proteins is the presence of large extracellular domains that add to the computational cost without being directly involved in pore formation. To speed up calculations, coarse-grain (CG) MD simula-

tions with the Martini 2 forcefield have been used on pore-forming proteins. This approach showed that PLY and gasdermin, both β-pore forming proteins, repel lipids in different ways depending on whether they are assembled into full pores or partial arcs, highlighting the importance of studying both pores and their precursors [153, 155].

CGMD is also a powerful approach to study protein-lipid interactions for both transmembrane and peripheral membrane proteins. CGMD simulations using the Martini 2 forcefield identified ceramide binding sites on the ion channel hERG1, which were confirmed by mutagenesis [156]. Similarly, CGMD simulations of the peripheral membrane-binding *tubby* domain revealed a previously unknown binding site for phosphatidylinositol 4,5-bisphosphate that works cooperatively with the canonical binding site [157].

MD simulations of GPI-anchored proteins

Glycosylphosphatidylinositols (GPIs) are glycolipids found on the surface of many eukaryotic cells, often attached to the C-terminus of proteins [158]. All GPI anchors share a common saccharide core (as shown in Figure 2.5 for CD59) connected to its respective protein *via* a phosphoethanolamine (EtNP) linker. Studying the structure of GPI anchors using experimental methods has been challenging due to the flexibility of the glycosydic linkages and the need for a hydrophobic environment around the lipid tails. MD simulations have been used to characterise the conformational flexibility of GPI-anchored proteins, revealing that the anchor exchanges between two main conformational states [159, 160]. In its extended conformation (termed the *lollipop* conformation), the sugars point away from the membrane whereas in the *flop-down* conformation, the sugar moieties interact with the lipid headgroups, bringing the protein closer to the membrane (Figure 2.9). Recent atomistic simulations of CD59 and structurally related proteins have suggested that it preferentially adopts the *flop-down* conformation [160]. However, the link between the orientation of CD59 relative to the

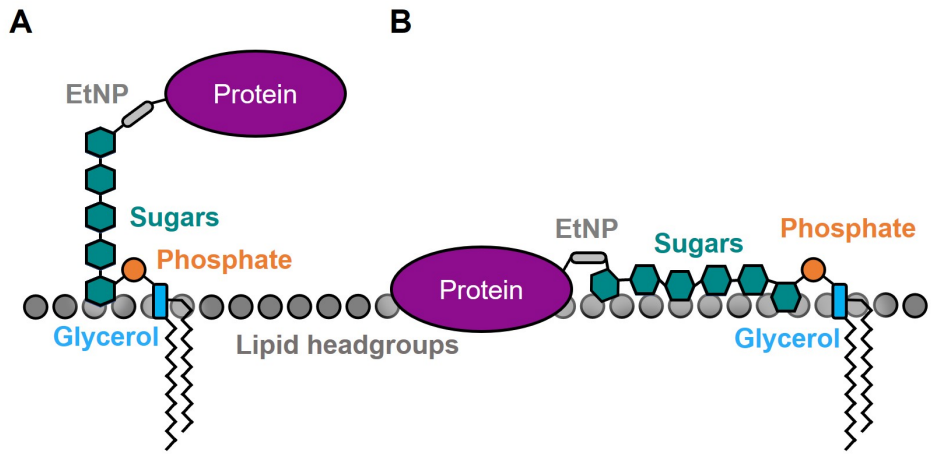


Figure 2.9 | GPI anchors exchange between extended, *lollipop* conformations **(A)** and *flop-down* conformations **(B)**, where the protein is brought close to the membrane.

membrane and its ability to bind C8α and ILY remains unknown.

One of the long-standing challenges of simulating GPI-anchored proteins has been the lack of accurate CG representations of GPIs. Banerjee *et al.* recently developed a CG GPI model based on the Martini 2 forcefield [161]. The authors parametrised the sugar moieties and lipid tails based on atomistic simulations using the GLYCAM06h [162] and Lipid14 [163] forcefields. The result is a CG molecule with 15 beads corresponding to the sugar residues, one bead for the phosphate group, and eight beads describing two lipid tails. This model is not specific for CD59: it lacks the EtNP and GalNAc decorations on Man_2 and the third lipid tail on carbon C3 of the Ino residue (Figure 2.5). Moreover, the two lipid tails are represented by three apolar beads each, which would be more suited to lipids with a C_{12} or C_{14} chain than to the longer chains found in the CD59 GPI anchor. While the overall model is a simplified version of a GPI core, it remains a useful tool for simulating GPI-anchored proteins.

2.1.4 Aims

The work presented in this chapter uses MD simulations to expand on the structural information about CD59 inhibition of MAC and to explore the hypotheses derived from the cryoEM reconstructions. To assess whether CD59 readily exchanges between its CDC-bound and

MAC-bound orientations, I performed atomistic simulations of GPI-anchored CD59 in a membrane environment. To determine how inhibited MAC affects its local lipid environment, I simulated a CG model of C5b8-CD59, which first required the parametrisation of the EtNP linker between CD59 and its GPI anchor.

2.2 Computational Approach

2.2.1 Atomistic simulations of GPI-anchored CD59

The NMR structure of CD59 (PDB 1CDR) [146] was uploaded to the Membrane Builder utility of the CHARMM-GUI server [164, 165]. The final Asn-77 residue was linked to the GPI anchor core with a phosphoethanolamine (EtNP) moiety and an extra EtNP modification was added to the second mannose residue. The CHARMM-GUI server could not model the bond between carbon C3 of the inositol residue and the third lipid tail, so a simplified GPI anchor was built with two acyl tails connected to the phosphoglycerol, each with 17 aliphatic carbons. The N-linked glycan at Asn-18 was included as an N-acetylglucosamine (GlcNAc) residue decorated with a second GlcNAc and a fucose (Fuc) residue; the IUPAC nomenclature of the full glycan is GlcNAc(1→4)[Fuc(1→6)]GlcNAc(1→)Asn-18. The protein was placed in a DOPC bilayer and solvated in TIP3P water [166] including 150 mM NaCl (with two extra Na^+ ions for neutralisation), within a 12 x 12 x 13.8 nm^3 box.

Three independent replicate simulations were performed in Gromacs 2021.3 [79] using the CHARMM36m forcefield [167]. In each case, the system was minimised with the steepest descent algorithm to a maximum force of 1000 kJ/mol or until convergence. The output from minimisation was then equilibrated for 250 ps in an NVT ensemble at 310 K (1 fs timestep) with position restraints on the protein and lipids, before a further 1 ns equilibration run in NPT ensemble at 310 K (2 fs timestep) with backbone restraints only. 500 ns

production runs were performed at 310 K with a Nosé-Hoover thermostat (τ_T = 1.0 ps) and a semi-isotropic Parrinello-Rahman barostat maintaining a pressure of 1 bar (τ_P = 5.0 ps). The LINCS algorithm [168] was used to fix bonds to hydrogen atoms.

The trajectories were visualised in VMD [169] and selected frames were exported as PDB files for figure generation in ChimeraX [170]. The angle between the *z*-axis of the box and the vector defined by the backbone nitrogens of Leu-1 and Asn-77 was measured with the *gangle* tool in Gromacs. The angle between the *z*-axis and the vector formed by C6 of Ino and C4 of Man_3 was similarly analysed. The root mean squared fluctuation (RMSF) of the protein was calculated over the three trajectories with the MDAnalysis.analysis.rms module of MDAnalysis [171, 172, 173, 174], and was compared to the root mean squared deviation (RMSD) of the NMR ensemble [146] calculated in Chimera [175]. Potential hydrogen bond pairs between CD59 and DOPC were identified and the average number of hydrogen bonds established by each residue over the course of the three replicates was calculated using the HydrogenBondAnalysis module of MDAnalysis [173, 174]. To generate an Euler plot of CD59 orientations relative to the membrane, each frame was exported to a PDB file, which was aligned to the same reference configuration (the output of the equilibration run for the first replicate) using the *align* command in ChimeraX [170]. The resulting rotation matrices were used to calculate the Euler angles θ and ψ of rotation (about the *y*- and *x*-axes, respectively). Values for the ϕ angle were not taken into account as the *z*-axis was considered to be normal to the plane of the membrane. The Euler angles were plotted using seaborn [176] and Matplotlib [177] in Python.

2.2.2 Parametrisation of the CD59-GPI EtNP linker

To model coarse-grain GPI-anchored CD59, the EtNP linker between the protein and the anchor was parametrised following the procedure suggested by Banerjee *et al.* [161]. A

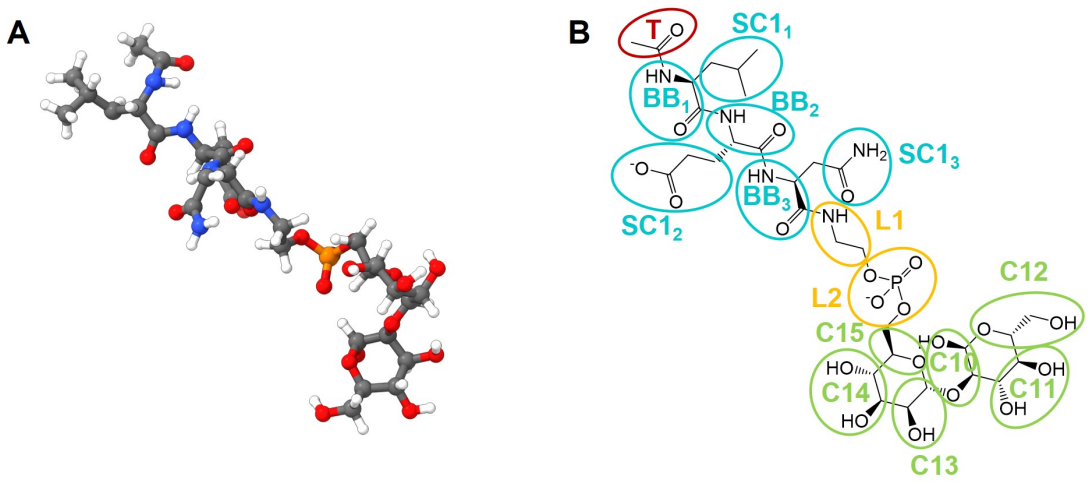


Figure 2.10 | Parametrisation of the EtNP linker. **A.** Atomistic molecule used for the parametrisation. **B.** Coarse-grain mapping of the molecule in panel A. This figure was used for publication in [143] under a CCBY 4.0 license.

molecule was built with the LEaP facility in AmberTools21 [66] comprising the three C-terminal residues of CD59 (with an N-terminal acetyl cap) bonded to the EtNP linker, itself linked to the first two mannose moieties of the anchor (Figure 2.10). The molecule was placed in a box with 2,466 TIP3P water molecules and two neutralising Na^+ ions. The system was converted for use in Gromacs 2021.3 with ACPYPE [178] and all simulations were run with the GLYCAM06h forcefield [162].

Independent triplicates were run to sample the behaviour of the atomistic molecule. For each replicate, energy minimisation was performed using the steepest-descent algorithm until a maximum force of 10 kJ/mol, followed by an NPT equilibration for 100 ps and a 200 ns run, both using the sd integrator [179].

A CG version of the same molecule was built using the mapping proposed in Banerjee *et al.*, adapted for the CD59 residues (Figure 2.10). The EtNP linker was modelled as two beads (L1 and L2), the CD59 residues were modelled according to the Martini 2.2 forcefield parameters for proteins, while the acetyl cap and the sugar parameters were taken from Banerjee *et al.* The CG molecule was hydrated with polarisable water [96] and two neutralising Na^+ ions in a 5 x 5 x 5 nm^3 box. A modified version of the Martini 2.2 forcefield

where the well depth of the Lennard-Jones potential for sugar-sugar interactions was scaled down to avoid aggregation was used in the original study [161] and the same forcefield was used for the work in this chapter. The system was minimised for 1,000,000 steps using the steepest-descent algorithm before a 500 ps NPT equilibration and a 200 ns run with the sd integrator. The bonded and non-bonded parameters were iteratively adjusted to match the behaviour of the atomistic molecule.

2.2.3 Coarse-grain simulations of C5b8-CD59

In the cryoEM structures of inhibited MAC, C9 does not interact with the membrane, so the C5b8-CD59 structure solved by Dr. Scott Gardner was chosen for CG simulations. Dr. Emma Couves prepared the structure for MD simulations by splitting C5b into two chains (accounting for the chain break resulting from C5a cleavage) and building the missing residues from the open MAC structure (PDB 6H03) [38]. The transmembrane hairpins of C8α were connected by triple-glycine linkers and all glycans were removed.

I converted the structure generated by Dr. Emma Couves to a CG model using the *martinize.py* script with the ElNeDyn framework [92] for Martini 2.2 with an elastic bond force constant of 500 kJ/mol/nm^2 and upper and lower elastic bond cut-offs of 1.0 nm and 0.5 nm, respectively. The GPI anchor was placed in close proximity of the CD59 C-terminus in Chimera and the topology (itp) file was modified to include the GPI anchor parameters from Banerjee *et al.* [161] and the linker parameters generated as detailed above (cf. Section 2.3.4). The complex was placed in a 7:3 (molar ratio) DOPC:cholesterol membrane in a 22.5 x 22.5 x 30 nm^3 box using the *insane.py* script [180] with polarisable water [96] and 150 mM of neutralising NaCl. Independent triplicates were run in Gromacs 2021.3. The system was minimised with the steepest descent algorithm to a maximum force of 100 kJ/mol or until convergence. The output was equilibrated for 2 ns (1 fs timestep) in an NPT ensemble

with a pressure of 1 bar maintained by a semi-istropic Berendsen barostat (τ_P = 5.0 ps). The equilibrated system was used for a 2 µs production run (5 fs timestep) in an NPT ensemble maintained at a pressure of 1 bar by a semi-isotropic Parinello-Rahman barostat (τ_P = 12.0 ps). For both the equilibration and production runs, the leapfrog stochastic dynamics (sd) integrator was used to match the original study [161] and the temperature was kept at 310 K by a V-rescale thermostat (τ_T = 1.0 ps). The small timesteps used were necessary to avoid numerical instability, and the Lennard-Jones potentials were scaled down as described above and in [161].

2.3 Results

2.3.1 CD59 samples orientations allowing it to bind C8β and ILY

To test whether CD59 samples the ILY-bound and C8α-bound orientations in the absence of either binding partner, I ran atomistic MD simulations of CD59 attached to a simplified GPI anchor, in a DOPC membrane. Aligning the respective complex structures (ILY-CD59: PDB 5IMT [181]; C5b8-CD59: PDB 8H0F [143]) reveals that CD59 samples both orientations in the simulations (Figure 2.11). In the frame most similar to the C8α-binding pose, the C-terminus of CD59 is pointing down towards the membrane, in the opposite direction to the crystal structure fitted in the cryoEM density.

To assess the flexibility of CD59, the angle between the *z*-axis of the box (normal to the membrane) and a vector connecting the C- and N-termini was measured over the course of the three trajectories (Figure 2.12). This suggests that CD59 samples a large orientation space, with the angle between the two vectors adopting values between 55° and 180°.

To determine whether this level of movement arises from flexibility inherent to the protein moiety of CD59, the root mean squared fluctuation (RMSF) of each residue was calcu-

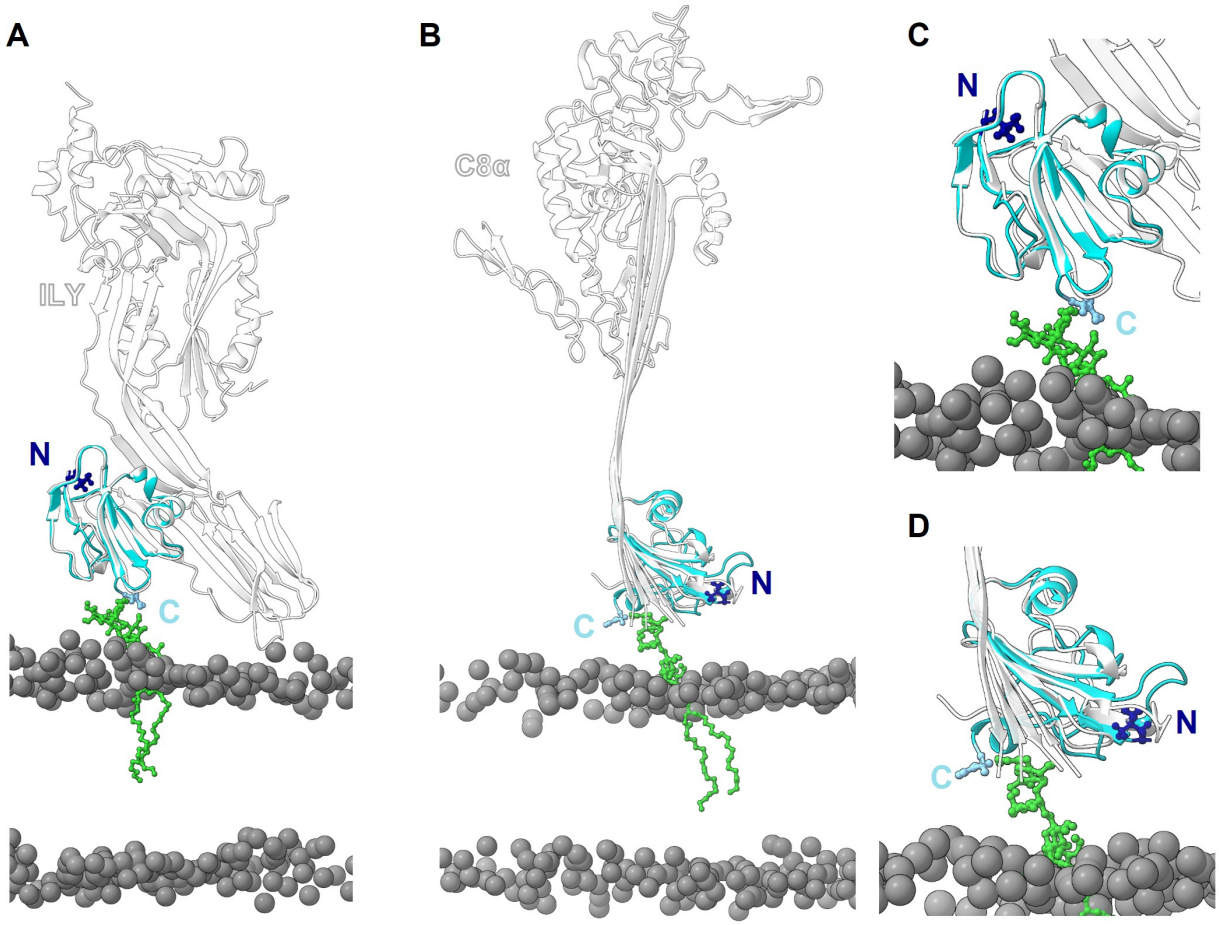


Figure 2.11 | GPI-anchored CD59 samples the ILY-bound and C8α-bound orientations relative to the membrane. **A.** ILY-CD59 structure (white, PDB 5IMT) superimposed with a frame from the GPI-anchored CD59 simulations. **B.** C8α-CD59 structure (white, PDB 8H0F) superimposed with a frame from the GPI-anchored CD59 simulations. **C.** Close-up of the CD59 alignment in panel A. **D.** Close-up of the CD59 alignment in panel B. Grey: lipid headgroups; green: GPI anchor; cyan: CD59 (atomistic simulation); dark blue: N-terminal residue of CD59; light blue: C-terminal residue of CD59; white: reference structure.

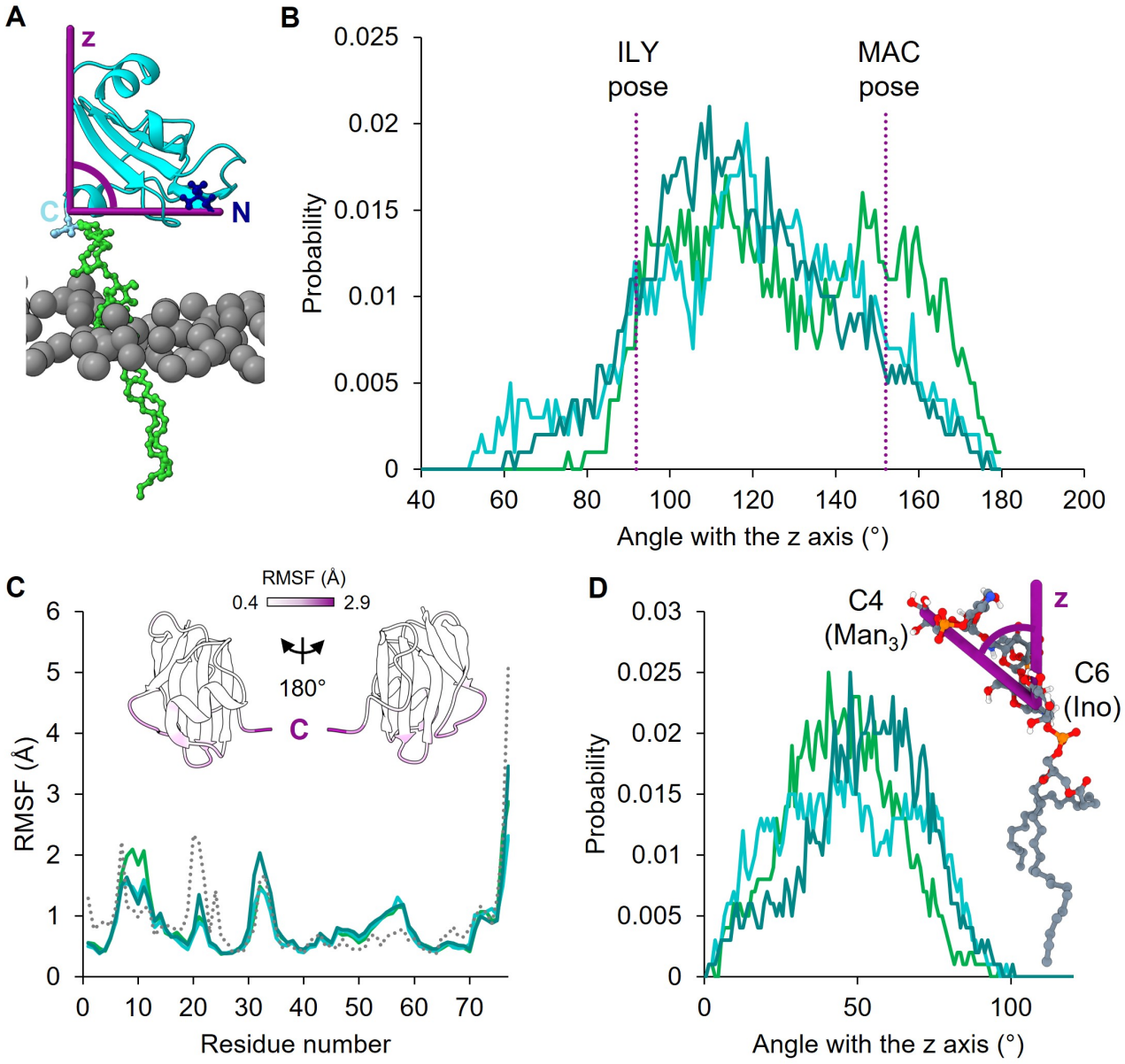


Figure 2.12 | CD59 flexibility in atomistic MD simulations. **A.** Representation of the angle measured between the z-axis and a vector connecting the C- and N-termini (passing through their backbone nitrogen atoms). Lipids headgroups are in grey. **B.** Distribution of the angle shown in panel A. The values of the angle corresponding to the ILY-bound and MAC-bound orientations are shown by the purple dotted lines. **C.** Per-residue RMSF of CD59 in the atomistic simulations (shades of green). The RMSD derived from the NMR ensemble [146] is shown in dotted grey for comparison. Inset: CD59 coloured by RMSF averaged over the independent triplicates. The C-terminus is indicated by the letter C. **D.** Distribution of the angle between the z-axis and the vector connecting the first and last sugar residues of the GPI anchor. Inset: representation of the angle measured (purple). The different shades of green in panels B, C and D correspond to independent triplicates.

lated over the course of the simulations, and compared to the root mean squared deviation (RMSD) calculated from the NMR ensemble [146] (Figure 2.12C). Both the MD and the NMR data highlight regions of flexibility in loops connecting rigid secondary structure elements of the protein, but the highest level of flexibility is found in the C-terminus of CD59. To assess the flexibility of the anchor, the angle between the *z*-axis and the vector passing through carbons C6 of Ino and C4 of Man_3 was monitored over the course of the simulations (Figure 2.12D). This shows a large extent of flexibility, suggesting that the sugar moieties of the GPI anchor act as a hinge, allowing CD59 to sample a wide variety of orientations.

2.3.2 Interactions between CD59 and the membrane

To determine whether the protein moiety of GPI-anchored CD59 interacts with the membrane, potential hydrogen (H-) bonds with DOPC were analysed using the HydrogenBondAnalysis module of MDAnalysis [173, 174]. As DOPC does not contain any hydrogen atom covalently bonded to an electronegative atom (Figure 2.13C), it can only act as an acceptor, through its head group nitrogen, phosphate oxygens or carbonyl oxygens in the fatty acid tail.

Plotting the average number of H-bonds established by each residue over the course of the three replicates highlights a cluster of residues located around the edge β-strand of CD59: Tyr-61, Tyr-62, Lys-65 and Lys-66 (Figure 2.13). Those residues are involved in both C8β and ILY binding [143, 182], suggesting that, when they form H-bonds with the membrane, CD59 may not be available to engage either protein.

2.3.3 Orientation sampling of GPI-anchored CD59

GPI-anchored CD59 shows a high level of flexibility due to its C-terminus and the sugar moieties of the anchor (cf. Figure 2.12). To visualise how this translates to orientations of

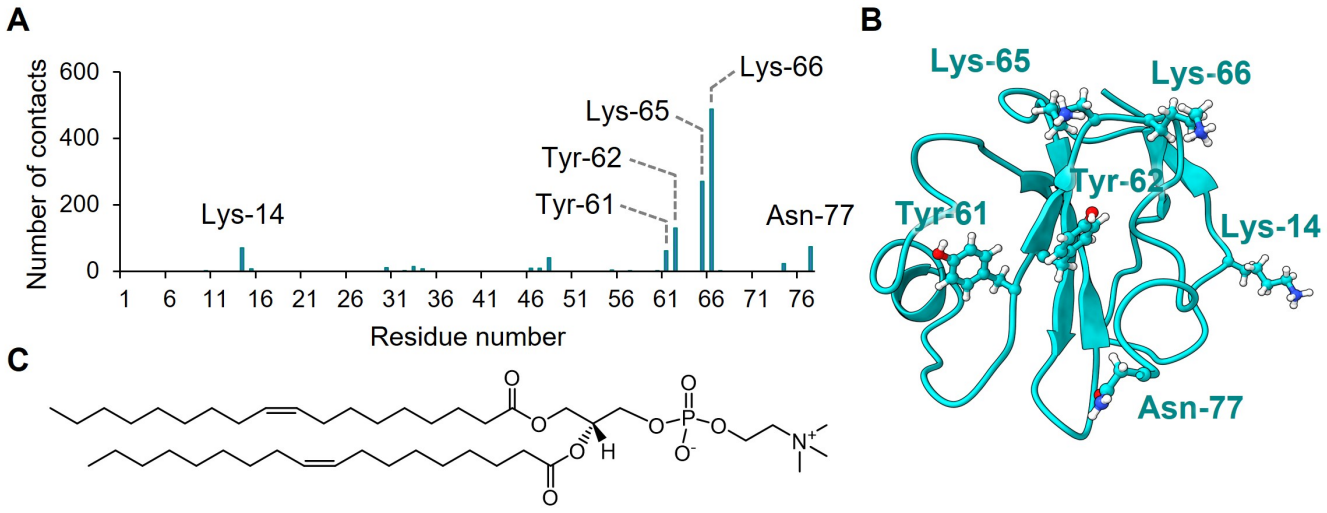


Figure 2.13 | CD59 residues involved in hydrogen bonds with DOPC. **A.** Average number of hydrogen bonds established with DOPC by each CD59 residue over the course of three independent simulations. **B.** Structure of CD59 (PDB 1CDR) with residues highlighted in panel A shown as sticks. **C.** Chemical structure of DOPC.

CD59 relative to the membrane, each frame was aligned to the same reference (the last frame from the equilibration run of the first replicate) in ChimeraX [170] and the rotation matrix was extracted. A rotation matrix of the form shown in Equation 2.1 can be used to derive Euler angles describing the rotation around each of the three axes of the box.

$$R = \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{bmatrix} \tag{2.1}$$

The angles of rotation θ (about the *y*-axis) and ψ (about the *x*-axis) can be calculated with Equations 2.2 and 2.3 as long as $\cos(\theta) \neq 0$. There are two solutions for θ, and thus two solutions for ψ but as they are simply shifted by π, only one solution is needed for an angular orientation plot. As the *z*-axis is normal to the plane of the membrane, rotations around it do not represent different orientations relative to the membrane and the corresponding rotation angle ϕ was not used.

$$\theta = -\sin^{-1}(R_{31}) \tag{2.2}$$

$$\psi = \text{atan2}\left(\frac{R_{32}}{\cos(\theta)}, \frac{R_{33}}{\cos(\theta)}\right) \tag{2.3}$$

The rotation angles for each frame (compared to the same reference) were plotted in Figure 2.14A. GPI-anchored CD59 samples a wide variety of orientations relative to the membrane, although defined clusters suggest that some orientations may be favoured.

Frames corresponding to the numbered points in Figure 2.14A were visualised in ChimeraX. To test whether each orientation is compatible with C8α or ILY binding, the structures of CD59-bound C8α (PDB 8H0F) [143] and CD59-bound ILY (PDB 5IMT) [181] were aligned to the protein moiety from each selected frame (Figure 2.14B). One of the orientation clusters corresponds to a pose in which the final β-strand of CD59 is positioned such that it can engage the hairpin of C8α as it unfurls towards the membrane. Various different orientations of CD59 can accommodate monomeric ILY, as it has more degrees of freedom than C8α, which is held in place by the rest of the MAC proteins. One of the clusters corresponds to an orientation in which the last strand of CD59 is not available for binding either protein, but instead lies flat on the membrane. Hydrogen bond analysis in ChimeraX for a frame from that cluster revealed that Tyr-61, Tyr-62 and Lys-66 are involved in H-bonds with the DOPC phosphates, consistent with the results in Figure 2.13.

2.3.4 Parametrising the CD59-GPI EtNP linker

Coarse-grain (CG) simulations of C5b8-CD59 required a CG model of GPI-anchored CD59. Banerjee *et al.* developed the first GPI anchor representation compatible with the Martini 2.2 forcefield [161]. The protein moiety of CD59 did not require parametrisation as the Martini 2.2 forcefield is parametrised for all standard amino acids [91] and the parameters for the sugars and lipid tails were taken from the original study [161]. The phosphoethanolamine (EtNP) linker between the sugar residues and the protein, however, was parametrised specif-

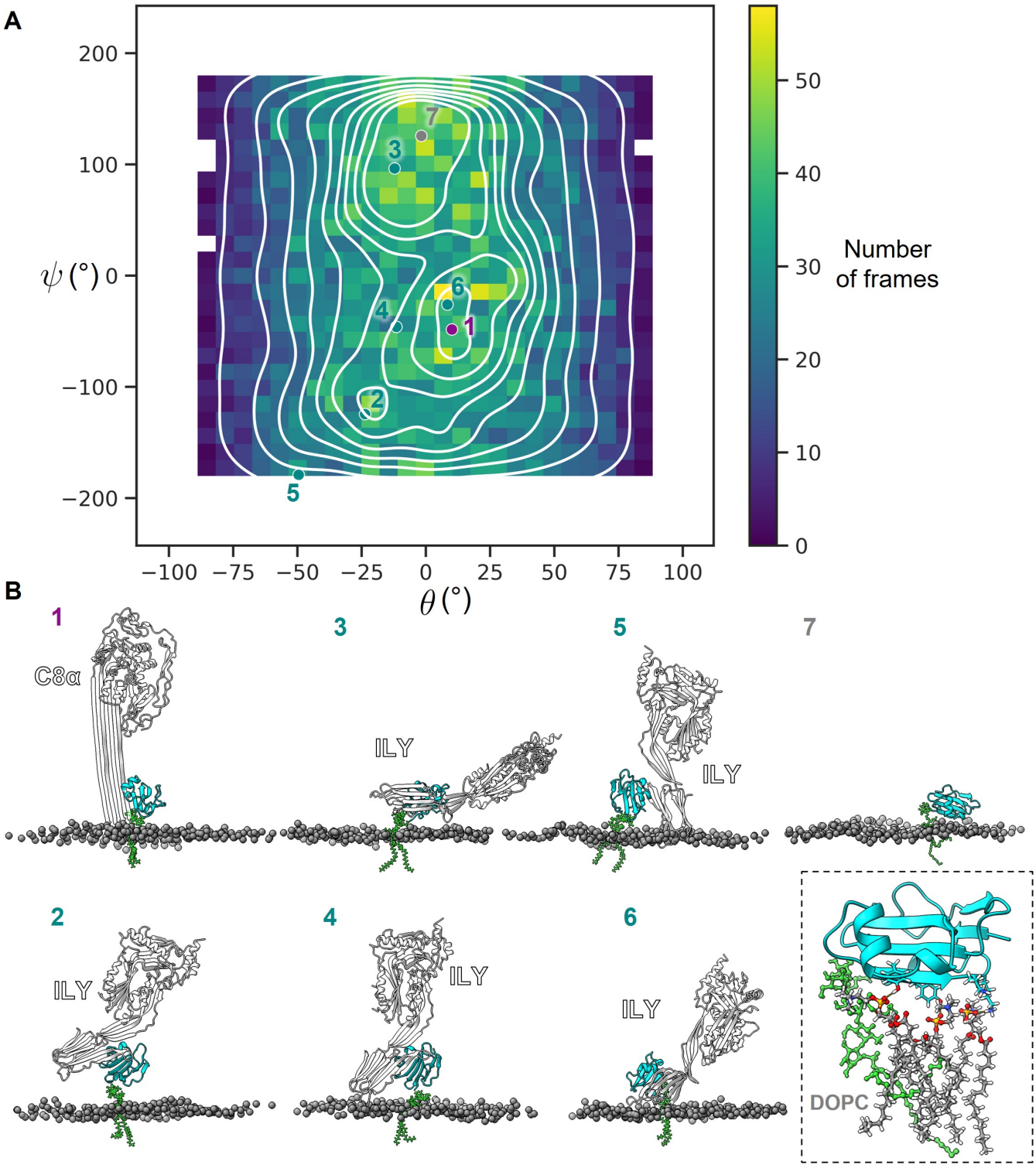


Figure 2.14 | Orientations of CD59 relative to the membrane. **A.** Angular distribution plot of CD59 θ and ψ rotation angles (around the *y*- and *x*-axes, respectively) from all trajectories compared to the final step of equilibration for the first replicate. The lines represent the contours of a kernel density estimation performed by seaborn [176]. Selected orientations are shown by numbered points. Purple: orientation compatible with C8α binding; turquoise: orientations compatible with ILY binding; grey: orientation compatible with H-bond formation with the membrane. **B.** Selected CD59 orientations from panel A. 1: The C8α-CD59 complex was aligned to the frame and C8α is shown in white. 2-7: The ILY-CD59 complex was aligned to the frame and ILY is shown in white. 8: CD59 cannot accommodate C8α or ILY, but instead forms hydrogen bonds with the membrane. Inset: Close-up view of CD59 with the DOPC molecules engaging in hydrogen bonds with the protein shown as sticks.

ically for CD59 following the strategy suggested by Banerjee *et al*. This was achieved by parametrising a molecule containing the last three residues of CD59, the EtNP linker, and the first two sugar residues of the anchor (Figure 2.10). The parameters derived for the linker molecule are given in Table 2.1 and the comparison between the atomistic and CG bonds, angles and dihedrals involving the EtNP beads (L1 and L2) is shown in Figure 2.15.

Table 2.1: Coarse-grain parameters of the molecule in Figure 2.10. Beads L1 and L2 represent the EtNP linker itself.

Bead name	Bead type	Bond	r_0 (nm)	K_b (kJ/mol)	Angle	θ_0 (°)	K_a (kJ/mol)
T	SN0	T — BB_1	0.3	20000	T — BB1 — BB2	128	20.0
BB_1	P5	BB_1 — $SC1_1$	0.33	7500	T — BB_1 — $SC1_1$	108	30.0
$SC1_1$	C1	BB_1 — BB_2	0.39	20000	BB_1 — BB_2 — BB_3	132	30.0
BB_2	P5	BB_2 — $SC1_2$	0.4	5000	$SC1_2$ — BB_2 — BB_3	108	30.0
$SC1_2$	Qa	BB_2 — BB_3	0.35	20000	BB_2 — BB_3 — $SC1_3$	96	80.0
BB_3	P5	BB_3 — $SC1_3$	0.32	5000	BB_2 — BB_3 — L1	105	2.5
$SC1_3$	P5	BB_3 — L1	0.33	10000	$SC1_3$ — BB_3 — L1	90	1.0
L1	GSNda	L1 — L2	0.35	5000	BB_3 — L1 — L2	130	60.0
L2	GQa	L2 — C15	0.365	18000	L1 — L2 — C15	45	5.0
C15	GNa	C15 — C14	0.33	20000	L2 — C15 — C13	180	15.0
C14	GP3	C15 — C13	0.35	20000	L2 — C15 — C14	70	20.0
C13	GSP1	C14 — C13	0.28	40000	C15 — C13 — C10	65	45.0
C12	GP2	C13 — C10	0.36	20000	C14 — C13 — C10	95	100.0
C11	GP3	C12 — C11	0.33	30000			
C10	GSN0	C12 — C10	0.35	30000			
		C11 — C10	0.28	40000			

Dihedral	Type	Φ_0 (°)	K_d (kJ/mol)
BB_2 — BB_3 — L1 — L2	Proper	220	1.5
BB_3 — L1 — L2 — C15	Proper	-195	4.0
L1 — L2 — C15 — C13	Proper	320	3.5
L2 — C15 — C13 — C10	Proper	30	3.0
BB_1 — BB_2 — BB_3 — L1	Improper	-200	3.0

2.3.5 Building C5b8-CD59 and placing it in the membrane

To model C5b8-CD59, the atomistic structure of the complex was prepared by Dr. Emma Couves. I used the *martinize.py* script with the ElNeDyn framework for Martini 2.2 to create a CG representation of the protein subunits [92]. Adding the beads for the GPI anchor and modifying the itp file accordingly yielded the full structure used for simulations.

The Memembed script [183] was initially used to predict the position of atomistic C5b8-CD59 in the membrane. C5b8-CD59 alone was not predicted to stand upright above the membrane, so the full open conformation of MAC (PDB 6H03 [38]) was used and C5b8-

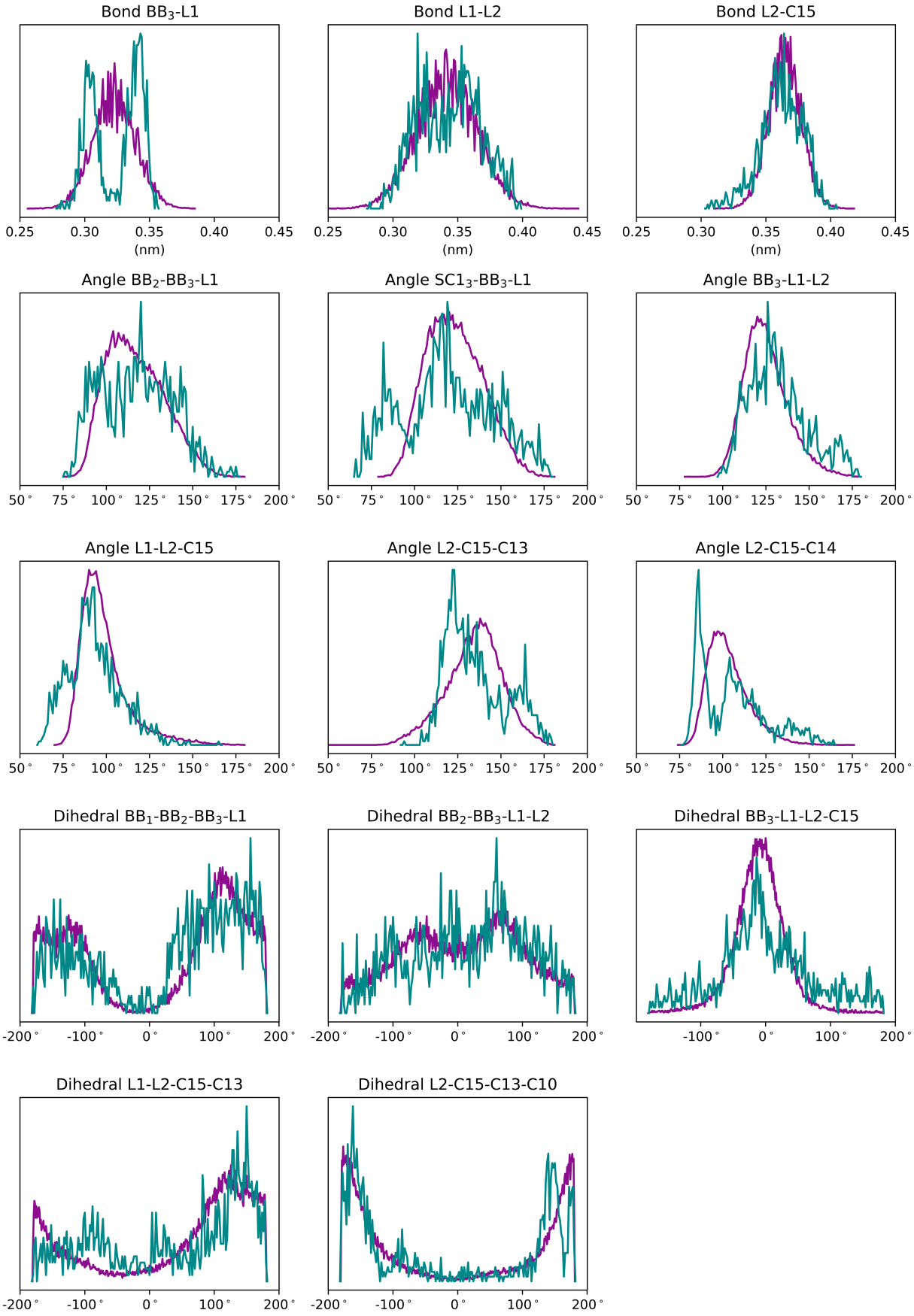


Figure 2.15 | Comparison of atomistic (turquoise) and coarse-grain (purple) profiles of bonds, angles and dihedrals for the molecule in Figure 2.10. Only parameters involving the EtNP beads (L1 and L2) are shown.

CD59 was aligned to the predicted coordinates. However, within only 100 ns, the complex moved upwards relative to the membrane and C8β was pushed out of the bilayer, with its hydrophobic face resting on top of the lipids (Figure 2.16A). The same result was obtained by using the Positioning of Proteins in Membrane (PPM) 2.0 orientation predictor [184].

To ensure C5b8 remained associated with the membrane, an arbitrarily low starting position was chosen, where C8β spanned the membrane entirely. Even after 1 µs, the complex was stably associated with the membrane (Figure 2.16B). Tracking the position of the C8α tip (*via* Phe-328) shows that, in all three replicates, the complex moves up relative to the membrane for the first 200 ns of the simulations, after which it remains relatively stable (Figure 2.16C), suggesting that it has reached its equilibrium position. The simulations were then run for a total of 2 ns and only the last 1.5 ns were used for analysis.

2.3.6 Lipid binding

To detect potential protein-lipid interactions, the trajectories were analysed with the python toolkit PyLipID [185]. The programme uses a distance-based approach: a protein-lipid interaction starts when a lipid molecule moves closer to a protein residue than a defined lower cut-off, and it ends when the lipid moves further than a defined upper cut-off. The use of a dual cut-off compensates for the 'rattling in a cage' effect whereby lipids can rapidly change protein contacts without fully dissociating, resulting in sudden jumps in the minimum distance between the two interacting molecules [185]. By plotting the occupancy of each protein residue (i.e. the proportion of the time it is within interacting distance of a lipid molecule) and its residence time (i.e. how long a specific interaction lasts on average), potential lipid binding sites can be detected (Figure 2.17).

The PyLipID analysis revealed that most of the longer-lived protein-lipid interactions are between the complex and DOPC: seven residues show a DOPC residence time of over

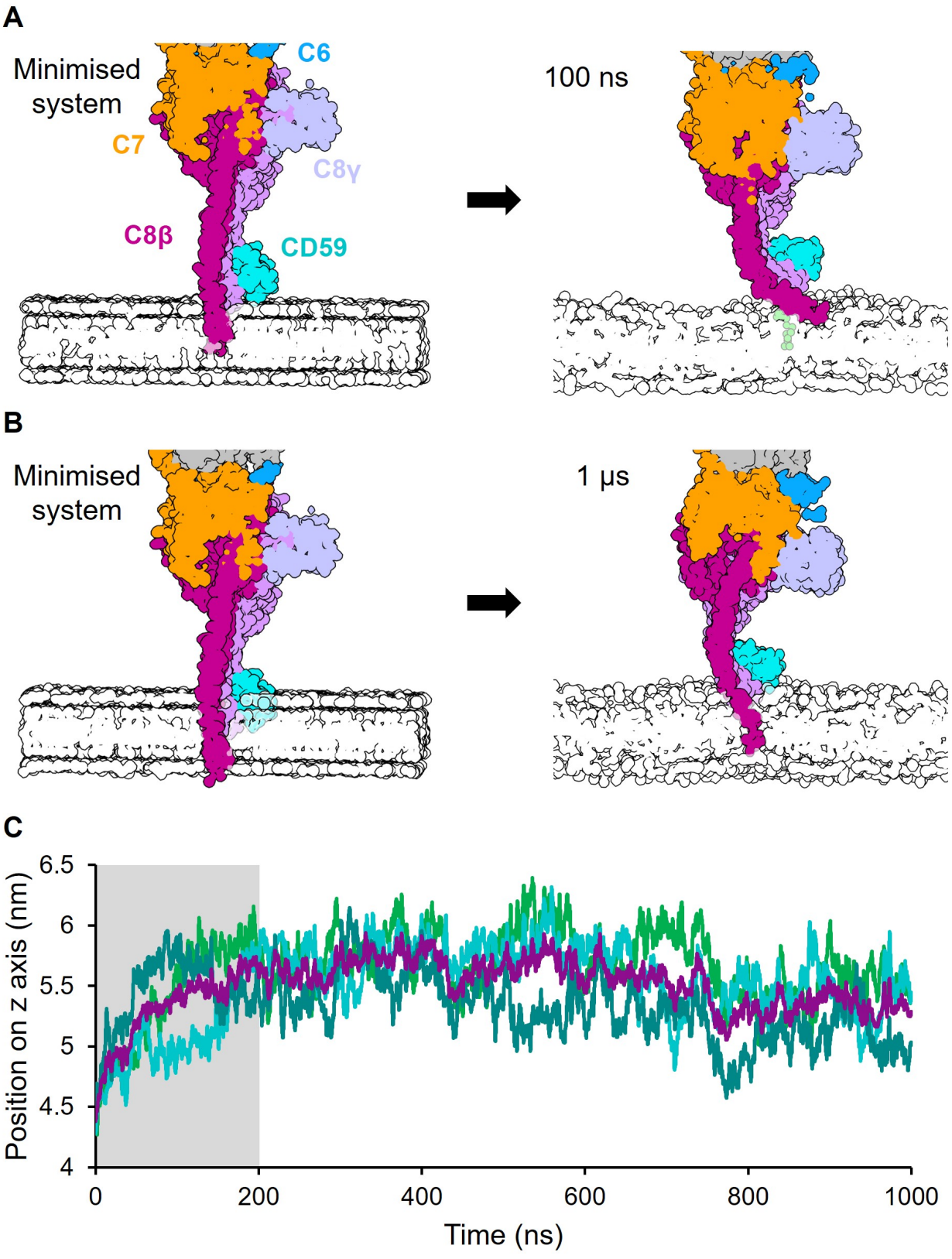


Figure 2.16 | Placing the C5b8-CD59 complex in the membrane. **A.** The starting position predicted by Memembed [183] (left) quickly results in C8β being pushed out of the membrane (right, after 100 ns). **B.** Placing the complex arbitrarily low in the membrane (left) leads to C8β remaining in the membrane even after 1 µs (right). **C.** Tracking the position along the z axis of Phe-328 on the tip of C8α shows that the complex moves to its equilibrium position within 200 ns and remains stable over the course of the simulation. Each replicate is shown in a shade of green and the average is in purple.

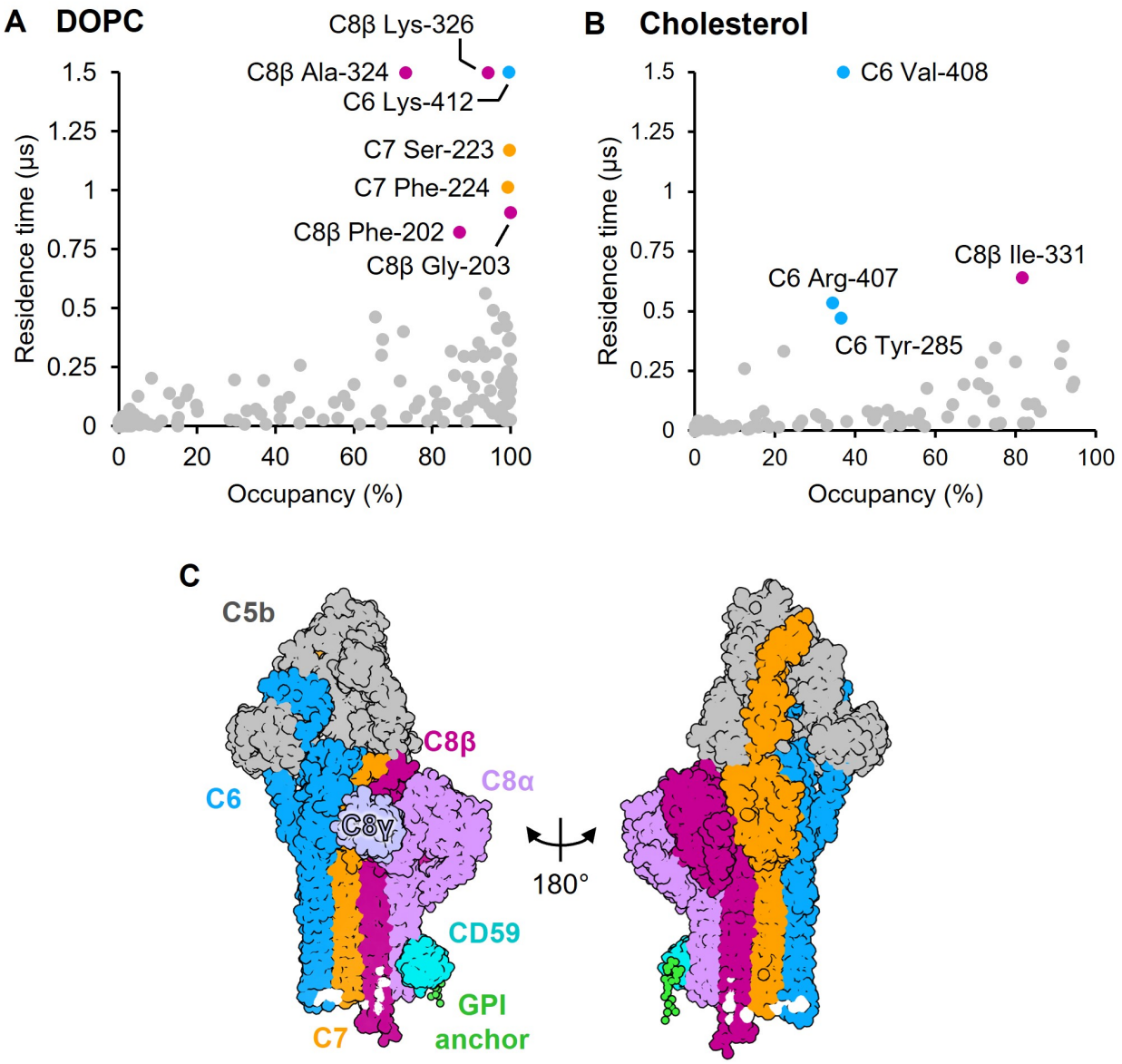


Figure 2.17 | PyLipID highlights potential lipid binding site. **A.** Occupancy and residence time of DOPC for each residue. **B.** Occupancy and residence time of cholesterol for each residue. As cholesterol interactions are weaker, the cut-off residence time and occupancy values for a hotspot were lower than for DOPC. Potential binding sites in panels A and B are coloured according to the MAC protein the residues belong to. Results averaged over three replicates. **C.** Mapping of the potential hotspots in panels A and B on the CG structure, highlighted in white.

750 ns (half the length of time analysed) while only one residue (Val-408 of C6) shows a cholesterol residence time over 750 ns. Overall, C6, C7 and C8β contain potential lipid interaction sites near the tips of their β-hairpins and C8β has the most hotspots with five residues having an occupancy over 80% and a residence time over 500 ns. Lipid-interacting residues show a variety of physico-chemical properties, from aromatic (e.g. C7 Phe-224) to charged side chains (e.g. C6 Lys-412). The location of the potential lipid-binding hotspots, spread over C6, C7 and C8β, is in agreement with the CryoDRGN map in Figure 2.8.

2.3.7 Membrane thinning

Visual inspection of the trajectories revealed a local thinning of the membrane, near the tip of the C8β hairpins, which was observed in all three replicates (Figure 2.18A). C8β spans the membrane and is hydrated at its extremity, but does not form a pore in the bilayer as it does not repel lipids. Overall, these observations are consistent with the cryoEM reconstruction in Figure 2.8.

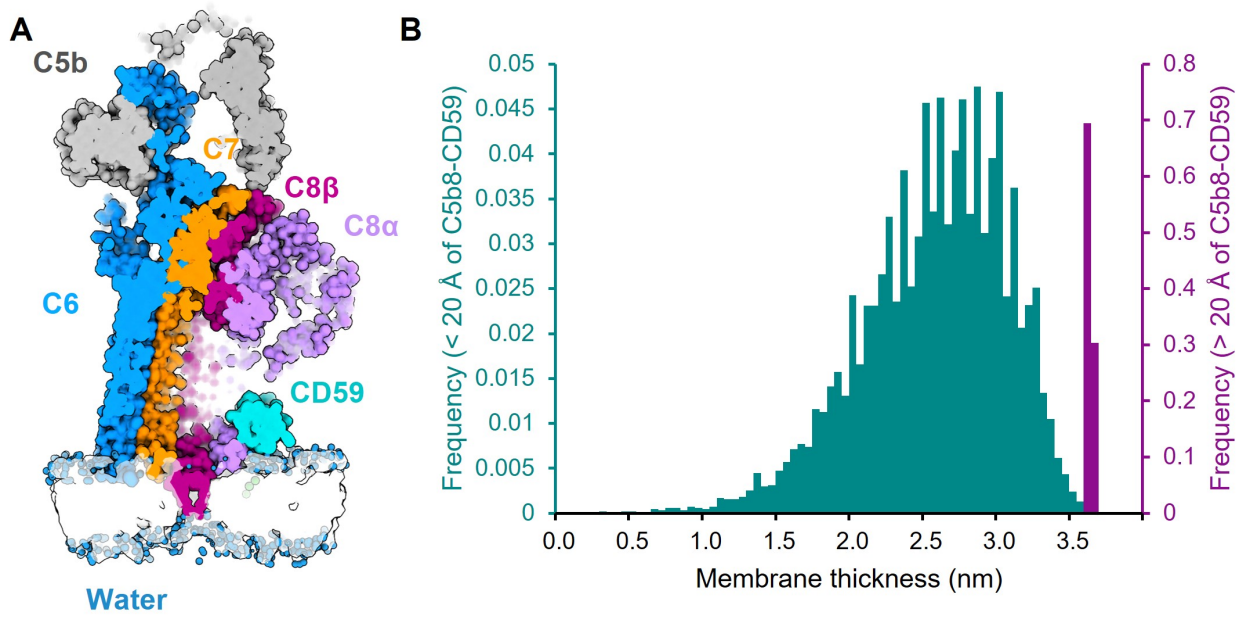


Figure 2.18 | C5b8-CD59 induces a local thinning of the membrane. **A.** Cross-section of a representative frame from the C5b8-CD59 simulations, with the nearest lipids shown in transparent white. The bilayer is thinner near the tip of the C8β hairpins, which are hydrated. **B** Thickness of the bilayer within 20 Å of the protein (turquoise), and further than 20 Å away from the protein (purple).

To monitor the effect of C5b8-CD59 on the membrane, the thickness of the bilayer within 20 Å of the protein was measured with FATSLiM [186] for each frame after the initial 500 ns equilibration period. This was compared to the thickness outside that 20 Å radius (Figure 2.18B). Away from the protein, the membrane has an average thickness of 36 Å. The lipids within 20 Å of the complex show a reduction in thickness of 10 Å on average, but with a wide range of thickness values. The variability in thickness is probably due to the smaller number of lipids and artefacts in thickness evaluation arising from distortions induced by C5b8-CD59. These results show that the thinning effect of the complex is localised to the immediate vicinity of the protein..

2.4 Discussion

Atomistic MD simulations of GPI-anchored CD59 in a DOPC membrane revealed a high level of flexibility in the C-terminus of the protein, and in the sugars moieties of the GPI anchor (Figure 2.12). This is consistent both with the NMR ensemble of CD59 [146] and with previous MD simulations of GPI anchors, which have shown that they act as flexible hinges [159]. Computing the rotation angles of CD59 shows that this flexibility results in a wide range of orientations relative to the membrane (Figure 2.14A).

The angular rotation plot contains defined clusters, suggesting that CD59 may preferentially adopt specific orientations. Visual analysis of selected frames reveals that, in one of the most populated clusters, the final β-strand of CD59 is in the correct position to engage the hairpins of C8α. Preferential sampling of this orientation is likely important to the function of CD59 as an effective inhibitor of MAC, even though it is not the most populated state. Monomeric GPI-anchored CD59 was simulated in a pure DOPC bilayer, which is a simplified proxy for the crowded environment of the plasma membrane. At the cell surface, CD59 forms transient homodimers whose half-life is dependent on the presence of the GPI anchor

and cholesterol [187]. As a result, intermolecular interactions between the ectodomains of two CD59 molecules may hold the protein in a specific orientation while interactions between cholesterol and the GPI anchor may affect its flexibility. Simulations including multiple CD59 molecules in a cholesterol-containing membrane may reveal a different orientation plot where the MAC-binding orientation is sampled at a different frequency. Moreover, MAC forms in clusters on cells [143], consistent with the localised activation of complement (cf. Section 2.1.1) and the lipid environment around complement proteins may influence the orientation of CD59. However, limited information about lipid-protein interactions in the complement system make it challenging to reproduce accurately the context of MAC formation and inhibition in MD simulations.

Another major cluster of CD59 orientations is incompatible with either MAC or ILY binding as the final β-strand is facing the membrane and some of the key residues involved in protein binding (Tyr-61, Tyr-62, Lys-65, Lys-66) are engaged in hydrogen bonds with DOPC. In this orientation, the protein lies flat atop the membrane as the sugar moieties of the GPI anchor are bent towards the bilayer. This corresponds to the *flop-down* conformation observed in previous MD simulations of GPI-anchored proteins [159]. A recently published study based on atomistic simulations of GPI-anchored proteins in lipid bilayer suggested that this orientation is the preferred state of CD59 and did not observe other orientation clusters [160]. As the authors used a more complex membrane, differences in lipid composition may explain the different orientation landscape derived from the simulations in this Chapter.

The precise sequence of events dictating ILY association with the membrane is unclear. Its ability to bind CD59 in solution [181, 182] suggests that it may recognise the protein receptor before engaging lipids. However, at high concentration, ILY can oligomerise on membranes in the absence of CD59 [188], supporting a mechanism where lipid binding precedes CD59 recruitment. The frames from atomistic simulations in Figure 2.14B show

that, if ILY binds CD59 first, it is potentially able to engage the GPI-anchored receptor in a wide variety of orientations. This is in contrast with C8α, which is held in place by the rest of the C5b8 complex and therefore can only bind CD59 in a defined orientation. ILY may have evolved to hijack the inherent flexibility of GPI-anchored CD59 in the plasma membrane. The order of association events is likely driven by differential affinity of ILY for CD59 and cholesterol-containing lipid bilayers. Binding assays have reported an affinity for CD59 in the order of $4-5$ nM [189, 190]. However, its affinity for lipid bilayers (in the absence of CD59) is still unknown. Experimental methods such as surface plasmon resonance with immobilised liposomes [191] or one-dimensional NMR in the presence of lipid vesicles [192] may help disentangle the order of binding events.

The computational cost of atomistic simulations means that the three production runs only represent 500 ns each, which may have limited the orientation sampling. While running longer simulations would improve sampling, another, less computationally expensive approach would be to use frames representing different orientations as starting points for new simulations. The GPI anchor used in this work is a simplified model of the native anchor, but most differences are confined to the lipid tails (i.e. their lengths and number), which is expected to have a limited influence on the flexibility of the sugar core. Despite their limitations, the atomistic simulations in this Chapter have shown that CD59 samples a variety of orientations, including those observed in the experimental structures of the C5b8-CD59 [143] and ILY-CD59 complexes [181, 182].

Building a coarse-grain (CG) model of CD59 allowed the study of inhibited MAC by MD in the microsecond timescale. The standard Martini 2.2 parameters for amino acids were used for the protein moiety and the GPI parameters were taken from the study by Banerjee *et al.* [161]. The EtNP linker was parametrised specifically for CD59, using atomistic simulations as a reference. The overall match between the atomistic and CG parameters

was satisfactory (Figure 2.15) but some discrepancies could not be resolved. For instance, the equivalent of the bond between beads BB_3 (the backbone of Asn-77) and L1 (the first EtNP bead) shows a bimodal length distribution in the atomistic simulations, which cannot be reproduced in CG as it is described as a single harmonic oscillator. Moreover, the bottom-up parametrisation approach has limitations as it requires reliable atomistic simulations and does not take into account experimental data. Given the importance of flexibility in the GPI anchor for the orientation of CD59, any results from CG simulations should be interpreted carefully with regards to the movement of CD59. Finally, the GPI anchor used here is a simplified version of the GPI core, which may not accurately represent the behaviour of the protein in cell membranes. However, this is unlikely to affect the interactions between CG C5b8 and the membrane, making the system a useful tool to study inhibited MAC.

The placement of C5b8-CD59 in the membrane was made challenging by the structure of the complex. The stretch of hydrophobic residues at the tip of the β-hairpins is shorter than the thickness of most biological lipid bilayers (∼30-50 nm), so neither Memembed [183] nor the Positioning of Proteins in Membrane (PPM) 2.0 server [184] could accurately position it in the bilayer. While they generated reasonable predictions for the full MAC assembly, they treated the membrane as perfectly planar. However, the hydrophobic belt on the outer face of the pore breaks planarity around C5b8 (Figure 2.3), so C9 largely drove all orientation predictions. As a result, aligning C5b8-CD59 to the predicted MAC position in the membrane led to the complex being ejected from the bilayer within 100 ns. To overcome this, the complex was placed at an arbitrarily low position, and it quickly (within ∼200 ns) equilibrated to a stable position. Non-planar interfaces between membrane proteins and lipids are common in biology, and new methods, which were not available at the time this work was conducted, can now predict their position more accurately. For instance, the PPM 3.0 is able to predict bilayer curvature for both transmembrane and peripheral membrane proteins,

which simplifies the study of membrane-deforming proteins [193].

While the full membrane attack complex forms a pore, C5b8-CD59 does not repel lipids in the simulations, consistent with the hypothesis that C8α penetrating the bilayer is necessary to form an ion channel [150]. Instead, C5b8-CD59 binds lipids, as revealed by the PyLipID analysis (Figure 2.17). Interactions with lipids on the outer face of the complex are expected due to its hydrophobic belt. However, C6, C7 and C8β also bind lipids *via* residues in what would become the lumen of the pore. This supports the hypothesis that the extra density in Figure 2.8A corresponds to partially ordered lipids.

The three DOPC-binding residues in C6 and C7 all show 100% occupancy, which may be related to the membrane-anchoring function of these proteins [38, 136]. C5b8 shows far weaker interactions with cholesterol than with DOPC. MAC has evolved to assemble on and lyse a wide variety of membranes and, while many pathogen targets contain phospholipids, cholesterol is more rarely found in non-self cells. This may explain the preferential binding to DOPC compared to cholesterol. However, complement proteins may interact with cholesterol in specific ordered domains, which would not have formed due to membrane composition and simulation length [89]. The interpretation of protein-lipid interactions is further limited by the CG nature of Martini 2.2, which does not represent all constituent atoms. The new parameters in Martini 3 have led to more accurate interactions between proteins, lipids and solvent, including binding within protein cavities [102], but the GPI anchor was parametrised for Martini 2.2 and therefore C5b8-CD59 could not be simulated with Martini 3 without a complete re-parametrisation strategy.

C5b8-CD59 induces local membrane thinning in CGMD simulations, as evidenced by the snapshot in Figure 2.18A. In the simulations, the membrane is thin enough for C8β to span it entirely and have its hairpins hydrated by water molecules. Monitoring the thickness of the bilayer within 20 Å of the complex and in the rest of the box shows that the thinning

effect of C5b8-CD59 is localised near the protein (Figure 2.18B). The initial placement of the complex in the bilayer exposes the tip of the C8β hairpins to the solvent, so the pinching and thinning of the membrane near C8β may be an artefact arising from the assembly of the system. However, Martini lipids are known to self-assemble around membrane proteins to shield their hydrophobic regions [194] and the cryoEM map in Figure 2.8B shows a thinning in the same position as the CGMD simulations, supporting the idea that it is a physiological process. Better membrane protein orientation predictors and the use of atomistic simulations may help determine whether thinning is an artefact of the simulation setup, or a biologically relevant phenomenon.

The CG and atomistic simulations presented in this Chapter have contributed to our understanding of CD59 function on the surface of human cells (Figure 2.19). The atomistic simulations revealed that GPI-anchored CD59 samples a wide range of orientations relative to the membrane, due to intrinsic flexibility in its C-terminus and in the GPI anchor itself. CD59 shows a preference for some orientations, including the one in which it engages C8α to inhibit MAC, and one where it establishes hydrogen bonds with membrane lipids. A wide variety of orientations may be compatible with binding to ILY, suggesting that the toxin has evolved to hijack the high degree of flexibility of CD59 to anchor itself to target membranes. However, the impact of lipid composition and local protein environment on CD59 orientation remains unknown. CGMD simulations have shown that inhibited MAC affects its local lipid environment through membrane thinning and lipid binding. This may arise before the unfurling of C8α and contribute to recruiting CD59 to the site of MAC deposition. CD59 recruitment may also be driven by changes in membrane properties due to the deposition of upstream complement proteins (cf. Sections 2.1.1 and 2.1.2). However, the effects observed here may be specific to C5b8-CD59 complexes and act as a signal for recruiting the molecular machinery needed for MAC clearance. Future studies should focus on the responses

to MAC deposition and inhibition at the scale of the membrane and the whole cell, using cryo-electron tomography and proteomics to identify the effects of MAC on self-cells.

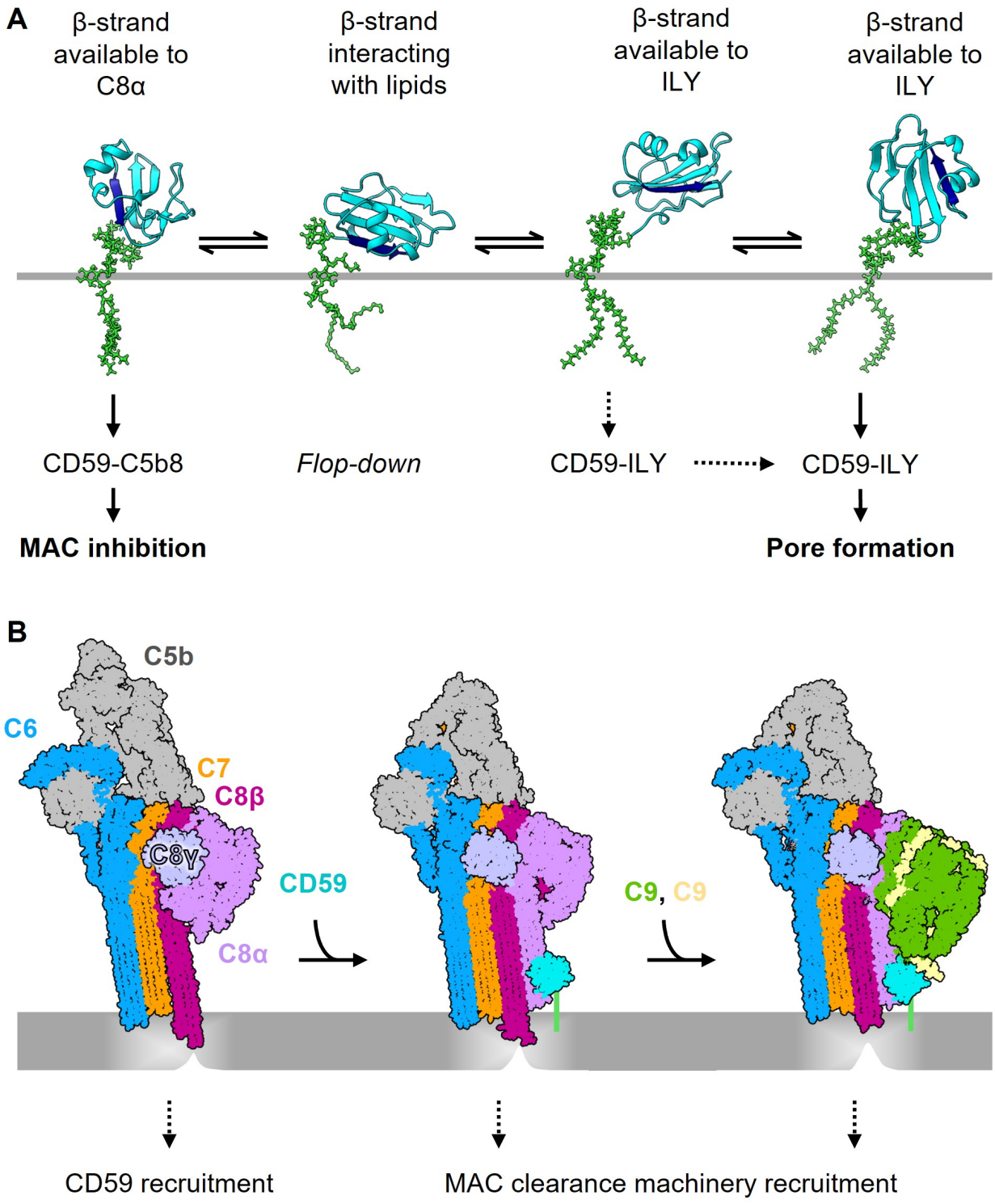


Figure 2.19 | Schematic summary of CD59 interactions. **A.** CD59 exchanges between various orientations relative to the membrane, where its final β-strand (dark blue) can interact with lipids or be available for binding to either C8α or ILY. It is unclear whether ILY binds a specific orientation of CD59 or whether it can associate with other states before re-orienting the complex to generate the geometry needed for oligomerisation. **B.** Local changes to the membrane environment (e.g. lipid binding, membrane thinning) induced by the C5b8 complex before unfurling of the C8α MACPF/CDC hairpins may recruit CD59 to the site of MAC deposition. Alternatively, changes in the local lipid environment around the CD59-C5b8 and CD59-C5b9 inhibited complexes may trigger recruitment of the MAC clearance machinery.

3 || Functional characterisation of the horizontal helix of cholesterol-dependent cytolysins

Chapter summary

Cholesterol-dependent cytolysins (CDCs) are bacterial pore-forming proteins which specifically target cholesterol-rich membranes. They undergo a series of complex and sequential conformational changes to form a β-barrel pore. Dr. Nita Shah, in the Bubeck lab, solved the structure of the CDC intermedilysin (ILY) early prepore by cryo-electron microscopy. This revealed the movement of a key amphipathic helix, termed the *horizontal (h-) helix*. The results presented in this chapter build on her work to characterise the h-helix.

ILY intermediates can be trapped by engineering disulfide bonds to block specific conformational changes. To confirm the suitability of such mutants, the Hoogenboom lab (UCL) recorded atomic force microscopy data, whose analysis is presented here. Mutagenesis and functional assays show that the h-helix is key to the function of the ILY late prepore, and that its role is conserved among CDCs. These results suggest that the CDC h-helix, by its amphipathic nature, lowers the free energy energy barrier for rupturing the bilayer by disturbing the membrane.

Related publication: Nita R. Shah, Tomas B. Voisin, Edward S. Parsons, Courtney M. Boyd, Bart W. Hoogenboom, Doryen Bubeck. (2020) Structural basis for tuning activity and membrane ty of bacterial cytolysins. *Nature Communications* 11(1):5818.

3.1 Introduction

3.1.1 Pore formation by cholesterol-dependent cytolysins

Cholesterol-dependent cytolysins (CDCs) are bacterial toxins from the MACPF superfamily of pore-forming proteins [32]. They are secreted as monomers and oligomerise on the surface of cholesterol-rich membranes, where they undergo a series of conformational changes to form pores. The crystal structures of nine CDCs have been solved to date: perfringolysin O [35], intermedilysin [181, 182, 195], anthrolysin O [196], suilysin [197], listeriolysin O [198], streptolysin O [199], pneumolysin [51, 200], vaginolysin [181] and desulfolysin [201]. Sequence analysis and structural data show that all CDCs share a conserved architecture [32] comprised of four domains (Figure 3.1), with the exception of lectinolysin, which contains an extra C-terminal sugar-binding domain [202]. Domains 1 and 3 (D1 and D3) contain the pore-forming MACPF/CDC fold, while domain 4 (D4) mediates membrane binding. Domain 2 (D2) connects D1-D3 to D4 and acts as a hinge during pore formation.

D4 is made up of two antiparallel β-sheets connected by seven loops, with its four bottom loops inserting into the membrane (Figure 3.2). Loop 1 contains the cholesterol recognition motif: a conserved threonine-leucine pair that confers CDCs part of their specificity to cholesterol-containing membranes [188, 203]. Cholesterol recognition by the threonine-leucine pair leads to the insertion of loops 2 and 3, as well as the tryptophan-rich undecapeptide (UDP) in loop 4.

While most CDCs directly recognise cholesterol in lipid membranes, *Streptococcus intermedius* intermedilysin (ILY), *Gardenerella vaginalis* vaginolysin (VLY) and *Streptococcus mitis* lectinolysin (LLY) bind a protein receptor [190, 204, 205]. All three CDCs have evolved to hijack the complement inhibitor CD59 (also known as protectin), an 11-kDa glycosylphosphatidylinositol (GPI)-anchored protein [206]. The crystal structures of ILY-CD59 and VLY-

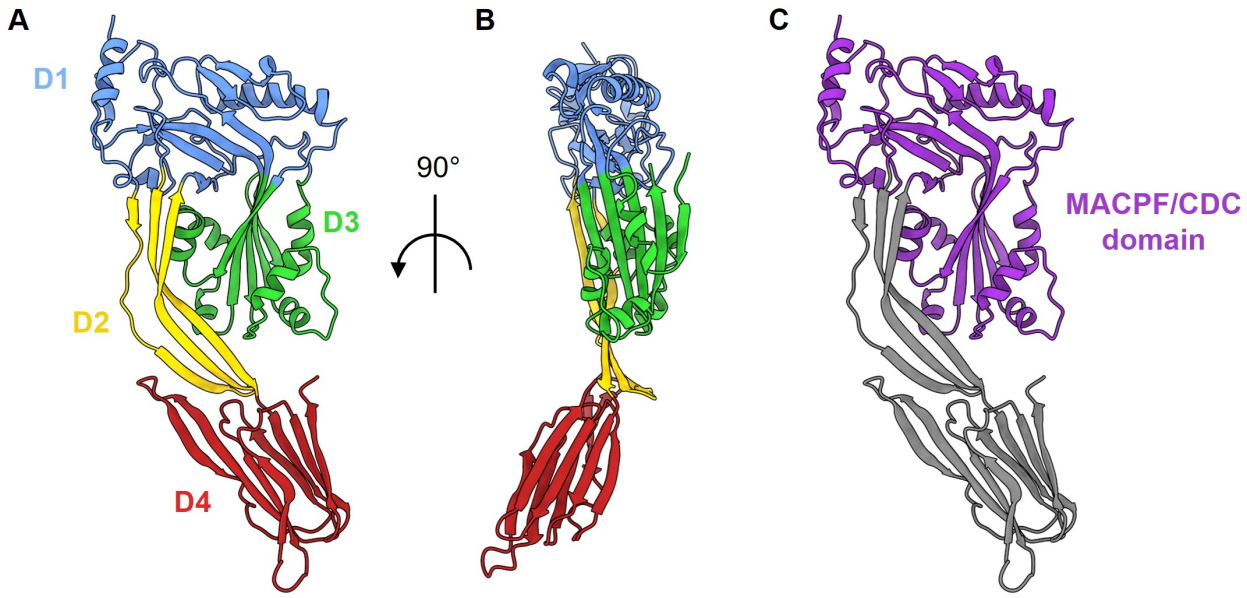


Figure 3.1 | Domain composition of the archetypal CDC pneumolysin (PLY). **A** and **B.** Crystal structure of soluble monomeric PLY (PDB 5CR6) highlighting domains D1 (blue), D2 (yellow), D3 (green) and D4 (red). **C.** Together, D1 and D3 make up the MACPF/CDC domain (purple) responsible for membrane perforation.

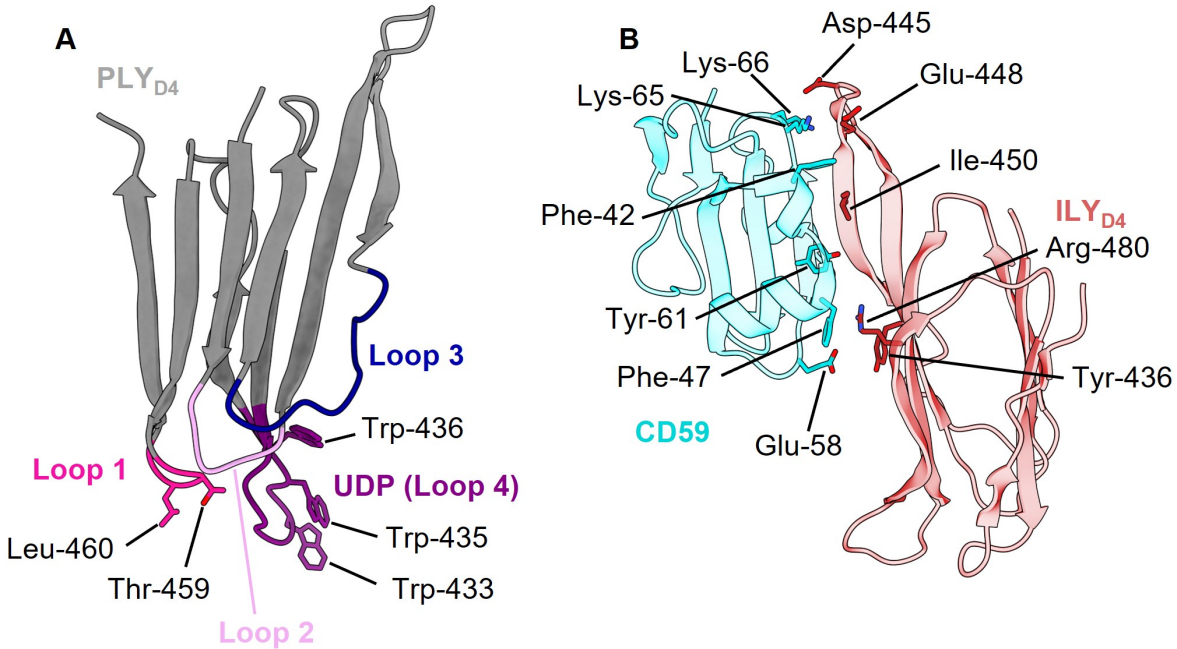


Figure 3.2 | Structure of the CDC domain 4 (D4). **A.** Crystal structure of D4 from PLY (PDB 5CR6). Loops 1 (magenta), 2 (light pink) and 3 (dark blue) as well as the Trp-rich undecapeptide (UDP, purple) interact with membrane lipids. Thr-459 and Leu-460 form the cholesterol-recognition motif. **B.** Crystal structure of ILY D4 (red) bound to its receptor CD59 (cyan) (PDB 4BIK). The interaction leads to an extension of the D4 β-sheet, stabilised by salt bridges (Arg-480^{ILY}•Glu-58^{CD59}, Glu-448^{ILY}•Lys-65^{CD59} and Asp-445^{ILY}•Lys-66^{CD59})

.

CD59 complexes [181, 182] reveal a similar binding pose for the two CDCs, with the CD59-ILY structure shown in Figure 3.2. Although the specific residues binding CD59 are different in ILY and VLY, both complexes are formed by the alignment of β-strands from D4 and CD59 to create a large β-sheet spread across the two binding partners. ILY can bind membranes without cholesterol if they are decorated with CD59, making it a powerful model system to study pore formation [188].

Membrane cholesterol and CD59 are the two canonical receptors for CDCs, but there is mounting evidence that, in a cellular context, CDCs bind other cell-surface molecules. Both pneumolysin (PLY) and streptolysin O require binding to glycans on the surface of red blood cells to be haemolytic [207]. Glycan arrays involving eight major CDCs revealed that they are all able to bind various sugars with different affinities [208]. While non-glycosylated CD59 can bind ILY with high enough affinity for structure determination [50, 182], its O-linked glycans further stabilise the interaction [208]. This suggests that, although cholesterol (and CD59 in some cases) is usually enough for CDC membrane binding in simple systems such as liposomes, a more complex set of interactions with cell-surface glycans dictates membrane recognition and CDC cell specificity.

Once CDC monomers have bound their target membrane, they form a homo-oligomeric early prepore. The size of the assembly can vary from small arcs of about five monomers to complete rings of 30 to 45 subunits. The structure of the ILY early prepore on lipid nanodiscs was recently solved by cryoEM, revealing the conformational changes accompanying oligomerisation (Figure 2.3) [50]. The fifth β-strand (β5) of the MACPF/CDC domain refolds into a helix, revealing the edge of the previous strand (β4), which in turn swings backwards to interact with β1 from the neighbouring monomer (Figure 2.3B). The oligomer is likely stabilised by charge complementarity between adjacent subunits [50]. Another notable motion is the movement of an α-helix within the MACPF/CDC domain (Figure 3.3). We have termed

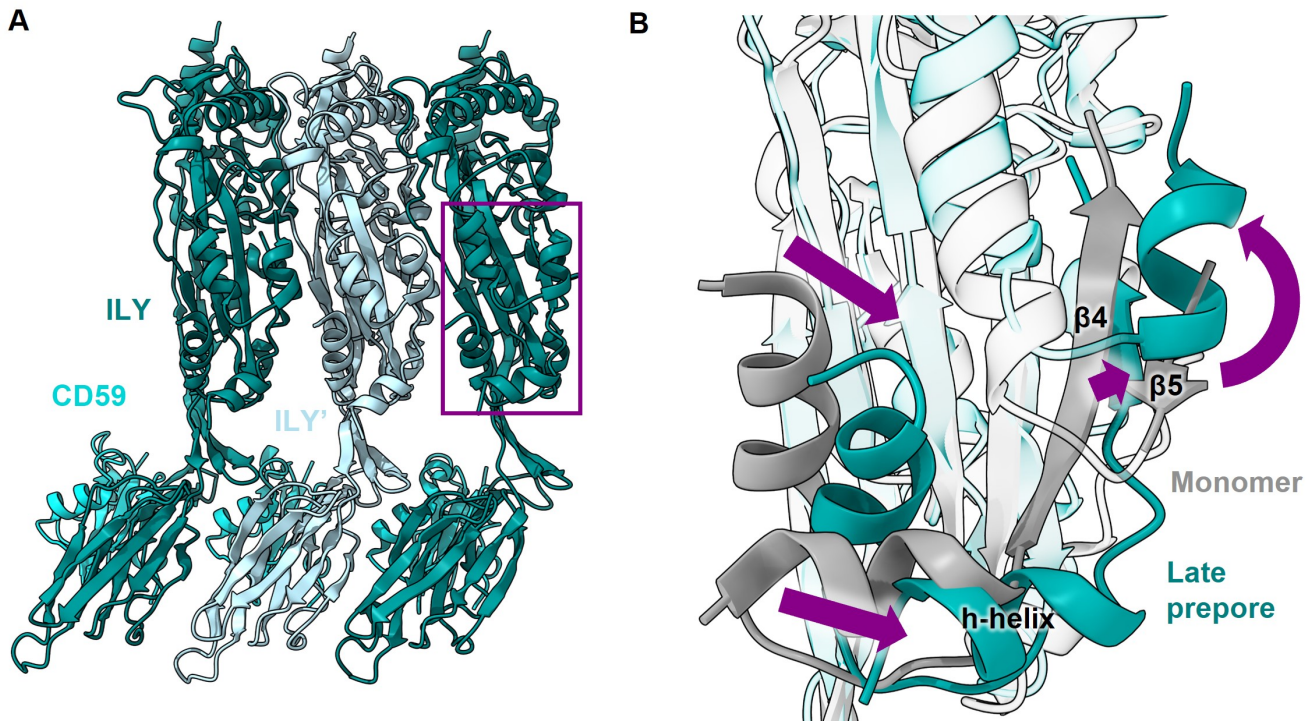


Figure 3.3 | ILY forms an oligomeric early prepore. **A.** Structure of the ILY early prepore solved by cryoEM on lipid nanodiscs (PDB 6ZD0). **B.** Close-up view of the region highlighted by the purple box in A. Oligomerisation is associated with conformational changes from the membrane-monomer (grey, PDB 4BIK) to the early prepore (teal, PDB 6ZD0). Both helices in helical bundle 2 (HB2) move, bringing the h-helix towards the membrane. The β4 strand moves backwards to straighten the MACPF/CDC core, and it is exposed to the next monomer by the transition of β5 into a helix.

this helix the *horizontal (h-) helix* as it comes to lie parallel to the membrane plane.

Once assembled, the early prepore undergoes large conformational changes to form a collapsed, late prepore. While there is no experimental structure of a CDC late prepore, atomic force microscopy (AFM) has shown that it is associated with a reduction in height of ∼2-4 nm, depending on the exact protein [188, 209, 210]. The cryoEM structure of the PLY pore revealed that collapse is the result of a 90° rotation of D2 [51], bringing down the MACPF/CDC domain (D1-D3) close to the membrane (Figure 3.4). This movement is necessary as the MACPF/CDC hairpins would be too short to span a bilayer if D1 and D3 remained in their early prepore position, high above the membrane. Unlike the early prepore, the late prepore is resistant to SDS treatement [211, 212], suggesting stronger interactions between monomers.

The final conformational change in the CDC pore formation pathway is conserved in the MACPF superfamily: the helical bundles in D3 (part of the MACPF/CDC domain) unfurl into transmembrane β-hairpins (Figure 3.4). Each monomer contributes 2 hairpins to a membrane-spanning β-barrel pore. The PLY pore structure solved by cryoEM on liposomes revealed a large assembly made up of 42 monomers, with an internal diameter of 250 Å. The outside of the pore contains a hydrophobic belt, which was shown by molecular dynamics simulations to interact with lipids [213]. The lumen of the pore, however, is lined with hydrophilic residues (Figure 3.4). Those residues repel lipids to form a hydrated zone across the bilayer, eventually leading to total depletion of lipids inside the pore [213].

The predominant picture of CDC pore formation in the literature is that of a sequential process, where oligomerisation cannot proceed after collapse. This prepore-to-pore mechanism is in contrast with that of the membrane attack complex (MAC), where each subunit associates with the previous in the membrane-inserted form, resulting in a growing pore [214]. However, recent single-molecule fluorescence data have suggested that perfringolysin O monomers can be added to incomplete pores, even after membrane rupture [215]. This contradicts previous AFM data of suilysin, which showed that oligomerisation ends before membrane perforation [210]. These conflicting results suggest that the sequence and timing of events may differ between CDCs.

3.1.2 The h-helix and disulfide-locked mutants

Dr. Nita Shah in the Bubeck lab solved the structure of the ILY early prepore on lipid nanodiscs using cryoEM [50]. This work revealed the conformational changes associated with CDC oligomerisation described above, in particular the movement of the h-helix (Figure 3.3). The h-helix is an amphipathic α-helix with polar and charged residues pointing towards the membrane (Figure 3.5). Mutagenesis and liposome lysis assays performed by Dr. Shah

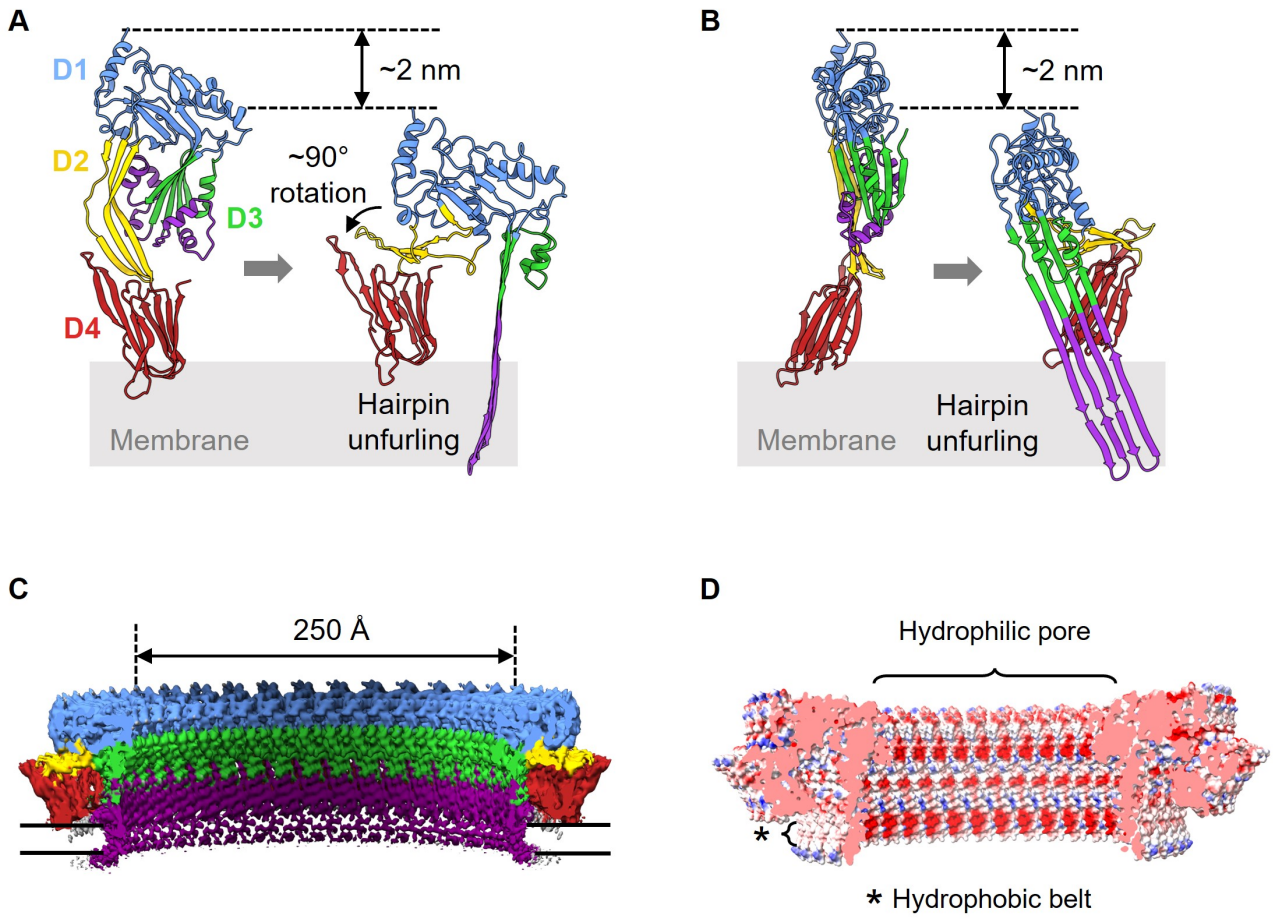


Figure 3.4 | Structure of the PLY pore. **A** and **B.** Structural differences between the membrane-bound monomer (left of each panel) and the pore (right of each panel). First, a 90°rotation of D2 leads to a ∼2 nm reduction in assembly height. The second change is the unfurling of the helical bundles (left, purple) into transmembrane hairpins (right, purple). Blue: D1; yellow: D2; green: D3 excluding helical bundles/transmembrane hairpins; red: D4. As there is no structure of the PLY membrane-bound monomer, the crystal structure of the soluble monomer (PDB 5CR6) was used as a proxy to compare to the pore (PDB 5LY6). **C.** CryoEM map of the PLY pore (EMD-4118) coloured by domain as in A and B. **D.** Electrostatic potential of the pore as calculated by ChimeraX. Red: negative potential, blue: positive potential.

showed that the h-helix is responsible for tuning the lytic activity of ILY (Figure 3.5 and [50]). Mutating the hydrophilic residues on the bottom face to alanine abrogated lysis, whereas subsitution of two apolar residues with lysine on the top face led to an increase in lytic activity. However, the step of the pore formation pathway in which the h-helix performs its function remained unknown.

To study the function of the h-helix, we used disulfide-locked mutants of ILY, whereby disulfide bonds are engineered to hold the protein trapped in a certain conformation. By

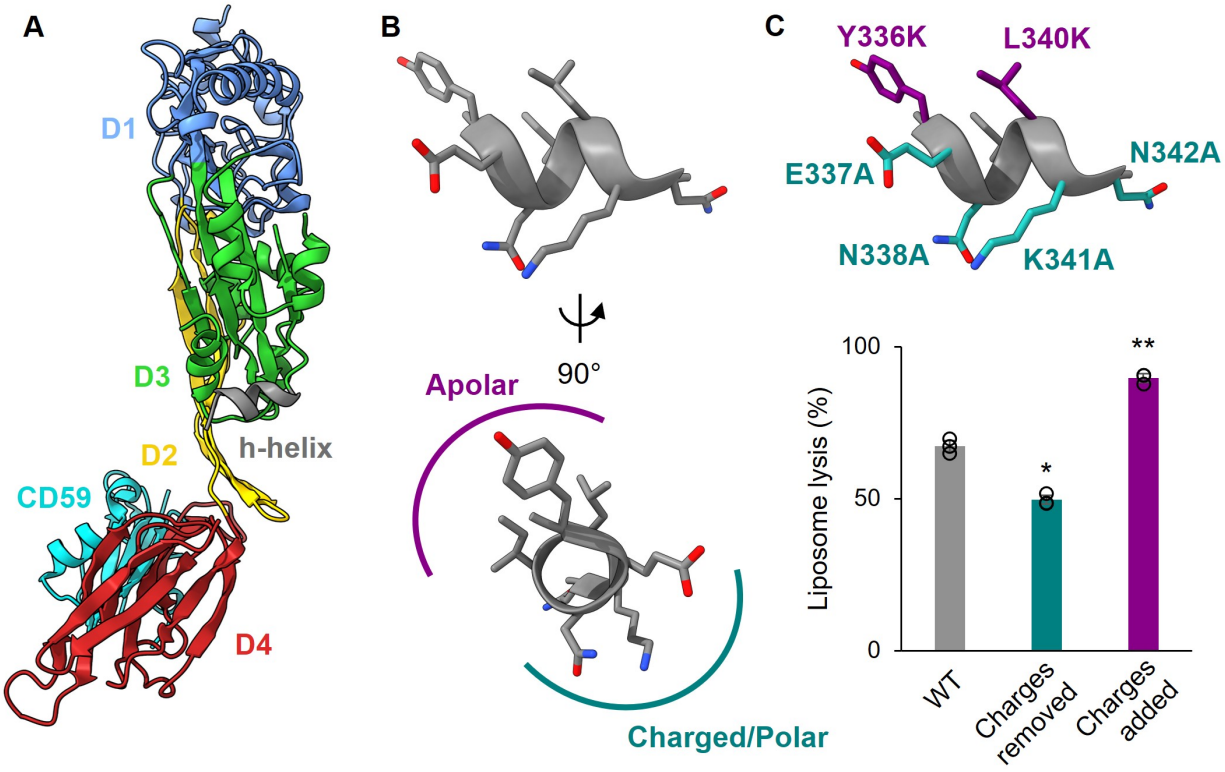


Figure 3.5 | The h-helix in the ILY early prepore. **A.** ILY early prepore (shown as a monomer) bound to CD59, coloured by domain. Blue: D1, yellow: D2, green: D3, red: D4, cyan: CD59. The h-helix within D3 is shown in grey. **B.** The h-helix is amphipathic and presents apolar residues (purple) on its top face and charged residues (teal) on its bottom face. **C.** The amphipathic nature of the h-helix is key to determining lytic activity. Removing charges from the bottom face of the helix (teal) impairs liposome lysis, whereas adding charges to the top face (purple) makes ILY more active. Data generated by Dr. Nita Shah [50]. Pairwise comparisons are between wild-type and each mutant: *$p < 0.05$, **$p < 0.01$. N = 3.

preventing the conformational changes needed for ILY to progress along the pore formation pathway, we were able to investigate specific steps of pore formation. The early prepore (ILY^{EP}) was trapped by locking D2 and D3 (I104C, G244C) as described first in [188], and was used for structure determination [50]. To trap the late prepore, two mutants were generated by Dr. Shah, locking HB1 (ILY^{HB1}: A221C, N404C) and HB2 (ILY^{HB2}: Y208C, D322C). However, as no structure of HB-locked ILY has been solved, the mutants first needed validation by atomic force microscopy (AFM).

3.1.3 Aims

The work presented in this chapter builds on Dr. Shah's structural observations and aims to characterise the function of the h-helix. To assess whether ILY^{HB1} and ILY^{HB2} are able to form collapsed prepores, I analysed AFM topographs recorded by the Hoogenboom lab (UCL). To place the role of the h-helix in the context of the pore formation pathway, I introduced the overactivating h-helix mutations in ILY^{EP} and measured lytic activity in a liposome lysis assay. To determine whether the function of the h-helix is conserved in CDCs, I recombinantly expressed and purified pneumolysin (PLY), introduced mutations analogous to those characterised in ILY, and recorded the effects on lytic activity. Finally, to gain insight in the molecular basis of h-helix function, I generated hybrid PLY prepores, based on the soluble and membrane-inserted structures, and used them for coarse-grain molecular dynamics simulations.

3.2 Materials and Methods

3.2.1 Atomic force microscopy data analysis

Atomic force microscopy (AFM) data were generated by Dr. Edward Parsons in the Hoogenboom lab (University College London) as described in [50]. Wild-type ILY (ILY^{WT}) and HB-locked mutants (ILY^{HB1} and ILY^{HB2}), all three with the only native cysteine in D4 mutated to serine, were incubated on DOPC:cholesterol (1:1 molar ratio) supported lipid bilayers decorated with CD59 at 37°C. Room-temperature measurements were recorded by Dr. Parsons with a MultiMode IV (Bruker) atomic force microscope to generate 500 x 500 nm^2 images.

I analysed the AFM images in the Gwyddion software [216]. The images were levelled by first-order plane fit, which fits a plane from all the data points and subtracts it as background.

Line-by-line flattening was then applied to correct for artefactual differences between scanning lines. The images were finally levelled by setting the membrane height to zero as a reference. For each ILY variant, ten height profiles were extracted from different assemblies and the average height was calculated.

3.2.2 PLY and ILY cloning and mutagenesis

The gene for wild-type pneumolysin (PLY-*WT*) was codon-optimised for recombinant expression in *Escherichia coli* and purchased from Twist Bioscience. I cloned it into a pTrcHisA plasmid (ThermoFisher) with an ampicillin resistance gene by Gibson assembly using Q5 DNA polymerase (New England Biolabs) and Gibson Assembly Master Mix (New England Biolabs) according to the manufacturer's instructions. After transforming into *E. coli* DH5α cells (ThermoFisher), an ampicillin-resistant colony was grown for plasmid extraction using the QIAprep Spin Miniprep Kit (Qiagen) and the plasmid was checked by sequencing (Eurofins). The pTrcHisA plasmid containing the gene for ILY^{EP} [188] was provided by Dr. Nita Shah.

Q5 DNA polymerase (New Englad Biolabs) was used for site-directed mutagenesis to alter the charge properties of the h-helix in ILY^{EP} and PLY. Charges were added to the top face of the helix in ILY^{EP} to generate ILY^{EP}-*charge+* (Y336K, L340K). PLY-*WT* was similarly mutated to generate PLY-*charge+* (W278K, L282K) and PLY-*nocharge* (K279A, Q280A, D283A, N284A). The primers used for mutagenesis are in Table 3.1 and were purchased from Eurofins. Mutated plasmids were transformed into *E. coli* DH5α cells (ThermoFisher) and checked by sequencing (Eurofins). All mutants were deposited to Addgene (ILY^{EP}-*charge+* #160522, PLY-*charge+* #160538 and PLY-*nocharge* #160539).

Table 3.1: List of primers used for site-directed mutagenesis. The mutated nucleotides are shown in bold.

Mutant name	**Mutations**	**Primer pair** (top: forward; bottom: reverse)
ILY^{EP}-*charge+*	Y336K, L340K	GCTGGTACGGAA**A**A**G**GAAAACATC**AA**GAAAAACACCAAAATCACGG CCGTGATTTTGGTGTTTTTC**TT**GATGTTTTC**C**T**T**TTCCGTACCAGC
PLY-*charge+*	W278K, L282K	CCACAGACCGAG**GC**GAAGCAGATT**GC**GGACAATACGGAAG CTTCCGTATTGTC**CG**CAATCTGCTTC**GC**CTCGGTCTGTGG
PLY-*nocharge*	K279A, Q280A, D283A, N284A	CAGACCGAGTGG**GC**G**GC**GATTTTG**GC**C**GC**TACGGAAGTGAAGG CCTTCACTTCCGTA**GC**G**GC**CAAAATC**GC**C**GC**CCACTCGGTCTG

3.2.3 ILY and PLY variants expression and purification

The ILY and PLY variants were expressed recombinantly in *E. coli* BL21 (DE3) cells (ThermoFisher). Transformed bacteria were grown in Luria Broth in the presence of 100 ng/ml ampicillin at 37°C to an optical density at 600 nm of 0.6-0.8. Protein expression was induced by adding a final concentration of 50 µg/ml isopropyl β-d-1-thiogalactopyranoside (IPTG) (Sigma) and shaking the cells at 18°C for 18 hours.

Cells were harvested by centrifugation and lysed by sonication in CDC buffer (200 mM NaCl, 20 mM Tris-HCl, pH 7.5) in the presence of DNAse I (New England Biolabs) and cOmplete Protease Inhibitors (Roche). Cell debris were removed by centrifugation and the clear supernatant was incubated with TALON cobalt affinity resin (Clontech) pre-equilibrated in CDC buffer. The resin was washed first with CDC buffer, then with CDC buffer supplemented with 10 mM imidazole. Bound proteins were first eluted with 100 mM, then 500 mM imidazole in CDC buffer. Further purification was achieved by size-exclusion chromatography on a Superdex 200 Increase 10/300 GL column (GE Healthcare) before assessing purity by SDS-PAGE. ILY^{EP} with its wild-type h-helix (ILY^{EP}-*WT*) was provided by Dr. Nita Shah.

3.2.4 Lipid monolayer negative stain electron microscopy

The ability of PLY-*WT* to form pores was checked by negative stain electron microscopy on lipid monolayers. Lipid monolayers containing a 1:1 molar ratio of 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC) (Avanti Polar Lipids) and cholesterol (Avanti Polar Lipids) were assembled in the wells of a Teflon plate. The wells were filled with CDC buffer, which was then covered with a 1 mg/ml lipid mixture in chloroform. The chloroform was evaporated and carbon-plated copper grids (Electron Microscopy Science) were placed on top. PLY-*WT* was added to the cell to a final concentration of 7 ng/µl before a 20 minute incubation at 37°C in a sealed petri dish. Grids were negatively stained with 2% (w/v) uranyl acetate. Grids were imaged on an FEI Tecnai G2 Spirit TWIN 120 kV instrument equipped with a LaB6 filament and an FEI 2K Eagle CCD camera at a magnification of 52,000×.

3.2.5 Liposome lysis assays

DOPC and cholesterol dissolved in chloroform were mixed to a 1:1 molar ratio and dried under nitrogen gas. The resulting lipid films were rehydrated in calcein buffer (50 mM calcein (Merck), phosphate-buffered saline, pH 7.5) and underwent four rounds of freezing in liquid nitrogen and thawing at 65°C. Unilamellar liposomes were then formed by extrusion through a 0.1 µm membrane. Liposomes loaded with calcein were separated from free calcein in solution by size exclusion chromatography using Sephadex G-50 resin (ThermoFisher) in liposome buffer (150 mM NaCl, 500 mM sucrose, 20 mM Tris-HCl, pH 7.5).

For lysis assays involving ILY, liposomes were incubated for 5 min at room temperature with 1 µg/ml (final concentration) of a CD59 construct covalently attached to a lipid-anchoring peptide to mimic the native GPI anchor (gifted by Dr. Richard Smith, King's College London, cf. Section 4.1.2). For PLY lysis assays, the liposomes were used undecorated. In all cases, baseline calcein fluorescence in liposomes was measured for 10 min at room temperature

using a CLARIOStar plate reader (BMG Labtech) with excitation and emission wavelengths of 490 nm and 520 nm, respectively. ILY variants were added to a final concentration of 0.9 µg/ml. Different concentrations of PLY were tested (cf. Section 3.3.2) and a final concentration of 48.4 µg/ml was used to compare the different mutants. Fluorescence was measured for 45 min at 90 second intervals. Full lysis was achieved with 0.87% n-dodecyl-beta-maltoside (DDM) and a freeze-thaw cycle at -80°C before a final fluorescence reading.

Lytic activity was calculated by subtracting the baseline fluorescence value of each well from its reading at a single time point after protein addition, and dividing the result by the post-full lysis value. Three replicates were measured in different wells on the same plate. The significance of variance within datasets was assessed by a single-factor ANOVA test and pairwise comparisons were made with a Bonferroni post-test.

3.2.6 Coarse-grain MD simulations of a hybrid PLY conformation

In the absence of a high-resolution experimental structure for a CDC late prepore, a hybrid model was built using the PLY pore structure [51] with the transmembrane hairpins substituted with the helical bundles from the soluble crystal structure [217]. The MACPF/CDC β-strands were trimmed to exclude residues 154 to 188 and residues 253 to 284, corresponding to HB1 and HB2, respectively. HB1 from the crystal structure of soluble PLY [200] was split in two segments (residues 154 to 174 and residues 180 to 188), which were moved as rigid bodies and covalently linked to their corresponding β-strands (β-1 and β-2, respectively). The residues between these two segments (residues 181 to 187) were modelled with SWISS-MODEL [218] to ensure continuity. HB2 was moved as a single rigid body and connected to β-3 and β-4. To avoid steric clashes with HB2, residues 305 to 309 were deleted and then modelled with SWISS-MODEL [218]. To relax any clashes, the resulting hybrid structure was solvated in TIP3 water with neutralising Na^+ ions and energy-minimised

in Gromacs 2020.3 [219] with the OPLS/AA forcefield [69, 70] using the steepest descent algorithm, to a maximum force of 1000 kJ/mol. A hexamer of the PLY hybrid structure was then built from the relaxed model, based on the symmetry of the cryoEM structure of the pore [51], before energy minimisation using the same protocol as above, to remove potential clashes between adjacent PLY molecules. The *swapaa* command of Chimera [175] was used to reproduce the *h-helix* mutations characterised experimentally, resulting in three constructs: PLY^{hybrid}-*WT*, PLY^{hybrid}-*nocharge* (K279A, Q280A, D283A, N284A) and PLY^{hybrid}-*charge+* (W278K, L282K).

Each hexameric PLY hybrid was converted to a coarse-grain representation with the *martinize.py* script for the Martini 2.2 forcefield [90, 91]. It was then placed on a 25 $\times$ 25 nm^2, 1:1 (molar ratio) DOPC:cholesterol membrane using the *insane.py* script [180] and based on the prediction of its orientation from the Memembed software [183]. The system was solvated in Martini water and 150 nM of NaCl (with neutralising ions). For each PLY^{hybrid} variant, three independent replicates were simulated in Gromacs 2020.3 [219]. First, the system was energy-minimised using the steepest-descent algorithm until a maximum force of 1000 kJ/mol. The resulting energy-minimised configuration was used for a 1-µs production run using the md integrator (20 ps timestep) in an NPT ensemble maintained at a pressure of 1 bar by a semi-isotropic Parinello-Rahman barostat (τ_p = 12.0 ps) and a temperature of 310 K by a V-rescale thermostat (τ_T = 1.0 ps).

The first 200 ns of each production run, corresponding to an equilibration stage, were excluded from the analysis. Average and local membrane thickness over the course of each trajectory were calculated with FATSLiM [186]. The minimum distance between the *h-helix* and the bilayer was monitored with the *mindist* function of Gromacs 2020.3. Protein-lipid interactions were analysed using PyLipID [185].

3.3 Results

3.3.1 The HB-locked ILY mutants form collapsed assemblies

For CDCs to form pores, the oligomeric early prepore must first collapse into the late prepore, decreasing the height of the assembly by around 2 nm, which can be tracked by atomic force microscopy (AFM) [188, 209, 210]. Dr. Shah showed that the HB-locked mutants of ILY (shown in Figure 3.6) are able to form oligomers visualisable by negative stain electron microscopy, but cannot lyse liposomes [50]. To discriminate between the early and late prepore states, the HB-locked mutants were incubated on supported lipid bilayers and AFM data were recorded by Dr. Ed Parsons in the Hoogenboom lab (UCL).

Processing of the AFM topographs shows that ILY^{WT}, ILY^{HB1} and ILY^{HB2} are all able to form oligomeric assemblies (Figure 3.6B). Both ILY^{WT} and ILY^{HB1} form full rings and lower-order oligomeric arcs, as has been described in the literature for ILY and other CDCs [188, 210, 220, 221, 222, 223]. In contrast, ILY^{HB2} only forms small arcs and no full ring was observed in the entirety of the topograph, which may be an artefact arising from the placement of the disulfide lock. The defect in geometry does not fully prevent oligomerisation as ILY monomers are known to be too mobile to be distinguished individually by AFM [188].

Ten height profiles were measured from different assemblies in each of the AFM images and the local height maxima were extracted. All three variants formed assemblies of comparable heights: 8.4 $\pm$ 0.6 nm (mean $\pm$ standard deviation) for ILY^{WT}, 7.9 $\pm$ 0.7 nm for ILY^{HB1} and 7.9 $\pm$ 0.4 nm for ILY^{HB2}. Example height profiles are shown in Figure 3.6C. All three ILY variants form oligomers $\sim$8 nm above the membrane, consistent with collapsed assemblies. This is expected for ILY^{WT} because there are no mutations preventing it from reaching the final pore conformation. As ILY^{HB1} and ILY^{HB2} cannot lyse liposomes (data generated by

Dr. Shah and published in [50]), their ∼8 nm height indicates that they are trapped in the collapsed, late-prepore conformation. This shows that the ILY^{HB1} and ILY^{HB2} mutants are appropriate tools to study the ILY late prepore.

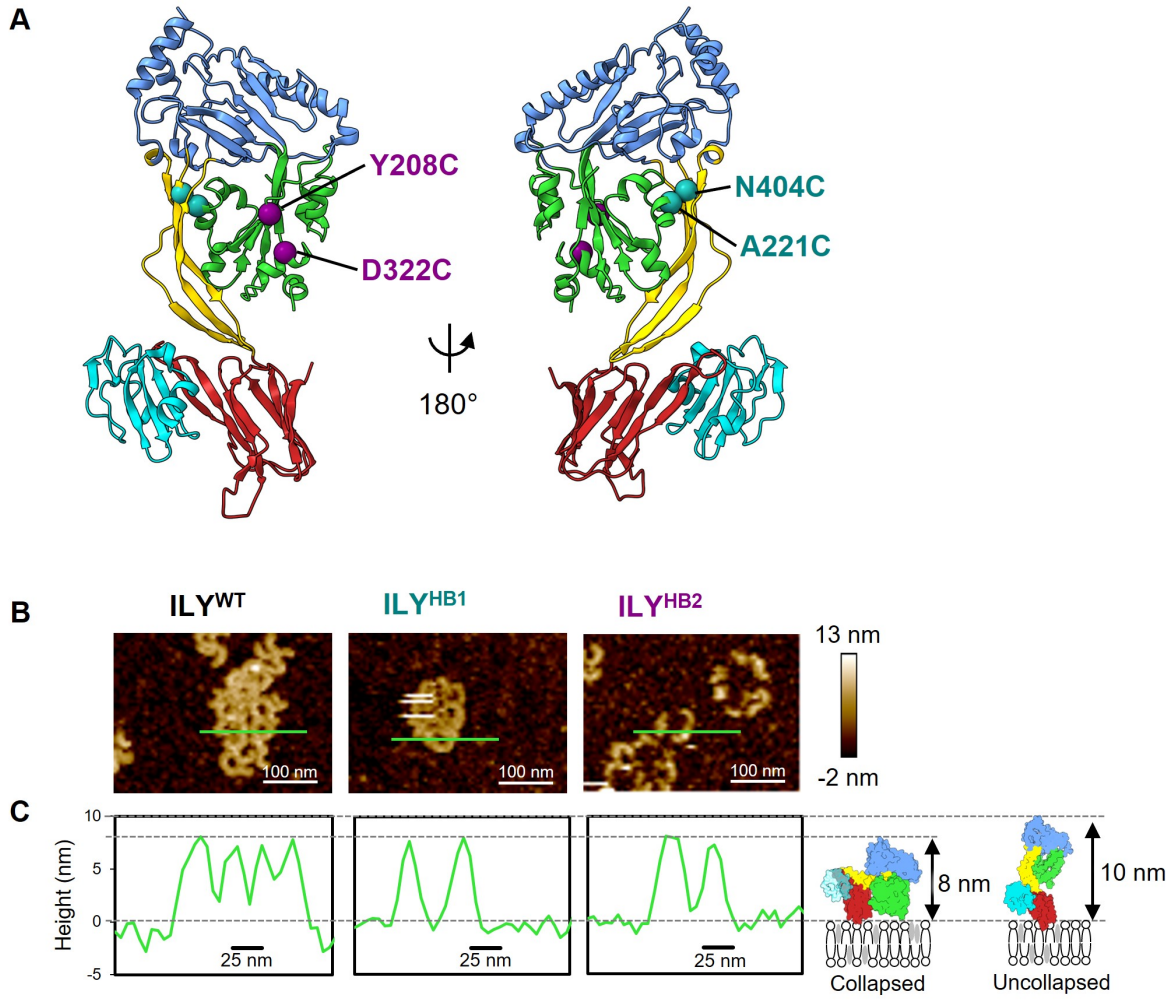


Figure 3.6 | The HB-locked mutants of ILY. **A.** Crystal structure of monomeric, CD59-bound ILY (PDB 4BIK) with the residues mutated to generate the HB-locked mutants. ILY^{HB1}: A221C, N404C. ILY^{HB2}: Y208C, D322C. Blue: D1; yellow: D2; green: D3; red: D4; cyan: CD59. **B.** Representative portions of AFM images of ILY variants on cholesterol-containing supported bilayers decorated with CD59. Each green line represents the position of one of ten height profiles taken for each variant. The proteins were purified by Dr. Shah and the AFM data were recorded by Dr. Parsons. **C.** Height profiles extracted from the corresponding lines in B. The schematics on the right represent the expected domain organisation of the collapsed late prepore (left) and that of the uncollapsed early prepore (right). All ILY variants form assemblies with heights indicative of the collapsed late prepore or pore. Colouring as in panel A. CD59 in the collapsed schematic is transparent as it is unclear whether it is still bound to the complex.

3.3.2 Recombinantly-expressed PLY is active

A histidine-tagged PLY construct was expressed in *E. coli* and purified by cobalt affinity followed by size exclusion chromatography (Figure 3.7) The resulting sample eluted as a monodisperse peak and its purity was confirmed by SDS-PAGE.

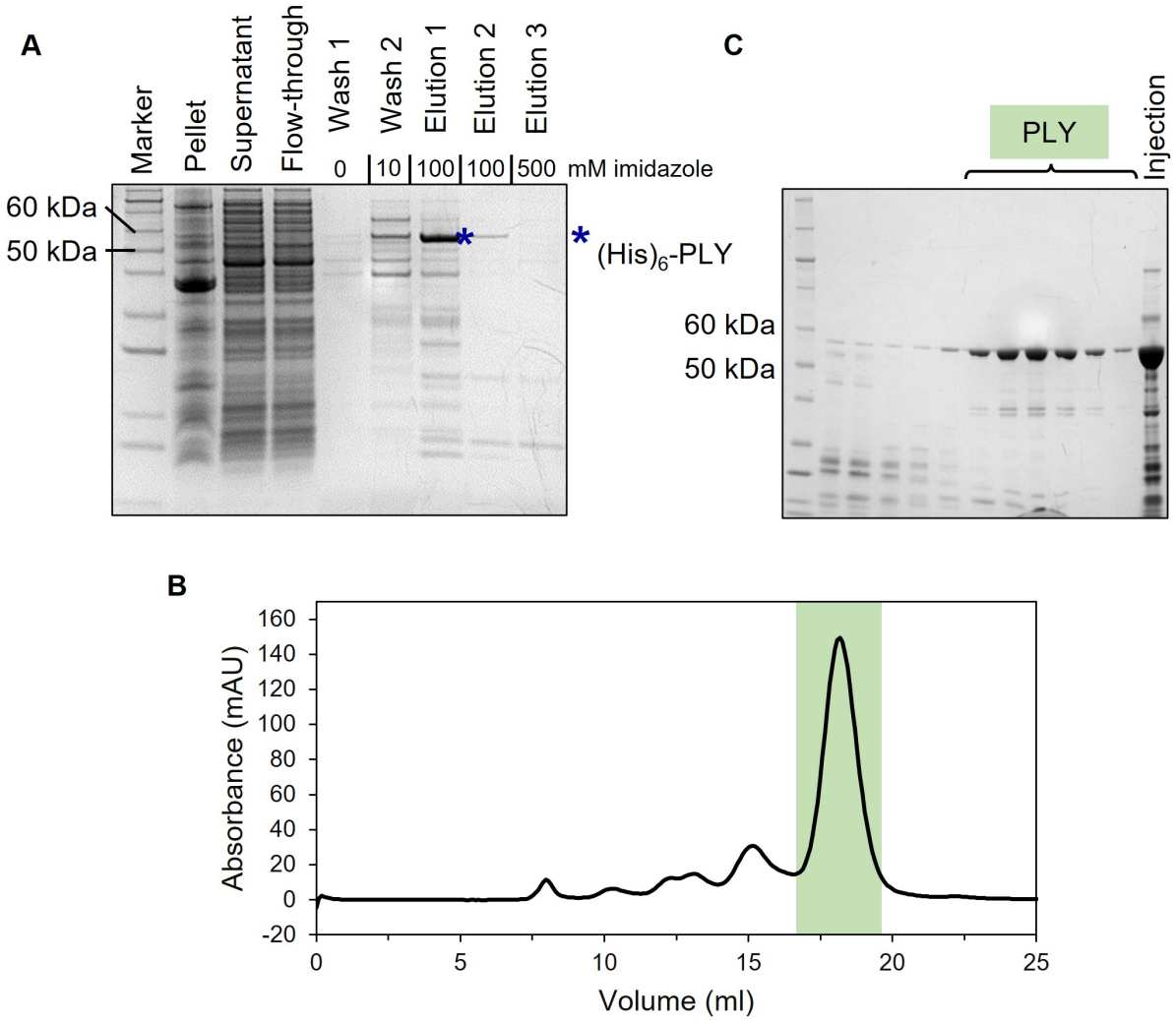


Figure 3.7 | Purification of recombinant histidine-tagged pneumolysin (PLY). **A.** SDS polyacrylamide gel of the cobalt affinity purification. The protein was eluted with imidazole (concentration given above the relevant lanes). The blue asterisk shows the band corresponding to PLY. **B.** Size exclusion chromatography (SEC) profile of PLY. Elutions 1–3 from panel A were pooled and purified with a Superdex 200 increase 10/300 GL column. PLY elutes as a monodisperse peak (green box). **C.** SDS polyacrylamide gel of SEC fractions. The lanes corresponding to the green box in panel B show pure PLY.

To test whether PLY was correctly folded, its ability to oligomerise on lipid monolayers was assessed. The negative stain electron micrograph in Figure 3.8 shows PLY is able to

form both full rings and arcs, as is the case for ILY (Figure 3.6) and other CDCs [188, 210, 220, 221, 222, 223]. The functionality of PLY was further assessed in a liposome lysis assay. Different final concentrations of PLY (ranging from 3 µg/ml to 48.4 µg/ml) were compared to the absence of protein. The results in Figure 3.8 show that recombinant PLY is able to lyse cholesterol-rich liposomes. Lytic activity is dependent on PLY concentration, and a final concentration of 48.4 µg/ml was used for further experiments (Section 3.3.3). The apparent negative lysis in the absence of PLY is likely a result of bleaching of the fluorescent calcein dye over time.

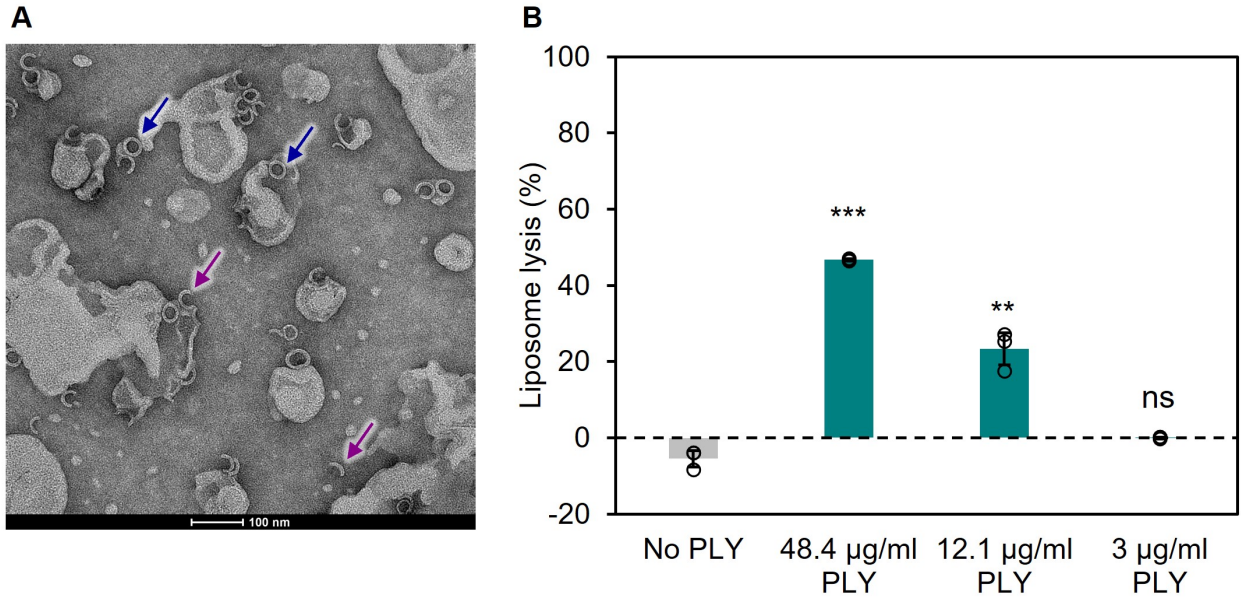


Figure 3.8 | Recombinant PLY is active. **A.** PLY is able to form rings (blue arrows) and arcs (purple arrows) on lipid monolayers. **B.** PLY lyses cholesterol-containing liposomes in a concentration-dependent manner. The negative lysis value in the absence of PLY is probably due to bleaching of the fluorescent dye over time. P-value significance determined by one-way ANOVA and a Bonferroni post-test: **p $<$ 0.01, ***p $<$ 0.001, ns: non-significant. Pairwise comparisons are between the measurements without PLY and the measurements for a given concentration of PLY.

3.3.3 The h-helix tunes the lytic activity of CDCs

The cryoEM structure of the ILY early prepore (ILY^{EP}) revealed that oligomerisation is accompanied by movement of an amphipathic α-helix, the h-helix. Liposome lysis assays realised by Dr. Shah showed that mutating two hydrophobic residues on the top side of the h-helix

(Tyr-366 and Leu-340) to lysine resulted in increased lysis (see Figure 3.5) [50]. The same mutations in the HB-locked mutants (ILY^{HB1} and ILY^{HB2}) enable them to lyse liposomes even though they cannot form the β-barrel pore [50], suggesting that the h-helix disturbs the membrane in the late prepore, before the helical bundles unfurl into β-hairpins. To check whether the h-helix needs to be brought close to the membrane for the mutations to have an effect on lytic activity, the same overactivating mutations were introduced in the early prepore-locked ILY^{EP} and liposome lysis was monitored (Figure 3.9).

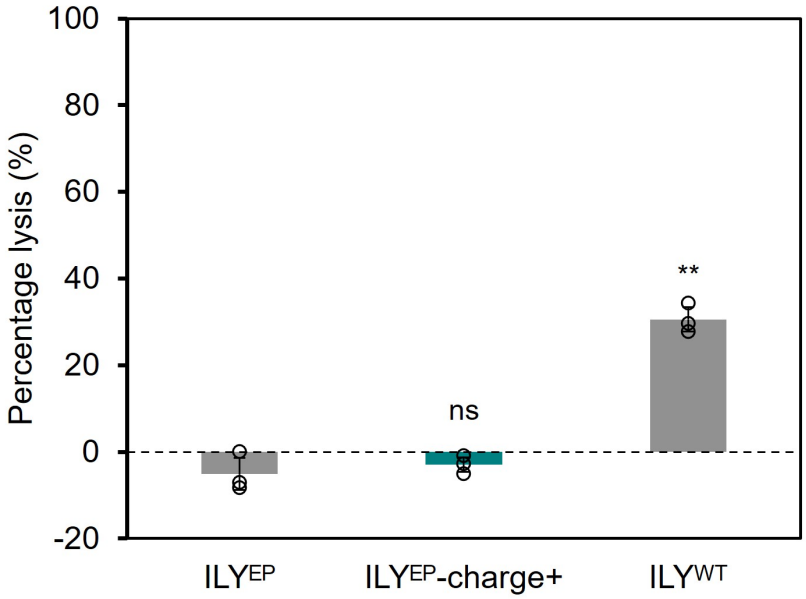


Figure 3.9 | Mutating the h-helix has no effect on the lytic activity of ILY^{EP}. P-value significance determined by one-way ANOVA and a Bonferroni post-test: **$p < 0.01$, ns: non-significant. Pairwise comparisons are with ILY^{EP}. ILY^{WT} was used as a positive control for the plate.

ILY^{EP} is unable to lyse liposomes, as had been reported in the literature [188]. The double h-helix mutation that increases lysis in ILY^{WT}, ILY^{HB1} and ILY^{HB2} do not confer lytic activity to ILY^{EP}. This is consistent with the hypothesis that the h-helix disturbs the membrane at the late prepore stage, once the collapse of the assembly brings it into contact with the lipids.

To assess whether the function of the h-helix is conserved across CDCs, the equivalent mutations to those in Figure 3.5 were introduced in PLY and their effect on lytic activity was measured in a liposome lysis assay (Figure 3.10). When compared to the wild-type,

mutating the polar and charged residues on the bottom face to alanine (PLY-*nocharge*) leads to a statistically significant decrease in lytic activity. Introducing charged residues to the top face (PLY-*charge+*), on the other hand, significantly increases the level of lysis. This pattern is identical to the one observed for ILY (Figure 3.5), suggesting that the role of the h-helix in tuning lytic activity is conserved among CDCs.

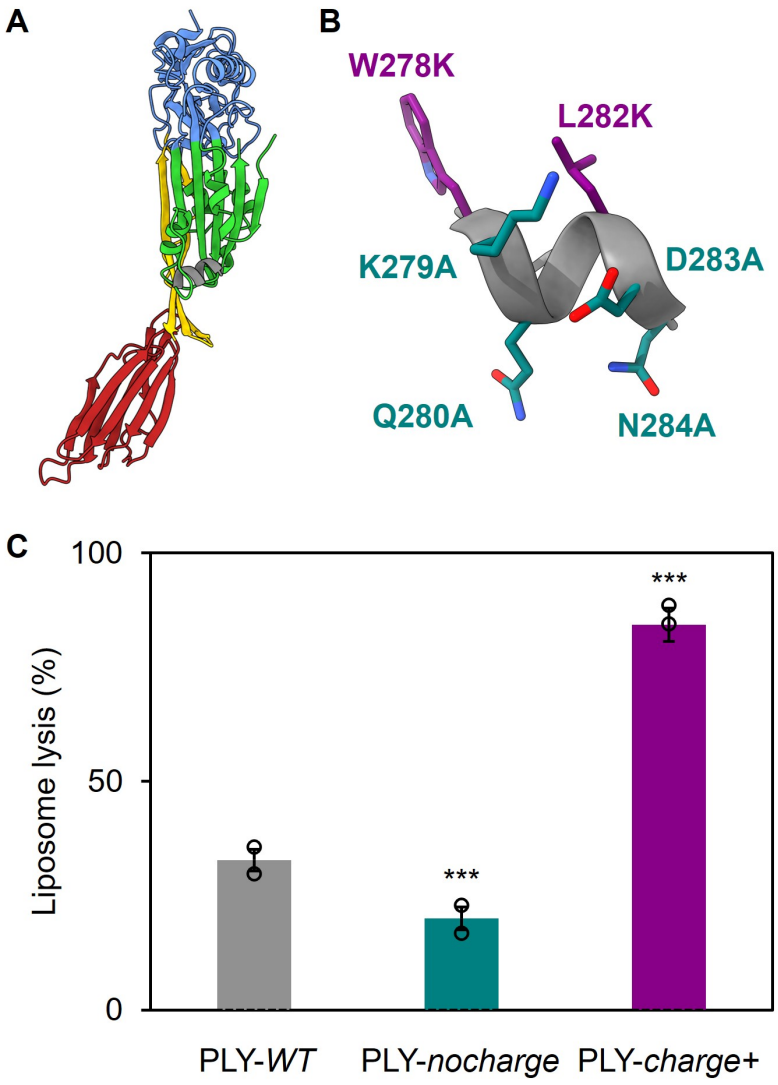


Figure 3.10 | The role of the h-helix is conserved. **A.** Crystal structure of pneumolysin (PDB 5CR6) with the h-helix shown in grey. Blue: D1; yellow: D2; green: D3; red: D4. **B.** Mutations introduced in the h-helix to generate PLY-*nocharge* (K279A, Q280A, D283A, N284A) and PLY-*charge+* (W278K, L282K). **C.** Compared to PLY-*WT*, removing hydrophilic and charged residues from the bottom face of the h-helix (PLY-*nocharge*) lowers the lytic activity. P-value significance determined by one-way ANOVA and a Bonferroni post-test: ***$p < 0.001$. Pairwise comparisons are with PLY-*WT*.

3.3.4 The PLY^{hybrid} model does not explain the function of the h-helix

To explore the structural basis of membrane disruption by the h-helix, coarse-grain models of the three PLY variants described above were simulated on lipid bilayers. As there is no experimental structure of a CDC late prepore, a conformational hybrid model was generated. PLY is the only CDC for which the pore conformation was solved by cryoEM [51], making it an ideal basis for generating the hybrid. The CDC late prepore assembly is shorter than the early prepore, as revealed by AFM [188, 210], so the conformation of D2 in the pore structure is likely a good approximation. Late prepores are resistant to SDS treatment [211, 212], suggesting that the individual CDC monomers establish stronger contacts with adjacent monomers than in the early prepore. This is likely due to the MACPF/CDC domain forming a flat β-sheet similar to the pore conformation, rather than the twisted β-sheet observed in the early prepore. To represent this, the PLY pore structure was modified to remove the residues corresponding to the helical bundles. The helical bundles from the soluble PLY crystal structure were then moved as rigid bodies to replace them. Loops clashing with the new positions of the helical bundles were removed and modelled with SWISS-MODEL [218]. The resulting structure, PLY^{hybrid}, is shown in Figure 3.11. This model was used to generate hexameric assemblies for coarse-grain MD simulations.

To match the biochemical experiments in Section 3.3.3, the h-helix mutations were reproduced and three variants were generated: PLY^{hybrid}-*WT*, PLY^{hybrid}-*nocharge* and PLY^{hybrid}-*charge+*. They were converted to coarse-grain representations using the Martini 2.2 forcefield and simulated on 1:1 (molar ratio) DOPC:cholesterol lipid bilayers to replicate the conditions of the liposome lysis assays. Visual inspection of the trajectories in VMD [169] did not reveal signs of pore formation or lipid disruption, even for the PLY^{hybrid}-*charge+* system, whereas the corresponding ILY^{HB1}-*charge+* and ILY^{HB2}-*charge+* late prepore mutants are able to lyse liposomes [50]. No difference was observed between the three PLY^{hybrid} variants

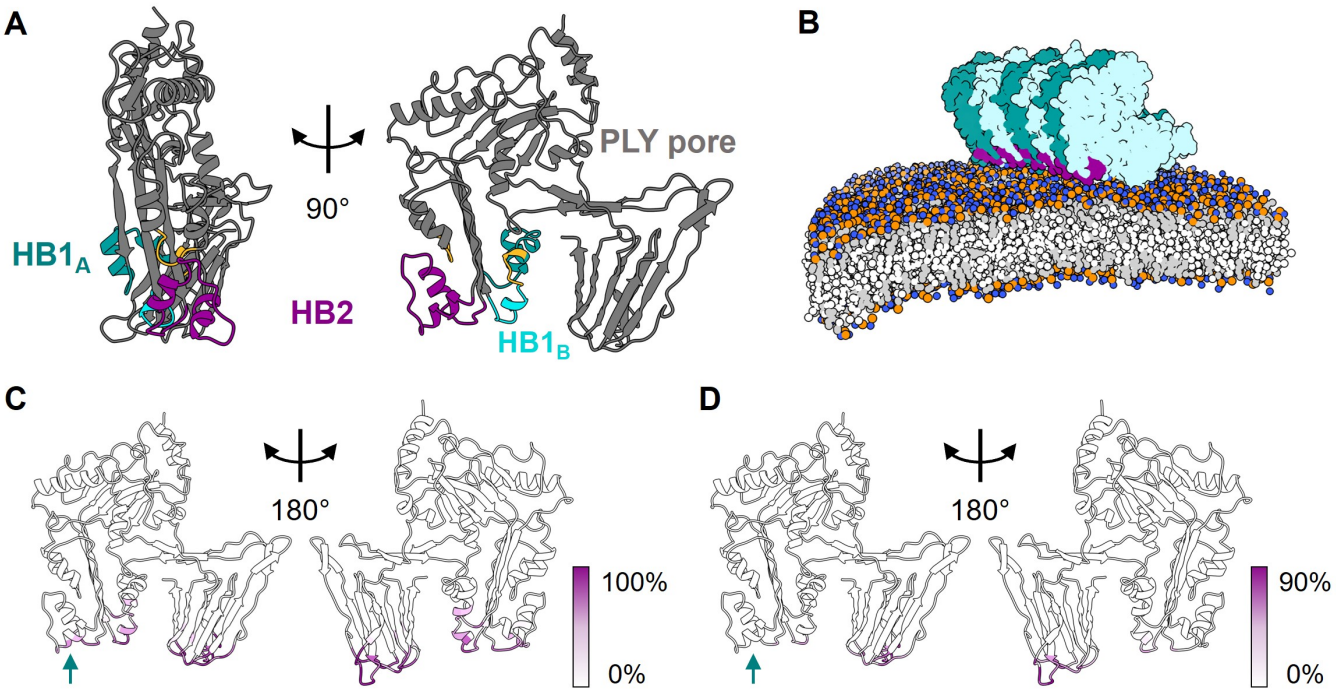


Figure 3.11 | Coarse-grain MD simulations of PLY^{hybrid}. **A.** The PLY^{hybrid} model was generated by combining the PLY pore structure (grey, PDB 5LY6) with helical bundles from the soluble structure (turquoise, cyan and purple, PDB 5CR6). HB1 was split into two segments: $HB1_A$ and $HB1_B$ (turquoise and cyan, respectively) which were moved as independent rigid bodies. HB2 (purple) was moved as a single rigid body. The residues in orange were removed from the structure and modelled using SWISS-MODEL [218] to ensure continuity and prevent steric clashes with the new HB2 position. **B.** Hexameric PLY^{hybrid}-*WT* (turquoise and blue) on a lipid bilayer after a 1-µs production run. **C.** Average DOPC occupancy mapped by residue of PLY^{hybrid}-*WT*, as calculated by PyLipID [185]. **D.** Average cholesterol occupancy mapped by residue of PLY^{hybrid}-*WT*, as calculated by PyLipID [185]. Turquoise arrow in panels C and D point to the h-helix.

in terms of membrane thickness (either global or local) or distance between the h-helix and the lipid bilayer. Protein-lipid interaction analysis with PyLipID [185] revealed very low levels of DOPC occupancy near HB2 and the h-helix, and no interaction with cholesterol. This was true of all h-helix variants, and Figure 3.11C-D shows the results for PLY^{hybrid}-*WT* averaged over three replicates. These results suggest that moving the helical bundles as rigid bodies and fusing them to the pore conformation does not place the h-helix in the correct position to disturb the membrane.

3.4 Discussion

Pore formation by CDCs is a complex process involving several intermediates. The conformational changes necessary to progress along the pore formation pathway can be blocked by engineering disulfide bonds in specific sites. This approach has been used extensively for the study of CDCs and has provided the structures of monomeric ILY bound to CD59 [182] and the oligomeric ILY early prepore [50]. Dr. Shah generated two disulfide-locked mutants (ILY^{HB1} and ILY^{HB2}) to trap the late prepore by preventing the unfurling of the MACPF/CDC helical bundles into transmembrane hairpins. AFM data recorded by the Hoogenboom lab and analysed in this chapter show that those mutants are able to form the collapsed late prepore. Other work has shown that AFM can be used to distinguish prepores from pores: the membrane patch inside the prepore is the same height as around the assembly, whereas a pore is typically associated with a decrease in height when the lipids are removed or a jump of several nm corresponding to a lipid vesicle being ejected [153, 210]. The datasets presented here, however, were not recorded at a high enough resolution to distinguish late prepores from transmembrane pores, but liposome lysis assays carried out by Dr. Nita Shah showed that the HB-locked mutants are non-lytic, suggesting they are trapped in the late-prepore conformation.

ILY^{HB1} and ILY^{HB2} are stalled before unfurling of the MACPF/CDC helices into transmembrane hairpins. While both mutants are able to oligomerise, ILY^{HB2} does not form the fully circular rings observed with ILY^{WT} (Figure 3.6). The disulfide bond in ILY^{HB2} involves a residue in the β-1 strand of the MACPF/CDC domain, which is responsible for establishing contacts between adjacent monomers in the oligomer [50]. This disulfide bond may prevent movement of the β-strand necessary for the prepore to adopt its proper geometry. ILY^{HB1}, however, is able to form full rings with the same geometry as ILY^{WT}, probably because the

engineered disulfide bond links the first helical bundle to D2, neither of which makes contacts with adjacent monomers.

The cryoEM structure of the ILY early prepore solved by Dr. Nita Shah revealed the movement of an amphipathic helix (the h-helix) in one of the helical bundles, which comes to lie parallel to the membrane. Mutagenesis showed that the amphipathicity of the helix determines the lytic activity of ILY: adding positively-charged lysine residues to the hydrophobic face of the helix makes the protein overactive, while replacing the hydrophilic residues with alanine impairs lysis. The addition of charges to the h-helix even rescues the non-lytic phenotype of ILY^{HB1} and ILY^{HB2} despite their inability to form the β-barrel pore (data generated by Dr. Shah and published in [50]). I have shown here that these same mutations have no effect on the lytic activity of the early prepore (Figure 3.9). This suggests that the h-helix plays its role in pore formation at the late-prepore stage, once it is brought close to the membrane.

Molecular dynamics (MD) studies of lipid bilayer electroporation have measured the free energy barrier for pore formation in the range of ∼17–78 kJ/mol, depending on the lipid composition [224, 225]. MD simulations of the melittin peptide, an amphipathic helix that forms pores in membranes, have shown that when several helices deposit on the surface of the membrane, they lower the free energy barrier of pore formation in a concentration-dependent way [226]. When four melittin peptides are adsorbed onto the membrane, the free energy barrier for pore formation is >50 kJ/mol, whereas the addition of a fifth melittin lowers the energy barrier to only ∼15 kJ/mol and a sixth monomer lowers it even further. The biochemical work detailed in this chapter suggests that the h-helix may perform a similar function in the ILY late prepore: its polar and charged residues probably disturb the membrane in a way that lowers the energy barrier for insertion of the transmembrane hairpins and, as a result, formation of a hydrophilic pore across the bilayer. This is consistent with a kinetic model of perfringolysin O pore formation derived from single-molecule data, where a

minimum of four subunits are required for efficient membrane perforation [215].

All CDC crystal structures contain an amphipathic h-helix in HB2 [35, 50, 181, 182, 195, 196, 197, 198, 199, 200, 200] and similar amphipathic helices are found in other pore-forming proteins, e.g. β-defensin 1 [227] and gasdermin [228] (Figure 3.12). This suggests a common mechanism where various pore-forming proteins rely on amphipathic helices to drive lipid bilayer perforation.

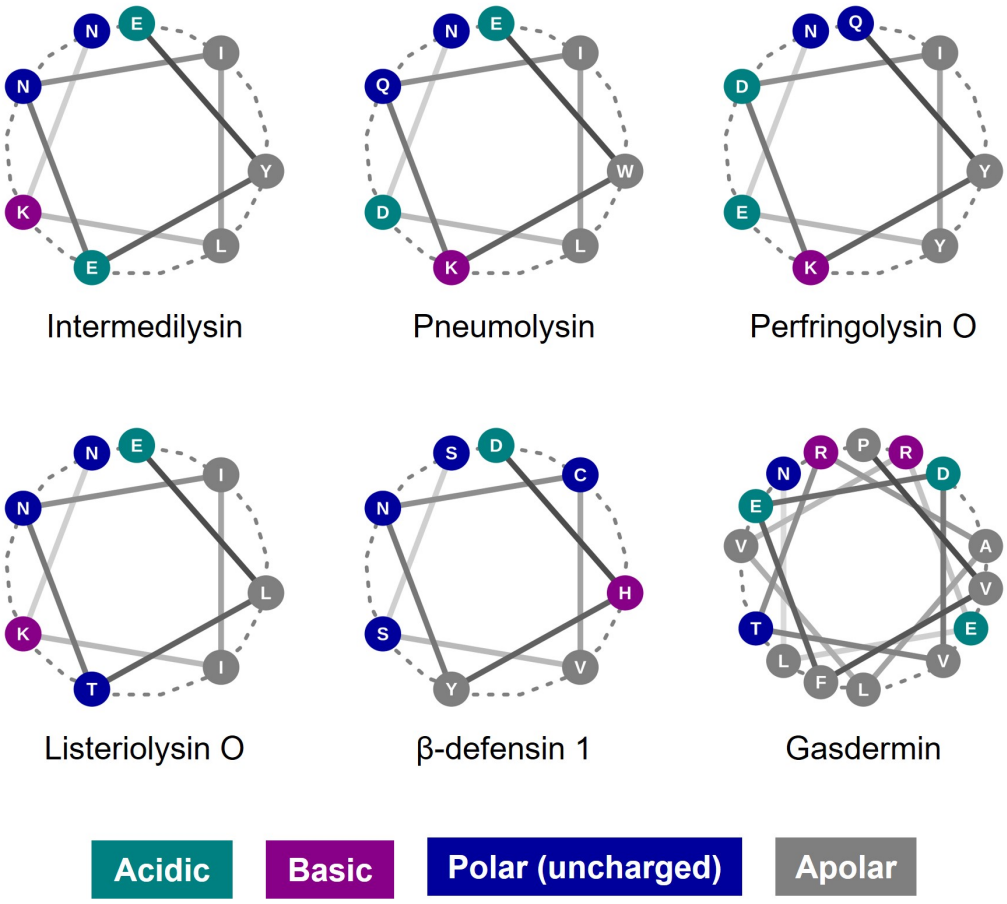


Figure 3.12 | Amphipathic helices are found in various membrane-lysing proteins. Helical wheel representation of the conserved h-helix in CDCs (intermedilysin, pneumolysin, perfringolysin O, listeriolysin O) and equivalent helices in human β-defensin 1 and gasdermin. Plots generated with the NetWheels server [229].

To test whether the function of the h-helix in regulating lytic activity is conserved among CDCs, I purified recombinantly-expressed pneumolysin. Imaging PLY on lipid monolayers by electron microscopy confirmed that it is able to form oligomers with the correct geometry. The activity of those oligomers was then assessed by monitoring their ability to lyse

liposomes filled with a fluorescent dye. Recording PLY lysis at different concentrations (Section 3.3.2) showed that, to achieve comparable levels of activity, the amount of PLY used must be ∼50 times higher than the amount of ILY. Data reported in the literature show that PLY has a much lower lytic activity than the widely-studied CDC perfringolysin O (PFO), and that this effect is even more pronounced at temperatures below 30°C. This may explain the need for a high concentration of PLY, especially since all lysis assays were performed at room temperature.

The PLY h-helix was mutated similarly to that of ILY. Adding charged residues to the apolar side of the h-helix increases the lytic activity of either protein, whereas removing polar and charged residues from the hydrophilic face lowers it. This demonstrates that the proposed role of the h-helix in priming the membrane for hairpin insertion is not confined to ILY. As amphipathic helices are found in other pore-forming proteins, this may be a general strategy for rupturing lipid bilayers.

The structure of the ILY early prepore and the biochemical characterisation of the late prepore, using disulfide-locked variants and h-helix mutants, have added to our understanding of pore formation by CDCs. Our results show that the h-helix, which is believed to be brought close to the membrane upon collapse, disturbs the bilayer. The overall pore formation pathway is shown in Figure 3.13: a soluble, monomeric CDC binds the target membrane (by recognising cholesterol directly or *via* a receptor such as CD59) before oligomerising into the early prepore. The assembly then collapses into the late prepore and the MACPF/CDC domain is brought close to the membrane, allowing the h-helix to lower the energy barrier for pore formation. The helical bundles refold into transmembrane β-hairpins and the β-barrel pore is formed.

To gain structural insight into membrane priming by the h-helix, a hybrid prepore model was generated by grafting the helical bundles of the soluble PLY structure onto the pore

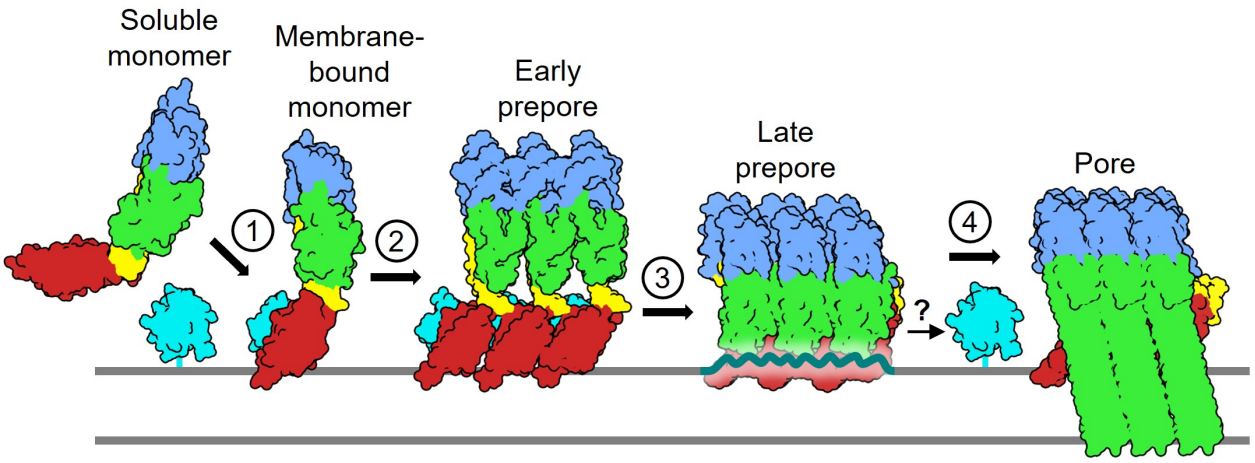


Figure 3.13 | Overview of the CDC pore formation pathway. A soluble monomer (PDB 1S3R) binds a cholesterol-containing membrane (1) *via* either a receptor (e.g. CD59, cyan) or cholesterol directly. The membrane-bound monomer oligomerises (2) to form the early prepore, which then collapses (3) into the late prepore. At this stage, the h-helix is brought close to the membrane and starts disturbing the bilayer (turquoise), priming it for the unfurling of the MACPF/CDC domain helical bundles into transmembrane β-hairpins (4). CD59 is believed to be absent from the final pore but the precise timing of its release is unclear. Blue: D1; yellow: D2; green: D3; red: D4.

conformation, as rigid bodies. The h-helix was then mutated to match the functional assays and hexameric hybrid constructs were simulated using the Martini 2.2 forcefield. This approach did not reproduce the experimental results from the liposome lysis assays: membrane disruption was never observed, and there was no noticeable difference between the h-helix variants. Mapping lipid occupancy per residue (Figure 3.11C-D) reveals that the h-helix establishes very few contacts with the lipid bilayer in the hybrid model. This suggests that simply moving the helical bundles as rigid bodies is not enough to model the PLY late prepore, and the h-helix may adopt an intermediate conformation. Understanding the structural basis of h-helix function will require solving the high-resolution structure of a CDC late prepore.

4 || Towards the structure of the intermedilysin late prepore by cryoEM

Chapter summary

Pore formation by cholesterol-dependent cytolysins (CDCs) is a complex, stepwise process. The work in Chapter 3 contributed to our understanding of the mechanism of action of CDCs by highlighting the role of the late prepore in priming the membrane for pore formation, *via* its amphipathic h-helix. However, the lack of high-resolution structural information regarding the late prepore limits our understanding of this process, as the structural basis for the function of the h-helix remains unknown.

To tackle this question, I carried out some of the first steps towards solving the high-resolution structure of the ILY late prepore by cryoEM. By combining an engineered disulfide lock and the inactivating mutations described in Chapter 3, I assembled the ILY late prepore on lipid nanodiscs. A sucrose density gradient coupled with chemical fixation by two crosslinkers yielded a pure and stable complex for structure determination. I optimised cryoEM grids and recorded datasets at 200 kV and 300 kV. Initial processing of the 300 kV datasets resulted in two 3D reconstructions, potentially corresponding to the early and the late prepores, respectively. The data datasets show high levels of compositional and conformational heterogeneity, limiting the resolution of the maps. However, the work in this Chapter is a significant step towards a structural understanding of the function of the CDC late prepore.

4.1 Introduction

4.1.1 The intermedilysin early and late prepores

Cholesterol-dependent cytolysins (CDCs) are bacterial toxins that oligomerise on the surface of cholesterol-containing membranes to form lytic pores (for a more detailed introduction to CDCs, cf. Section 3.1.1). CDCs are secreted as soluble monomers and undergo a series of complex and irreversible conformational changes to form oligomeric pores. The structure of the first oligomeric species, the early prepore, was solved by cryoEM [50] to a global resolution of 4.6 Å. This structure revealed which conformational changes are associated with oligomerisation, in particular the movement of an amphipathic helix within helical bundle (HB)-2 of the MACPF/CDC domain, called the *h-helix* (Figure 3.3). However, preferred orientation limited the resolution of the cryoEM reconstruction, preventing the inclusion of side chains in the final model. As a result, there is a lack of high-resolution information regarding the early steps of CDC oligomerisation, including the orientations of side chains in the h-helix. The biochemical characterisation of the h-helix in Chapter 3 shows that it is responsible for tuning the lytic activity of CDCs, probably by disturbing the membrane at the late prepore stage, priming it for hairpin insertion. This likely happens once the complex has undergone collapse, where the MACPF/CDC domain is brought close to the membrane. However, coarse-grain simulations of a hybrid late prepore-like conformation where domains were moved as rigid bodies (cf. Section 3.3.4) did not show any interactions between the h-helix and the membrane. This may be because unknown conformational changes must happen within the MACPF/CDC domain to bring the h-helix to its correct position. Fully understanding the mechanism of membrane perturbation by the h-helix requires solving the structure of a CDC late prepore.

In the ILY early prepore structure, CD59 contacts two adjacent ILY molecules. The in-

teraction surface between CD59 and the first ILY subunit is similar to that in the monomeric ILY-CD59 complex (Figure 3.2), where an extended β-sheet is formed across both proteins. In the second interface, an α-helix of CD59 is sandwiched by β-hairpins of the adjacent ILY molecule [50]. However, the resolution of the reconstruction precludes defining the interaction surface at the atomic level. CD59 is believed to be absent from the final pore, but the precise timing of dissociation remains unclear. LaChapelle *et al.* used antibody labelling and FRET data to show that CD59 remains bound to a disulfide-locked ILY prepore [230]. The location of the engineered disulfide bond (linking HB1 to D2) suggests that it would prevent unfurling of the HBs but not collapse, indicating that the authors may have captured the late prepore. However, they failed to confirm this, which complicates the interpretation of their results. In another study, Boyd *et al.* incubated wild-type ILY with CD59-decorated liposomes in the absence of cholesterol [188]. As ILY is able to collapse on, but not lyse liposomes without cholesterol, they considered this system to correspond to the late prepore. In a flotation assay, they observed that some ILY rings dissociated from liposomes, while others remained bound. As both species were SDS-resistant, they were both likely late prepores. The authors conclude that the released ILY rings have dissociated from CD59, while the others are either still bound to it or interacting with the membrane in a CD59-independent way, suggesting that CD59 is released by the late prepore. The results from LaChapelle *et al.* and Boyd *et al.* may be reconciled by the existence of multiple late prepore assemblies, with and without CD59. The ambiguity highlights the difficulty of interpreting biochemical data corresponding to an intermediate in the absence of a high-resolution structure. Moreover, the conformations trapped by disulfide locks are proxies of the intermediates in the pore formation pathway, which may not accurately represent the structural transitions undergone by the wild-type protein. In the future, time-resolved cryoEM may help uncover new, more physiological conformations of pore-forming proteins [231].

Structural studies of CDC oligomers are challenging because they need a membrane on which to assemble. The pneumolysin (PLY) pore structure was solved by cryoEM by forming the complex on liposomes and solubilising it with a detergent before replacing the detergent with an amphipol [51]. The ILY early prepore structure was solved on lipid nanodiscs, disc-like membrane mimetics where lipids are held together by a scaffold protein (cf. Section 4.1.2). This approach was necessary as CDCs are effectively peripheral membrane proteins once they oligomerise, but only become transmembrane proteins when the final pore is formed. Incubating the disulfide-locked ILY early prepore mutant (ILY^{EP}) on nanodiscs decorated with CD59 yielded a complex suitable for structure determination. To solve the structure of the ILY late prepore, a detergent-based strategy may succeed at solubilising the late prepore formed on liposomes, as late prepores are SDS-resistant [211, 212]. However, detergent treatment would likely disrupt interactions between ILY and CD59, preventing any insights in the molecular basis of CD59 release in the pore formation pathway. Previous attempts to capture a disulfide-locked late prepore on nanodiscs in the Bubeck lab were unsuccessful as the sample was prone to aggregation. This is probably a result of the h-helix disrupting the lipid bilayer. Indeed, while the disulfide-locked late prepores (ILY^{HB1} and ILY^{HB2}) do not lyse liposomes, the lipids enclosed in nanodiscs are likely too constrained to accomodate the disruption caused by the h-helix. This may expose hydrophobic regions of the protein, consistent with the sample aggregation observed by previous lab members (data not shown). However, the work in Chapter 3 and published in [50] presents new mutations that impair the function of the h-helix. These inactivating mutations resulted in stable late prepore complexes on lipid nanodiscs, providing a new opportunity to solve the structure of the penultimate CDC assembly, the last step of pore formation for which there is no structural information.

4.1.2 Sample preparation for cryoEM

Membrane-anchored CD59

As CD59 binding is one of the earliest steps of ILY pore formation, a membrane mimetic decorated with CD59 was necessary. *In vivo*, CD59 is attached to the plasma membrane *via* a GPI anchor, but GPI-anchored proteins are challenging to purify. Instead, Hill *et al.* [232] developed a protocol whereby CD59 with a C-terminal cysteine is expressed in *E. coli* and refolded from inclusion bodies. The C-terminal cysteine is then reacted with a 17-residue cytotopic peptide (KSSKKSPSKKDDKKPGDC), modified with two myristoyl moieties (Figure 4.1). The resulting chemically modified protein (referred to as $CD59^{cyto}$ henceforth) was a generous gift from Prof. Richard Smith (King's College London).

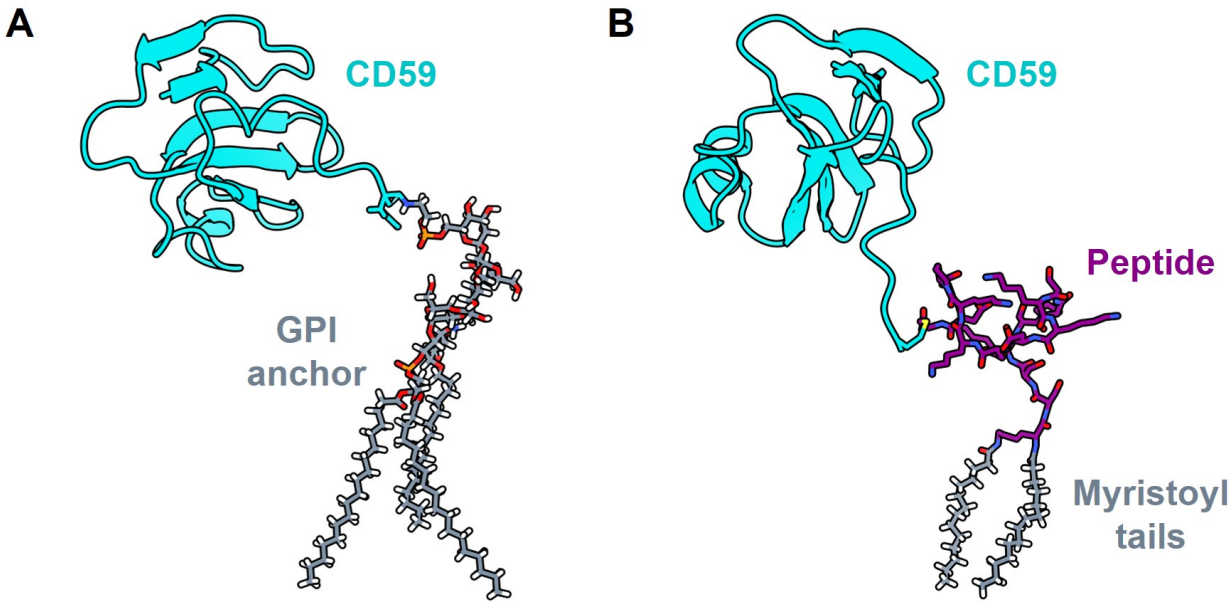


Figure 4.1 | **A.** CD59 covalently attached to its native GPI anchor with three lipid tails (cf. Section 2.1.2 and Figure 2.5). This model was built with the LEaP facility in AmberTools21 [66] and is not an experimentally solved structure. **B.** CD59 construct developed by Hill *et al.* [232] and gifted by Prof. Richard Smith (King's College London). Recombinantly-expressed CD59 is linked to a cytotopic peptide (purple) decorated by two myristoyl moieties (grey) *via* a disulfide bond involving a C-terminal cysteine. The structure of the peptide was predicted with PEP-FOLD3 [233] for illustrative purposes only.

The myristoylated cytotopic peptide inserts spontaneously in lipid membranes by simple incubation at room temperture. This reagent was used to solve the structures of the ILY

early prepore [50] and of the inhibited membrane attack complex [143], indicating that the peptide is flexible enough to allow CD59 to adopt two very different orientations relative to the membrane (cf. Section 2.3.1). The construct enables the formation of functional ILY pores, as it was used in all liposome lysis assays presented in Chapter 3 and published in [50].

Lipid nanodiscs as model membranes

One of the challenges of studying peripheral membrane proteins is the need for appropriate membrane mimetics. Lipid nanodiscs are a popular solution as they can be generated to contain lipids of a chosen composition, before incubating them with the protein of interest. Nanodiscs are disc-like bilayers surrounded by a membrane scaffold protein (MSP), which shields the hydrophobic tails of lipids from the solvent (Figure 4.2A).

Nanodiscs can be made by incubating the MSP of choice, which dictates the size of the final discs, with lipids solubilised in detergent [234, 235, 236]. The mixture also includes BioBeads (Bio-Rad), non-polar adsorbent beads, which chelate the detergent (Figure 4.2B). The MSP then spontaneously wraps around the lipids, resulting in assembled nanodiscs that can be purified by size-exclusion chromatography.

Negative stain EM

Electron microscopy (EM) can be performed at cryogenic temperatures, for high-resolution structure determination, or at room temperature. The latter generally takes the form of negative stain EM, whereby the sample is applied onto a grid and stained with heavy metal, e.g. uranyl acetate. As the grid dries, the original protein sample is destroyed, but a cast of it remains in the stain, which can be imaged in the electron microscope. The resolution is limited by the grain size of the stain, around 20 Å for uranyl acetate, and high-resolution information is lost. However, the speed of making and imaging negative stain grids and the high contrast

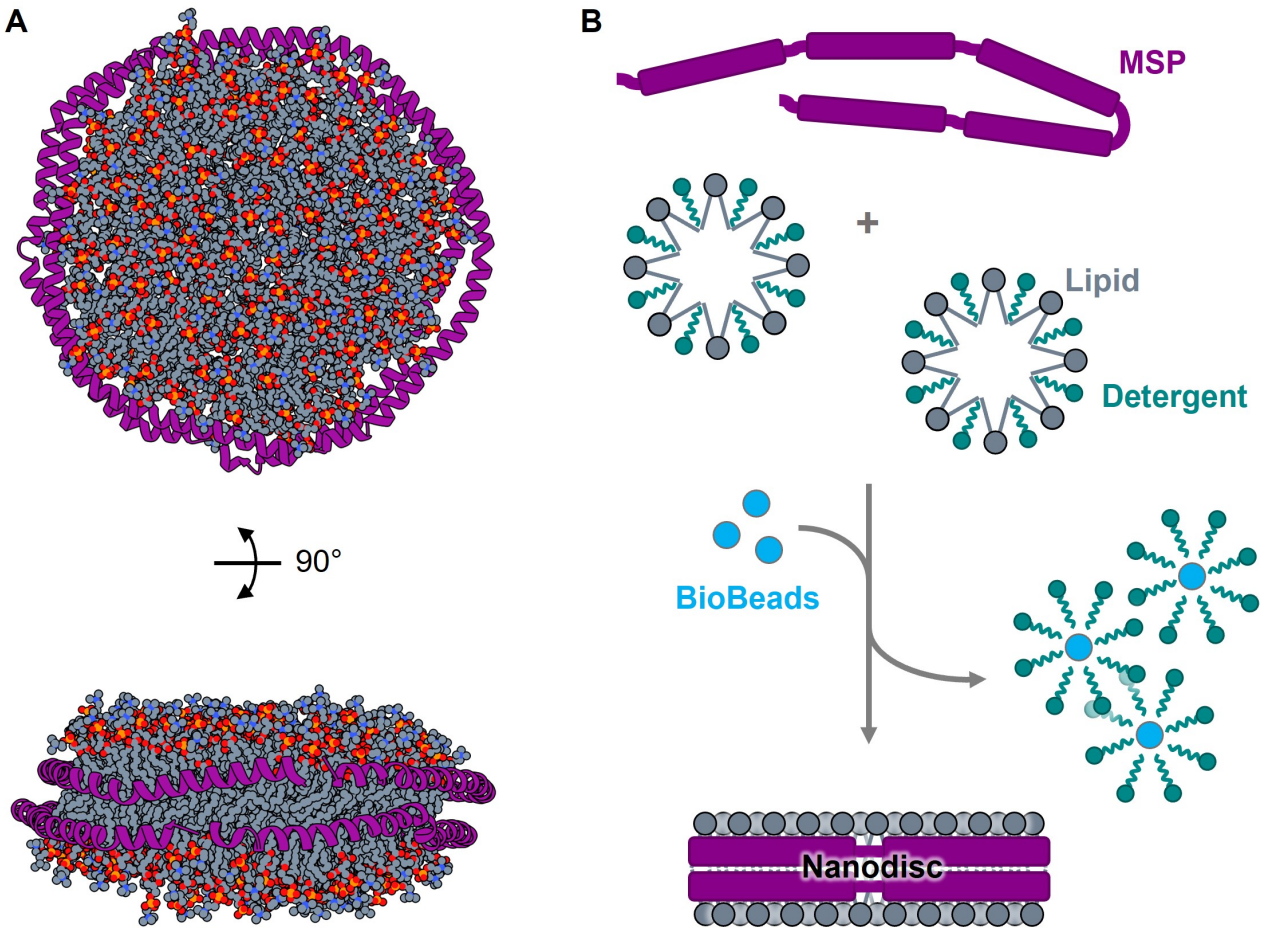


Figure 4.2 | Lipid nanodiscs as a model membrane system. **A.** Nanodiscs are lipid bilayers (grey, orange, red and blue) surrounded by a membrane scaffold protein (MSP, purple). PDB 6CM1 with only the MSP and lipid molecules shown. **B.** Nanodiscs can be assembled by incubating the chosen MSP with detergent-solubilised lipids. As the detergent is removed from solution with BioBeads, the MSP spontaneously wraps around the lipids to form nanodiscs.

resulting from the stain make it a powerful diagnostic tool for sample preparation.

Cryo-grid preparation and optimisation

The first step of sample preparation for cryoEM is generating homogeneous material in sufficient quantities for imaging. Negative stain EM can be used to check the quality of the sample by assessing the stability and integrity of complexes in various conditions, oligomeric homogeneity, and concentration. The appropriate sample can then be vitrified on cryo-grids for cryoEM structure determination. A variety of methods are available for preparing cryo-grids [237, 238, 239, 240], and the one used in this project is shown in Figure 4.3. A metal grid (usually made of copper or gold with a carbon film) is 'glow-discharged': it is placed in

a chamber with a modest vacuum and the remaining gas is ionised, forming radicals that interact with the grid to make its surface hydrophilic. The sample is then placed on the grid and the excess liquid is blotted away, before rapidly plunging the grid in liquid ethane kept at liquid nitrogen temperature. The speed of the freezing process allows the formation of a thin layer of vitreous ice, rather than ice crystals.

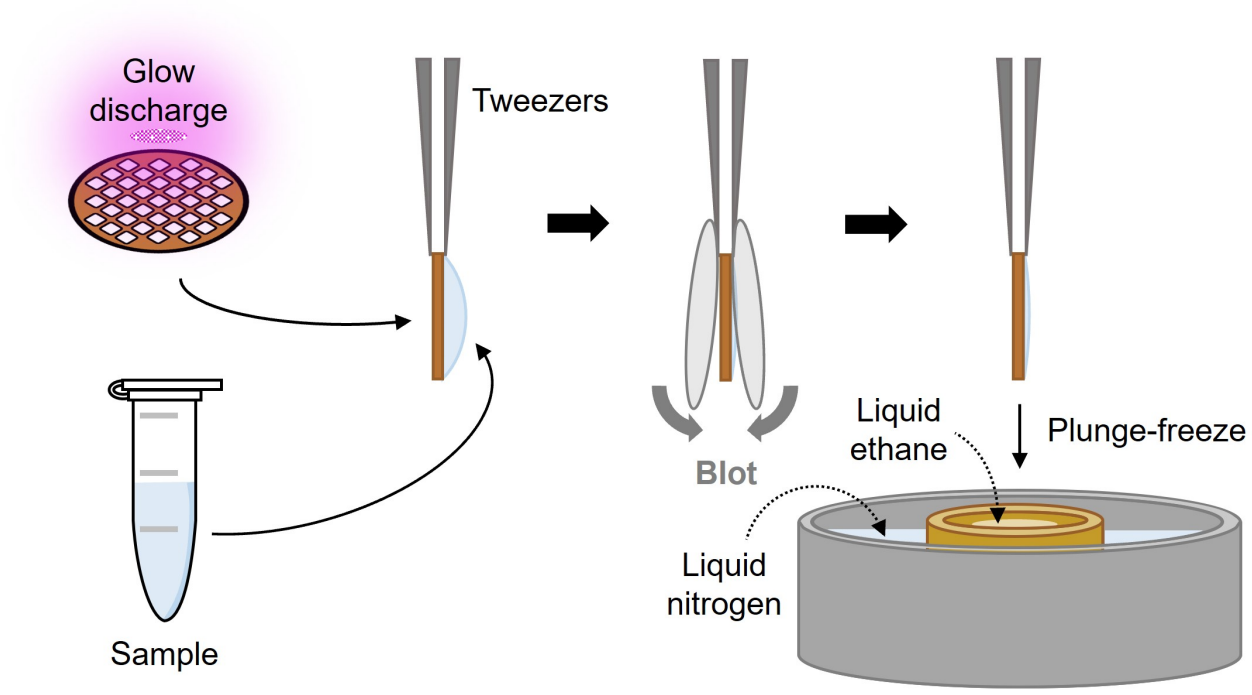


Figure 4.3 | CryoEM grid preparation. A cryo-grid is glow-discharged to make its surface hydrophilic. A sample is then applied to one side of the grid, before blotting away the excess liquid. The resulting grid with a thin layer of sample in buffer is rapidly plunge-frozen in liquid ethane maintained at liquid nitrogen temperature before storing or imaging.

Many parameters in grid preparation (e.g. type of grid, sample concentration, buffer composition, presence or absence of grid support, blot time and force) influence the behaviour of particles in the vitreous ice [241]. Ideally, the protein concentration should be high to maximise the number of particles imaged during each exposure, while avoiding particle overlap. Particles should be found inside the grid holes rather than adsorbed to the carbon edge. Particles should also ideally show random orientations to fill in the entirety of the 3D Fourier transform of the sample. Moreover, the ice should be thin to reduce the background, but thick enough to embed the particles fully and keep them from the air-water interface, which

could induce preferred orientation or even complete protein denaturation. The wide variety of parameters dictating particle distribution in cryoEM grids is matched by a large number of strategies to overcome the challenges of sample preparation, including crosslinking the sample to alter its charge properties, changing the buffer composition [241], coating grids with a physical substrate for particles to adsorb on [242], adding antibodies, nanobodies or megabodies [243], using self-wicking grids or a different sample delivery method when freezing [240], etc.

4.1.3 Data collection and image processing

Once sample preparation has been optimised and high-quality grids have been generated, they can be imaged in a transmission electron microscope. Thousands of micrographs are collected and processed. Direct electron detectors record multiple frames during a single exposure, capturing the movement of particles induced by a combination of factors, including stage instability, buckling of the ice and charging of the support film [244]. This is corrected computationally by tracking the particle motions and averaging frames to compensate for it, in a process called *motion correction*. The next step of image processing is fitting the parameters of the CTF to each motion-corrected micrograph. Micrographs with poorly estimated CTFs are generally excluded from downstream processing steps, as errors in CTF estimation propagate to high spatial frequencies, where oscillations in the CTF accelerate.

Pre-processed micrographs can then be used to pick particles. A variety of approaches can be taken at this stage, for instance manual picking requires the user to identify each particle on all the micrographs. With the recent advances in data collection speed, the number of micrographs routinely recorded have made this approach difficult and computational algorithms have been devised to automate particle picking. For instance the blob picker utility of cryoEM processing software Cryosparc recognises Gaussian signals in micrographs

as particles [60], while the program Topaz uses neural networks trained on manually picked particles (either directly from the dataset or from a general training set) [245]. As the signal-to-noise ratio of cryoEM data is low, it is often challenging to distinguish real particles from spurious picks on micrographs, which may correspond to ice contamination, broken particles, aggregates, etc. This is remedied by 2D classification, whereby particles are classified by similarity, aligned and averaged to increase signal. This allows the removal of unwanted 'junk' particles, which are grouped in classes with poorly defined features or lacking protein features entirely. This can also provide information on the particles as different views can be identified and subunits or domains may are often distinguishable.

'Good' particles are those which belong to 2D classes with well defined high-resolution features, for instance visible secondary structure elements, and a flat background, indicative that the alignment is driven by protein signal rather than noise. Once 'good' particles have been selected, they can be used to generate one or more *ab initio* reconstructions as explained in Section 1.2. These initial models are then refined: the angles between projections are changed in small increments until convergence. To minimise the template bias and the impact of noise build-up, the data are randomly split in two, and two half-maps are refined independently [55, 246]. The correlation between the two maps in Fourier space is calculated as a function of spatial frequency, resulting a the 'gold standard' Fourier Shell Correlation (FSC) curve. At higher spatial frequencies, the data are noisier and alignments are less accurate, resulting in a decrease in FSC. In cryoEM, the spatial frequency at which the FSC reaches the value of 0.143 is generally taken to be the overall resolution of a reconstruction [247].

Heterogeneity is a common limiting factor in cryoEM structure determination. It can arise from differences in the composition of each particle in the case of a complex (e.g. number of subunits for an oligomeric assembly, presence or absence of a binding partner), differ-

ences in conformation between particles (e.g. several intermediates may be present in the dataset) or inherent flexibility of the macromolecules. CryoEM reconstructions require accurate alignment of particles, which is challenging when they are heterogeneous. To overcome this problem, several approaches have been devised to separate discrete particle populations, a process sometimes referred to as *in silico* purification. 3D classification is one such method, where particles are iteratively separated into classes based on their similarities, either with or without refining angular assignments.

4.1.4 Aims

The work presented in this chapter aims to provide the first steps towards solving the structure of the ILY late prepore on lipid nanodiscs using the ILY^{HB1}-*nocharge* mutant. The specific goals were to assemble and purify the complex for cryoEM structure determination. Sample quality and integrity were checked by negative stain EM imaging. Two data collections on a 200 kV instrument were used to generate 3D reconstructions and assess particle distribution. Finally, two datasets were collected at 300 kV and the initial steps of image processing were performed to determine whether the assembly corresponded to the late prepore.

4.2 Materials and Methods

4.2.1 Protein expression and purification

The ILY^{HB1}-*nocharge* mutant is described in Chapter 3 and its expression plasmid was a gift from Dr. Nita Shah. The protein was expressed in *E. coli* BL21 (DE3) and purified following the same protocol as in Section 3.2.3.

The expression plasmid for the membrane scaffold protein 2N2 (MSP2N2) [236] was purchased from Addgene (Plasmid #29520). The protein was expressed in *E. coli* BL21 (DE3)

cells and purified following a modified version of the protocol in [50]. Cells were grown at 37°C in 2X YT media to an OD_{600} of 1.0 to 1.2. Expression was induced by adding 0.4 mM (final concentration) of IPTG and lowering the temperature to 28°C for 4 hours. The cells were then pelleted and resuspended in 40 mM Tris-HCl pH 7.4 supplemented with cOmplete Protease Inhibitors (Roche), before the addition of 1% (final concentration) TritonX-100 and lysis by sonication. The lysate was cleared by centrifugation and the supernatant was supplemented with 300 mM NaCl prior to incubating it with TALON Cobalt resin (Clontech), pre-equilibrated in 40 mM Tris-HCl pH 8.0 and 300 mM NaCl. Protein bound to the resin was washed first with 40 mM Tris-HCl pH 8.0, 300 mM NaCl and 0.5% TritonX-100, then with 40 mM Tris-HCl pH 8.0, 300 mM NaCl and 50 mM sodium cholate. MSP2N2 was eluted in three 4-ml fractions with 40 mM Tris-HCl pH 8.0, 300 mM NaCl and 100 mM imidazole, then in a further three 4-ml fractions with 300 mM imidazole. The eluted fractions were pooled, concentrated and further purified by size-exclusion chromatography with a Superdex 200 Increase 10/300 GL column (GE Healthcare) in 40 mM Tris-HCl pH 8.0, 100 mM NaCl and 0.5 mM EDTA.

4.2.2 Nanodisc preparation and ILY complex assembly

Nanodiscs were generated according to the protocol in [50]. 1.875 mg of lipids (DOPC and cholesterol in a 1:1 molar ratio) were dried and resuspended in 40 mM Tris-HCl pH 8.0, 100 mM NaCl, 0.5 mM EDTA and 64 mM sodium cholate, with four cycles of freeze-thawing followed by 20 min of sonication. Solubilised lipids were mixed with MSP2N2 at 10:1 lipid:protein molar ratio and incubated on ice for 30 min. Sodium cholate was removed by adding 108 mg of BioBeads (Bio-Rad) pre-activated in methanol according to the manufacturer's instructions and incubating overnight at 4°C. Nanodiscs were purified by gel filtration using a Superose 6 Increase 10/300 GL column (GE Healthcare) in 40 mM Tris-HCl pH 8.0, 100 mM

NaCl, 0.5 mM EDTA.

The ILY-CD59 oligomer was assembled on nanodiscs by first incubating ∼50 nM of MSP2N2 nanodiscs (as estimated by absorbance at 280 nm) with 330 nM $CD59^{cyto}$ at room temperature for 20 min. ILY^{HB1}-*nocharge* was added to a final concentration of 20 µg/ml and incubated at room temperature for 30 min.

4.2.3 Sucrose density centrifugation and sucrose removal

The assembled complexes on nanodiscs were purified using sucrose density gradient centrifugation, following a protocol adapted from [248] and [143]. The complexes were loaded onto linear 10-30% (w/v) sucrose gradients in a HEPES buffer (20 mM HEPES pH 7.4, 120 nM NaCl), which was later modified to include a 0-0.15 mM glutaraldehyde (ThermoFisher) gradient or a combination of 0-0.15 mM glutaraldehyde and 0-0.3 mM bis(sulfosuccinimidyl) suberate (BS^3) (ThermoFisher) gradients. The gradients were centrifuged for 18 hrs at 4°C and 39,000 rpm in an SW60-Ti rotor with minimal acceleration and no break. The gradients were manually fractionated into 150-µl fractions.

Several approaches were attempted to remove the sucrose from the fractions of interest. Fractions were passed through a 0.5-ml 40 kDa MW cut-off Zeba spin column (ThermoFisher) pre-equilibrated in either HEPES buffer (20 mM HEPES pH 7.4, 120 nM NaCl) or Tris buffer (200 mM NaCl, 20 mM Tris-HCl) according to the manufacturer's instructions. Alternatively, they were dialysed in a 0.5-ml Slide-A-Lyzer MINI 20 kDa MW cut-off dialysis cup (ThermoFisher) against HEPES buffer (20 mM HEPES pH 7.4, 120 nM NaCl) according to manufacturer's instructions for 2 hrs at room temperature or 2 hrs followed by an overnight incubation in fresh buffer at 4°C. The last strategy was to concentrate the fractions at 2,000 g for 30 min in a 0.5-ml 30 kDa MW cut-off Pierce concentrator (ThermoFisher) before 3 rounds of dilution-concentration in 20 mM HEPES pH 7.4, 120 nM NaCl.

4.2.4 Negative stain EM imaging and image processing

For negative stain EM imaging, samples were applied onto glow-discharged carbon-coated copper grids (Agar Scientific) and stained with 2% uranyl acetate. For screening, grids were imaged on a Tecnai G2 Spirit Twin electron microscope (120 kV) equipped with a LaB6 filament and a 2 k $\times$ 2 k EAGLE CCD camera (FEI). The negative stain data collection was run on a Tecnai T12 Twin electron microscope (120 kV) equipped with a LaB6 filament and a 2 k $\times$ 2 k F216 CCD camera (TVIPS). 550 micrographs were collected at a 50,000$\times$ magnification with a pixel size of 2.313 px/Å, in spot scan mode with a 10 s stage shift delay and a defocus range of -3.5 µm to -1.0 µm.

Micrographs were imported to Cryosparc v2.13.2 [60] and CTF-estimated with CTFFIND4 [249]. The blob picker utility was used to pick particles before removing 'junk' particles with several rounds of 2D classification. Further rounds of 2D classification separated complexes with trimers and tetramers of ILY^{HB1}-*nocharge*. *Ab-initio* reconstruction was performed using the tetramer particles with four classes, and the particles from the best class were refined against the corresponding volume.

4.2.5 CryoEM grid preparation and imaging on a 120 kV instrument

Graphene oxide (GrOx)-coated grids were prepared by plasma cleaning copper Quantifoil 1.2/1.3 grids (Agar Scientific) for 40 s before applying 3 µl of 0.1 mg/ml GrOx (ThermoFisher) to the carbon mesh side. After a 60 s incubation, grids were washed in water and the application-wash process was repeated twice, before two final washes. Grids were left to dry in air for 60 min before sample application.

Unsupported grids (Quantifoil 1.2/1.3, UltrAuFoil 1.2/1.3 [Agar Scientific] or Gold Lacey [Agar Scientifc]) were glow-discharged for 40 s before sample application. All grids were prepared using an FEI Vitrobot at 22°C and 95% humidity. Freezing parameters were optimised

for each grid type. The GrOx-coated grids used for data collection on a 200 kV instrument were generated by applying 3.5 µl of sample (at around twice the ideal concentration for negative stain), waiting for 60 s and blotting for 2.5 s with a blot force of -2 before plunging in liquid ethane. The Gold Lacey grids used for data collection (both at 200 kV and 300 kV) were generated by applying 4 µl of sample at a concentration around 5 times higher than for the GrOx-coated grids. Grids were then immediately blotted for 2 s with a blot force of -2 before plunging in liquid ethane. Grids were screened in low-dose mode on a Tecnai T12 Twin electron microscope (120 kV) equipped with a LaB6 filament and a 2 k $\times$ 2 k F216 CCD camera (TVIPS), or on a Glacios (200 kV) instrument equipped with a field emission gun and a FALCON 4 direct electron detector.

4.2.6 200 kV and 300 kV cryoEM data collections

200 kV data were collected on a Glacios transmission electron microscope equipped with a field emission gun and a FALCON 4 direct electron detector at 100,000$\times$ nominal magnification (pixel size: 1.1 Å/px).

300 kV data were recorded at the electron Biological Imaging Centre (eBIC), with a Titan Krios equipped with a Gatan BioQuantum K3 direct electron detector and energy filter, at 81,000$\times$ magnification (pixel size 1.072 Å/px). 11,855 micrographs were collected without tilting the grid, and 7,535 micrographs were collected at a -30° tilt.

4.2.7 CryoEM image processing

Pre-processing

Micrographs from each dataset were imported into Cryosparc v3.3.2 [60] before motion correction [250] and patch CTF estimation with CTFFIND4 [249], with each dataset being pre-processed independently. Micrographs were curated based on the quality of the CTF esti-

mation and those with CTF fits reaching resolutions below 8 Å (for the 200 kV datasets) or 6 Å (for the 300 kV datasets) were excluded.

200 kV Gold Lacey dataset

For the 200 kV Gold Lacey dataset, particles were initially picked with the blob picker utility of Cryosparc and extracted with a binning factor of 4. 'Junk' particles were removed with several rounds of 2D classification and side views were used for template picking, before extracting new picks binned by 4. Particle picks were merged, duplicates were removed based on distance and 'junk' particles were removed by 2D classification. Particles classifying as pentamers were used for *ab initio* reconstruction with four classes, the class most resembling the early prepore was excluded and the other particles were used to generate four new *ab initio* models. The best class (i.e. the class with the best-defined features) was improved by a first round of homogeneous refinement followed by a second round using all particles from the second *ab initio* reconstruction job. Particles were 3D classified with principle component analysis initialisation and the best class was taken forward for homogeneous refinement, followed by local refinement using a mask around the three central monomers.

200 kV GrOx dataset

2D classes from the 200 kV Gold Lacey dataset were used for template picking with the GrOx dataset, particles were extracted binned by 4 and 'junk' particles were removed by 2D classification. Three processing strategies were attempted. Four *ab initio* classes were generated and the best one was used for homogeneous refinement. Alternatively, the particles were homogeneously refined against the final reconstruction from the 200 kV Gold Lacey dataset (low-pass filtered to 40 Å). Finally, particles from both the 200 kV Gold Lacey and GrOx datasets were merged and used for *ab initio* reconstruction with two classes and the best class was used for homogeneous refinement.

300 kV Gold Lacey dataset

The processing workflow for the 300 kV Gold Lacey dataset is detailed in Section 4.3.5 and Figure 4.18. Briefly, tilted and untilted micrographs were pre-processed separately. Particles were picked from the untilted dataset with the blob picker utility. Manual picking was also performed to select for side views of the complex before using those views to train a Topaz model. Particles from the blob picker and Topaz were then extracted binned by 4, merged and duplicates were removed. 'Junk' particles were removed by 2D classification and 'good' particles were used to train another Topaz model for picking from the tilted dataset. All particles were merged and cleaned by 2D classification. Four *ab initio* reconstructions were generated (from 321,472 particles) and the best two classes were used for homogeneous refinement separately. The resulting models were used as input volumes (twice each) for heterogeneous refinement of all particles to yield four reconstructions. The two models with the best-defined features and which showed the least contribution from noise (with 111,319 and 117,371 particles respectively) were then processed separately. In each case, homogeneous refinement preceded particle unbinning, followed by a round of homogeneous refinement of the unbinned particles against the previous, binned reconstruction. Per-particle CTF refinement was performed before non-uniform refinement [251] against the previously generated model.

To assign a conformational state to each reconstruction, the structure of the ILY early prepore (PDB 6ZD0) [50] was rigid-body fitted into the density with ChimeraX [170] and the local map-model correlation was calculated with Phenix [252]. For the second reconstruction (made from 117,371 particles), a trimer of the MACPF/CDC domain and a trimer of domain 4 (D4)-CD59 were generated from the early prepore structure. They were either aligned to the entire prepore model fitted into the density, or individually rigid-body fitted in ChimeraX. The local map-model correlation was calculated with Phenix in both cases.

4.3 Results

4.3.1 Purifying ILY^{HB1}-*nocharge*, MSP2N2 and nanodiscs

The ILY^{HB1} mutant (A221C, N404C) contains a disulfide bond locking the first helical bundle (HB1) and domain 2 (cf. Section 3.3.1 and Figure 3.6). Atomic force microscopy shows that it can form 8-nm-tall assemblies, consistent with the collapsed late prepore. As the h-helix is part of HB2 and the ILY^{HB2} mutant does not form full rings on supported bilayers (Figure 3.6), it was not an ideal target for studying the late prepore. The inactivating h-helix mutations presented in Chapther 3 (E337A, N338A, K341A, N342A) were also expected to increase the stability of the complex by minimising nanodisc lysis. The The ILY^{HB1}-*nocharge* was expressed in *E. coli* and purified by cobalt affinity chromatography and size exclusion chromatography, resulting in a pure and monodisperse protein sample appropriate for complex reconstitution (Figure 4.4).

Recombinant MSP2N2 was expressed in *E. coli*, purified and used to produce 1:1 (molar ratio) DOPC:cholesterol nanodiscs. MSP2N2 forms nanodiscs with a diameter of $\sim$17 nm [236], which allows multiple copies of ILY to oligomerise on the membrane disc [50]. Size exclusion chromatography was used to separate nanodiscs from lipid/protein aggregates and unused protein or lipids (Figure 4.5A). The integrity of the nanodiscs was checked by negative stain EM imaging (Figure 4.5B), which confirmed that they were pure and of a roughly homogeneous size.

4.3.2 Assembling the complex and sucrose gradient

The CD59^{cyto} construct is able to insert spontaneously in lipid bilayers, so nanodiscs were first incubated with CD59^{cyto} for 20 minutes at room temperature. ILY^{HB1}-*nocharge* was added and incubated for a further 30 minutes to form the complexes. The ratio between

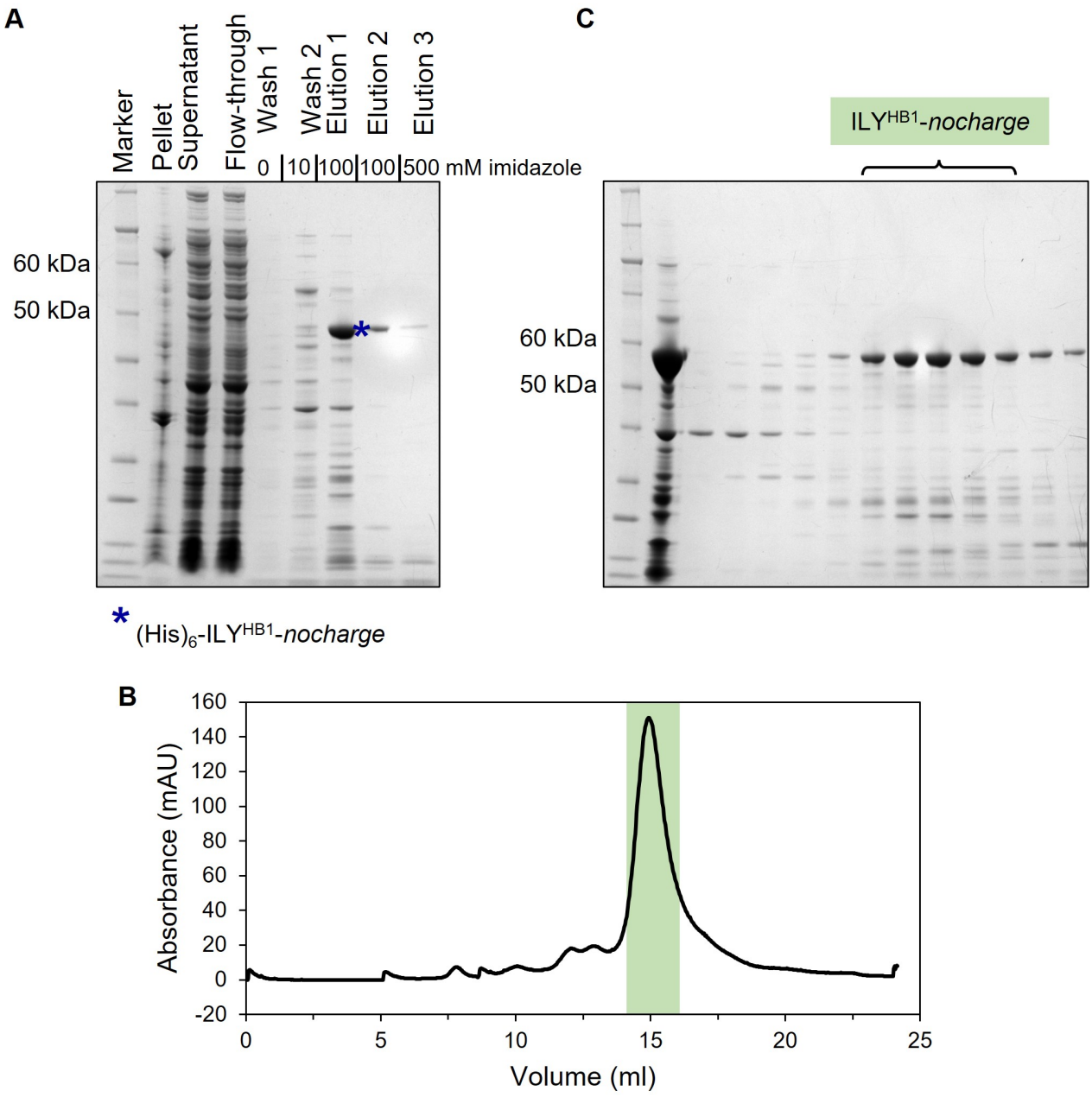


Figure 4.4 | ILY^{HB1}-*nocharge* purification. **A.** SDS-PAGE of the cobalt affinity chromatography step. The protein was eluted with imidazole (concentration given above the corresponding lanes). The blue asterisk shows the band corresponding to the protein of interest. **B.** Size-exclusion chromatography profile of ILY^{HB1}-*nocharge*, with the monodisperse peak of interest in the green box. **C.** SDS-PAGE of the gel filtration fractions. The lanes corresponding to the green box in panel B show pure ILY^{HB1}-*nocharge*.

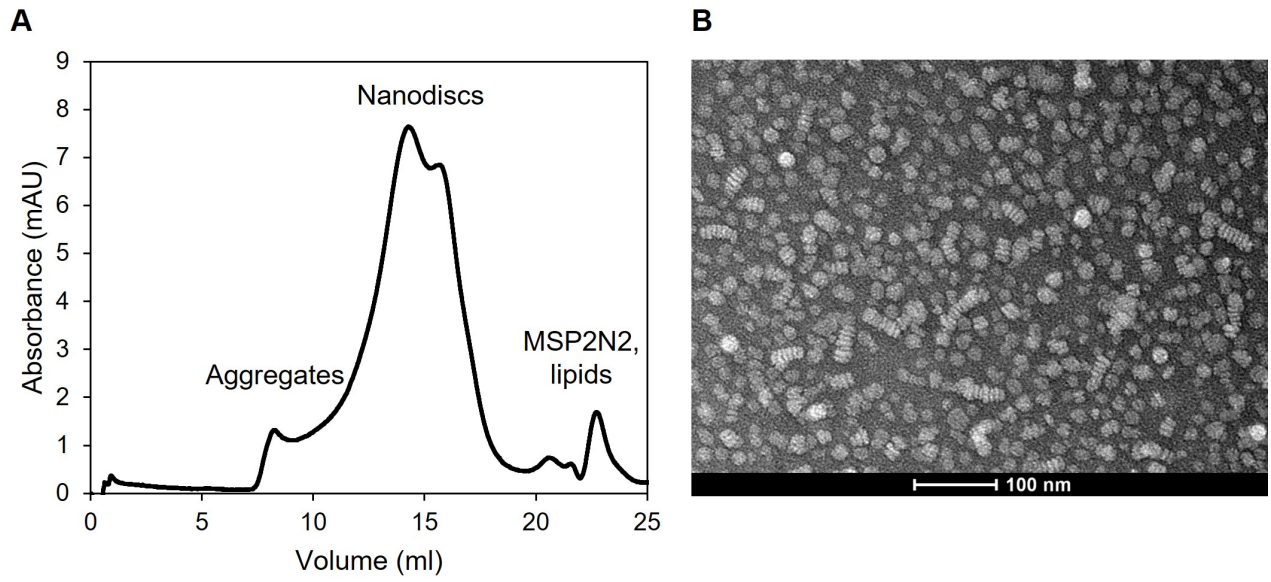


Figure 4.5 | Generation of DOPC-cholesterol nanodiscs. **A.** Gel filtration profile showing nanodiscs separated from aggregates in the void volume and unused MSP2N2 and lipids. **B.** Negative stain EM imaging of DOPC-cholesterol nanodiscs. Undecorated nanodiscs tend to form lipid-to-lipid stacks.

nanodiscs, $CD59^{cyto}$ and ILY^{HB1}-*nocharge* was kept close to the one used for the early pre-pore structure, where nanodiscs were mostly decorated with pentamers [50].

Negative stain EM reveals that ILY^{HB1}-*nocharge* forms assemblies on $CD59^{cyto}$-decorated nanodiscs (Figure 4.6). However, 'empty' nanodiscs (i.e. undecorated with ILY^{HB1}-*nocharge*) were also observed, along with large aggregates. While it is difficult to count the number of ILY^{HB1}-*nocharge* monomers on each nanodisc, a high level of heterogeneity was observed in the early prepore dataset [50], suggesting that such heterogeneity may also be present here, which would complicate image processing for high-resolution cryoEM.

To separate occupied and 'empty' nanodiscs and remove other contaminants (e.g. aggregates, off-target assemblies), the sample was purified by sucrose density gradient centrifugation. Linear 10-30% (w/v) sucrose gradients have been used to purify membrane proteins in MSP2N2 nanodiscs [143, 248] so the same approach was used. SDS-PAGE was used to check for the presence of the different proteins (MSP2N2, $CD59^{cyto}$ and ILY^{HB1}-*nocharge*) in each fraction of the gradient (Figure 4.7A). This showed that most assemblies were in

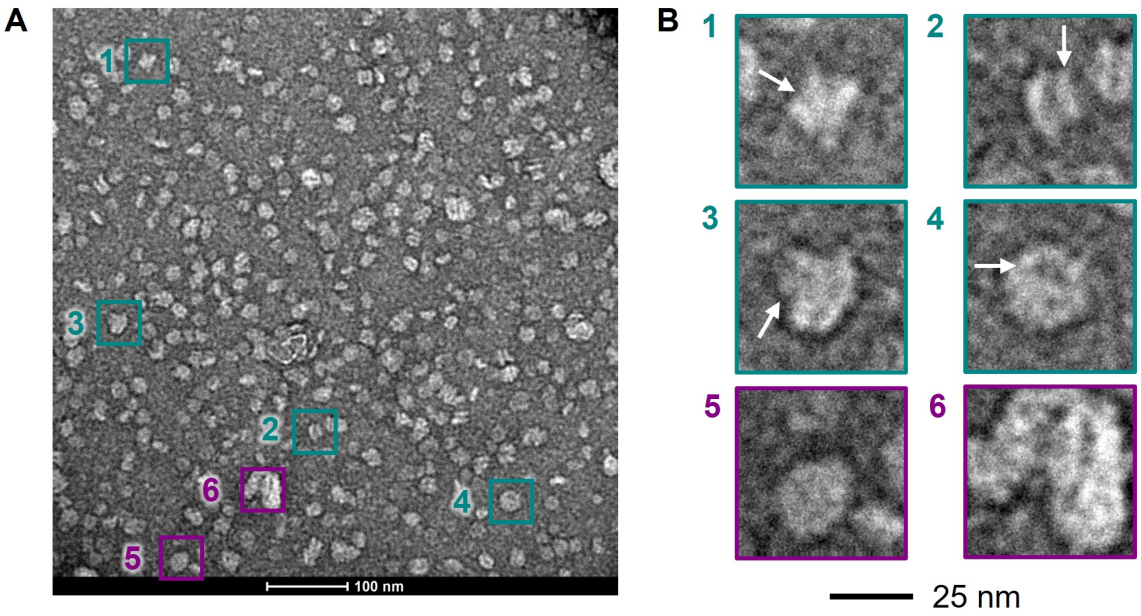


Figure 4.6 | Negative stain EM imaging of the ILY^{HB1}-*nocharge* complex on lipid nanodiscs. **A.** Negative stain micrograph of the complex made formed by incubating CD59^{cyto}-decorated nanodiscs with ILY^{HB1}-*nocharge*. **B.** Close-up views of selected areas shown in panel A. 1-4: Assembled complexes with white arrows pointing to ILY^{HB1}-*nocharge*. 5: 'Empty' nanodisc, not decorated with ILY^{HB1}-*nocharge*. 6: Aggregate.

fractions 5 to 18, but gave no information about the number of ILY^{HB1}-*nocharge* subunits. CD59^{cyto} could not be resolved on the gel, but negative stain EM imaging confirmed that the ILY^{HB1}-*nocharge* complex was correctly formed on the nanodiscs, indicating that CD59^{cyto} was present (Figure 4.7B).

Screening the fractions by negative stain EM identified which was best for structure determination. Early fractions contain mostly 'empty' nanodiscs, as indicated by their tendency to form stacks, a behaviour also observed prior to decorating them with any protein (Figure 4.5B and Figure 4.7B). Fractions near the middle of the gradient contain nanodiscs decorated with ILY^{HB1}-*nocharge*, and the number of subunits in each complex seems to increase with the fraction number. However, signs of aggregation can be seen from fraction 15 onwards, so fraction 14 was taken forward.

To characterise the complex further, I solved a low-resolution negative stain structure of the sample in fraction 14. 550 micrographs were collected on a 120 kV instrument and

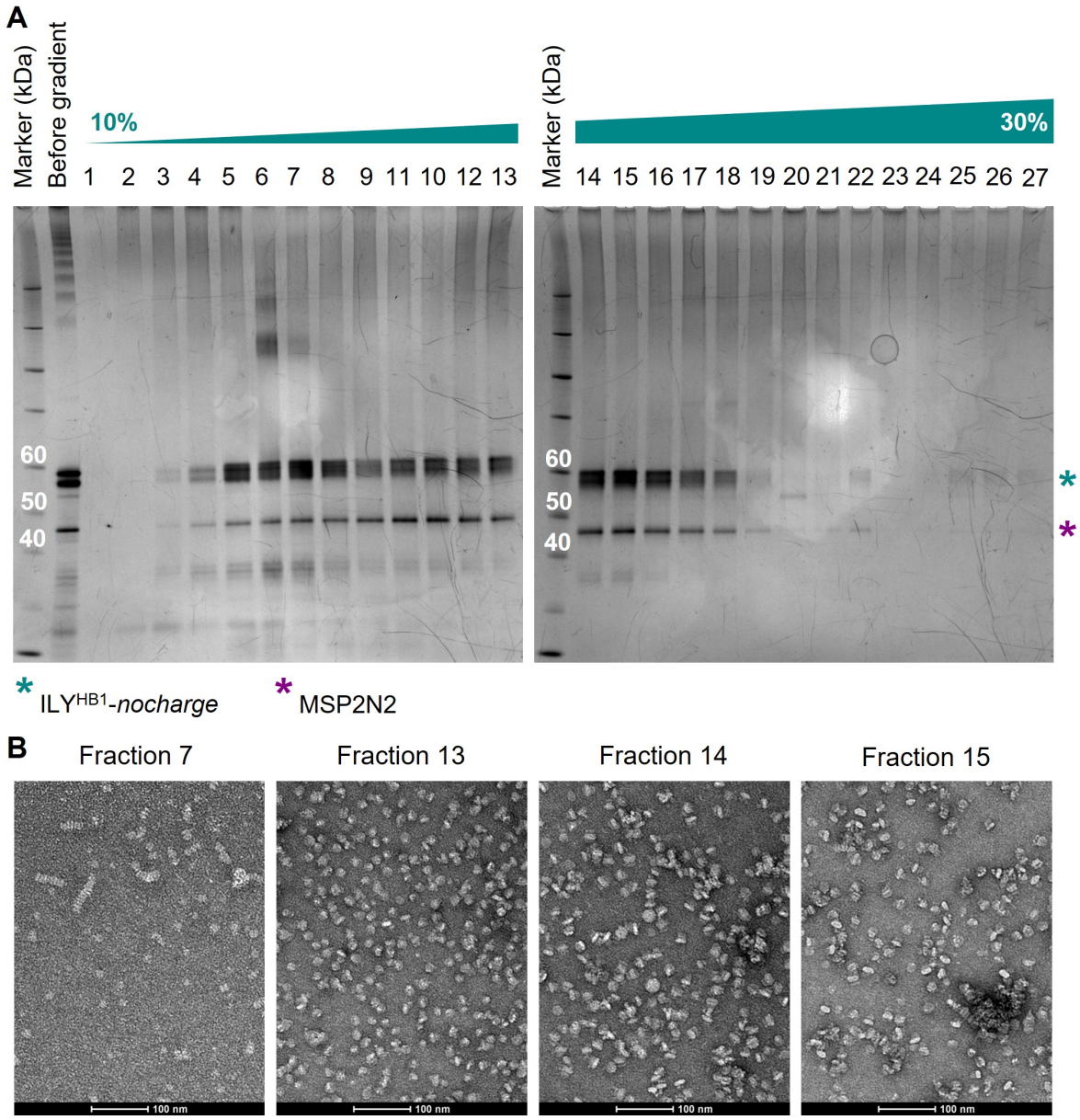


Figure 4.7 | Complex purification by sucrose density gradient centrifugation. **A.** Silver-stained SDS-PAGE of fractions 1 to 27 of the 10-30% (w/v) sucrose gradient. The double band corresponding to ILY^{HB1}-*nocharge* indicates partial formation of the engineered disulfide bond, consistent with what has been observed for ILY disulfide-locked mutants [50]. **B.** Negative stain EM images of selected fractions. Fraction 7 shows mostly stacks of 'empty' nanodiscs. Fraction 13 shows complexes with few ILY subunits. Fraction 14 is enriched in higher-order oligomers. Fraction 15 also contains higher-order oligomers, as well as aggregates.

processed in Cryosparc v2.14.2 [60]. Particles were picked with the built-in blob picker utility before several rounds of 2D classification to remove 'junk' particles. The final 2D classes (Figure 4.8B) show clear density for ILY^{HB1}-*nocharge* and the nanodisc, and a variety of orientations. The 2D classes contained both trimers and tetramers of ILY^{HB1}-*nocharge*, and the tetramers were taken forward for 3D reconstruction. The final structure, with a resolution of 20 Å, shows separation between the four monomers (Figure 4.8A).

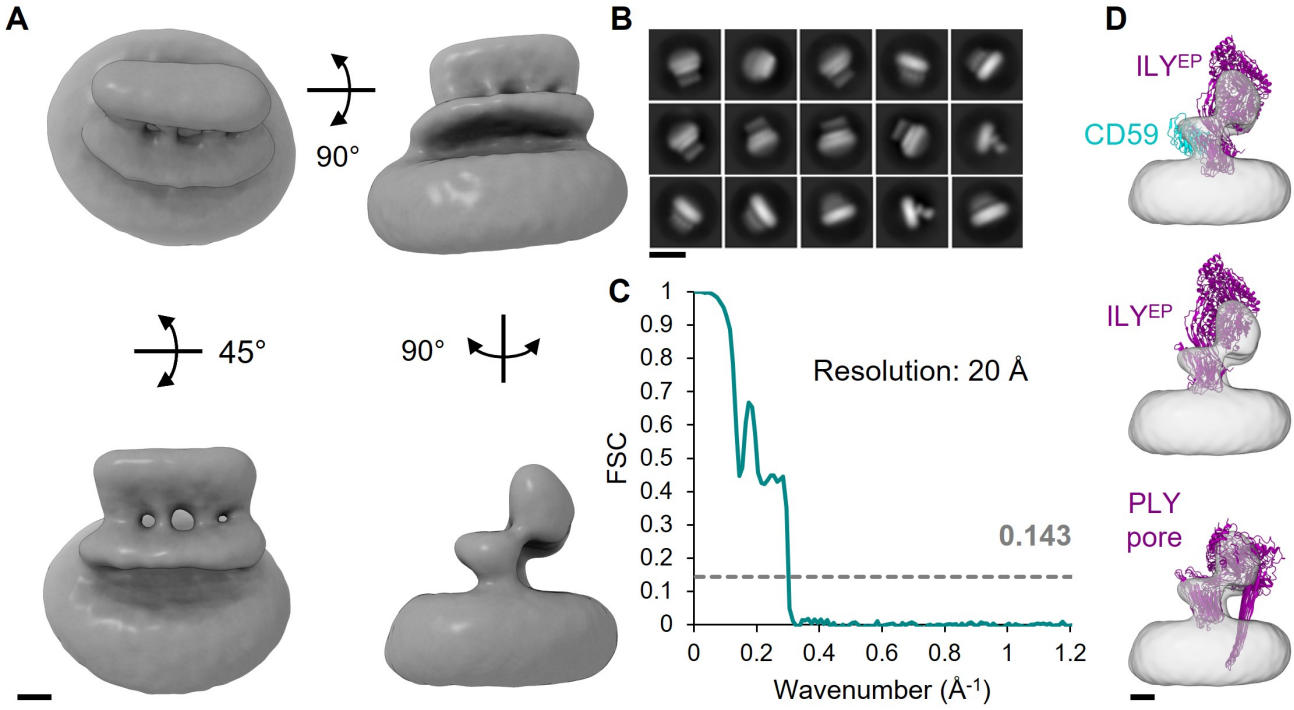


Figure 4.8 | Negative stain reconstruction of ILY^{HB1}-*nocharge* on lipid nanodiscs. **A.** 3D negative stain structure of tetrameric ILY^{HB1}-*nocharge*. Scale bar: 20 Å. **B.** Selected 2D classes showing a variety of orientations. Scale bar: 30 nm. **C.** FSC curve of the negative stain reconstruction. **D.** Overlaying the negative stain reconstruction with different structures: the ILY early prepore (ILY^{EP}, PDB 6ZD0) including CD59 (top), the ILY early prepore without CD59 (middle) and the PLY pore (PDB 5LY6, bottom). Scale bar: 20 Å.

To determine whether the complex was collapsed, it was overlayed with three different structures (Figure 4.8D). Manually fitting the ILY early prepore (PDB 6ZD0) with CD59 shows that the density does not fully match with the model, but this may be due to the low resolution of the map. If CD59 is removed and the model is moved to place domain 4 in the density directly above the nanodisc, the MACPF/CDC domain is out of density. When manually fitting a tetrameric version of the PLY pore structure (PDB 5LY6), both domain 4 and the

MACPF/CDC domain match the density (except for the unfurled β-hairpins) better than either fitting of the ILY early prepore. Together, this suggests that the complex corresponds to the collapsed prepore, although the low resolution of the reconstruction precludes unambiguous assignment of a conformational state.

4.3.3 Sucrose removal and cryoEM grid optimisation

The negative stain reconstruction indicates that the complex was successfully formed and purified. However, the sample contained a large concentration of sucrose, which would have prevented high-resolution structure determination by contributing background noise [54]. The same fraction that was used for the negative structure was run through a Zeba spin desalting column pre-equilibrated in HEPES buffer (20 mM HEPES pH 7.4, 120 nM NaCl) according to the manufacturer's instructions. Negative stain EM imaging revealed that the complex was destroyed in the process (Figure 4.9B). To test whether overnight dialysis was a viable strategy, the sample was kept at 4°C overnight and checked with negative stain EM, which showed aggregates and a lower overall concentration of particles (Figure 4.9C), probably due to aggregated complexes coming out of solution.

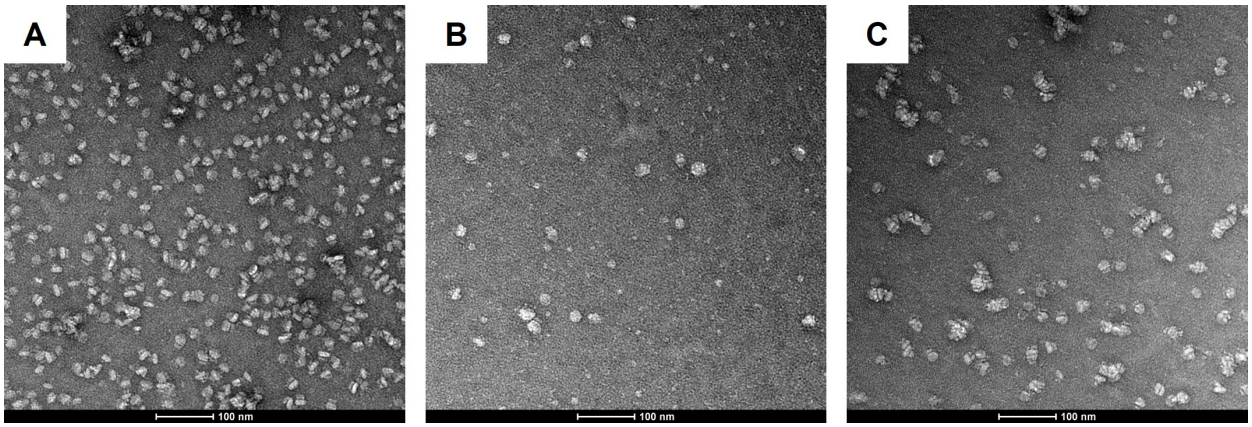


Figure 4.9 | Negative stain EM imaging to check sample stability. **A.** Negative stain micrograph of fraction 14 directly after gradient fractionation. **B.** Negative stain micrograph of fraction 14 after Zeba spin treatment for sucrose removal. **C.** Negative stain micrograph of fraction 14 after overnight incubation at 4°C. All grids were made using the same protocol, without sample dilution.

Since sample instability hindered sucrose removal, it was fixed by adding glutaraldehyde, a mild chemical crosslinker, to the sucrose density gradient. By adding glutaraldehyde to the high-sucrose solution, a gradient of crosslinker was established at the same time as the sucrose gradient, an approach that has been termed 'GraFix' [253, 254]. Glutaraldehyde is an organic compound made of a five-carbon chain with a formyl group at either end. It forms covalent bonds with the primary amines of lysine residues in proximity with each other, crosslinking the residues (Figure 4.10A). This has been shown to improve the stability of some protein samples, as well as increasing their contrast in the electron microscope [143, 253, 254].

The fixed sample was able to withstand sucrose removal by Zeba spin much more readily than the unfixed complex, but some lipid aggregates could still be seen (Figure 4.10B). After overnight dialysis in a dialysis cup, the sample aggregated (Figure 4.10C), whereas a 2-hour dialysis led to less aggregation (Figure 4.10D).

The sample after a 2-hour dialysis was taken forward for cryoEM grid preparation. On standard carbon-film Quantifoil grids with a copper mesh, particles aggregated on the carbon and did not enter the holes (Figure 4.11A). To overcome this, I adopted two approaches. The sample was frozen on similar copper Quantifoil grids, which were first coated in graphene oxide (GrOx) to provide a support in which particles were adsorbed (Figure 4.11B). The particles on GrOx-coated grids lacked sharpness, but it was unclear whether this was a result of partial embedding in too thin ice, or a low signal-to-noise ratio due to remaining sucrose in the solution. The sample was also frozen on UltrAuFoil grids, where both the grid and the foil are made of gold to avoid aggregation on a carbon foil. This led to particles inside holes (Figure 4.11C), but imaging the grid at low magnification showed inconsistent ice thickness over neighbouring squares (Figure 4.11D). This may be a result of heterogeneous wetting of the grid arising from the presence of sucrose in the buffer, suggesting that the 2-hour

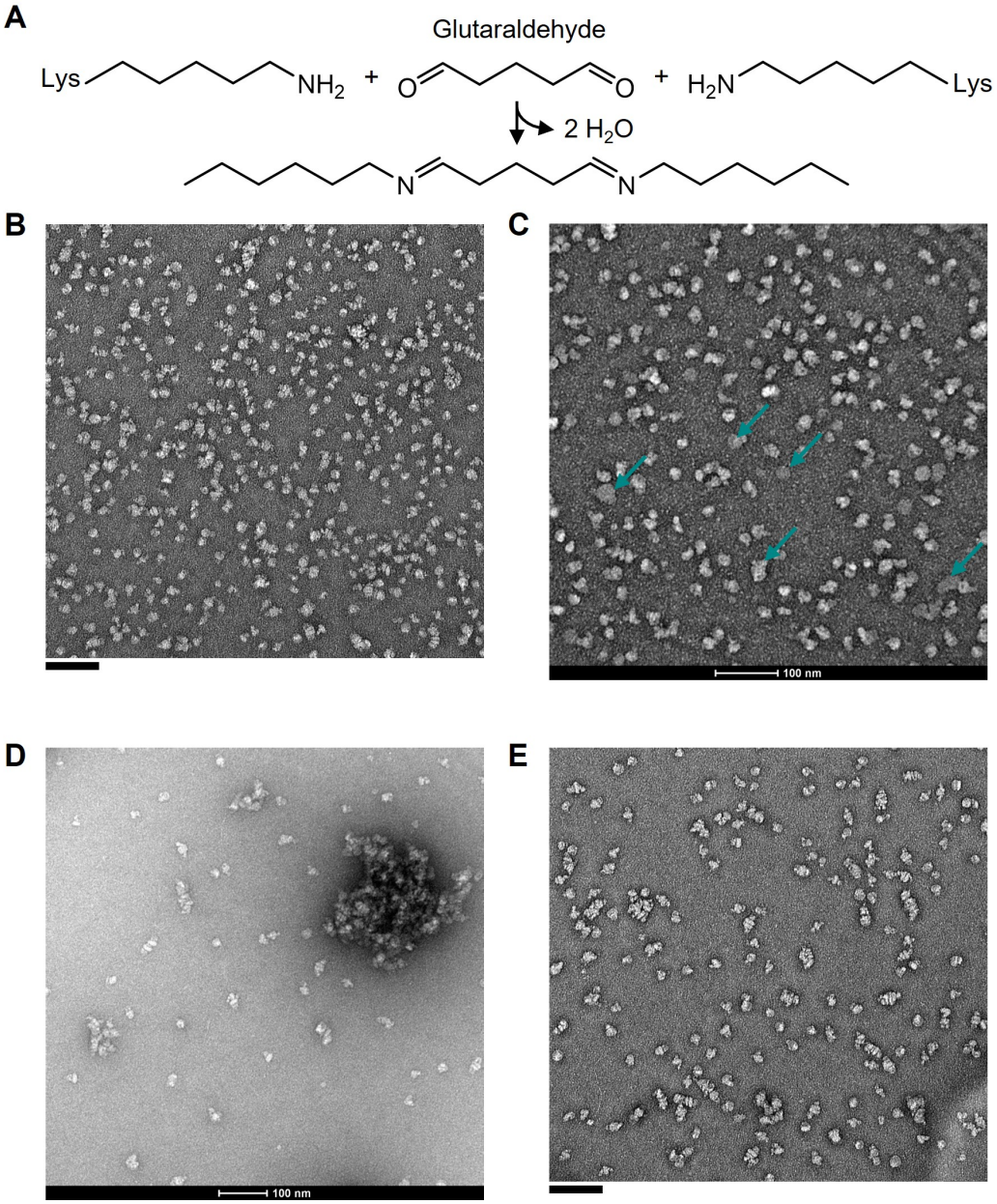


Figure 4.10 | GraFix crosslinking of the sample. **A.** Glutaraldehyde reacts with the primary amines of lysine residues to crosslink them. **B.** Negative stain micrograph of fixed fraction 14. Scale bar: 100 nm. **C.** Negative stain micrograph of fixed fraction 14 after Zeba spin treatment. Arrows point to lipid aggregates. **D.** Negative stain micrograph of fixed fraction 14 after overnight dialysis. **E.** Negative stain micrograph of fixed fraction 14 after 2-hour dialysis. Scale bar: 100 nm.

dialysis step did not succeed in removing the entirety of the sucrose from the sample.

As the sample could not withstand Zeba spin centrifugation or overnight dialysis, and a 2-hour dialysis did not remove all sucrose, it was buffer-exchanged into HEPES buffer (20 mM HEPES pH 7.4, 120 nM NaCl) with four cycles of dilution and slow concentration. Negative stain EM imaging showed that the procedure did not destroy the complex (Figure 4.12A), so

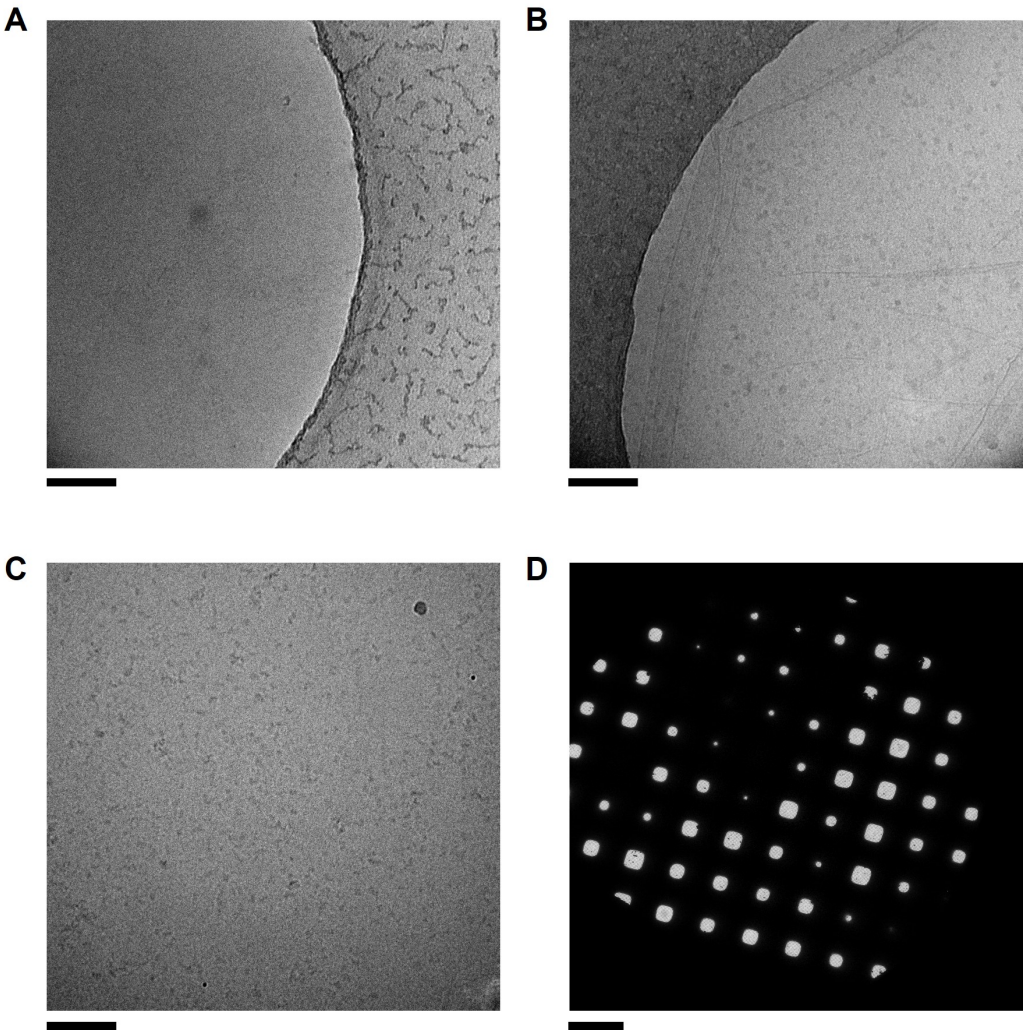


Figure 4.11 | Screening of cryoEM grid types. **A.** Sample aggregates on the carbon foil of copper Quantifoil grids. Scale bar: 100 nm. **B.** The addition of a graphene oxide support leads to particles present in the holes. The lack of sharpness of the particles may be due to improper ice thickness or the presence of residual sucrose. Scale bar: 100 nm. **C.** On UltrAuFoil grids, the sample is frozen in holes. Scale bar: 100 nm. **D.** Inconsistent ice thickness over the grid suggests that there may be sucrose present in the sample after a 2-hour dialysis. Scale bar: 100 µm.

it was used to prepare cryoEM grids. As GrOx-coated copper Quantifoil grids and UltrAuFoil grids had previously shown the most promising results, the sample was frozen on those. The GrOx-coated grids showed good particle distribution and concentration (Figure 4.12B-C) but no particles were observed in holes on the UltrAuFoil grids (not shown). As aggregation on carbon had been observed before (Figure 4.11A), the sample was also frozen without support on gold grids with a lacey carbon film, which have a reduced amount of carbon compared to standard copper Quantifoil grids. Due to the limited amount of sample available,

fraction 15 was used for freezing on Gold Lacey grids. This also resulted in good particle distribution and concentration (Figure 4.12D-F).

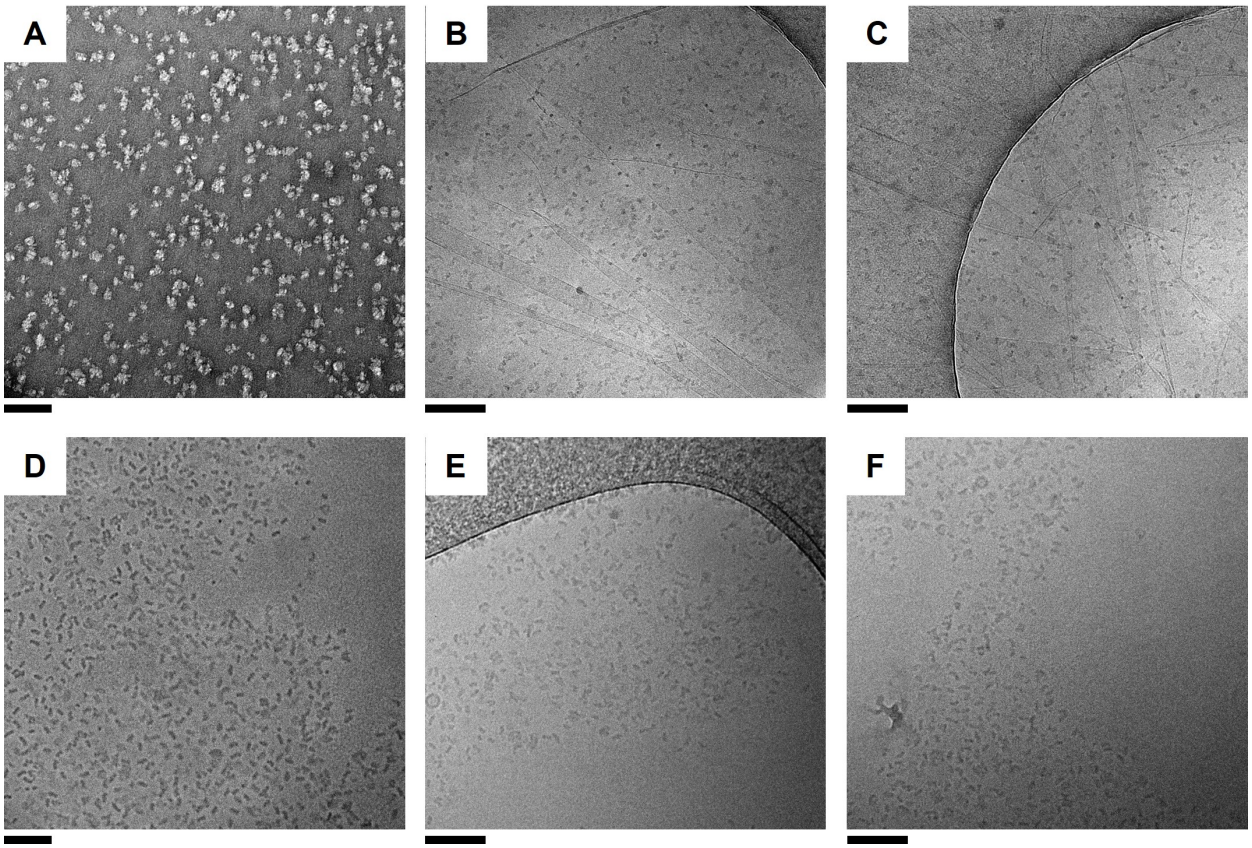


Figure 4.12 | Sucrose removal by dilution and concentration yields a sample suitable for structure determination. **A.** Negative stain micrograph of fixed sample after four rounds of dilution and concentration to remove the sucrose. **B-C.** Micrographs of GrOx-coated copper Quantifoil grids. **D-E.** Micrographs of lacey carbon on gold grids. Scale bar: 100 nm (all panels).

4.3.4 Glacios data collections

As particles showed a good distribution on both gold lacey and GrOx-coated copper Quantifoil grids, data were collected from both types of grids on a 200 kV Glacios instrument. All data collections were performed using a FALCON 4 direct electron detector and the data collection parameters are shown in Table 4.1. Motion correction [250] and CTF estimation [249] were done for each dataset separately, in Cryosparc v3.3.2 [60]. Micrographs were then curated based on the quality of the CTF estimation, and micrographs with Thon rings that did not go beyond 8 Å were excluded. For the Gold Lacey dataset, micrographs were

also sorted by average intensity, and those with high intensity were excluded as they were mostly occupied by the lacey carbon film.

Table 4.1: Data collection details for the 200 kV Glacios datasets.

Grid type	Gold Lacey	GrOx copper Quantifoil	
Voltage (kV)	200	200	
Magnification	100,000×	100,000×	
Detector	FALCON 4	FALCON 4	
Pixel size (Å/px)	1.1	1.1	
Defocus range (μm)	−1.1 to −3.2	−0.9 to −3.0	
Defocus step (μm)	0.3	0.3	
Electron dose (e^-/Å^2)	50	40	30
Micrographs	3,244	11,709	2,897

Processing the Gold Lacey dataset

The Gold Lacey dataset and the two GrOx Quantifoil datasets were processed separately. For the Gold Lacey dataset, particles were picked with the blob picker utility of Cryosparc and extracted binned by 4. 'Junk' particles were removed iteratively with several rounds of 2D classification. Most 2D classes corresponded to top or front views of the particles, i.e. views where the different ILY^{HB1}-*nocharge* monomers could be resolved (Figure 4.13). To enhance for side views, classes where signal could be seen from the nanodisc were used as templates for template picking. Those particles were extracted, binned by 4, and merged with the original 'clean' particles before removing duplicates and further round of 2D classification to remove newly extracted 'junk' particles. To get a better understanding of the oligomeric state of the complex, 2D classes were then split into groups based on how many ILY^{HB1}-*nocharge* subunits could be counted (Figure 4.13). Most complexes were pentamers, consistent with what was observed for the early prepore [50]. Tetramers, hexamers and heptamers were also observed, but only pentamers showed front views. Overall, only ∼1% of particles classified into side views with clear protein signal.

As pentameric assemblies dominated the dataset and showed the most diverse views,

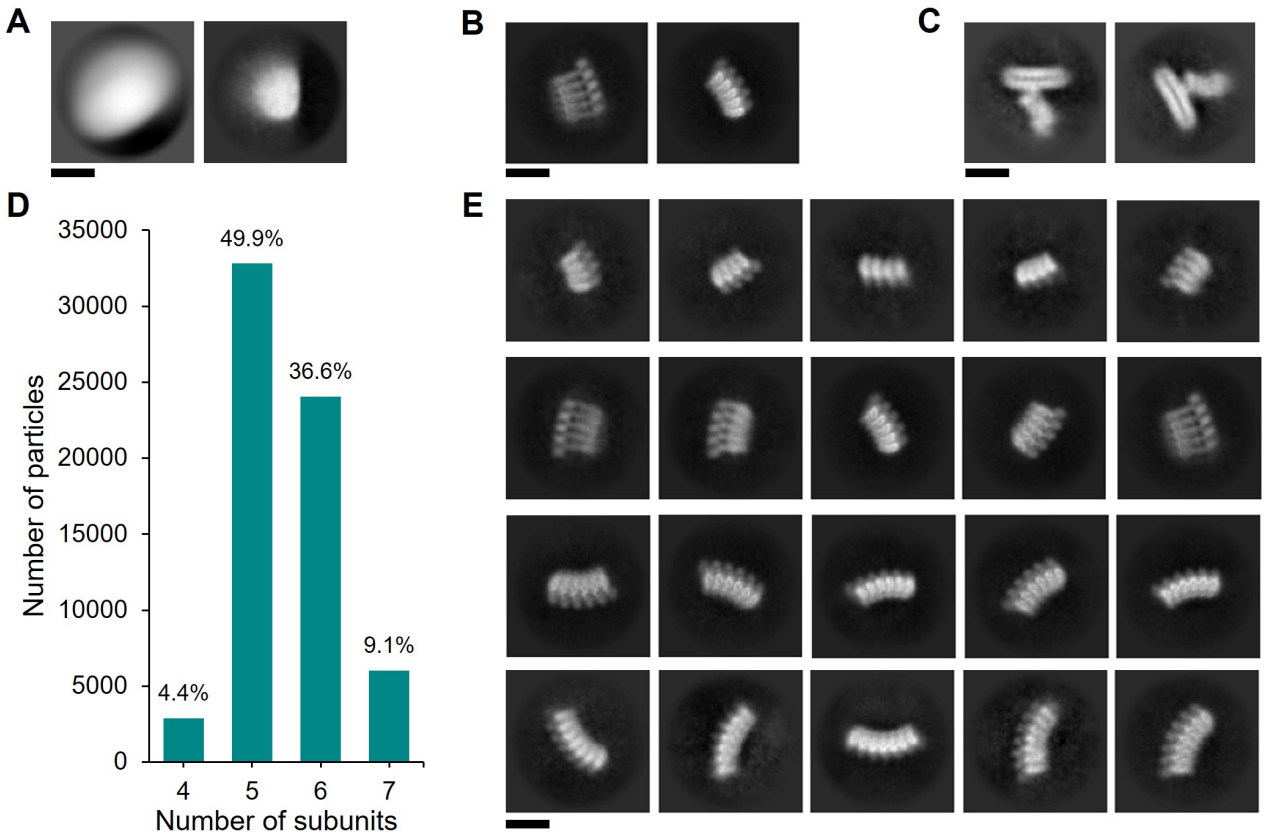


Figure 4.13 | 2D classification of particles from the Gold Lacey dataset. **A.** Example 'junk' classes that were removed with several rounds of 2D classes. **B.** Example front (left) and top (right) views, where ILY^{HB1}-*nocharge* subunits can be counted. **C.** Side views where protein density can be seen, representing ~1% of 'good' particles. **D.** Particle count for each oligomeric state (excluding side views where subunits cannot be counted). **E.** Example 2D classes of ILY^{HB1}-*nocharge* tetramers (top), pentamers (second from top), hexamers (second from bottom) and heptamers (bottom). Scale bar: 20 nm (panels A-C and E).

they were combined with the side views and used for 3D reconstruction. First, an *ab initio* reconstruction was performed with four classes (Figure 4.14A). One of them was reminiscent of the uncollapsed, early prepore, which was confirmed by running a homogoneous refinement of the corresponding particles (Figure 4.14B). Several approaches were attempted to deal with conformational heterogeneity between the uncollapsed and collapsed classes (including several rounds of heterogeneous refinement and 3D classification) but the best results came from a second round of *ab initio* reconstruction excluding the clearly uncollapsed particles. The best class was used for homogeneous refinement (with 7,717 particles) and yielded an 18 Å reconstruction with strong preferred orientation (Figure 4.14C).

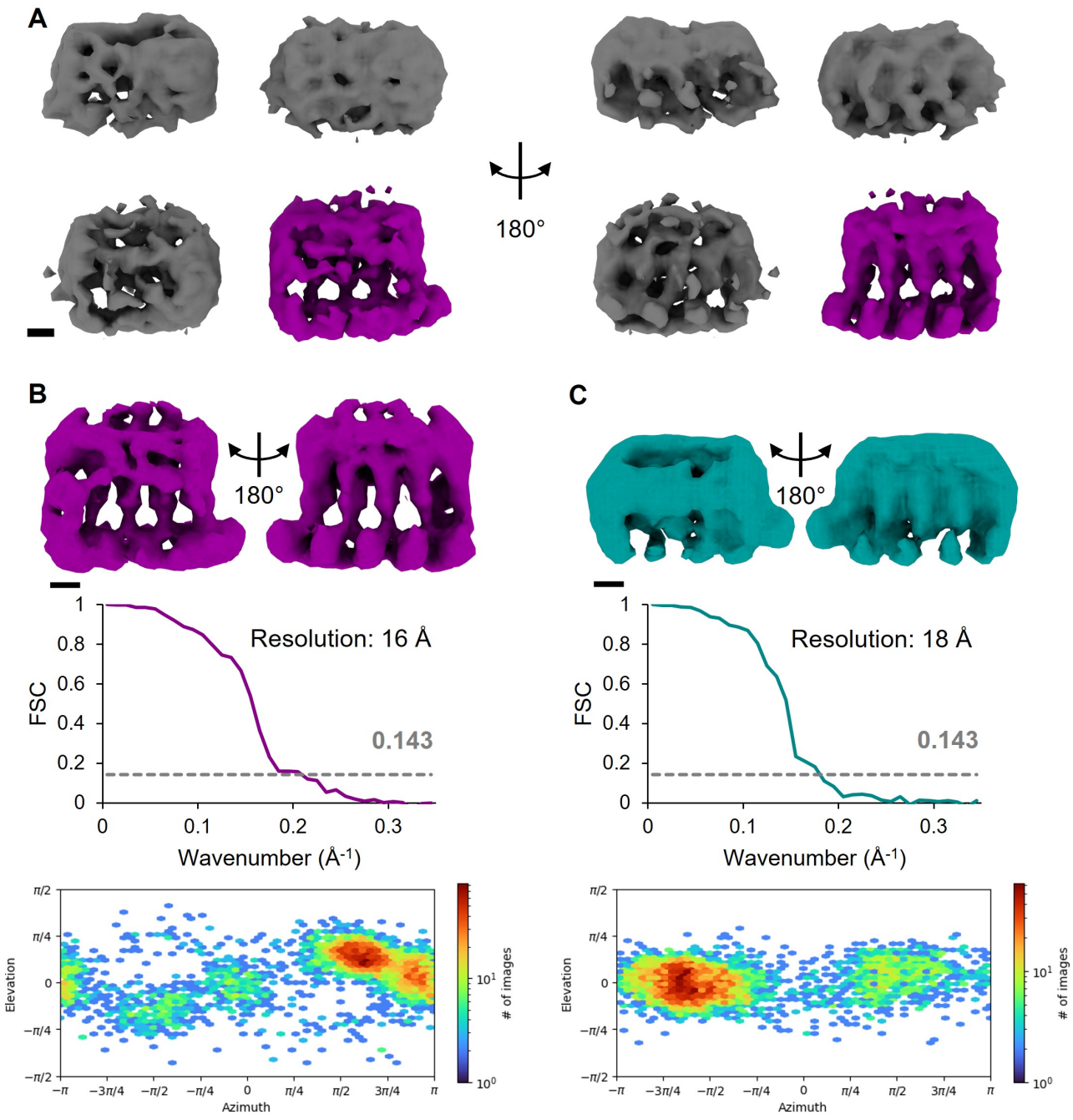


Figure 4.14 | Initial processing to deal with conformational heterogeneity of the Gold Lacey dataset. **A.** Four classes generated by the first *ab initio* reconstruction. The purple class is reminiscent of the early prepore structure [50]. **B.** Homogeneous refinement of uncollapsed particles from the purple class in panel A. Top: consensus map; middle: FSC curve; bottom: orientation plot. **C.** Homogeneous refinement of 7,717 particles from the best class of the second round of *ab initio* reconstruction. Top: consensus map; middle: FSC curve; bottom: orientation plot. Scale bar: 20 Å (all panels).

To increase the particle number, all the particles that had not been classified into the uncollapsed class were used again. Particles from 2D classes where the nanodisc was visible but protein density was not (which had been excluded so far) were also added to the

pool in an attempt to recover more views. These particles were used for homogeneous refinement against the map in Figure 4.14C, then classified in 3D into ten classes. This did not highlight oligomeric or conformational heterogeneity but generated classes with different levels of feature definition. The best class (containing 12,683 particles) was used for a first homogeneous refinement job, followed by local refinement with a static mask around the three central monomers (Figure 4.15A). This yielded a 17 Å map of a pentameric assembly with severe preferred orientation (Figure 4.15B-D). However, the five subunits can be clearly distinguished, with the three central monomers showing the best-defined features. Unbinning the data did not improve the reconstruction, even with the addition of more particles excluded in earlier steps, suggesting that the resolution is limited by heterogeneity and preferential orientation. Attempts at separating different conformations by 3D classification, heterogeneous refinement using the maps in Figure 4.14B and Figure 4.15B, or more rounds of *ab initio* reconstruction with the unbinned data, did not improve the map.

To determine the conformational state of the reconstruction, it was overlayed with a PLY trimer it its pore conformation (PDB 5LY6) and with the ILY early prepore (PDB 6ZD0). In the case of the PLY pore, the best match came from fitting the structure on one side of the map rather than the centre, but only domain 4 (D4) from the peripheral monomer fits inside the density (Figure 4.15E). The ILY early prepore structure fits better in the map if the handedness is inverted (Figure 4.15F). The map shows density which matches well with CD59 and with the MACPF/CDC domains, but this fitting places D4 outside of any density, suggesting that the density attributed to CD59 may in fact correspond to D4, moved to a position similar to the one found in the PLY pore conformation. As neither structure fits the map entirely, it may represent an intermediate conformation, the early prepore. However, the resolution is too low to rule out the early prepore conformation.

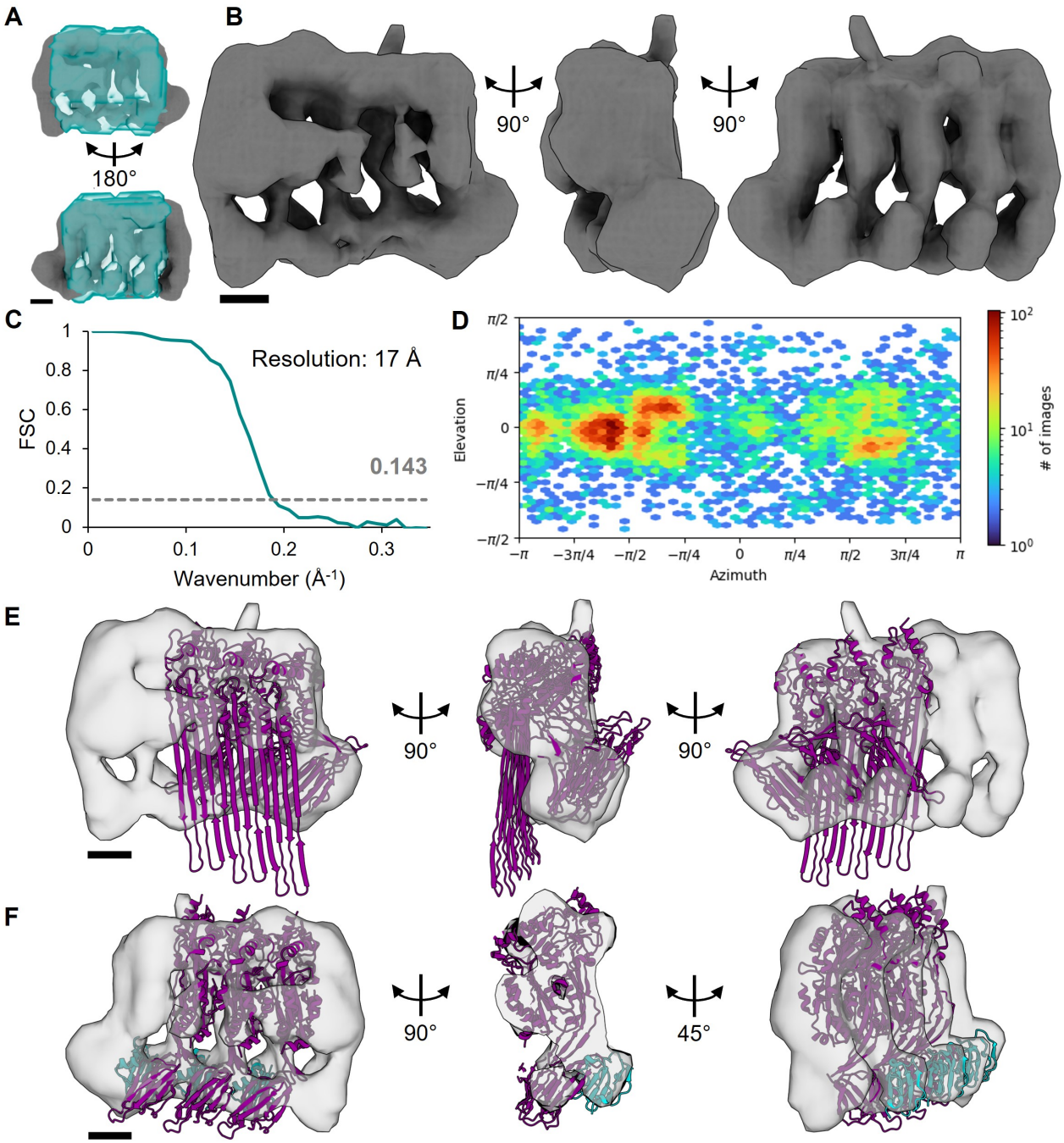


Figure 4.15 | 3D reconstruction from the Gold Lacey dataset (binned by 4). **A.** Input volume (grey) and static mask (turquoise) used for homogeneous refinement. **B.** Consensus map generated by homogeneous refinement with the static mask shown in panel A. **C.** FSC curve for the map in panel B. **D.** Orientation plot of the map in panel B. **E.** Fitting of trimeric PLY in the pore conformation (PDB 5LY6) into the map in panel B. **F.** Fitting of trimeric ILY in the early prepore conformation (PDB 6ZD0) into the map in panel B after inverting the handedness. Scale bar: 20 Å (all panels).

Processing the GrOx Quantifoil datasets

Particle picking from the GrOx Quantifoil micrographs was performed using 2D classes from the Gold Lacey dataset for template picking. 'Junk' particles were removed in several rounds of 2D classification binned by 4 as for the Gold Lacey dataset. 2D classification revealed that particles adopted different orientations on the GrOx Quantifoil grids (Figure 4.16A). Most particles (49.0%) belong to classes corresponding to side views where the alignment is driven by the protein and only the nanodisc signal directly under ILY^{HB1}-*nocharge* is visible. These are most similar to the rare side views in the Gold Lacey dataset, but they are tilted by different angles compared to them. A large proportion of particles (30.9%) classified into side views where the nanodisc drives the alignment and the protein signal is blurred on either side of it. Finally, only 20.1% of particles correspond to top views where the ILY^{HB1}-*nocharge* subunits can be counted, revealing mostly trimers, tetramers and pentamers. This is consistent with the tetrameric negative stain reconstruction, which was also generated from fraction 14 of a gradient, whereas the Gold Lacey datasets used fraction 15.

Three main processing strategies were trialled on the GrOx Quantifoil datasets. An *ab initio* reconstruction job yielded four classes and the best was taken forward for homogeneous refinement, resulting in the map in Figure 4.16B. The reconstruction includes density corresponding to the nanodisc but it only contains a dimer of ILY^{HB1}-*nocharge* with poorly defined features. The second approach was to low-pass filter the gold lacey reconstruction in Figure 4.15B to 40 Åand use it for homogeneous refinement of the particles from the GrOx Quantifoil datasets. This resulted in a poorly refined map, suggesting that the algorithm failed to align particles in 3D (Figure 4.16C). Similarly, merging 'good' particles (as determined by 2D classification) from all datasets and running an *ab initio* reconstruction followed by homogeneous refinement of the best class did not yield satisfactory results (Figure 4.16D). In all three cases, rounds of heterogeneous refinements and 3D classification failed to improve

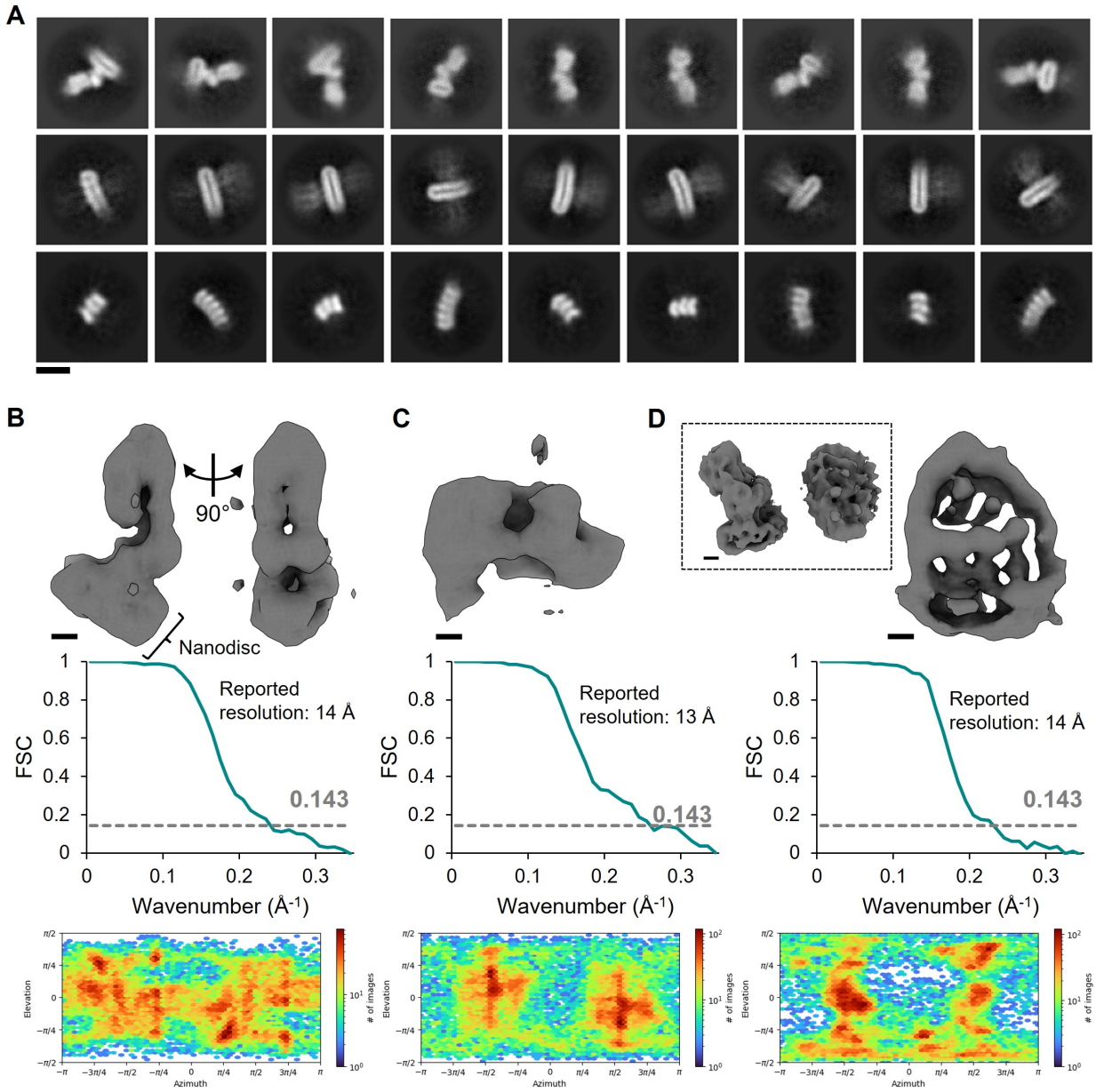


Figure 4.16 | Image processing for the GrOx Quantifoil datasets. **A.** Selected 2D classes showing tilted side views (top), side views aligned on the nanodisc (middle) and top views (bottom). Scale bar: 20 nm. **B.** 3D reconstruction resulting from homogeneous refinement of the best *ab initio* class (top), with its FSC curve (middle) and orientation plot (bottom). **C.** 3D reconstruction resulting from homogeneous refinement of the GrOx Quantifoil particles against the consensus map generated with the Gold Lacey dataset (top), with its FSC curve (middle) and orientation plot (bottom). **D.** Map resulting from merging particles from all datasets and refining the best class from an *ab initio* reconstruction (top), with its FSC curve (middle) and orientation plot (bottom). Inset: classes from the *ab initio* reconstruction, the left-hand side class was used for homogeneous refinement. Scale bar: 20 Å (panels B-D and inset in panel D).

the quality of the map. All three reconstructions in Figure 4.16 have a reported resolution below 15 Å, but visual inspection of the maps suggests that they are of poor quality and the resolution is overestimated, possibly due to over-refinement of noise. Moreover, the orien-

tation plots all show a much wider distribution of views than would be expected from the 2D classes, which suggests that the particles are misaligned and the angular assignments are incorrect.

4.3.5 Double-crosslinking and eBIC data collection

The 200 kV datasets revealed that the particles suffer from preferential orientation on both Gold Lacey and GrOx-coated Quantifoil grids. As the charge properties of the particles influence their behaviour on cryoEM grids, the GraFix protocol was modified to alter the extent of crosslinking. The length of a crosslinker determines which amino acid residue pairs it can crosslink, so using reagents with different chain lengths will result in differential crosslinking. This has been observed for a combination of glutaraldehyde and the longer-chain chemical bis(sulfosuccinimidyl) suberate (BS^3) [255]. Combining glutaraldehyde and BS^3 also results in more extensive crosslinking, potentially stabilising the complex further. BS^3 was added to the sucrose density gradient, which was run and harvested as before.

To remove the sucrose from the doubly-crosslinked sample, it was treated with a Zeba spin column following the manufacturer's instructions. The desalting column was pre-equilibrated in either HEPES buffer (20 mM HEPES pH 7.4, 120 nM NaCl) or Tris buffer (200 mM NaCl, 20 mM Tris-HCl). Negative stain EM imaging showed that the sample could withstand the Zeba spin treatment (Figure 4.17A), and that more protein was recovered with the Tris buffer, probably because the residual crosslinkers were quenched by the primary amine of Tris. Moreover, the sample could withstand being concentrated faster and to higher concentrations, facilitating the preparation of cryoEM grids. As fraction 15 contains higher-order oligomers than fraction 14 and did not show obvious signs of aggregation, it was vitrified on both GrOx-coated copper Quantifoil and Gold Lacey grids (Figure 4.17B-C).

As the Gold Lacey dataset contained more top views, resulting in a better reconstruction,

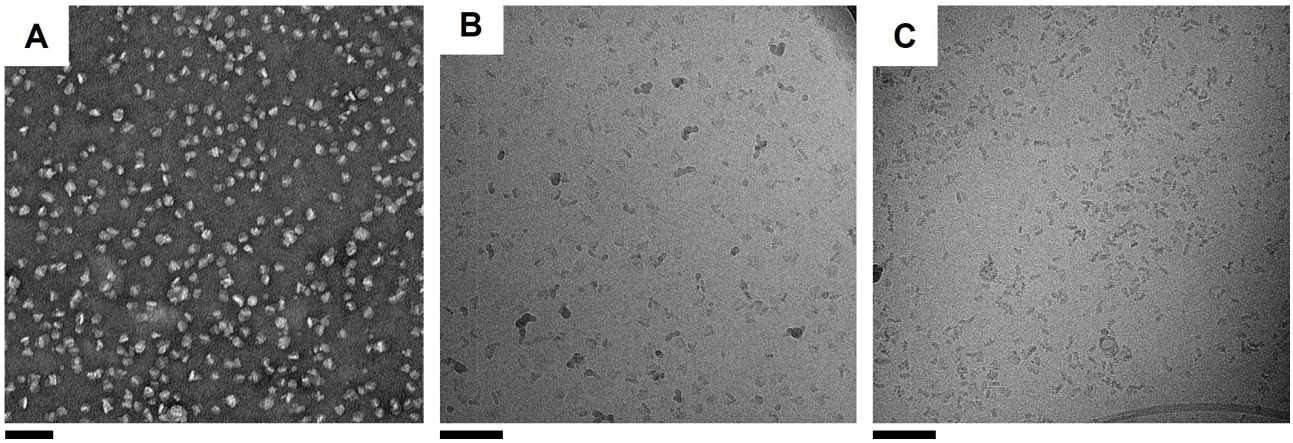


Figure 4.17 | Sample after double-crosslinking with glutaraldehyde and BS^3. **A.** Negative stain micrograph of fraction 15 after sucrose removal with a Zeba spin column. **B.** Sample vitrified on a GrOx-coated copper Quantifoil grid. **C.** Sample vitrified on a Gold Lacey grid. Scale bar: 100 nm (all panels).

Gold Lacey grids of double-crosslinked sample were taken for data collection on a 300 kV Titan Krios instrument at the Electron Bio-Imaging Centre (eBIC). To capture more views and overcome the preferred orientation issues observed previously, two datasets were recorded: at 0° and at a 30° tilt. The details of the data collection are in Table 4.2.

Table 4.2: Data collection details for the 300 kV eBIC datasets.

Tilt angle (°)	0	−30
Voltage (kV)	300	300
Magnification	81,000×	81,000×
Detector and energy filter	Gatan BioQuantum K3	Gatan BioQuantum K3
Pixel size (Å/px)	1.072	1.072
Defocus range (μm)	−1.1 to −3.2	−1.4 to −2.6
Defocus step (μm)	0.3	0.3
Electron dose ($e^-/Å^2$)	50	50
Micrographs	11,855	7,535

Micrographs were imported into Cryosparc v3.3.2 and pre-processed (each dataset separately) for motion correction and CTF estimation. Micrographs were curated based on the quality of the CTF estimation and, in the case of the tilted dataset, overfocused and excessively underfocused micrographs were removed. Due to time constraints, I was only able to perform initial steps of image processing on these two datasets. The overall processing workflow is shown in Figure 4.18 and detailed below.

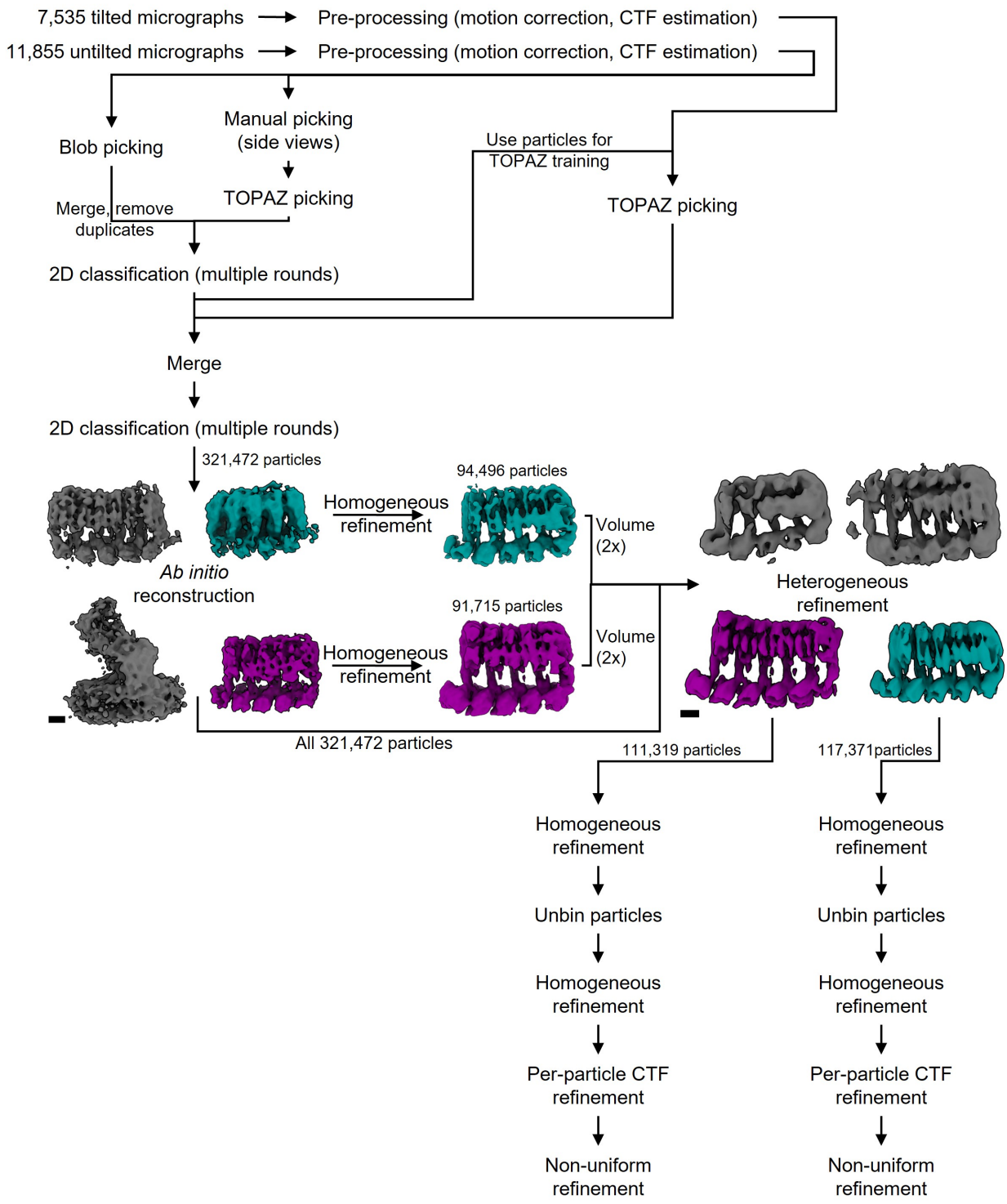


Figure 4.18 | Overall processing workflow for the eBIC datasets. Scale bars: 20 Å.

Particles were picked from the untilted dataset using the blob picker, extracted (binned by 4) and subjected to rounds of 2D classification to remove 'junk' particles. Manual picking was also performed on a selection of micrographs to target specifically side views where the nanodisc was visible (Figure 4.19). Those picks were used to train a Topaz model [245] and

subsequently extract particles (binned by 4) before cleaning them with several rounds of 2D classification. Both sets of particles were then merged and duplicates were removed before another round of 2D classification. Side views represented ∼3.5% of particles taken forward for processing. Oligomers ranging from tetramers to heptamers were observed and all were included in subsequent processing steps. The particle picks from the untilted dataset was used to train a Topaz model, which picked particles from the tilted dataset.

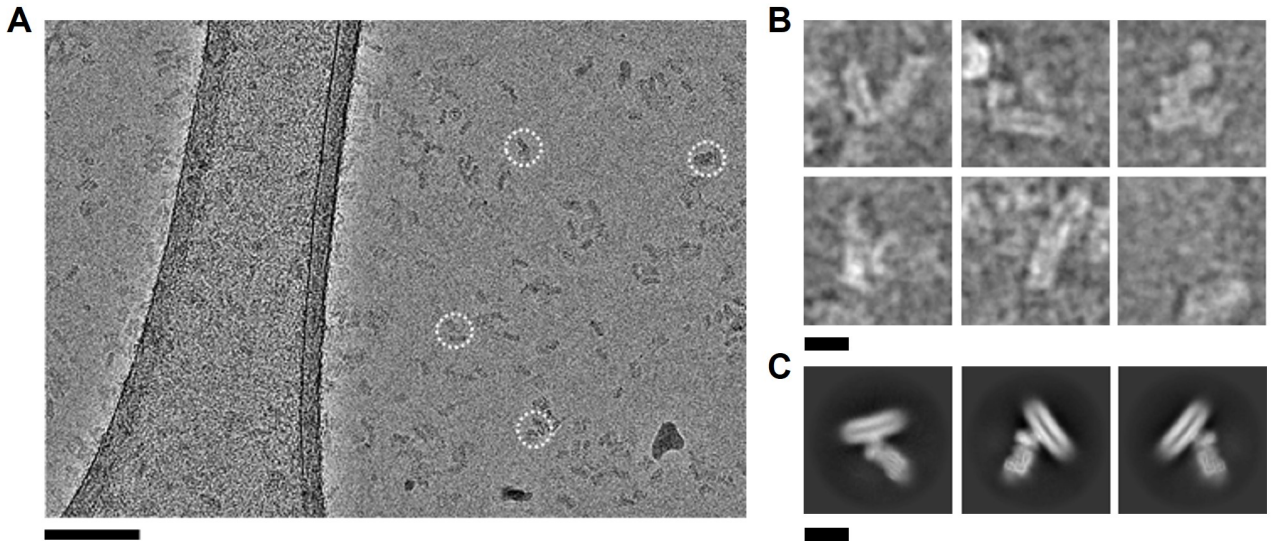


Figure 4.19 | Using Topaz to recover side views from the untilted data. **A.** Example micrograph with particle picks used for Topaz training circled in white. Scale bar: 100 nm. **B.** Example particle picks from the Topaz model. Scale bar: 100 Å. **C.** Final 2D classes of Topaz picks. Scale bar: 100 Å.

Four *ab initio* reconstructions were generated from the pooled particles (Figure 4.20), with one class resembling the early prepore conformation, and another similar, but smaller in height. Each of those two classes was used for homogenous refinement to give an initial consensus map of the early prepore, and one which may correspond to the late prepore. All particles were subjected to heterogeneous refinement with two copies of each map as input volumes.

The best class corresponding to the early prepore and the best class of the potential new conformation were taken forward for further processing separately. For the early prepore, the map was first improved by homogeneous refinement, before re-extracting the particles

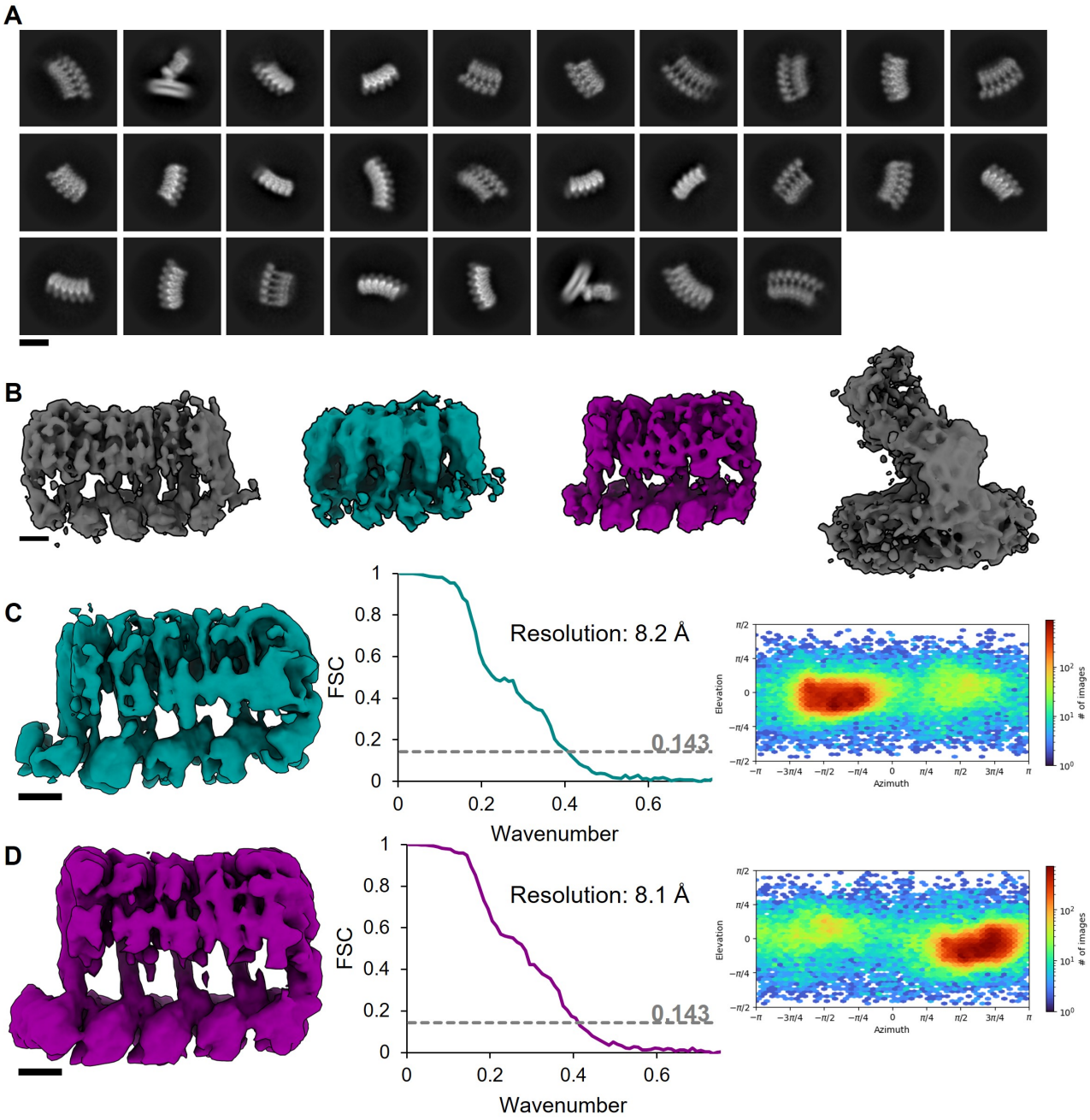


Figure 4.20 | Initial reconstructions from the eBIC datasets. **A.** 2D classes of the particles used for the initial reconstructions. **B.** Four classes from the *ab initio* reconstruction. The turquoise and purple classes were refined separately. **C.** Initial consensus map of the new ILY conformation (left) with its FSC curve (centre) and orientation plot (right). **D.** Initial consensus map of the ILY early prepore (left) with its FSC curve (centre) and orientation plot (right). Scale bar: 20 Å (all panels).

unbinned. The unbinned particles were homogeneously refined against the previously generated map. CTF estimation on micrographs collected at a tilt angle is often inaccurate as the grid is not flat relative to the electron beam, resulting in parts of the micrograph being at different defoci. To correct for this, per-particle CTF estimation was performed, followed by non-

uniform refinement. The resulting reconstruction (from 111,319 particles) was a hexamer, has a reported resolution of 4.31 Å and shows the best density for the central monomers (Figure 4.21). It includes density attributable to the nanodisc, but the protein density does not lie flat on the nanodisc, which suggests errors in the alignment (Figure 4.21D). This is supported by the S-shape of angular assignments in the orientation plot (Figure 4.21C), which is uncharacteristic of well-refined cryoEM maps.

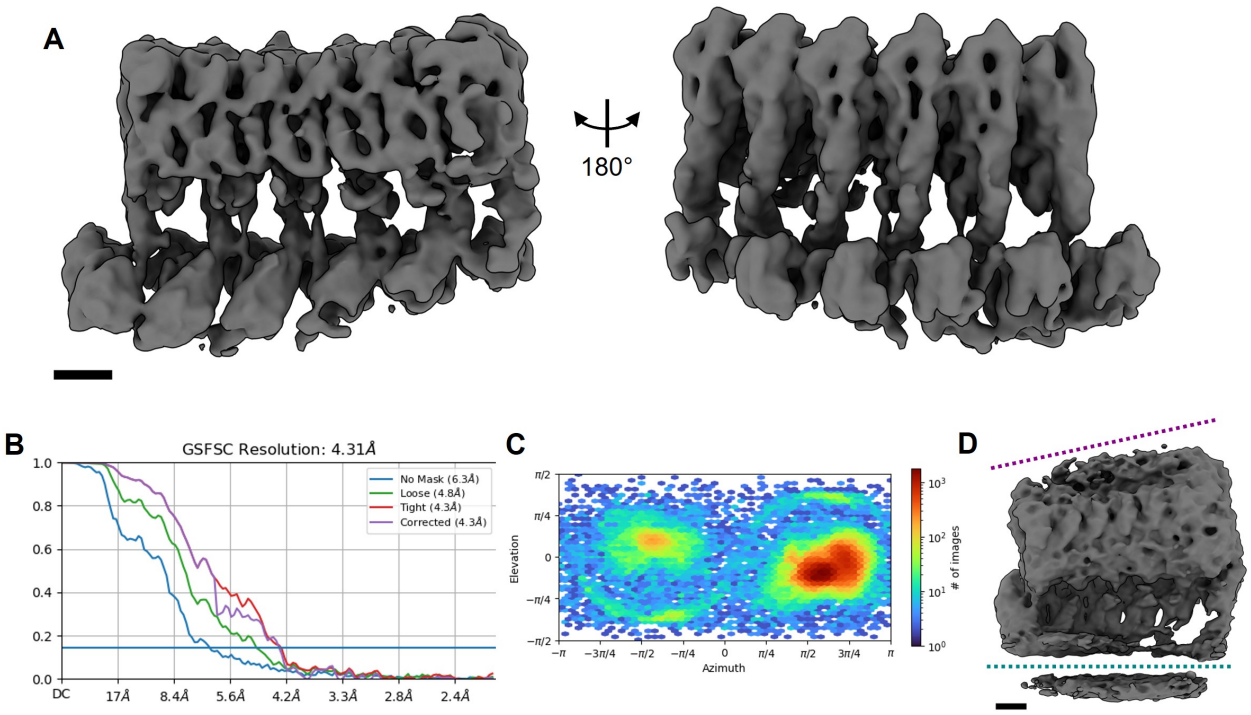


Figure 4.21 | Final reconstruction of the early prepore. **A.** Consensus map of the early prepore. **B.** FSC curve of the reconstruction in panel A. **C.** Orientation plot of the reconstruction in panel A. **D.** A high visualisation threshold reveals signal corresponding to the nanodisc at the bottom of the protein density. The plane of the top of the MACPF/CDC domains (purple line) is not parallel with the plane of the nanodisc (turquoise line) suggesting errors in the alignment. Scale bar: 10 Å (panels A and D).

The particles that yielded the reconstruction in Figure 4.20C were processed similarly to those corresponding to the early prepore. The final consensus map refined from 117,371 particles reached a reported resolution of 4.3 Å and represents a pentamer of ILY^{HB1}-*nocharge* (Figure 4.22). The orientation plot indicates that the structure suffers from preferential orientation. This was confirmed by running a 2D classification job with the final particle set, which revealed the absence of side views in this reconstruction. However, it is unclear whether

the side views observed in the dataset correspond to a different conformation or whether the nanodisc density prevented the algorithm from correctly aligning side views with other views. For both the hexameric and pentameric maps, 3D classification, heterogeneous refinement and local refinement with masks around the central monomers all failed to improve the resolution.

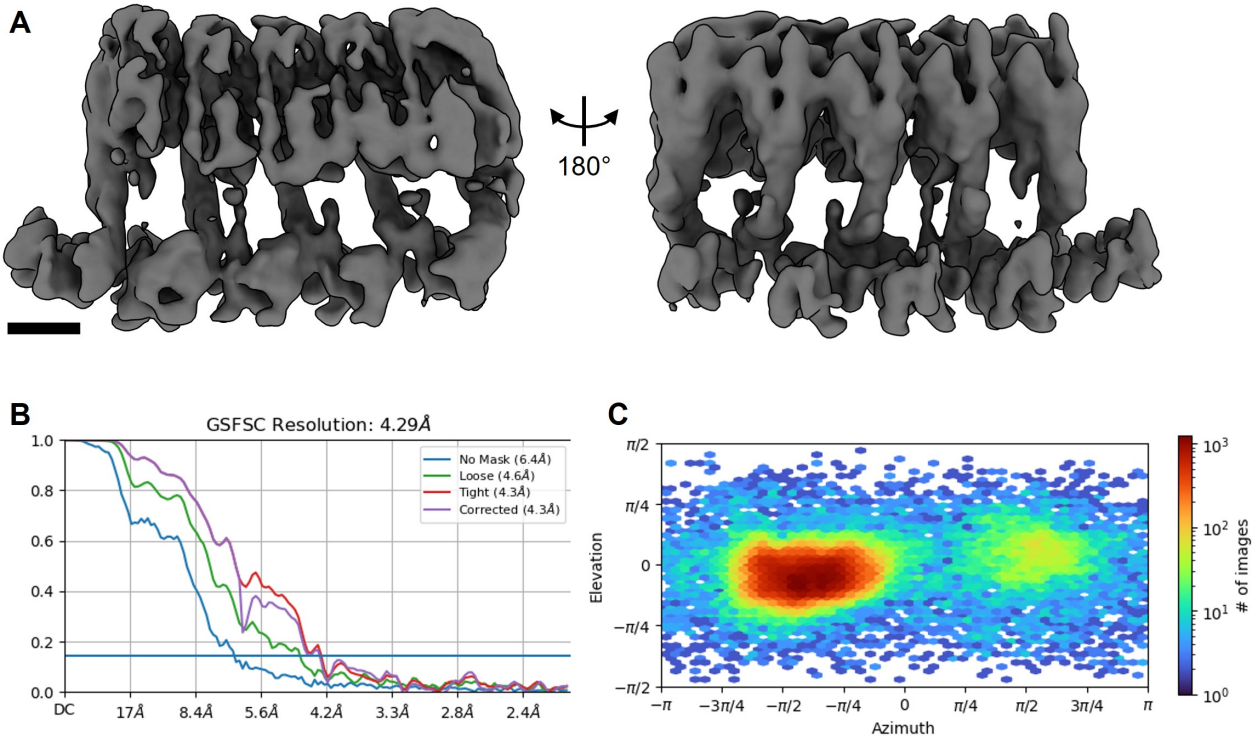


Figure 4.22 | Final reconstruction of the potential new ILY conformation. **A.** Consensus map resulting from non-uniform refinement. Scale bar: 10 Å. **B.** FSC curve of the reconstruction in panel A. **C.** Orientation plot of the reconstruction in panel A.

To assign a conformational state to each of the two reconstructions (the hexameric map in Figures 4.21 and the pentameric map in Figure 4.22, respectively), the early prepore structure (PDB 6ZD0) was fitted as a rigid body into each map in ChimeraX [170] and the local map-model correlation coefficient (CC) was calculated in Phenix [252] (Figure 4.23). The hexameric map shows a better fit with the early prepore model (CC = 0.528) than the pentameric map (CC = 0.326). To improve the fit, the trimeric structure was split in two parts: D4-CD59 and D1-D3 (corresponding to the MACPF/CDC domain), and each was rigid-body fitted individually into the pentameric map. This results in D4-CD59 being shifted

up to fill the density and an increase in local map-model correlation coefficient from 0.387 to 0.467 (Figure 4.23C). Similarly, D1-D3 moves down and the CC increases from 0.244 to 0.341. The maps presented here are initial reconstructions which can still be improved on, resulting in rather poor correlation coefficients overall, but the improvements in fitting suggest that the pentameric reconstruction may represent a different conformation of ILY, potentially corresponding to the late prepore. Measuring the distance between the top of the MACPF/CDC domain and the bottom of D4 suggests a decrease in height of $\sim$1.2 nm from the early prepore to the new conformation (Figure 4.23F). The pentameric map also shows density for $CD59^{cyto}$, suggesting that it remains bound to the late prepore.

4.4 Discussion

Pore-forming proteins of the MACPF/CDC superfamily are challenging structural targets because of their need for a lipid bilayer on which to oligomerise. Nanodiscs are a powerful tool as they provide a substrate for them, yielding complexes that can be imaged by cryoEM. However, the assembly pathway of those proteins, and of CDCs in particular, pose several challenges for structure determination. Resolving intermediates requires trapping the protein in a specific state, which can be achieved by engineering disulfide bonds [50, 188]. As some of those intermediates can still disrupt membranes, further inactivating mutations are sometimes needed to stabilise the complex long enough for structure determination. While incubating ILY^{HB1}-*nocharge* on $CD59^{cyto}$-decorated nanodiscs is enough for the complex to form, this results in high levels of heterogeneity as some nanodiscs remain unoccupied, and those on which ILY^{HB1}-*nocharge* deposits show a range of subunit number. Sucrose density gradient centrifugation was successful in purifying assembled complexes from unoccupied nanodiscs and, to a lesser extent, in separating complexes according to oligomeric order. This approach is particularly powerful for separating populations when affinity chromatogra-

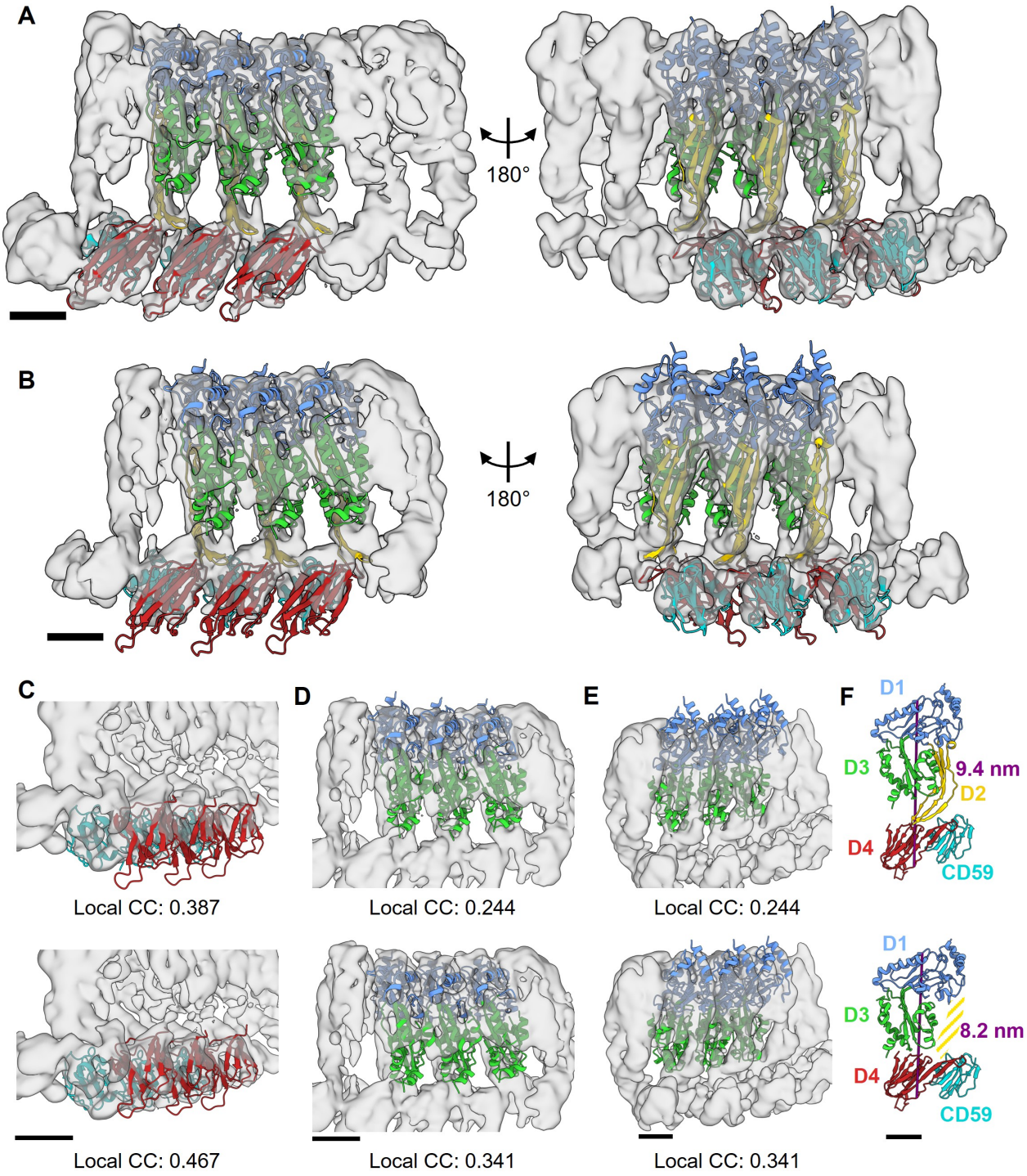


Figure 4.23 | Fitting the early prepore structure into the cryoEM reconstructions. **A.** Hexameric map overlayed with the early prepore model (PDB 6ZD0). **B.** Pentameric map overlayed with the early prepore model (PDB 6ZD0). **C.** D4-CD59 aligned to the fit of the entire early prepore into the pentameric map (top) compared to D4-CD59 rigid-body fitted alone (bottom). **D.** D1-D3 from aligned to the fit of the entire early prepore into the pentameric map (top) compared to D1-D3 rigid-body fitted alone (bottom) **E.** Different view of the fit in panel D. **F.** Relative position of the domains in the early prepore structure (top) and resulting from individually fitting D4-CD59 and D1-D3 into the pentameric map (bottom). In both cases, the distance between the Cα carbons of Val-173 and Gly-459 is shown in purple. Blue: D1, yellow: D2, green: D3, red: D4, turquoise: CD59. Scale bar: 20 Å (all panels).

phy or size-exclusion chromatography are not viable options.

While high-resolution reconstructions can only be achieved with cryoEM, electron microscopy (EM) at non-cryogenic temperatures is a powerful tool for sample preparation and characterisation. In this project, negative stain EM was used to visualise complex formation and to check the integrity of the sample at each step, in particular during buffer exchange after the sucrose density centrifugation. Analysing the gradient fractions by SDS-PAGE suggested the complex of interest was present in many fractions, while negative stain EM imaging rapidly identified the ideal fractions for structure determination by cryoEM. A small data collection, run over a few hours, can be enough to yield a low-resolution structure for initial sample characterisation, which can be helpful to guide sample preparation.

A downside of sucrose density gradients is the need to remove the sucrose after the purification, as it would lower the signal-to-noise ratio and prevent high resolution reconstruction by cryoEM. Sucrose removal can be a bottleneck in cases, such as this project, where the complex is too unstable to withstand dialysis, treatment with desalting columns, or even extensive concentration. Chemical fixation with a mild crosslinker, such as glataraldehyde or BS^3, can improve sample stability and facilitate handling in the wet lab. This approach was instrumental in generating sucrose-free samples compatible with cryoEM and the results presented in this chapter highlight some of the advantages of using different crosslinkers to achieve different levels of fixation. The GraFix approach is particularly powerful in that it combines the purification and resolving power of a density gradient with the stabilising effects of chemical fixation [253]. Moreover, crosslinkers change the charge properties of biological macromolecules by reacting with amino acid side chains, preferentially positively-charged lysines in the case of glutaraldehyde and BS^3. Other crosslinkers target specifically acidic residues [256], and combining reagents of different lengths and different specificities is likely to affect a given protein complex differently. As the charge properties of particles

can influence their distribution on grids, altering them with various crosslinkers may alleviate preferential orientation [257].

Preferential orientation can also be tackled with other approaches at the sample preparation stage. The two 200 kV datasets presented above highlight how particles can behave differently on different grid types. In the Gold Lacey dataset, only $\sim$1% of 'good' particles classified into side views, while a variety of top and front views could be observed in two dimensions. When frozen on a graphene oxide (GrOx) support, $\sim$80% of particles classified into side views, which showed a much wider variety of orientations than the side views observed on Gold Lacey grids. While there are many ways to influence particle orientation on grids (e.g. crosslinking, addition of surfactants, adsorption on support, different grid types or hole sizes, binding to megabobies) [241, 243], no systematic solution has yet been found to work on all samples. When preferred orientation cannot be solved at the sample preparation stage, collecting data at a tilt angle may help alleviate its impact on 3D reconstructions [258]. Collecting two datasets in-house on a 200 kV instrument revealed that the sample suffered from preferential orientation, which guided the data collection strategy on the higher-end 300 kV Titan Krios, where micrographs were recorded at 0° and $-30°$tilt.

The data collected on a 300 kV Titan Krios microscope showed a better orientation distribution than the initial two 200 kV datasets, as evidenced by the appearance of new views in two dimension. The inclusion of tilted data is likely contributing many of these views, but the higher proportion of side views in the untilted dataset ($\sim$3.5% compared to $\sim$1% in the 200 kV Gold Lacey dataset) suggests that the double crosslinking with glutaraldehyde and BS^3 may have also played a part. However, it may be that training a Topaz model simply recovered a higher proportion of side views from that dataset than the blob picker had been able to detect in the 200 kV Gold Lacey dataset.

The image processing presented in this Chapter only represents the early steps of struc-

ture determination. Two reconstructions were generated, of a hexamer and of a pentamer, but due to time constraints, they are of limited resolution. Beyond their number of subunits, the more biologically relevant difference between the two maps is that they may represent different conformational states. Indeed, the early prepore structure fits well in the hexameric map, but the fit is poorer in the pentameric one. In the latter, D4 and the MACPF/CDC domain (D1-D3) both seem to be shifted partly out of density (Figure 4.23). Removing D2 from the structure and rigid-body fitting a trimer of D4-CD59 and a trimer of D1-D3 separately resulted in an increase in local map-model correlation coefficients for the pentameric map. This suggests that the relative positions of the domains is different from that of the early prepore, possibly indicating a different conformation. As the resulting structure is shorter in height than the early prepore, it may correspond to the late prepore. Measuring the distance between domains in both fits (the full early prepore structure and the individually-fitted domains) suggests a decrease of only $\sim$1.2 nm, rather than the 2 nm expected from AFM data (cf. Chapter 3). However, assigning a precise height to a structure is challenging as the exact orientation of ILY on the membrane is unknown. The nanodisc density visible in Figure 4.21D suggests that the data from this project may shed a light on this, although it is clear that there are still errors in the alignment and more processing would be required to place the ILY oligomer in the context of the membrane with confidence.

While the two reconstructions from the 300 kV datasets may correspond to the early and the late prepore, respectively, the quality of the map is not high enough to assign a new conformational state to the pentameric model with confidence. For instance, although rigid-body fitting D4-CD59 and D1-D3 independently led to an increase in local map-model correlation, the correlation coefficients (CCs) are still relatively low. In the case of D4-CD59, the CC is probably limited mostly by the quality of the map. In the case of D1-D3, it is also possible that the MAPCF/CDC domain adopts a different conformation in this structure. If the map truly

corresponds to the late prepore, it is likely that the β-strands of the MACPF/CDC domain are shifted backwards to form a more planar sheet across ILY subunits, which would explain why CDC late prepores are insensitive to SDS [211, 212]. Although the two reconstructions have a reported resolution around 4.3 Å, the map features are more consistent with a resolution around 5-6 Å as α-helices are visible, while the separation between adjacent β-strands is not [54].

Several factors are likely limiting the resolution of the reconstructions presented here. Despite the presence of new views in 2D classes, the orientation plot in Figure 4.22C shows that the pentameric map stil suffers from preferred orientation. This could induce artefacts in 3D reconstructions, including the elongation of the map in one direction, or its compression in another, which could explain the height difference compared to the hexameric reconstruction. However, more views may still be recovered from the data. Indeed, 2D classification of the particles from each of the two maps revealed that only the hexameric reconstruction included side views. As both the early prepore and the late prepore are peripheral membrane proteins, correct alignment of the nanodisc with the rest of the complex can be challenging. This is evidenced by the angle between the nanodisc and ILY in Figure 4.21D and the fact that, to obtain the original early prepore structure, particles had to be picked manually to avoid signal from the nanodisc [50]. Therefore, there may be side views corresponding to the pentameric map which were not assigned to it during heterogeneous refinement due to the difficulty of aligning them with other views. This may be remedied in the future by subtracting the nanodisc signal from the particles so that only the proteins of interest drive the alignment.

Another probable limiting factor to the quality of the maps presented here is heterogeneity in the data. The height difference between the two reconstructions is small, and capturing it during image processing is challenging. This is further complicated by the fact that the pre-

dominant form of heterogeneity is not conformational, but compositional: it arises from the range of subunit numbers in the complexes. The most obvious difference between the two reconstruction is that one is a hexamer, and the other a pentamer. However, subunit dissociation upon collapse has not been observed in CDCs, so there is no biological reason why the early prepore should be predominantly hexameric while the late prepore is pentameric. This suggests that, during heterogeneous refinement, the number of subunits was a larger driving force for classification than the conformational state was. Applying more advanced computational methods to deal with heterogeneity (e.g. the new 3D classification algorithm in Cryosparc v4, 3D variability analysis [259], CryoDRGN [64]) may help separate conformationally different populations, which would determine with more confidence whether the ILY late prepore was successfully resolved.

While more work is needed to obtain a high-resolution structure of the ILY late prepore, the data presented here are significant steps towards it. A sucrose density gradient purification combined with chemical fixation by two mild crosslinkers yielded a pure and stable complex in high enough quantities for cryoEM grid preparation. Optimisation of freezing conditions resulted in good particle distribution and ice thickness on Gold Lacey grids. The initial data processing in this Chapter suggests that there may be two conformationally different populations in the datasets, potentially corresponding to the early and the late prepore. Future efforts should focus on disentangling both conformational and compositional heterogeneity, which are likely limiting the resolution of the reconstructions. Preferred orientation and incorrect angular assignment are also currently limiting the quality of the maps. However, resolving heterogeneity in the data may help align particles better and recover views which have so far been lost during 3D reconstruction as the algorithm failed to match them with other views of the same sample. Solving the first high-resolution structure of a CDC late prepore will provide us with more information on how the h-helix carries out its membrane-

priming function, significantly contributing to our understanding of pore formation by proteins from the MACPF superfamily. It will also reveal whether structural changes in the ILY-CD59 interface prime the complex for CD59 release upon insertion of the MACPF/CDC β-hairpins.

5 || Conclusions and future directions

5.1 Thesis summary

The work presented in this thesis combines molecular dynamics (MD) simulations, functional assays and cryo-electron microscopy (cryoEM) to gain new structural and mechanistic insights into the membrane attack complex/perforin (MACPF) superfamily of pore-forming proteins. This work focuses on MAC, a human immune pore, and bacteria cholesterol-dependent cytolysins (CDCs), which both form β-barrelled pores.

Atomistic simulations of GPI-anchored CD59 in Chapter 2 revealed that flexibility in the GPI sugars and the C-terminus of the protein causes CD59 to adopt a wide range of orientations relative to the membrane. The resulting orientational landscape is linked to the function of CD59. Among the preferentially sampled poses are some where the final β-strand is in the correct position to engage C8α and inhibit MAC pore formation. Other orientations make the same β-strand available for binding the CDC intermedilysin (ILY) and aiding pore formation. Finally, CD59 can adopt a *flop-down* orientation where its final β-strand contacts the membrane and is unavailable to engage C8α and ILY. As protein-lipid interactions play a role in keeping CD59 in its *flop-down* orientation, the composition of the membrane, which varies with cell type, likely contributes to determining CD59 activity [260, 261].

The sequential assembly of MAC on lipid bilayers triggers changes in the physical properties of the membrane [38].To study the impact of inhibited MAC on lipids in the microsecond timescale, I simulated a coarse-grain model of the CD59-C5b8 complex. This revealed protein-lipid interactions, which may be involved in anchoring the complex to the membrane,

as has been suggested in the literature [38, 136]. The complex also induces membrane thinning near the tip of the C8β MACPF/CDC hairpins. These observations are consistent with the cryoEM maps reconstructed by Dr. Emma Couves [143]. Local changes in the lipid environment of complement proteins may be involved in recruiting CD59 and/or the molecular machinery for clearing inhibited MAC from the cell membrane.

Bacterial CDCs follow a stepwise pore-formation pathway involving several conformational changes. The cryoEM structure of the first oligomeric intermediate (the early prepore) solved by Dr. Nita Shah revealed the movement of a key amphipathic α-helix (the h-helix) in the MACPF/CDC domain: the horizontal (h-) helix [50]. Site-directed mutagenesis of the h-helix reported in Chapter 3 showed that it is responsible for tuning the lytic activity of CDCs. The second oligomeric assembly of CDCs (the late prepore), likely brings the h-helix close to the bilayer, where it disturbs lipids before insertion of the transmembrane β-hairpins, thus lowering the energy barrier for membrane hydration. The presence of similar amphipathic helices in other membrane-lysing proteins suggests that this may be a common mechanism of pore formation. Other members of the MACPF superfamily, such as MAC and perforin-1, do not form a prepore with the MACPF/CDC domain helical bundles near the membrane. Upstream complement pathways or, in the case of perforin-1, initial binding to the membrane may disturb the bilayer to facilitate pore formation. Moreover, the extensive network of hydrogen bonds formed by the comparatively longer β-hairpins may provide additional stabilising energy.

A complete understanding of how the h-helix performs its function will require solving the high-resolution structure of the CDC late prepore on a membrane mimetic. Chapter 4 provides the first steps towards this goal by using a disulfide-locked mutant of ILY characterised in Chapter 3. Further inactivating mutations in the h-helix helped stabilise the complex assembled on lipid nanodiscs decorated with a cytotopically labelled CD59 construct. Sucrose

density gradient centrifugation yielded a pure sample for structure determination by cryoEM. I optimised freezing conditions for cryoEM grids and recorded data on 200 kV and 300 kV instruments. The initial steps of image processing in Chapter 4 highlight compositional and conformational heterogeneity in the sample, as well as preferred orientation, which limit the interpretability of the reconstructions. However, there is evidence for populations of early and late prepore complexes on the grids, suggesting that further processing may generate a high-resolution structure of the ILY late prepore.

5.2 Future directions

Atomistic MD simulations of CD59 suggest that the orientation of the protein may affect its function by modulating the availability of its protein-binding site. While only three CDCs are known to require CD59 for pore formation [190, 204, 205], there is growing evidence that other CDCs interact with various molecules on the surface of cells, including specific carbohydrates [207, 208]. A complete understanding of pathogenicity and virulence of CDC-secreting bacteria will require further investigations into potential cell surface receptors dictating target specificity and cell tropism. Elucidating the molecular basis of interactions driving CDC function may help engineer constructs for various applications. For instance, the CDC perfringolysin O has been used to create various biosensors to study the organisation and dynamics of cholesterol in cell membranes [262, 263, 264, 265, 266]. Listeriolysin O has been used for gene delivery to breast cancer cells, where specificity was achieved by using the antibody Trastuzumab [267]. The overactivating h-helix mutations described in Chapter 3 enable ILY to lyse liposomes in the absence of cholesterol (data generated by Dr. Nita Shah, published in [50]). These mutations may provide a basis for engineering CDCs with new lipid specificity, potentially helping target pathogens or cancer cells.

A number of outstanding questions remain on the structural basis of pore formation by

CDCs, including the existence of one or multiple late prepore-like conformations, the mechanism of action of the h-helix, and the timing of CD59 release. Disulfide-locked mutants have often been used to trap specific steps of the pore-forming pathway [32, 50, 182, 188, 230], but they are based on prior knowledge of expected conformations. As a result, they may not correspond to physiological intermediates and may overlook previously unknown states. The challenges of studying conformational intermediates are not limited to CDCs, but are common to many biological systems. Efforts in developing new approaches for time-resolved cryoEM are likely to shine a new light on conformational transitions of proteins [231, 268, 269, 270, 271]. Single-molecule fluorescence imaging of perfringolysin O and atomic force microscopy of suilysin indicate that CDC pore formation occurs on a timescale of minutes after oligomerisation [210, 215]. A potential time-resolved cryoEM strategy to solve the structure of the CDC late prepore could involve assembling and purifying a disulfide-locked early prepore following the protocol in Chapter 4. Applying the sample on a cryoEM grid and subsequently adding a reducing agent would allow collapse and generate a more native-like late prepore, which could be trapped by freezing the grid before the MACPF/CDC helices can unfurl.

Understanding the full impacts of pore formation on cells will require *in situ* structural biology approaches. Cryo-electron tomography (cryoET) and subtomogram averaging are powerful techniques for studying biological macromolecules in their native cellular environment [272, 273]. CryoET imaging of MAC on liposomes has highlighted membrane deformation as a result of MAC challenge, showing that it is a viable strategy to study the impact of complement proteins on their target membranes [274]. Imaging MAC assemblies on bacteria and human cells is likely to generate new insights into the mechanisms of MAC lysis, inhibition and clearance.

References

[1] Mirko Bischofberger, Ioan Iacovache, and F. Gisou van der Goot. Pathogenic Pore-Forming Proteins: Function and Host Response. *Cell Host & Microbe*, 12(3):266–275, 2012.

[2] Matteo Dal Peraro and F. Gisou van der Goot. Pore-forming toxins: Ancient, but never really out of fashion. *Nature Reviews Microbiology*, 14(2):77–92, 2016.

[3] Uris Ros, Lohans Pedrera, and Ana J. Garcia-Saez. Partners in Crime: The Interplay of Proteins and Membranes in Regulated Necrosis. *International Journal of Molecular Sciences*, 21(7):2412, 2020.

[4] Haydee Mesa-Galloso, Lohans Pedrera, and Uris Ros. Pore-forming proteins: From defense factors to endogenous executors of cell death. *Chemistry and Physics of Lipids*, 234:105026, 2021.

[5] Peter Vandenabeele, Geert Bultynck, and Savvas N. Savvides. Pore-forming proteins as drivers of membrane permeabilization in cell death pathways. *Nature Reviews Molecular Cell Biology*, pages doi: 10.1038/s41580–022–00564–w, 2022.

[6] Pradeep Sathyanarayana, Sandhya S. Visweswariah, and K. Ganapathy Ayappa. Mechanistic Insights into Pore Formation by an α-Pore Forming Toxin: Protein and Lipid Bilayer Interactions of Cytolysin A. *Accounts of Chemical Research*, 54(1):120–131, 2021.

[7] Mahendra Madegowda, Subramaniam Eswaramoorthy, Stephen K. Burley, and Subramanyam Swaminathan. X-ray crystal structure of the B component of Hemolysin BL fromBacillus cereus. *Proteins: Structure, Function, and Bioinformatics*, 71(2):534–540, 2008.

[8] Stuart Hunt, Jeffrey Green, and Peter J. Artymiuk. Hemolysin E (HlyE, ClyA, SheA) and Related Toxins. In Nathan Back, Irun R. Cohen, Abel Lajtha, John D. Lambris, Rodolfo Paoletti, Gregor Anderluh, and Jeremy Lakey, editors, *Proteins Membrane Binding and Pore Formation*, volume 677, pages 116–126. Springer New York, New York, 2010. ISBN 978-1-4419-6326-0 978-1-4419-6327-7.

[9] Alistair J. Wallace, Timothy J Stillman, Angela Atkins, Stuart J. Jamieson, Per A. Bullough, Jeffrey Green, and Peter J. Artymiuk. E. coli Hemolysin E (HlyE, ClyA, SheA): X-Ray Crystal Structure of the Toxin and Observation of Membrane Pores by Electron Microscopy. *Cell*, 100:265–276, 2000.

[10] Marcus Mueller, Ulla Grauschopf, Timm Maier, Rudi Glockshuber, and Nenad Ban. The structure of a cytolytic α-helical toxin pore reveals its assembly mechanism. *Nature*, 459(7247):726–730, 2009.

[11] Eric Cascales, Susan K. Buchanan, Denis Duché, Colin Kleanthous, Roland Lloubès, Kathleen Postle, Margaret Riley, Stephen Slatin, and Danièle Cavard. Colicin Biology. *Microbiology and Molecular Biology Reviews*, 71(1):158–229, 2007.

[12] Seunghyon Choe, Melanie J. Bennett, Gary Fujii, Paul M. G. Curmi, Katherine A. Kantardjieff, R. John Collier, and David Eisenberg. The crystal structure of diphtheria toxin. *Nature*, 357(6375):216–222, 1992.

[13] Chengchen Xu, Bi-Cheng Wang, Ziniu Yu, and Ming Sun. Structural Insights into Bacillus thuringiensis Cry, Cyt and Parasporin Toxins. *Toxins*, 6(9):2732–2770, 2014.

[14] Katarina Črnigoj Kristan, Gabriella Viero, Mauro Dalla Serra, Peter Maček, and Gregor Anderluh. Molecular mechanism of pore formation by actinoporins. *Toxicon*, 54(8): 1125–1134, 2009.

[15] Koji Tanaka, Jose M.M. Caaveiro, Koldo Morante, Juan Manuel González-Mañas, and Kouhei Tsumoto. Structural basis for self-assembly of a cytolytic pore lined by protein and lipid. *Nature Communications*, 6(1):6337, 2015.

[16] Nejc Rojko, Katarina Č. Kristan, Gabriella Viero, Eva Žerovnik, Peter Maček, Mauro Dalla Serra, and Gregor Anderluh. Membrane Damage by an α-Helical Pore-forming Protein, Equinatoxin II, Proceeds through a Succession of Ordered Steps. *Journal of Biological Chemistry*, 288(33):23704–23715, 2013.

[17] Christos G. Savva, Sérgio P. Fernandes da Costa, Monika Bokori-Brown, Claire E. Naylor, Ambrose R. Cole, David S. Moss, Richard W. Titball, and Ajit K. Basak. Molecular Architecture and Functional Analysis of NetB, a Pore-forming Toxin from Clostridium perfringens. *Journal of Biological Chemistry*, 288(5):3512–3522, 2013.

[18] Maria Manich, Oliver Knapp, Maryse Gibert, Elke Maier, Colette Jolivet-Reynaud, Blandine Geny, Roland Benz, and Michel R. Popoff. Clostridium perfringens Delta Toxin Is Sequence Related to Beta Toxin, NetB, and Staphylococcus Pore-Forming Toxins, but Shows Functional Differences. *PLoS ONE*, 3(11):e3764, 2008.

[19] Anthony L. Keyburn, Trudi L. Bannam, Robert J. Moore, and Julian I. Rood. NetB, a Pore-Forming Toxin from Necrotic Enteritis Strains of Clostridium perfringens. *Toxins*, 2(7):1913–1927, 2010.

[20] Swastik De and Rich Olson. Crystal structure of the *Vibrio cholerae* cytolysin heptamer reveals common features among disparate pore-forming toxins. *Proceedings of the National Academy of Sciences*, 108(18):7385–7390, 2011.

[21] Daichi Yamashita, Takaki Sugawara, Miyu Takeshita, Jun Kaneko, Yoshiyuki Kamio, Isao Tanaka, Yoshikazu Tanaka, and Min Yao. Molecular basis of transmembrane beta-barrel formation of staphylococcal pore-forming toxins. *Nature Communications*, 5(1):4897, 2014.

[22] Langzhou Song, Michael R. Hobaugh, Christopher Shustak, Stephen Cheley, Hagan Bayley, and J. Eric Gouaux. Structure of Staphylococcal α-Hemolysin, a Heptameric Transmembrane Pore. *Science*, 274(5294):1859–1865, 1996.

[23] Pawel Szczesny, Ioan Iacovache, Anna Muszewska, Krzysztof Ginalski, F. Gisou van der Goot, and Marcin Grynberg. Extending the Aerolysin Family: From Bacteria to Vertebrates. *PLoS ONE*, 6(6):e20349, 2011.

[24] Onya Opota, Isabelle Vallet-Gély, Renaud Vincentelli, Christine Kellenberger, Ioan Iacovache, Manuel Rodrigo Gonzalez, Alain Roussel, F. Gisou van der Goot, and Bruno Lemaitre. Monalysin, a Novel β-Pore-Forming Toxin from the Drosophila Pathogen

Pseudomonas entomophila, Contributes to Host Intestinal Damage and Lethality. *PLoS Pathogens*, 7(9):e1002259, 2011.

[25] Ioan Iacovache, Mirko Bischofberger, and F Gisou van der Goot. Structure and assembly of pore-forming proteins. *Current Opinion in Structural Biology*, 20(2):241–246, 2010.

[26] Richard Galinier, Julien Portela, Yves Moné, Jean François Allienne, Hélène Henri, Stéphane Delbecq, Guillaume Mitta, Benjamin Gourbal, and David Duval. Biomphalysin, a New β Pore-forming Toxin Involved in Biomphalaria glabrata Immune Defense against Schistosoma mansoni. *PLoS Pathogens*, 9(3):e1003216, 2013.

[27] Yang Xiang, Chao Yan, Xiaolong Guo, Kaifeng Zhou, Sheng'an Li, Qian Gao, Xuan Wang, Feng Zhao, Jie Liu, Wen-Hui Lee, and Yun Zhang. Host-derived, pore-forming toxin–like protein and trefoil factor complex protects the host against microbial infection. *Proceedings of the National Academy of Sciences*, 111(18):6702–6707, 2014.

[28] Matteo T Degiacomi, Ioan Iacovache, Lucile Pernot, Mohamed Chami, Misha Kudryashev, Henning Stahlberg, F. Gisou van der Goot, and Matteo Dal Peraro. Molecular assembly of the aerolysin pore reveals a swirling membrane-insertion mechanism. *Nature Chemical Biology*, 9(10):623–629, 2013.

[29] Carlos J. Rosado, Stephanie Kondos, Tara E. Bull, Michael J. Kuiper, Ruby H. P. Law, Ashley M. Buckle, Ilia Voskoboinik, Phillip I. Bird, Joseph A. Trapani, James C. Whisstock, and Michelle A. Dunstone. The MACPF/CDC family of pore-forming toxins. *Cellular Microbiology*, 10(9):1765–1774, 2008.

[30] Michael A. Hadders, Dennis X. Beringer, and Piet Gros. Structure of C8α-MACPF Reveals Mechanism of Membrane Attack in Complement Immune Defense. *Science*, 317(5844):1552–1554, 2007.

[31] Michael E. D'Angelo, Michelle A. Dunstone, James C. Whisstock, Joseph A. Trapani, and Phillip I. Bird. Perforin evolved from a gene duplication of MPEG1, followed by a complex pattern of gene gain and loss within Euteleostomi. *BMC Evolutionary Biology*, 12(1):59, 2012.

[32] Rodney K. Tweten, Eileen M. Hotze, and Kristin R. Wade. The Unique Molecular Choreography of Giant Pore Formation by the Cholesterol-Dependent Cytolysins of Gram-Positive Bacteria. *Annual Review of Microbiology*, 69(1):323–340, 2015.

[33] Natalya Lukoyanova, Stephanie C. Kondos, Irene Farabella, Ruby H. P. Law, Cyril F. Reboul, Tom T. Caradoc-Davies, Bradley A. Spicer, Oded Kleifeld, Daouda A. K. Traore, Susan M. Ekkel, Ilia Voskoboinik, Joseph A. Trapani, Tamas Hatfaludi, Katherine Oliver, Eileen M. Hotze, Rodney K. Tweten, James C. Whisstock, Maya Topf, Helen R. Saibil, and Michelle A. Dunstone. Conformational Changes during Pore Formation by the Perforin-Related Protein Pleurotolysin. *PLOS Biology*, 13(2):e1002049, 2015.

[34] Kristin R. Wade and Rodney K. Tweten. The Apicomplexan CDC/MACPF-like pore-forming proteins. *Current Opinion in Microbiology*, 26:48–52, 2015.

[35] Jamie Rossjohn, Susanne C. Feil, William J. McKinstry, Rodney K. Tweten, and Michael W. Parker. Structure of a Cholesterol-Binding, Thiol-Activated Cytolysin and a Model of Its Membrane Form. *Cell*, 89(5):685–692, 1997.

[36] Jasmine K. Bickel, Tomas B. Voisin, Edward W. Tate, and Doryen Bubeck. How Structures of Complement Complexes Guide Therapeutic Design. In J. Robin Harris and Jon Marles-Wright, editors, *Macromolecular Protein Complexes III: Structure and Function*, volume 96, pages 273–295. Springer International Publishing, Cham, 2021. ISBN 978-3-030-58970-7.

[37] Daniel Ricklin, George Hajishengallis, Kun Yang, and John D. Lambris. Complement: A key system for immune surveillance and homeostasis. *Nature Immunology*, 11(9): 785–797, 2010.

[38] Anaïs Menny, Marina Serna, Courtney M. Boyd, Scott Gardner, Agnel Praveen Joseph, B. Paul Morgan, Maya Topf, Nicholas J. Brooks, and Doryen Bubeck. CryoEM reveals how the complement membrane attack complex ruptures lipid bilayers. *Nature Communications*, 9(1):5316, 2018.

[39] Eckhard R. Podack, John Ding-E Young, and Zanvil A. Cohn. Isolation and biochemical and functional characterization of perforin 1 from cytolytic T-cell granules. *Proceedings of the National Academy of Sciences*, 82:8629–8633, 1985.

[40] John Ding-E Young, Hans Hengartner, Eckhard R. Podack, and Zanvil A. Cohn. Purification and characterization of a cytolytic pore-forming protein from granules of cloned lymphocytes with natural killer activity. *Cell*, 44(6):849–859, 1986.

[41] John Ding-E Young, Zanvil A. Cohn, and Eckhard R. Podack. The Ninth Component of Complement and the Pore-Forming Protein (Perforin 1) from Cytotoxic T Cells: Structural, Immunological, and Functional Similarities. *Science*, 233(4760):184–190, 1986.

[42] Carl Leung, Adrian W. Hodel, Amelia J. Brennan, Natalya Lukoyanova, Sharon Tran, Colin M. House, Stephanie C. Kondos, James C. Whisstock, Michelle A. Dunstone, Joseph A. Trapani, Ilia Voskoboinik, Helen R. Saibil, and Bart W. Hoogenboom. Real-time visualization of perforin nanopore assembly. *Nature Nanotechnology*, 12(5):467–473, 2017.

[43] Marina E. Ivanova, Natalya Lukoyanova, Sony Malhotra, Maya Topf, Joseph A Trapani, Ilia Voskoboinik, and Helen R Saibil. The pore conformation of lymphocyte perforin. *Science Advances*, 8:eabk3147, 2022.

[44] Leidy C. Merselis, Zachary P. Rivas, and George P. Munson. Breaching the Bacterial Envelope: The Pivotal Role of Perforin-2 (MPEG1) Within Phagocytes. *Frontiers in Immunology*, 12:597951, 2021.

[45] Siew Siew Pang, Charles Bayly-Jones, Mazdak Radjainia, Bradley A. Spicer, Ruby H. P. Law, Adrian W. Hodel, Edward S. Parsons, Susan M. Ekkel, Paul J. Conroy, Georg Ramm, Hariprasad Venugopal, Phillip I. Bird, Bart W. Hoogenboom, Ilia Voskoboinik, Yann Gambin, Emma Sierecki, Michelle A. Dunstone, and James C. Whisstock. The cryo-EM structure of the acid activatable pore-forming immune effector Macrophage-expressed gene 1. *Nature Communications*, 10(1):4288, 2019.

[46] Tao Ni, Fang Jiao, Xiulian Yu, Saša Aden, Lucy Ginger, Sophie I. Williams, Fangfang Bai, Vojtěch Pražák, Dimple Karia, Phillip Stansfeld, Peijun Zhang, George Munson, Gregor Anderluh, Simon Scheuring, and Robert J. C. Gilbert. Structure and mechanism of bactericidal mammalian perforin-2, an ancient agent of innate immunity. *Science Advances*, 6(5):eaax8286, 2020.

[47] Xiulian Yu, Tao Ni, George Munson, Peijun Zhang, and Robert J C Gilbert. Cryo-EM structures of perforin-2 in isolation and assembled on a membrane suggest a mechanism for pore formation. *The EMBO Journal*, page e111857, 2022.

[48] Nobuki Sakurai, Jun Kaneko, Yoshiyuki Kamio, and Toshio Tomita. Cloning, expression, and pore-forming properties of mature and precursor forms of pleurotolysin, a sphingomyelin-specific two-component cytolysin from the edible mushroom Pleurotus ostreatus. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, 1679(1):65–73, 2004.

[49] Toshio Tomita, Kayoko Noguchi, Hitomi Mimuro, Fumio Ukaji, Kiyoshi Ito, Noriko Sugawara-Tomita, and Yohichi Hashimoto. Pleurotolysin, a Novel Sphingomyelin-specific Two-component Cytolysin from the Edible Mushroom Pleurotus ostreatus, Assembles into a Transmembrane Pore Complex. *Journal of Biological Chemistry*, 279(26):26975–26982, 2004.

[50] Nita R. Shah, Tomas B. Voisin, Edward S. Parsons, Courtney M. Boyd, Bart W. Hoogenboom, and Doryen Bubeck. Structural basis for tuning activity and membrane specificity of bacterial cytolysins. *Nature Communications*, 11(1):5818, 2020.

[51] Katharina van Pee, Alexander Neuhaus, Edoardo D'Imprima, Deryck J. Mills, Werner Kühlbrandt, and Özkan Yildiz. CryoEM structures of membrane pore and prepore complex reveal cytolytic mechanism of Pneumolysin. *eLife*, 6:e23644, 2017.

[52] Jordan C. Evans, Bronte A. Johnstone, Sara L. Lawrence, Craig J. Morton, Michelle P. Christie, Michael W. Parker, and Rodney K. Tweten. A Key Motif in the Cholesterol-Dependent Cytolysins Reveals a Large Family of Related Proteins. *mBio*, 11(5): e02351–20, 2020.

[53] David B. Williams and C. Barry Carter. *Transmission Electron Microscopy. Part I: Basics*, volume 1. Springer Science, 2nd edition edition, 2009. ISBN 978-0-387-76501-3.

[54] Robert M. Glaeser, Eva Nogales, and Wah Chiu, editors. *Single-Particle Cryo-EM of Biological Macromolecules*. IOP Publishing, 2021. ISBN 978-0-7503-3039-8.

[55] Eva Nogales and Sjors H.W. Scheres. Cryo-EM: A Unique Tool for the Visualization of Macromolecular Complexity. *Molecular Cell*, 58(4):677–689, 2015.

[56] G. McMullan, A.R. Faruqi, and R. Henderson. Direct Electron Detectors. In *Methods in Enzymology*, volume 579, pages 1–17. Elsevier, 2016. ISBN 978-0-12-805382-9.

[57] A.R. Faruqi and G. McMullan. Direct imaging detectors for electron microscopy. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 878:180–190, 2018.

[58] Takanori Nakane, Abhay Kotecha, Andrija Sente, Greg McMullan, Simonas Masiulis, Patricia M. G. E. Brown, Ioana T. Grigoras, Lina Malinauskaite, Tomas Malinauskas, Jonas Miehling, Tomasz Uchański, Lingbo Yu, Dimple Karia, Evgeniya V. Pechnikova, Erwin de Jong, Jeroen Keizer, Maarten Bischoff, Jamie McCormack, Peter Tiemeijer, Steven W. Hardwick, Dimitri Y. Chirgadze, Garib Murshudov, A. Radu Aricescu, and Sjors H. W. Scheres. Single-particle cryo-EM at atomic resolution. *Nature*, 587(7832): 152–156, 2020.

[59] Sjors H.W. Scheres. A Bayesian View on Cryo-EM Structure Determination. *Journal of Molecular Biology*, 415(2):406–418, 2012.

[60] Ali Punjani, John L. Rubinstein, David J. Fleet, and Marcus A. Brubaker. cryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods*, 14(3):290–296, 2017.

[61] Jeong Min Chung, Clarissa L. Durie, and Jinseok Lee. Artificial Intelligence in Cryo-Electron Microscopy. *Life*, 12(8):1267, 2022.

[62] Tristan Bepler, Andrew Morin, Micah Rapp, Julia Brasch, Lawrence Shapiro, Alex J. Noble, and Bonnie Berger. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nature Methods*, 16(11):1153–1160, 2019.

[63] Thorsten Wagner, Felipe Merino, Markus Stabrin, Toshio Moriya, Claudia Antoni, Amir Apelbaum, Philine Hagel, Oleg Sitsel, Tobias Raisch, Daniel Prumbaum, Dennis Quentin, Daniel Roderer, Sebastian Tacke, Birte Siebolds, Evelyn Schubert, Tanvir R. Shaikh, Pascal Lill, Christos Gatsogiannis, and Stefan Raunser. SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Communications Biology*, 2(1):218, 2019.

[64] Ellen D. Zhong, Tristan Bepler, Bonnie Berger, and Joseph H. Davis. CryoDRGN: Reconstruction of heterogeneous cryo-EM structures using neural networks. *Nature Methods*, 18(2):176–185, 2021.

[65] Jay W. Ponder and David A. Case. Force Fields for Protein Simulations. In *Advances in Protein Chemistry*, volume 66, pages 27–85. Elsevier, 2003. ISBN 978-0-12-034266-2.

[66] D. A. Case, H. M. Aktulga, K. Belfon, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. Cheatham, G. O Cisneros, V. W. D. Cruzerio, T. A. Darden, R. E. Duke, G. Giambasu, M. K. Gilson, H. Gohlke, A. W. Goetz, R. Harris, S. Izadi, S. A. Izmailov, C. Jin, K. Kasavajhala, Kaymak, E. King, A. Kovalenko, T. Kurtzman, T. S. Lee, S. LeGrand, S. Li, C. Lin, T. Luchko, R. Luo, M. Machado, V. Man, M. Manathunga, K. M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, K. A. O'Hearn, A. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, D. R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. L. Simmerling, N. R. Skrynnikov, J. Smith, J. Swails, R. C. Walker, J. Wang, H. Wei, R. M. Wolf, X. Wu, Y. Xue, D. M. York, S. Zhao, and P. A. Kollman. Amber 2021. University of California.

[67] Jing Huang and Alexander D. MacKerell. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *Journal of Computational Chemistry*, 34(25):2135–2145, 2013.

[68] Maria M. Reif, Philippe H. Hünenberger, and Chris Oostenbrink. New Interaction Parameters for Charged Amino Acid Side Chains in the GROMOS Force Field. *Journal of Chemical Theory and Computation*, 8(10):3705–3723, 2012.

[69] William L. Jorgensen, David S. Maxwell, and Julian Tirado-Rives. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *Journal of the American Chemical Society*, 118(45):11225–11236, 1996.

[70] Michael J. Robertson, Julian Tirado-Rives, and William L. Jorgensen. Improved Peptide and Protein Torsional Energetics with the OPLS-AA Force Field. *Journal of Chemical Theory and Computation*, 11(7):3499–3509, 2015.

[71] K. Vanommeslaeghe and A.D. MacKerell. CHARMM additive and polarizable force fields for biophysics and computer-aided drug design. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1850(5):861–871, 2015.

[72] Kresten Lindorff-Larsen, Paul Maragakis, Stefano Piana, Michael P. Eastwood, Ron O. Dror, and David E. Shaw. Systematic Validation of Protein Force Fields against Experimental Data. *PLoS ONE*, 7(2):e32131, 2012.

[73] Antonija Kuzmanic, Ruth B. Pritchard, D. Flemming Hansen, and Francesco L. Gervasio. Importance of the Force Field Choice in Capturing Functionally Relevant Dynamics in the von Willebrand Factor. *The Journal of Physical Chemistry Letters*, 10 (8):1928–1934, 2019.

[74] Mueed Ur Rahman, Ashfaq Ur Rehman, Hao Liu, and Hai-Feng Chen. Comparison and Evaluation of Force Fields for Intrinsically Disordered Proteins. *Journal of Chemical Information and Modeling*, 60(10):4912–4923, 2020.

[75] Vishal Minhas, Tiedong Sun, Alexander Mirzoev, Nikolay Korolev, Alexander P. Lyubartsev, and Lars Nordenskiöld. Modeling DNA Flexibility: Comparison of Force Fields from Atomistic to Multiscale Levels. *The Journal of Physical Chemistry B*, 124 (1):38–49, 2020.

[76] Anita Plazinska and Wojciech Plazinski. Comparison of Carbohydrate Force Fields in Molecular Dynamics Simulations of Protein–Carbohydrate Complexes. *Journal of Chemical Theory and Computation*, 17(4):2575–2585, 2021.

[77] H. Fehske, R. Schneider, A. Weiße, R. Beig, W. Beiglböck, W. Domcke, B.-G. Englert, U. Frisch, P. Hänggi, G. Hasinger, K. Hepp, W. Hillebrandt, D. Imboden, R. L. Jaffe, R. Lipowsky, H. v. Löhneysen, I. Ojima, D. Sornette, S. Theisen, W. Weise, J. Wess, and J. Zittartz, editors. *Computational Many-Particle Physics*, volume 739 of *Lecture Notes in Physics*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008. ISBN 978-3-540-74685-0 978-3-540-74686-7.

[78] Alessandra Nurisso, Antoine Daina, and Ross C. Walker. A Practical Introduction to Molecular Dynamics Simulations: Applications to Homology Modeling. In Andrew J. W. Orry and Ruben Abagyan, editors, *Homology Modeling*, volume 857, pages 137–173. Humana Press, Totowa, NJ, 2011. ISBN 978-1-61779-587-9 978-1-61779-588-6.

[79] Mark James Abraham, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C. Smith, Berk Hess, and Erik Lindahl. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1–2:19–25, 2015.

[80] Efrem Braun, Justin Gilmer, Heather B. Mayes, David L. Mobley, Jacob I. Monroe, Samarjeet Prasad, and Daniel M. Zuckerman. Best Practices for Foundations in Molecular Simulations. *Living Journal of Computational Molecular Science*, 1(1):5957, 2018.

[81] Scott A. Hollingsworth and Ron O. Dror. Molecular Dynamics Simulation for All. *Neuron*, 99(6):1129–1143, 2018.

[82] Siewert J. Marrink, Luca Monticelli, Manuel N. Melo, Riccardo Alessandri, D. Peter Tieleman, and Paulo C. T. Souza. Two decades of Martini: Better beads, broader scope. *WIREs Computational Molecular Science*, 2022.

[83] Pavel G. Khalatur. Molecular Dynamics Simulations in Polymer Science. In *Polymer Science: A Comprehensive Reference*, pages 417–460. Elsevier, 2012. ISBN 978-0-08-087862-1.

[84] Marc W. van der Kamp and Adrian J. Mulholland. Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods in Computational Enzymology. *Biochemistry*, 52(16):2708–2728, 2013.

[85] Tsjerk A. Wassenaar, Helgi I. Ingólfsson, Marten Prieß, Siewert J. Marrink, and Lars V. Schäfer. Mixing MARTINI: Electrostatic Coupling in Hybrid Atomistic–Coarse-Grained Biomolecular Simulations. *The Journal of Physical Chemistry B*, 117(13):3516–3530, 2013.

[86] Siewert J. Marrink and Alan E. Mark. Molecular Dynamics Simulation of the Formation, Structure, and Dynamics of Small Phospholipid Vesicles. *Journal of the American Chemical Society*, 125(49):15233–15242, 2003.

[87] Siewert J. Marrink and Alan E. Mark. The Mechanism of Vesicle Fusion as Revealed by Molecular Dynamics Simulations. *Journal of the American Chemical Society*, 125 (37):11144–11145, 2003.

[88] Siewert J. Marrink, H. Jelger Risselada, Serge Yefimov, D. Peter Tieleman, and Alex H. de Vries. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *The Journal of Physical Chemistry B*, 111(27):7812–7824, 2007.

[89] H. J. Risselada and S. J. Marrink. The molecular face of lipid rafts in model membranes. *Proceedings of the National Academy of Sciences*, 105(45):17367–17372, 2008.

[90] Luca Monticelli, Senthil K. Kandasamy, Xavier Periole, Ronald G. Larson, D. Peter Tieleman, and Siewert-Jan Marrink. The MARTINI Coarse-Grained Force Field: Extension to Proteins. *Journal of Chemical Theory and Computation*, 4(5):819–834, 2008.

[91] Djurre H. de Jong, Gurpreet Singh, W. F. Drew Bennett, Clement Arnarez, Tsjerk A. Wassenaar, Lars V. Schäfer, Xavier Periole, D. Peter Tieleman, and Siewert J. Marrink. Improved Parameters for the Martini Coarse-Grained Protein Force Field. *Journal of Chemical Theory and Computation*, 9(1):687–697, 2013.

[92] Xavier Periole, Marco Cavalli, Siewert-Jan Marrink, and Marco A. Ceruso. Combining an Elastic Network With a Coarse-Grained Molecular Force Field: Structure, Dynamics, and Intermolecular Recognition. *Journal of Chemical Theory and Computation*, 5 (9):2531–2543, 2009.

[93] Cesar A. López, Andrzej J. Rzepiela, Alex H. de Vries, Lubbert Dijkhuizen, Philippe H. Hünenberger, and Siewert J. Marrink. Martini Coarse-Grained Force Field: Extension to Carbohydrates. *Journal of Chemical Theory and Computation*, 5(12):3195–3210, 2009.

[94] Jaakko J. Uusitalo, Helgi I. Ingólfsson, Parisa Akhshi, D. Peter Tieleman, and Siewert J. Marrink. Martini Coarse-Grained Force Field: Extension to DNA. *Journal of Chemical Theory and Computation*, 11(8):3932–3945, 2015.

[95] Jaakko J. Uusitalo, Helgi I. Ingólfsson, Siewert J. Marrink, and Ignacio Faustino. Martini Coarse-Grained Force Field: Extension to RNA. *Biophysical Journal*, 113(2):246–256, 2017.

[96] Semen O. Yesylevskyy, Lars V. Schäfer, Durba Sengupta, and Siewert J. Marrink. Polarizable Water Model for the Coarse-Grained MARTINI Force Field. *PLoS Computational Biology*, 6(6):e1000810, 2010.

[97] Helgi I. Ingólfsson, Manuel N. Melo, Floris J. van Eerden, Clément Arnarez, Cesar A. Lopez, Tsjerk A. Wassenaar, Xavier Periole, Alex H. de Vries, D. Peter Tieleman, and Siewert J. Marrink. Lipid Organization of the Plasma Membrane. *Journal of the American Chemical Society*, 136(41):14554–14559, 2014.

[98] Simon R. Bushell, Ashley C. W. Pike, Maria E. Falzone, Nils J. G. Rorsman, Chau M. Ta, Robin A. Corey, Thomas D. Newport, John C. Christianson, Lara F. Scofano, Chitra A. Shintre, Annamaria Tessitore, Amy Chu, Qinrui Wang, Leela Shrestha, Shubhashish M. M. Mukhopadhyay, James D. Love, Nicola A. Burgess-Brown, Rebecca Sitsapesan, Phillip J. Stansfeld, Juha T. Huiskonen, Paolo Tammaro, Alessio Accardi, and Elisabeth P. Carpenter. The structural basis of lipid scrambling and inactivation in the endoplasmic reticulum scramblase TMEM16K. *Nature Communications*, 10(1): 3956, 2019.

[99] Eric L. Dane, Alexis Belessiotis-Richards, Coralie Backlund, Jianing Wang, Kousuke Hidaka, Lauren E. Milling, Sachin Bhagchandani, Mariane B. Melo, Shengwei Wu, Na Li, Nathan Donahue, Kaiyuan Ni, Leyuan Ma, Masanori Okaniwa, Molly M. Stevens, Alfredo Alexander-Katz, and Darrell J. Irvine. STING agonist delivery by tumour-penetrating PEG-lipid nanodiscs primes robust anticancer immunity. *Nature Materials*, 21(6):710–720, 2022.

[100] Kristyna Pluhackova, Florian M. Wilhelm, and Daniel J. Müller. Lipids and Phosphorylation Conjointly Modulate Complex Formation of B2-Adrenergic Receptor and β-arrestin2. *Frontiers in Cell and Developmental Biology*, 9:807913, 2021.

[101] Shyamal Mosalaganti, Agnieszka Obarska-Kosinska, Marc Siggel, Reiya Taniguchi, Beata Turoňová, Christian E. Zimmerli, Katarzyna Buczak, Florian H. Schmidt, Erica Margiotta, Marie-Therese Mackmull, Wim J. H. Hagen, Gerhard Hummer, Jan Kosinski, and Martin Beck. AI-based structure prediction empowers integrative structural analysis of human nuclear pores. *Science*, 376(6598):eabm9506, 2022.

[102] Paulo C. T. Souza, Riccardo Alessandri, Jonathan Barnoud, Sebastian Thallmair, Ignacio Faustino, Fabian Grünewald, Ilias Patmanidis, Haleh Abdizadeh, Bart M. H. Bruininks, Tsjerk A. Wassenaar, Peter C. Kroon, Josef Melcr, Vincent Nieto, Valentina Corradi, Hanif M. Khan, Jan Domański, Matti Javanainen, Hector Martinez-Seara, Nathalie Reuter, Robert B. Best, Ilpo Vattulainen, Luca Monticelli, Xavier Periole, D. Peter Tieleman, Alex H. de Vries, and Siewert J. Marrink. Martini 3: A general purpose force field for coarse-grained molecular dynamics. *Nature Methods*, 18(4): 382–388, 2021.

[103] Adolfo B. Poma, Marek Cieplak, and Panagiotis E. Theodorakis. Combining the MARTINI and Structure-Based Coarse-Grained Approaches for the Molecular Dynamics Studies of Conformational Transitions in Proteins. *Journal of Chemical Theory and Computation*, 13(3):1366–1374, 2017.

[104] Md. Iqbal Mahmood, Adolfo B. Poma, and Kei-ichi Okazaki. Optimizing Gō-MARTINI Coarse-Grained Model for F-BAR Protein on Lipid Membrane. *Frontiers in Molecular Biosciences*, 8:619381, 2021.

[105] Christos Lamprakis, Ioannis Andreadelis, John Manchester, Camilo Velez-Vega, José S. Duca, and Zoe Cournia. Evaluating the Efficiency of the Martini Force Field to Study Protein Dimerization in Aqueous and Membrane Environments. *Journal of Chemical Theory and Computation*, 17(5):3088–3102, 2021.

[106] Sriraksha Srinivasan, Valeria Zoni, and Stefano Vanni. Estimating the accuracy of the MARTINI model towards the investigation of peripheral protein–membrane interactions. *Faraday Discussions*, 232:131–148, 2021.

[107] Paulo C. T. Souza, Sebastian Thallmair, Paolo Conflitti, Carlos Ramírez-Palacios, Riccardo Alessandri, Stefano Raniolo, Vittorio Limongelli, and Siewert J. Marrink. Protein–ligand binding with the coarse-grained Martini model. *Nature Communications*, 11(1):3714, 2020.

[108] Fabian Grünewald, Mats H. Punt, Elizabeth E. Jefferys, Petteri A. Vainikka, Melanie König, Valtteri Virtanen, Travis A. Meyer, Weria Pezeshkian, Adam J. Gormley, Maarit Karonen, Mark S. P. Sansom, Paulo C. T. Souza, and Siewert J. Marrink. Martini 3 Coarse-Grained Force Field for Carbohydrates. *Journal of Chemical Theory and Computation*, 18(12):7555–7569, 2022.

[109] Valery Lutsyk, Pawel Wolski, and Wojciech Plazinski. Extending the Martini 3 Coarse-Grained Force Field to Carbohydrates. *Journal of Chemical Theory and Computation*, 18(8):5089–5107, 2022.

[110] Kaila M. Bennett, Suzan H. M. Rooijakkers, and Ronald D. Gorham. Let's Tie the Knot: Marriage of Complement and Adaptive Immunity in Pathogen Evasion, for Better or Worse. *Frontiers in Microbiology*, 8:Article 89, 2017.

[111] Christoph A. Diebolder, Frank J. Beurskens, Rob N. de Jong, Roman I. Koning, Kristin Strumane, Margaret A. Lindorfer, Marleen Voorhorst, Deniz Ugurlar, Sara Rosati, Albert J. R. Heck, Jan G. J. van de Winkel, Ian A. Wilson, Abraham J. Koster, Ronald P. Taylor, Erica Ollmann Saphire, Dennis R. Burton, Janine Schuurman, Piet Gros, and Paul W. H. I. Parren. Complement Is Activated by IgG Hexamers Assembled at the Cell Surface. *Science*, 343(6176):1260–1263, 2014.

[112] Deniz Ugurlar, Stuart C. Howes, Bart-Jan de Kreuk, Roman I. Koning, Rob N. de Jong, Frank J. Beurskens, Janine Schuurman, Abraham J. Koster, Thomas H. Sharp, Paul W. H. I. Parren, and Piet Gros. Structures of C1-IgG1 provide insights into how danger pattern recognition activates complement. *Science*, 359(6377):794–797, 2018.

[113] Gérard J. Arlaud, Christine Gaboriaud, Nicole M. Thielens, Monika Budayova-Spano, Véronique Rossi, and Juan Carlos Fontecilla-Camps. Structural biology of the C1 complex of complement unveils the mechanisms of its activation and proteolytic activity. *Molecular Immunology*, 39:383–394, 2002.

[114] Simon A. Mortensen, Bjoern Sander, Rasmus K. Jensen, Jan Skov Pedersen, Monika M. Golas, Jens C. Jensenius, Annette G. Hansen, Steffen Thiel, and Gregers R. Andersen. Structure and activation of C1, the complex initiating the classical pathway of the complement cascade. *Proceedings of the National Academy of Sciences*, 114(5):986–991, 2017.

[115] Anne Nicholson-Weller and Lloyd B. Klickstein. C1q-binding proteins and C1q receptors. *Current Opinion in Immunology*, 11(1):42–46, 1999.

[116] Ionita Ghiran, Shiv-Raj Tyagi, Lloyd B. Klickstein, and Anne Nicholson-Weller. Expression and Function of C1q Receptors and C1q Binding Proteins at the Cell Surface. *Immunobiology*, 205(4-5):407–420, 2002.

[117] Uday Kishore, Christine Gaboriaud, Patrick Waters, Annette K. Shrive, Trevor J. Greenhough, Kenneth B.M. Reid, Robert B. Sim, and Gerard J. Arlaud. C1q and tumor necrosis factor superfamily: Modularity and versatility. *Trends in Immunology*, 25(10):551–561, 2004.

[118] Amanda J. Bradley, Donald E. Brooks, Raymond Norris-Jones, and Dana V. Devine. C1q Binding to liposomes is surface charge dependent and is inhibited by peptides consisting of residues 14–26 of the human C1qA chain in a sequence independent manner. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1418(1):19–30, 1999.

[119] Lee Aun Tan, Andrew C. Yang, Uday Kishore, and Robert B. Sim. Interactions of complement proteins C1q and factor H with lipid A and Escherichia coli: Further evidence that factor H regulates the classical complement pathway. *Protein & Cell*, 2(4): 320–332, 2011.

[120] Helena Païdassi, Pascale Tacnet-Delorme, Virginie Garlatti, Claudine Darnault, Berhane Ghebrehiwet, Christine Gaboriaud, Gérard J. Arlaud, and Philippe Frachet. C1q Binds Phosphatidylserine and Likely Acts as a Multiligand-Bridging Molecule in Apoptotic Cell Recognition. *The Journal of Immunology*, 180(4):2329–2338, 2008.

[121] Ken Ikeda, Takayoshi Sannoh, Nobuko Kawasaki, Toshisuke Kawasaki, and Ikuo Yamashina. Serum Lectin with Known Structure Activates Complement through the Classical Pathway. *Journal of Biological Chemistry*, 262(16):7454–7454, 1987.

[122] Michele M. Estabrook, Dominic L. Jack, Nigel J. Klein, and Gary A. Jarvis. Mannose-Binding Lectin Binds to Two Major Outer Membrane Proteins, Opacity Protein and Porin, of *Neisseria meningitidis*. *The Journal of Immunology*, 172(6):3784–3792, 2004.

[123] Carol Anne Ogden, Aimee deCathelineau, Peter R. Hoffmann, Donna Bratton, Berhane Ghebrehiwet, Valerie A. Fadok, and Peter M. Henson. C1q and Mannose Binding Lectin Engagement of Cell Surface Calreticulin and Cd91 Initiates Macropinocytosis and Uptake of Apoptotic Cells. *Journal of Experimental Medicine*, 194(6):781–796, 2001.

[124] David C Kilpatrick. Phospholipid-binding activity of human mannan-binding lectin. *Immunology Letters*, 61(2-3):191–195, 1998.

[125] Misao Matsushita and Teizo Fujita. Activation of the Classical Complement Pathway by Mannose-binding Protein in Association with a Novel C1s-like Serine Protease. *Journal of Experimental Medicine*, 176:1497–1502, 1992.

[126] Steffen Thiel, Thomas Vorup-Jensen, Cordula M. Stover, Wilhelm Schwaeble, Steen B. Laursen, Knud Poulsen, Seren Hansen, and Uffe Holmskov. A second serine protease associated with mannan-binding lectin that activates complement. *Nature*, 386(6624):506–510, 1997.

[127] S. K. Law, N. A. Lichtenberg, F. H. Holcombe, and R. P. Levine. Interaction between the labile binding sites of the fourth (C4) and fifth (C5) human complement proteins and erythrocyte cell membranes. *The Journal of Immunology*, 125(2):634–639, 1980.

[128] Hans J. Müller-Eberhard, Margaret J. Polley, and Mary Ann Calcott. Formation and Functional Significance of a Molecular Complex Derived from the Second and Fourth Component of Human Complement. *The Journal of Experimental Medicine*, 125(2): 359–380, 1967.

[129] Armin Sepp, Alister W. Dodds, Michael J. Anderson, R. Duncan Campbell, Antony C. Willis, and S.K. Alex Law. Covalent binding properties of the human complement protein C4 and hydrolysis rate of the internal thioester upon activation. *Protein Science*, 2(5):706–716, 1993.

[130] Russell Wallis, Alister W. Dodds, Daniel A. Mitchell, Robert B. Sim, Kenneth B.M. Reid, and Wilhelm J. Schwaeble. Molecular Interactions between MASP-2, C4, and C2 and Their Activation Fragments Leading to Complement Activation via the Lectin Pathway. *Journal of Biological Chemistry*, 282(11):7844–7851, 2007.

[131] Nenoo Rawal and Michael K. Pangburn. C5 Convertase of the Alternative Pathway of Complement. *Journal of Biological Chemistry*, 273(27):16828–16835, 1998.

[132] Evelien T. M. Berends, Ronald D. Gorham, Maartje Ruyken, Jasper A. Soppe, Hatice Orhan, Piet C. Aerts, Carla J. C. de Haas, Piet Gros, and Suzan H. M. Rooijakkers. Molecular insights into the surface-specific arrangement of complement C5 convertase enzymes. *BMC Biology*, 13(1):93, 2015.

[133] Seline A. Zwarthoff, Evelien T. M. Berends, Sanne Mol, Maartje Ruyken, Piet C. Aerts, Mihály Józsi, Carla J. C. de Haas, Suzan H. M. Rooijakkers, and Ronald D. Gorham. Functional Characterization of Alternative and Classical Pathway C3/C5 Convertase Activity and Inhibition Using Purified Models. *Frontiers in Immunology*, 9:1691, 2018.

[134] J.O. Sunyer, H. Boshra, and J. Li. Evolution of anaphylatoxins, their diversity and novel roles in innate immunity: Insights from the study of fish complement. *Veterinary Immunology and Immunopathology*, 108(1-2):77–89, 2005.

[135] Michael A. Hadders, Doryen Bubeck, Pietro Roversi, Svetlana Hakobyan, Federico Forneris, B. Paul Morgan, Michael K. Pangburn, Oscar Llorca, Susan M. Lea, and Piet Gros. Assembly and Regulation of the Membrane Attack Complex Based on Structures of C5b6 and sC5b9. *Cell Reports*, 1(3):200–207, 2012.

[136] Richard G. DiScipio, Deb N. Chakravarti, Hans J. Muller-Eberhard, and Georg H. Fey. The structure of human complement component C7 and the C5b-7 complex. *Journal of Biological Chemistry*, 263(1):549–560, 1988.

[137] Leslie L. Lovelace, Christopher L. Cooper, James M. Sodetz, and Lukasz Lebioda. Structure of Human C8 Protein Provides Mechanistic Insight into Membrane Pore Formation by Complement. *Journal of Biological Chemistry*, 286(20):17585–17592, 2011.

[138] Eric W. Steckel, Barbara E. Welbaum, and James M. Sodetz. Evidence of direct insertion of terminal complement proteins into cell membrane bilayers during cytolysis. Labeling by a photosensitive membrane probe reveals a major role for the eighth and ninth components. *Journal of Biological Chemistry*, 258(7):4318–4324, 1983.

[139] Marina Serna, Joanna L. Giles, B. Paul Morgan, and Doryen Bubeck. Structural basis of complement membrane attack complex formation. *Nature Communications*, 7(1): 10587, 2016.

[140] Eckhard R. Podack, Jürg Tschoop, and Hans J. Muller-Eberhard. Molecular organization of C9 within the membrane attack complex of complement. Induction of circular C9 polymerization by the C5b-8 assembly. *Journal of Experimental Medicine*, 156(1): 268–282, 1982.

[141] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, 2021.

[142] Mihaly Varadi, Stephen Anyango, Mandar Deshpande, Sreenath Nair, Cindy Natassia, Galabina Yordanova, David Yuan, Oana Stroe, Gemma Wood, Agata Laydon, Augustin Žídek, Tim Green, Kathryn Tunyasuvunakool, Stig Petersen, John Jumper, Ellen Clancy, Richard Green, Ankur Vora, Mira Lutfi, Michael Figurnov, Andrew Cowie, Nicole Hobbs, Pushmeet Kohli, Gerard Kleywegt, Ewan Birney, Demis Hassabis, and Sameer Velankar. AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research*, 50(D1):D439–D444, 2022.

[143] Emma C. Couves, Scott Gardner, Tomas B. Voisin, Jasmine K. Bickel, Phillip J. Stansfeld, Edward W. Tate, and Doryen Bubeck. Structural basis for membrane attack complex inhibition by CD59. *Nature Communications*, [in press]:preprint doi: 10.1101/2022.09.05.506672, 2023.

[144] Daniel Ricklin, Edimara S. Reis, and John D. Lambris. Complement in disease: A defence system turning offensive. *Nature Reviews Nephrology*, 12(7):383–401, 2016.

[145] Pauline M. Rudd, B. Paul Morgan, Mark R. Wormald, David J. Harvey, Carmen W. van den Berg, Simon J. Davis, Michael A.J. Ferguson, and Raymond A. Dwek. The Glycosylation of the Complement Regulatory Protein, Human Erythrocyte CD59. *Journal of Biological Chemistry*, 272(11):7229–7244, 1997.

[146] C. Mark Fletcher, Richard A. Harrison, Peter J. Lachmann, and David Neuhausl. Structure of a soluble, glycosylated form of the human complement regulatory protein CD59. *Structure*, 2(3):185–199, 1994.

[147] Dara H. Lockert, Kenneth M. Kaufman, Chi-Pei Chang, Thomas Hüsler, James M. Sodetz, and Peter J. Sims. Identity of the Segment of Human Complement C8 Recognized by Complement Regulatory Protein CD59. *Journal of Biological Chemistry*, 270(34):19723–19728, 1995.

[148] Yuxiang Huang, Fei Qiao, Ruben Abagyan, Starr Hazard, and Stephen Tomlinson. Defining the CD59-C9 Binding Interaction. *Journal of Biological Chemistry*, 281(37): 27398–27404, 2006.

[149] Richard G. DiScipio, Deb N. Chakravarti, Hans J. Muller-Eberhard, and Georg H. Fey. The structure of human complement component C7 and the C5b-7 complex. *Journal of Biological Chemistry*, 263(1):549–560, 1988.

[150] Imre Farkas, Lajos Baranyi, Yasushige Ishikawa, Noriko Okada, Csaba Bohata, Denes Budai, Atsuo Fukuda, Masaki Imai, and Hidechika Okada. CD59 blocks not only the insertion of C9 into MAC but inhibits ion channel formation by homologous C5b-8 as well as C5b-9. *The Journal of Physiology*, 539(2):537–545, 2002.

[151] Rajat Desikan, Amit Behera, Prabal K. Maiti, and K. Ganapathy Ayappa. Using multiscale molecular dynamics simulations to obtain insights into pore forming toxin mechanisms. In *Methods in Enzymology*, volume 649, pages 461–502. Elsevier, 2021. ISBN 978-0-12-823858-5.

[152] Pradeep Sathyanarayana, Satyaghosh Maurya, Amit Behera, Monisha Ravichandran, Sandhya S. Visweswariah, K. Ganapathy Ayappa, and Rahul Roy. Cholesterol promotes Cytolysin A activity by stabilizing the intermediates during pore formation. *Proceedings of the National Academy of Sciences*, 115(31), 2018.

[153] Martin Vögele, Ramachandra M. Bhaskara, Estefania Mulvihill, Katharina van Pee, Özkan Yildiz, Werner Kühlbrandt, Daniel J. Müller, and Gerhard Hummer. Membrane perforation by the pore-forming toxin pneumolysin. *Proceedings of the National Academy of Sciences*, 116(27):13352–13357, 2019.

[154] Ramesh Cheerla and K. Ganapathy Ayappa. Molecular Dynamics Study of Lipid and Cholesterol Reorganization Due to Membrane Binding and Pore Formation by Listeriolysin O. *The Journal of Membrane Biology*, 253(6):535–550, 2020.

[155] Stefania A. Mari, Kristyna Pluhackova, Joka Pipercevic, Matthew Leipner, Sebastian Hiller, Andreas Engel, and Daniel J. Müller. Gasdermin-A3 pore formation propagates along variable pathways. *Nature Communications*, 13(1):2609, 2022.

[156] Williams E. Miranda, Jiqing Guo, Haydee Mesa-Galloso, Valentina Corradi, James P. Lees-Miller, D. Peter Tieleman, Henry J. Duff, and Sergei Yu. Noskov. Lipid regulation of hERG1 channel function. *Nature Communications*, 12(1):1409, 2021.

[157] Veronika Thallmair, Lea Schultz, Wencai Zhao, Siewert J. Marrink, Dominik Oliver, and Sebastian Thallmair. Two cooperative binding sites sensitize $PI(4,5)P_2$ recognition by the tubby domain. *Science Advances*, 8(36):eabp9471, 2022.

[158] Margot G. Paulick and Carolyn R. Bertozzi. The Glycosylphosphatidylinositol Anchor: A Complex Membrane-Anchoring Structure for Proteins. *Biochemistry*, 47(27):6991–7000, 2008.

[159] Pallavi Banerjee, Marko Wehle, Reinhard Lipowsky, and Mark Santer. A molecular dynamics model for glycosylphosphatidyl-inositol anchors: “flop down” or “lollipop”? *Physical Chemistry Chemical Physics*, 20(46):29314–29324, 2018.

[160] Maxim M. Zaigraev, Ekaterina N. Lyukmanova, Alexander S. Paramonov, Zakhar O. Shenkarev, and Anton O. Chugunov. Orientational Preferences of GPI-Anchored Ly6/uPAR Proteins. *International Journal of Molecular Sciences*, 24(1):11, 2022.

[161] Pallavi Banerjee, Reinhard Lipowsky, and Mark Santer. Coarse-Grained Molecular Model for the Glycosylphosphatidylinositol Anchor with and without Protein. *Journal of Chemical Theory and Computation*, 16(6):3889–3903, 2020.

[162] Karl N. Kirschner, Austin B. Yongye, Sarah M. Tschampel, Jorge González-Outeiriño, Charlisa R. Daniels, B. Lachele Foley, and Robert J. Woods. GLYCAM06: A generalizable biomolecular force field. Carbohydrates: GLYCAM06. *Journal of Computational Chemistry*, 29(4):622–655, 2008.

[163] Callum J. Dickson, Benjamin D. Madej, Åge A. Skjevik, Robin M. Betz, Knut Teigen, Ian R. Gould, and Ross C. Walker. Lipid14: The Amber Lipid Force Field. *Journal of Chemical Theory and Computation*, 10(2):865–879, 2014.

[164] Sunhwan Jo, Taehoon Kim, Vidyashankara G. Iyer, and Wonpil Im. CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry*, 29(11):1859–1865, 2008.

[165] Jumin Lee, Dhilon S. Patel, Jonas Ståhle, Sang-Jun Park, Nathan R. Kern, Seonghoon Kim, Joonseong Lee, Xi Cheng, Miguel A. Valvano, Otto Holst, Yuriy A. Knirel, Yifei Qi, Sunhwan Jo, Jeffery B. Klauda, Göran Widmalm, and Wonpil Im. CHARMM-GUI *Membrane Builder* for Complex Biological Membrane Simulations with Glycolipids and Lipoglycans. *Journal of Chemical Theory and Computation*, 15(1):775–786, 2019.

[166] William L. Jorgensen, Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, and Michael L. Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926–935, 1983.

[167] Jing Huang, Sarah Rauscher, Grzegorz Nawrocki, Ting Ran, Michael Feig, Bert L de Groot, Helmut Grubmüller, and Alexander D MacKerell. CHARMM36m: An improved force field for folded and intrinsically disordered proteins. *Nature Methods*, 14(1):71–73, 2017.

[168] 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. *The Journal of Chemical Physics*, 81(8):3684–3690, 1984.

[169] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD: Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.

[170] Eric F. Pettersen, Thomas D. Goddard, Conrad C. Huang, Elaine C. Meng, Gregory S. Couch, Tristan I. Croll, John H. Morris, and Thomas E. Ferrin. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science*, 30(1):70–82, 2021.

[171] Douglas L. Theobald. Rapid calculation of RMSDs using a quaternion-based characteristic polynomial. *Acta Crystallographica Section A Foundations of Crystallography*, 61(4):478–480, 2005.

[172] Pu Liu, Dimitris K. Agrafiotis, and Douglas L. Theobald. Fast determination of the optimal rotational matrix for macromolecular superpositions. *Journal of Computational Chemistry*, 31(7):1561–1563, 2010.

[173] N. Michaud-Agrawal, E. J. Denning, T. B. Woolf, and O. Beckstein. MDAnalysis: A Toolkit for the Analysis of MolecularDynamics Simulations. *Journal of Computational Chemistry*, 32:2319–2327, 2011.

[174] Richard Gowers, Max Linke, Jonathan Barnoud, Tyler Reddy, Manuel Melo, Sean Seyler, Jan Domański, David Dotson, Sébastien Buchoux, Ian Kenney, and Oliver Beckstein. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. In *Python in Science Conference*, pages 98–105, Austin, Texas, 2016.

[175] Eric F. Pettersen, Thomas D. Goddard, Conrad C. Huang, Gregory S. Couch, Daniel M. Greenblatt, Elaine C. Meng, and Thomas E. Ferrin. UCSF Chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13):1605–1612, 2004.

[176] Michael L. Waskom. Seaborn: Statistical data visualization. *Journal of Open Source Software*, 6(60):3021, 2021.

[177] John D. Hunter. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, 9(3):90–95, 2007.

[178] Alan W. Sousa da Silva and Wim F. Vranken. ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Research Notes*, 5(1):367, 2012.

[179] N. Goga, A. J. Rzepiela, A. H. de Vries, S. J. Marrink, and H. J. C. Berendsen. Efficient Algorithms for Langevin and DPD Dynamics. *Journal of Chemical Theory and Computation*, 8(10):3637–3649, 2012.

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

[181] Sara L. Lawrence, Michael A. Gorman, Susanne C. Feil, Terrence D. Mulhern, Michael J. Kuiper, Adam J. Ratner, Rodney K. Tweten, Craig J. Morton, and Michael W. Parker. Structural Basis for Receptor Recognition by the Human CD59-Responsive Cholesterol-Dependent Cytolysins. *Structure*, 24(9):1488–1498, 2016.

[182] Steven Johnson, Nicholas J. Brooks, Richard A.G. Smith, Susan M. Lea, and Doryen Bubeck. Structural Basis for Recognition of the Pore-Forming Toxin Intermedilysin by Human Complement Receptor CD59. *Cell Reports*, 3(5):1369–1377, 2013.

[183] Timothy Nugent and David T. Jones. Membrane protein orientation and refinement using a knowledge-based statistical potential. *BMC Bioinformatics*, 14(1):276, 2013.

[184] Mikhail A. Lomize, Irina D. Pogozheva, Hyeon Joo, Henry I. Mosberg, and Andrei L. Lomize. OPM database and PPM web server: Resources for positioning of proteins in membranes. *Nucleic Acids Research*, 40(D1):D370–D376, 2012.

[185] Wanling Song, Robin A. Corey, T. Bertie Ansell, C. Keith Cassidy, Michael R. Horrell, Anna L. Duncan, Phillip J. Stansfeld, and Mark S. P. Sansom. PyLipID: A Python Package for Analysis of Protein–Lipid Interactions from Molecular Dynamics Simulations. *Journal of Chemical Theory and Computation*, 18(2):1188–1201, 2022.

[186] Sébastien Buchoux. FATSLiM: A fast and robust software to analyze MD simulations of membranes. *Bioinformatics*, 33(1):133–134, 2017.

[187] Kenichi G. N. Suzuki, Rinshi S. Kasai, Koichiro M. Hirosawa, Yuri L. Nemoto, Munenori Ishibashi, Yoshihiro Miwa, Takahiro K. Fujiwara, and Akihiro Kusumi. Transient GPI-anchored protein homodimers are units for raft organization and function. *Nature Chemical Biology*, 8(9):774–783, 2012.

[188] Courtney M. Boyd, Edward S. Parsons, Richard A. G. Smith, John M. Seddon, Oscar Ces, and Doryen Bubeck. Disentangling the roles of cholesterol and CD59 in intermedilysin pore formation. *Scientific Reports*, 6(1):38446, 2016.

[189] Timothy R. Hughes, Kirsty S. Ross, Graeme J.M. Cowan, Baalasubramanian Sivasankar, Claire L. Harris, Timothy J. Mitchell, and B. Paul Morgan. Identification of the high affinity binding site in the Streptococcus intermedius toxin intermedilysin for its membrane receptor, the human complement regulator CD59. *Molecular Immunology*, 46(7):1561–1567, 2009.

[190] Stephanie E. Wickham, Eileen M. Hotze, Allison J. Farrand, Galina Polekhina, Tracy L. Nero, Stephen Tomlinson, Michael W. Parker, and Rodney K. Tweten. Mapping the Intermedilysin-Human CD59 Receptor Interface Reveals a Deep Correspondence with the Binding Site on CD59 for Complement Binding Proteins C8α and C9. *Journal of Biological Chemistry*, 286(23):20952–20962, 2011.

[191] Allison J. Farrand, Eileen M. Hotze, Takehiro K. Sato, Kristin R. Wade, William C. Wimley, Arthur E. Johnson, and Rodney K. Tweten. The Cholesterol-dependent Cytolysin Membrane-binding Interface Discriminates Lipid Environments of Cholesterol to Support β-Barrel Pore Insertion. *Journal of Biological Chemistry*, 290(29):17733–17744, 2015.

[192] Nick Darvill, David J. Dubois, Sarah L. Rouse, Pierre-Mehdi Hammoudi, Tom Blake, Stefi Benjamin, Bing Liu, Dominique Soldati-Favre, and Steve Matthews. Structural Basis of Phosphatidic Acid Sensing by APH in Apicomplexan Parasites. *Structure*, 26 (8):1059–1071.e6, 2018.

[193] Andrei L. Lomize, Spencer C. Todd, and Irina D. Pogozheva. Spatial arrangement of proteins in planar and curved membranes by PPM 3.0. *Protein Science*, 31(1): 209–220, 2022.

[194] Thomas D Newport, Mark S P Sansom, and Phillip J Stansfeld. The MemProtMD database: A resource for membrane-embedded protein structures and their lipid interactions. *Nucleic Acids Research*, 47(D1):D390–D397, 2019.

[195] G. Polekhina, K. S. Giddings, R. K. Tweten, and M. W. Parker. Insights into the action of the superfamily of cholesterol-dependent cytolysins from studies of intermedilysin. *Proceedings of the National Academy of Sciences*, 102(3):600–605, 2005.

[196] Raymond W. Bourdeau, Enrico Malito, Alexandre Chenal, Brian L. Bishop, Mark W. Musch, Mitch L. Villereal, Eugene B. Chang, Elise M. Mosser, Richard F. Rest, and Wei-Jen Tang. Cellular Functions and X-ray Structure of Anthrolysin O, a Cholesterol-dependent Cytolysin Secreted by Bacillus anthracis. *Journal of Biological Chemistry*, 284(21):14645–14656, 2009.

[197] Lingfeng Xu, Bo Huang, Huamao Du, Xuejun C. Zhang, Jianguo Xu, Xuemei Li, and Zihe Rao. Crystal structure of cytotoxin protein suilysin from Streptococcus suis. *Protein & Cell*, 1(1):96–105, 2010.

[198] Stefan Köster, Katharina van Pee, Martina Hudel, Martin Leustik, Daniel Rhinow, Werner Kühlbrandt, Trinad Chakraborty, and Özkan Yildiz. Crystal structure of listeriolysin O reveals molecular details of oligomerization and pore formation. *Nature Communications*, 5(1):3690, 2014.

[199] Susanne C. Feil, David B. Ascher, Michael J. Kuiper, Rodney K. Tweten, and Michael W. Parker. Structural Studies of Streptococcus pyogenes Streptolysin O Provide Insights into the Early Steps of Membrane Penetration. *Journal of Molecular Biology*, 426(4):785–792, 2014.

[200] Jamie E. Marshall, Bayan H. A. Faraj, Alexandre R. Gingras, Rana Lonnen, Md. Arif Sheikh, Mohammed El-Mezgueldi, Peter C. E. Moody, Peter W. Andrew, and Russell Wallis. The Crystal Structure of Pneumolysin at 2.0 Å Resolution Reveals the Molecular Packing of the Pre-pore Complex. *Scientific Reports*, 5(1):13293, 2015.

[201] Kristin R. Wade, Sara L. Lawrence, Allison J. Farrand, Eileen M. Hotze, Michael J. Kuiper, Michael A. Gorman, Michelle P. Christie, Santosh Panjikar, Craig J. Morton, Michael W. Parker, and Rodney K. Tweten. The Structural Basis for a Transition State That Regulates Pore Formation in a Bacterial Toxin. *mBio*, 10(2):e00538–19, 2019.

[202] Susanne C. Feil, Sara Lawrence, Terrence D. Mulhern, Jessica K. Holien, Eileen M. Hotze, Stephen Farrand, Rodney K. Tweten, and Michael W. Parker. Structure of the Lectin Regulatory Domain of the Cholesterol-Dependent Cytolysin Lectinolysin Reveals the Basis for Its Lewis Antigen Specificity. *Structure*, 20(2):248–258, 2012.

[203] A. J. Farrand, S. LaChapelle, E. M. Hotze, A. E. Johnson, and R. K. Tweten. Only two amino acids are essential for cytolytic toxin recognition of cholesterol at the membrane surface. *Proceedings of the National Academy of Sciences*, 107(9):4341–4346, 2010.

[204] Kara S. Giddings, Ji Zhao, Peter J. Sims, and Rodney K. Tweten. Human CD59 is a receptor for the cholesterol-dependent cytolysin intermedilysin. *Nature Structural & Molecular Biology*, 11(12):1173–1178, 2004.

[205] S. E. Gelber, J. L. Aguilar, K. L. T. Lewis, and A. J. Ratner. Functional and Phylogenetic Characterization of Vaginolysin, the Human-Specific Cytolysin from Gardnerella vaginalis. *Journal of Bacteriology*, 190(11):3896–3903, 2008.

[206] Alexandra Davies, David L Simmons, Geoff Hale, Richard A Harrison, Helen Tighe, S Peter J Lachmann, and Herman Waldmanns. CD59, an LY-6-like protein expressed in human lymphoid cells, regulates the action of the complement membrane attack complex on homologous cells. *Journal of Experimental Medicine*, 170:637–654, 1989.

[207] Lucy K. Shewell, Richard M. Harvey, Melanie A. Higgins, Christopher J. Day, Lauren E. Hartley-Tassell, Austen Y. Chen, Christine M. Gillen, David B. A. James, Francis Alonzo, Victor J. Torres, Mark J. Walker, Adrienne W. Paton, James C. Paton, and Michael P. Jennings. The cholesterol-dependent cytolysins pneumolysin and streptolysin O require binding to red blood cell glycans for hemolytic activity. *Proceedings of the National Academy of Sciences*, 111(49):E5312–E5320, 2014.

[208] Lucy K. Shewell, Christopher J. Day, Freda E.-C. Jen, Thomas Haselhorst, John M. Atack, Josephine F. Reijneveld, Arun Everest-Dass, David B. A. James, Kristina M. Boguslawski, Stephan Brouwer, Christine M. Gillen, Zhenyao Luo, Bostjan Kobe, Victor Nizet, Mark von Itzstein, Mark J. Walker, Adrienne W. Paton, James C. Paton,

Victor J. Torres, and Michael P. Jennings. All major cholesterol-dependent cytolysins use glycans as cellular receptors. *Science Advances*, 6(21):eaaz4926, 2020.

[209] Daniel M. Czajkowsky, Eileen M. Hotze, Zhifeng Shao, and Rodney K. Tweten. Vertical collapse of a cytolysin prepore moves its transmembrane β-hairpins to the membrane. *The EMBO Journal*, 23(16):3206–3215, 2004.

[210] Carl Leung, Natalya V. Dudkina, Natalya Lukoyanova, Adrian W. Hodel, Irene Farabella, Arun P. Pandurangan, Nasrin Jahan, Mafalda Pires Damaso, Dino Osmanović, Cyril F Reboul, Michelle A. Dunstone, Peter W. Andrew, Rana Lonnen, Maya Topf, Helen R. Saibil, and Bart W. Hoogenboom. Stepwise visualization of membrane pore formation by suilysin, a bacterial cholesterol-dependent cytolysin. *eLife*, 3:e04247, 2014.

[211] Laura A. Shepard, Oleg Shatursky, Arthur E. Johnson, and Rodney K. Tweten. The Mechanism of Pore Assembly for a Cholesterol-Dependent Cytolysin: Formation of a Large Prepore Complex Precedes the Insertion of the Transmembrane β-Hairpins. *Biochemistry*, 39(33):10284–10293, 2000.

[212] Casie E. Soltani, Eileen M. Hotze, Arthur E. Johnson, and Rodney K. Tweten. Specific Protein-Membrane Contacts Are Required for Prepore and Pore Assembly by a Cholesterol-dependent Cytolysin. *Journal of Biological Chemistry*, 282(21):15709–15716, 2007.

[213] Martin Vögele, Ramachandra M. Bhaskara, Estefania Mulvihill, Katharina van Pee, Özkan Yildiz, Werner Kühlbrandt, Daniel J. Müller, and Gerhard Hummer. Membrane perforation by the pore-forming toxin pneumolysin. *Proceedings of the National Academy of Sciences*, 116(27):13352–13357, 2019.

[214] Eckhard R. Podack. Molecular composition of the tubular structure of the membrane attack complex of complement. *Journal of Biological Chemistry*, 259(13):8641–8647, 1984.

[215] Conall McGuinness, James C. Walsh, Charles Bayly-Jones, Michelle A. Dunstone, Michelle P. Christie, Craig J. Morton, Michael W. Parker, and Till Böcking. Single-molecule analysis of the entire perfringolysin O pore formation pathway. *eLife*, 11: e74901, 2022.

[216] David Nečas and Petr Klapetek. Gwyddion: An open-source software for SPM data analysis. *Open Physics*, 10(1), 2012.

[217] Sara L. Lawrence, Susanne C. Feil, Craig J. Morton, Allison J. Farrand, Terrence D. Mulhern, Michael A. Gorman, Kristin R. Wade, Rodney K. Tweten, and Michael W. Parker. Crystal structure of Streptococcus pneumoniae pneumolysin provides key insights into early steps of pore formation. *Scientific Reports*, 5(1):14352, 2015.

[218] Andrew Waterhouse, Martino Bertoni, Stefan Bienert, Gabriel Studer, Gerardo Tauriello, Rafal Gumienny, Florian T. Heer, Tjaart A. P. de Beer, Christine Rempfer, Lorenza Bordoli, Rosalba Lepore, and Torsten Schwede. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research*, 46(W1): W296–W303, 2018.

[219] Mark James Abraham, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C. Smith, Berk Hess, and Erik Lindahl. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1–2:19–25, 2015.

[220] Anders Olofsson, Hans Hebert, and Monica Thelestam. The projection structure of Perfringolysin 0 (Clostridium perfringens &toxin). *FEBS Letters*, 319(1, 2):125–127, 1993.

[221] Peter J. Morgan, Stefan C. Hyman, Arthur J. Rowe, Timothy J. Mitchell, Peter W. Andrew, and Helen R. Saibil. Subunit organisation and symmetry of pore-forming, oligomeric pneumolysin. *FEBS Letters*, 371(1):77–80, 1995.

[222] Lidia Prieto, Yi He, and Themis Lazaridis. Protein Arcs May Form Stable Pores in Lipid Membranes. *Biophysical Journal*, 106(1):154–161, 2014.

[223] Estefania Mulvihill, Katharina van Pee, Stefania A. Mari, Daniel J. Müller, and Özkan Yildiz. Directly Observing the Lipid-Dependent Self-Assembly and Pore-Forming Mechanism of the Cytolytic Toxin Listeriolysin O. *Nano Letters*, 15(10):6965–6973, 2015.

[224] W. F. Drew Bennett, Nicolas Sapay, and D. Peter Tieleman. Atomistic Simulations of Pore Formation and Closure in Lipid Bilayers. *Biophysical Journal*, 106(1):210–219, 2014.

[225] Jason T. Sengel and Mark I. Wallace. Measuring the potential energy barrier to lipid bilayer electroporation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1726):20160227, 2017.

[226] Yusuke Miyazaki, Susumu Okazaki, and Wataru Shinoda. Free energy analysis of membrane pore formation process in the presence of multiple melittin peptides. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1861(7):1409–1419, 2019.

[227] David M. Hoover, Oleg Chertov, and Jacek Lubkowski. The Structure of Human β-Defensin-1. *Journal of Biological Chemistry*, 276(42):39021–39026, 2001.

[228] Jianbin Ruan, Shiyu Xia, Xing Liu, Judy Lieberman, and Hao Wu. Cryo-EM structure of the gasdermin A3 membrane pore. *Nature*, 557(7703):62–67, 2018.

[229] A.R. Mól, M.S. Castro, and W. Fontes. NetWheels: A web application to create high quality peptide helical wheel and net projections. page Preprint doi: 10.1101/416347, 2018.

[230] Stephanie LaChapelle, Rodney K. Tweten, and Eileen M. Hotze. Intermedilysin-Receptor Interactions during Assembly of the Pore Complex: Assembly Intermediates Increase Host Cell Susceptibility to Complement-Mediated Lysis. *Journal of Biological Chemistry*, 284(19):12719–12726, 2009.

[231] Märt-Erik Mäeots and Radoslav I. Enchev. Structural dynamics: Review of time-resolved cryo-EM. *Acta Crystallographica Section D Structural Biology*, 78(8):927–935, 2022.

[232] Anita Hill, Simon H. Ridley, Dirk Esser, Rodney G. Oldroyd, Matthew J. Cullen, Paula Kareclas, Seán Gallagher, Geoffrey P. Smith, Stephen J. Richards, Jennifer White,

Richard A. G. Smith, and Peter Hillmen. Protection of erythrocytes from human complement–mediated lysis by membrane-targeted recombinant soluble CD59: A new approach to PNH therapy. *Blood*, 107(5):2131–2137, 2006.

[233] Alexis Lamiable, Pierre Thévenet, Julien Rey, Marek Vavrusa, Philippe Derreumaux, and Pierre Tufféry. PEP-FOLD3: Faster *de novo* structure prediction for linear peptides in solution and in complex. *Nucleic Acids Research*, 44(W1):W449–W454, 2016.

[234] Timothy H. Bayburt, Yelena V. Grinkova, and Stephen G. Sligar. Self-Assembly of Discoidal Phospholipid Bilayer Nanoparticles with Membrane Scaffold Proteins. *Nano Letters*, 2(8):853–856, 2002.

[235] I. G. Denisov, Y. V. Grinkova, A. A. Lazarides, and S. G. Sligar. Directed Self-Assembly of Monodisperse Phospholipid Bilayer Nanodiscs with Controlled Size. *Journal of the American Chemical Society*, 126(11):3477–3487, 2004.

[236] Y. V. Grinkova, I. G. Denisov, and S. G. Sligar. Engineering extended membrane scaffold proteins for self-assembly of soluble nanoscale lipid bilayers. *Protein Engineering Design and Selection*, 23(11):843–848, 2010.

[237] John L. Rubinstein, Hui Guo, Zev A. Ripstein, Ali Haydaroglu, Aaron Au, Christopher M. Yip, Justin M. Di Trani, Samir Benlekbir, and Timothy Kwok. Shake-it-off: A simple ultrasonic cryo-EM specimen-preparation device. *Acta Crystallographica Section D Structural Biology*, 75(12):1063–1070, 2019.

[238] Yong Zi Tan and John L. Rubinstein. Through-grid wicking enables high-speed cryoEM specimen preparation. *Acta Crystallographica Section D Structural Biology*, 76(11): 1092–1103, 2020.

[239] Roman I. Koning, Hildo Vader, Martijn van Nugteren, Peter A. Grocutt, Wen Yang, Ludovic L. R. Renault, Abraham J. Koster, Arnold C. F. Kamp, and Michael Schwertner. Automated vitrification of cryo-EM samples with controllable sample thickness using suction and real-time optical inspection. *Nature Communications*, 13(1):2985, 2022.

[240] Talya S. Levitz, Miriam Weckener, Ivan Fong, James H. Naismith, Catherine L. Drennan, Edward J. Brignole, Daniel K. Clare, and Michele C. Darrow. Approaches to Using the Chameleon: Robust, Automated, Fast-Plunge cryoEM Specimen Preparation. *Frontiers in Molecular Biosciences*, 9:903148, 2022.

[241] Ieva Drulyte, Rachel M. Johnson, Emma L. Hesketh, Daniel L. Hurdiss, Charlotte A. Scarff, Sebastian A. Porav, Neil A. Ranson, Stephen P. Muench, and Rebecca F. Thompson. Approaches to altering particle distributions in cryo-electron microscopy sample preparation. *Acta Crystallographica Section D Structural Biology*, 74(6):560–571, 2018.

[242] Ke Wu, Di Wu, Li Zhu, and Yi Wu. Application of Monolayer Graphene and Its Derivative in Cryo-EM Sample Preparation. *International Journal of Molecular Sciences*, 22 (16):8940, 2021.

[243] Tomasz Uchański, Simonas Masiulis, Baptiste Fischer, Valentina Kalichuk, Uriel López-Sánchez, Eleftherios Zarkadas, Miriam Weckener, Andrija Sente, Philip Ward, Alexandre Wohlkönig, Thomas Zögg, Han Remaut, James H. Naismith, Hugues Nury, Wim Vranken, A. Radu Aricescu, Els Pardon, and Jan Steyaert. Megabodies expand the nanobody toolkit for protein structure determination by single-particle cryo-EM. *Nature Methods*, 18(1):60–68, 2021.

[244] Katerina Naydenova, Peipei Jia, and Christopher J. Russo. Cryo-EM with sub–1 Å specimen movement. *Science*, 370(6513):223–226, 2020.

[245] Tristan Bepler, Andrew Morin, Micah Rapp, Julia Brasch, Lawrence Shapiro, Alex J. Noble, and Bonnie Berger. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nature Methods*, 16(11):1153–1160, 2019.

[246] Sjors H W Scheres and Shaoxia Chen. Prevention of overfitting in cryo-EM structure determination. *Nature Methods*, 9(9):853–854, 2012.

[247] Peter B. Rosenthal and Richard Henderson. Optimal Determination of Particle Orientation, Absolute Hand, and Contrast Loss in Single-particle Electron Cryomicroscopy. *Journal of Molecular Biology*, 333(4):721–745, 2003.

[248] Timothy H. Bayburt, Andrew J. Leitz, Guifu Xie, Daniel D. Oprian, and Stephen G. Sligar. Transducin Activation by Nanoscale Lipid Bilayers Containing One and Two Rhodopsins. *Journal of Biological Chemistry*, 282(20):14875–14881, 2007.

[249] Alexis Rohou and Nikolaus Grigorieff. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of Structural Biology*, 192(2):216–221, 2015.

[250] Shawn Q. Zheng, Eugene Palovcak, Jean-Paul Armache, Kliment A. Verba, Yifan Cheng, and David A. Agard. MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature Methods*, 14(4):331–332, 2017.

[251] Ali Punjani, Haowei Zhang, and David J. Fleet. Non-uniform refinement: Adaptive regularization improves single-particle cryo-EM reconstruction. *Nature Methods*, 17 (12):1214–1221, 2020.

[252] Paul D. Adams, Pavel V. Afonine, Gábor Bunkóczi, Vincent B. Chen, Ian W. Davis, Nathaniel Echols, Jeffrey J. Headd, Li-Wei Hung, Gary J. Kapral, Ralf W. Grosse-Kunstleve, Airlie J. McCoy, Nigel W. Moriarty, Robert Oeffner, Randy J. Read, David C. Richardson, Jane S. Richardson, Thomas C. Terwilliger, and Peter H. Zwart. *PHENIX*: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D Biological Crystallography*, 66(2):213–221, 2010.

[253] Berthold Kastner, Niels Fischer, Monika Mariola Golas, Bjoern Sander, Prakash Dube, Daniel Boehringer, Klaus Hartmuth, Jochen Deckert, Florian Hauer, Elmar Wolf, Hannes Uchtenhagen, Henning Urlaub, Franz Herzog, Jan Michael Peters, Dietmar Poerschke, Reinhard Lührmann, and Holger Stark. GraFix: Sample preparation for single-particle electron cryomicroscopy. *Nature Methods*, 5(1):53–55, 2008.

[254] Holger Stark. GraFix: Stabilization of Fragile Macromolecular Complexes for Single Particle Cryo-EM. In *Methods in Enzymology*, volume 481, pages 109–126. Elsevier, 2010. ISBN 978-0-12-374906-2.

[255] Elisabetta Boeri Erba, Pierre Andre Klein, and Luca Signor. Combining a NHS ester and glutaraldehyde improves crosslinking prior to MALDI MS analysis of intact protein complexes: Dual crosslinking prior to MS of protein complexes. *Journal of Mass Spectrometry*, 50(10):1114–1119, 2015.

[256] Alexander Leitner, Lukasz A. Joachimiak, Pia Unverdorben, Thomas Walzthoeni, Judith Frydman, Friedrich Förster, and Ruedi Aebersold. Chemical cross-linking/mass spectrometry targeting acidic residues in proteins and protein complexes. *Proceedings of the National Academy of Sciences*, 111(26):9455–9460, 2014.

[257] Georgios Skiniotis and Daniel R. Southworth. Single-particle cryo-electron microscopy of macromolecular complexes. *Microscopy*, 65(1):9–22, 2016.

[258] Yong Zi Tan, Philip R. Baldwin, Joseph H. Davis, James R. Williamson, Clinton S. Potter, Bridget Carragher, and Dmitry Lyumkis. Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nature Methods*, 14(8):793–796, 2017.

[259] Ali Punjani and David J. Fleet. 3D variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *Journal of Structural Biology*, 213(2):107702, 2021.

[260] Irene Pradas, Kevin Huynh, Rosanna Cabré, Victòria Ayala, Peter J. Meikle, Mariona Jové, and Reinald Pamplona. Lipidomics Reveals a Tissue-Specific Fingerprint. *Frontiers in Physiology*, 9:1165, 2018.

[261] Dirk Fitzner, Jakob M. Bader, Horst Penkert, Caroline G. Bergner, Minhui Su, Marie-Theres Weil, Michal A. Surma, Matthias Mann, Christian Klose, and Mikael Simons. Cell-Type- and Brain-Region-Resolved Mouse Brain Lipidome. *Cell Reports*, 32(11): 108132, 2020.

[262] Akash Das, Joseph L. Goldstein, Donald D. Anderson, Michael S. Brown, and Arun Radhakrishnan. Use of mutant ^{125}I-Perfringolysin O to probe transport and organization of cholesterol in membranes of animal cells. *Proceedings of the National Academy of Sciences*, 110(26):10580–10585, 2013.

[263] Katarzyna Kwiatkowska, Ewelina Marszałek–Sadowska, Gabriela Traczyk, Piotr Koprowski, Małgorzata Musielak, Agnieszka Ługowska, Magdalena Kulma, Anna Grzelczyk, and Andrzej Sobota. Visualization of cholesterol deposits in lysosomes of Niemann-Pick type C fibroblasts using recombinant perfringolysin O. *Orphanet Journal of Rare Diseases*, 9(1):64, 2014.

[264] Cuiwen He, Xuchen Hu, Rachel S. Jung, Thomas A. Weston, Norma P. Sandoval, Peter Tontonoz, Matthew R. Kilburn, Loren G. Fong, Stephen G. Young, and Haibo Jiang. High-resolution imaging and quantification of plasma membrane cholesterol by NanoSIMS. *Proceedings of the National Academy of Sciences*, 114(8):2000–2005, 2017.

[265] Masashi Maekawa. Domain 4 (D4) of Perfringolysin O to Visualize Cholesterol in Cellular Membranes—The Update. *Sensors*, 17(3):504, 2017.

[266] Leire Goicoechea, Fabian Arenas, Fernanda Castro, Susana Nuñez, Sandra Torres, Carmen Garcia-Ruiz, and José C. Fernandez-Checa. GST-Perfringolysin O production for the localization and quantification of membrane cholesterol in human and mouse brain and liver. *STAR Protocols*, 3(1):101068, 2022.

[267] K. Mann and M. Kullberg. Trastuzumab-targeted gene delivery to Her2-overexpressing breast cancer cells. *Cancer Gene Therapy*, 23(7):221–228, 2016.

[268] Joachim Frank. Time-resolved cryo-electron microscopy: Recent progress. *Journal of Structural Biology*, 200(3):303–306, 2017.

[269] Venkata P. Dandey, William C. Budell, Hui Wei, Daija Bobe, Kashyap Maruthi, Mykhailo Kopylov, Edward T. Eng, Peter A. Kahn, Jenny E. Hinshaw, Nidhi Kundu, Crina M. Nimigean, Chen Fan, Nattakan Sukomon, Seth A. Darst, Ruth M. Saecker, James Chen, Brandon Malone, Clinton S. Potter, and Bridget Carragher. Time-resolved cryo-EM using Spotiton. *Nature Methods*, 17(9):897–900, 2020.

[270] Märt-Erik Mäeots, Byungjin Lee, Andrea Nans, Seung-Geun Jeong, Mohammad M. N. Esfahani, Shan Ding, Daniel J. Smith, Chang-Soo Lee, Sung Sik Lee, Matthias Peter, and Radoslav I. Enchev. Modular microfluidics enables kinetic insight from time-resolved cryo-EM. *Nature Communications*, 11(1):3465, 2020.

[271] Jonathan M. Voss, Oliver F. Harder, Pavel K. Olshin, Marcel Drabbels, and Ulrich J. Lorenz. Rapid melting and revitrification as an approach to microsecond time-resolved cryo-electron microscopy. *Chemical Physics Letters*, 778:138812, 2021.

[272] Xueming Li. Cryo-electron tomography: Observing the cell at the atomic level. *Nature Methods*, 18:440–441, 2021.

[273] Wolfgang Baumeister. Cryo-electron tomography: A long journey to the inner space of cells. *Cell*, 185(15):2649–2652, 2022.

[274] Thomas H. Sharp, Abraham J. Koster, and Piet Gros. Heterogeneous MAC Initiator and Pore Structures in a Lipid Bilayer by Phase-Plate Cryo-electron Tomography. *Cell Reports*, 15(1):1–8, 2016.