Membrane Protein Mechanotransduction: Computational Studies and Analytics Development

A. Caroline E. Dahl

Structural Bioinformatics and Computational Biochemistry
Department of Biochemistry
University of Oxford

and

St Cross College

Trinity 2014

A thesis submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy at the University of Oxford

A copy of this work has been deposited in the Bodleian Library, Oxford.
Copyright © 2014, 2015, 2016 A. Caroline E. Dahl.
First version 10.10.2014.
Second version 06.02.2015 with minor corrections.
Third version 06.07.2016 in hardback edition.

Viva voce examination held on the 6th of January 2015 by Prof. Charlotte Deane (University of Oxford) and Prof. Anthony Lee (University of Southampton).

To contact the author, email caroline.dahl AT clinicalinnovation.se.

This document was generated using the Lyx document processor v2.0.3.

To my mother and father
Ann-Marie Dahl and Peter Dahl
to my sister
Marie-Louise Dahl
and to my grandfather
Bengt Dahl

Membrane Protein Mechanotransduction: Computational Studies and Analytics Development

Anna Caroline Elisabeth Dahl

Structural Bioinformatics and Computational Biochemistry,
Department of Biochemistry
and St Cross College
Oxford

Abstract

Membrane protein mechanotransduction is the altered function of an integral membrane protein in response to mechanical force. Such mechanosensors are found in all kingdoms of life, and increasing numbers of membrane proteins have been found to exhibit mechanosensitivity. How they mechanotransduce is an active research area and the topic of this thesis.

The methodology employed is classical molecular dynamics (MD) simulations. MD systems are complex, and two programs were developed to reduce this apparent complexity in terms of both visual abstraction and statistical analysis. *Bendix* detects and visualises helices as cylinders that follow the helix axis, and quantifies helix distortion. The functionality of Bendix is demonstrated on the symporter Mhp1, where a state is identified that had hitherto only been proposed. *InterQuant* tracks, categorises and orders proximity between parts of an MD system. Results from multiple systems are statistically interrogated for reproducibility and significant differences at the resolution of protein chains, residues or atoms.

Using these tools, the interaction between membrane and the *Escherichia coli* mechanosensitive channel of small conductance, MscS, is investigated. Results are presented for crystal structures captured in different states, one of which features electron density proposed to be lipid. MD results supports this hypothesis, and identify differential lipid interaction between closed and open states. It is concluded that propensity for lipid to leave for membrane bulk drives MscS state stability.

In a subsequent study, MscS is opened by membrane surface tension for the first time in an MD setup. The gating mechanism of MscS is explored in terms of both membrane and protein deformation in response to membrane stretch. Using novel tension methodology and the longest MD simulations of MscS performed to date, a molecular basis for the *Dashpot* gating mechanism is proposed. Lipid emerges as an active structural element with the capacity to augment protein structure in the protein structure-function paradigm.

Membrane Protein Mechanotransduction: Computational Studies and Analytics Development

A Caroline E Dahl

Contents

Nomenclature

API	Application programming interface
APL	Area per lipid
AT	Atomistic
BC	Bonferroni correction
CG	Coarse-grained
CI	confidence interval
COM	Center of mass
df	degrees of freedom
DIMS	Dynamic importance sampling
DSSP	Define Secondary Structure of Proteins
E.coli	Escherichia coli
ENM	Elastic Network Model
EPR	Electron paramagnetic resonance
GUI	Graphical User interface
K+	Potassium ion
MD	Molecular Dynamics
MS	Mechanosensitivity
MSC	Mechanosensitive ion channel
MscS	Mechanosensitive channel of small conductance
MTD	Mechanotransduction
PDB	Protein Data Bank
PE	Phosphatidylethanolamine
PG	Phosphatidylglycerol
PME	particle-mesh-Ewald
POPE	Phosphatidylethanolamine (1-Palmitoyl-2-Oleoyl-sn-glycero-3-PhosphoEthanolamine)
POPG	Phosphatidylglycerol (1-Palmitoyl-2-Oleoyl-sn-glycero-3-PhosphoGlycerol)
SD	Standard deviation
TM	Transmembrane
TM1	Transmembrane helix 1 (other numbers denote subsequent helices)
wt	Wildtype
Å	Angstrom

1. Introduction to Mechanotransduction

1.1. Fundamental definitions

Mechanosensation is the measurable response of an object to a mechanical force. The object's response could simply be to deform, or it could respond by a change of function, e.g. a change to conductivity, as observed in electronic piezos. In that case, the input force is transduced, and the overall phenomenon is called *mechanotransduction* (MTD).

'Mechanics' is traditionally a term reserved for material sciences, and investigates matters like the application of macroscopic properties such as pressure, and analysis of its effects. *Mechanobiology* therefore concerns the role of mechanical forces in biological specimens.

1.2. The importance of mechanotransduction in biology

It is essential that living organisms are able to sense and react to stimuli in their environment. Mechanical forces produced by e.g. sound, gravity, and osmotic pressure are omnipresent and mechanosensors are found in all kingdoms of life. While individual organisms' signalling networks differ, they commonly involve mechanosensitive ion channels (MSCs) that fulfill a variety of functions [Hamill and Martinac, 2001, Corry and Martinac, 2008]. In plants, mechanosensation is critical for plastid size and shape maintenance [Haswell and Meyerowitz, 2006]. In animals, mechanosensation is prominently used for touch sensation and auditory function [Guharay and Sachs, 1984, Ohmori, 1984, Hudspeth, 1986] but also for more subtle functions such as bone tissue organisation and maintenance [Dahl and Thompson, 2011]. Bacterial turgor pressure control is done through mechanotransduction, which is important for cell survival under low osmotic pressure, and such fundamental activities as cell wall expansion and cell division [Koch, 1997]. MSCs have also been identified in yeast [Gustin et al., 1988] and play a role in protozoan metamorphosis in malaria as part of its lifecycle in the human host [Rayavara and Desai, 2008].

1.2.1. Mechanosensitive ion channels

While proteins that perform a certain function can display genetic sequence similarities [Mount, 2004], MSCs lack a signature genetic sequence, so are not identified by sequence alignment alone. Instead, researchers identify MSCs by their direct response to force in an electrophysiology setup, which makes it harder to identify putative MSCs than if they were identifiable by bioinformatics methods [Levina et al., 1999, Bass et al., 2002]. The first prokaryotic MSCs were identified in *Escherichia coli* (*E.coli*) [Martinac et al., 1987]. Since then, several MSCs have been identified and characterised, including the channel formed by the antibiotic gramicidin A [Martinac and Hamill, 2002], transient receptor potential (TRP) channels [Loukin et al., 2010, Hardie and Franze, 2012], inward rectifying two-pore-domain K^+-channels such as TRAAK and TREK-1 [Maingret et al., 1999, Patel et al., 1998, Brohawn et al., 2012] and lately the voltage-gated potassium channel [Schmidt and MacKinnon, 2008, Schmidt et al., 2012].

1.2.2. Molecular basis for mechanosensitivity

Hydrophobic regions of bilayer and membrane protein surface need to match. Similarly, lipid headgroups ought to be adjacent to hydrophilic residues that have the capacity to form hydrogen bonds to polar moieties of the glycerol, phosphate and whichever chemical group(s) define the rest of the headgroup. Second, the shape of the protein needs to be accommodated by surrounding membrane. To accomplish this, annular lipid curves, compresses and stretches about the protein, and the protein changes conformation by e.g. helix tilting and local optimisation of side-chain orientation [Killian and von Heijne, 2000, Lee, 2003].

Force application may increase the membrane surface tension, thinning the membrane, which causes hydrophobic mismatch at the protein-lipid interface. It can also alter the force vectors that an integral protein experiences, the so-called membrane pressure profile. Stretch-sensitive proteins have evolved to sense and react to such forces. Under the new circumstances of an altered pressure profile and potential hydrophobic mismatch, the energetic cost of a conformational status quo is higher than the cost of conformational change, which leads to a change in protein state [Gullingsrud and Schulten, 2004, Phillips et al., 2009, Ermakov et al., 2010]. In summary, the energetic balance between membrane and protein deformation drives membrane protein MTD. Naturally, the exact membrane composition influences this cost calculation, e.g the presence of cholesterol or sphingolipids [Ollila and Vattulainen, 2010, Anishkin and Kung, 2013]. Protein mechanosensitivity can also be mediated by accessory proteins such as cadherins or integrins, or by direct attachment to the cytoskeleton [Lauritzen et al., 2005] (see also [Dahl and Thompson, 2011] and references therein).

1.3. How osmoprotection works in bacteria

1.3.1. Hydrostatic pressure at equilibrium

In bacteria there is a hydrostatic pressure arising from the 3 atm inside the bacterium versus the lower atmospheric pressure outside. This pressure gradient is generated by the bacterium's internal hypertonicity relative to the outside, causing water to enter the cell by osmosis and push on the inner cell membrane. The rigid bacterial peptidoglycan cell wall prevents the cell from expansion, and if removed by e.g. addition of penicillin, the bacterium expands. [Naismith and Booth, 2012].

1.3.2. The effect of downshock

If the tonicity of the external solution is significantly lower than 3 atm, to the point where the cell wall can no longer sustain the difference, the bacterium is said to be subjected to *downshock*. Lipid bilayers are permeable to water, which results in water influx under the high water potential, causing the bacterium to expand, increasing the tension in the membrane and threatening to burst the host bacterium. In this situation, bacterial MSCs, that locate to the inner membrane, act as internal pressure release valves. These channels open at high membrane tension to let out *osmoprotectants*; small solutes that contribute to osmolarity and are membrane-impermeable, such as potassium ions and amino acids. As osmoprotectants exit the cell, cell osmolarity decreases, lowering the water potential across the membrane. Water molecules therefore cease to enter the cell and the bacterium avoids membrane rupture and cell death [Glaasker and Poolman, 1998].

1.3.3. E.coli mechanosensitive ion channels

Eight types of MSCs have been identified in *E.coli*. They are categorised into Large, Small and Mini subclasses according to their conductance at saturating surface tension [Berrier et al., 1996, Perozo and Rees, 2003, Naismith and Booth, 2012]. This offers the bacterium the possibility of a graded response to a water potential, depending on the potential's size. Each channel's conductivity is approximately proportional to the amount of force needed the open it. This reflects cost-benefit analysis

to the bacterium of the channel function, in terms of metabolite loss versus risk of rupture. For example, the largest channel only opens at near-lytic tension, at which point retention of osmolytes plausibly results in death. At lower tension, smaller MSCs gate to let out fewer osmolytes, which limits loss.

1.4. The mechanosensitive channel of small conductance, MscS

This thesis investigates the mechanosensitive channel of small conductance, *MscS*. While there is only one large conductance channel, *MscL*, the Small class contains six homologues, of which MscS has the largest conductance. MscS orthologs are found in all phyla of walled organisms; bacteria, archae, fungi and plants, so its function is not only of interest to microbiology [Pivetti et al., 2003, Haswell and Meyerowitz, 2006, Balleza and Gomez-Lagunas, 2009]. For a list of homologues in prokaryotes and archaea, see Bass et al. [2002].

1.4.1. MscS as a drug target

A drug could hypothetically be designed to prevent MscS from either opening or closing. Mutant strains that lack MscL or MscS lack phenotype at downshock, while knock-out of both channels results in 80-90% cell death [Levina et al., 1999, Nomura et al., 2006, 2008]. A drug discovery effort that attempts to prevent MscS from opening (which would be functionally equivalent to an MscS knock-out) would therefore also have to target MscL in order to have an antibiotic effect. However, over-expression of other Msc family member channels, that open at lower membrane surface tension than do MscS and MscL, can rescue the double knock-out of MscS and MscL [Naismith and Booth, 2012]. It is plausible that strains could evolve that master such over-expression in order to survive. Therefore, a more economical drug discovery strategy is to design a drug that forces MscS into the open state. A single *E.coli* has 20–30 functional MscS [Stokes et al., 2003] that are found in the inner membrane. Energy production is also located to the inner membrane, and it relies on impermeability to protons, so the membrane must not leak unless absolutely necessary. A drug that keeps MscS open would effectively create a leak in the membrane, so is therefore likely to prevent a chemical gradient, thus killing the cell. This makes MscS an interesting drug target for novel antibiotics. Insight into its structure, and the conformational change that MscS undergoes upon gating, aids such drug discovery efforts.

1.4.2. MscS structure overview

The MscS structure has been resolved in two published crystal structures. They reveal that MscS is 12 nm long with 8 nm diameter (c.f. Figure 1.4.2) [Bass et al., 2002]. The channel is heptameric [Miller et al., 2003a] and made up of chains that are 286 amino acids in length, whereof the most N-terminal 26 residues are not resolved in either crystal structure. These missing residues form another helix according to the structure annotation software *Define Secondary Structure of Proteins* (DSSP) [PDB, 2014b], while electron paramagnetic resonance (EPR) measurements suggest a less structured domain [Vásquez et al., 2008a]. MscS consists of a transmembrane (TM) domain, which makes up the N-terminus, followed by a large cytosolic cage domain that is made up of a middle-β domain that consists primarily of β-sheets, followed by an α/β-domain. Most C-terminal is a small β-barrel where side-chains obscure passage. The most C-terminal 6 residues are also missing from the crystal structures. The protein carries a strong positive charge, +42, the majority of which locates to the TM domain.

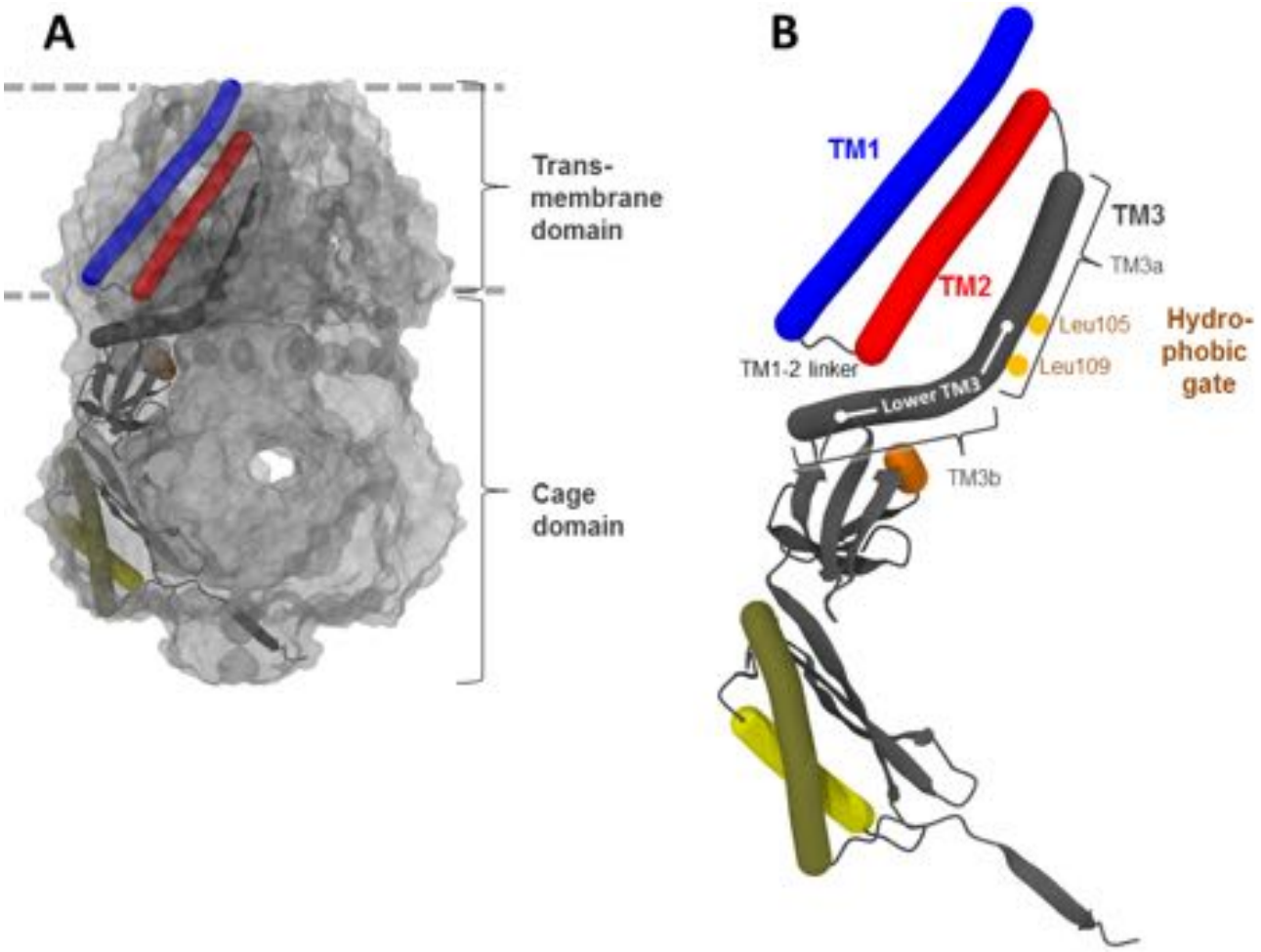

Figure 1.4.1.: Notable parts of the MscS structure.
A. MscS overview, showing a single subunit in colour and the other six chains as transparent gray surface. Its orientation in the bilayer is indicated by dashed lines that line the TM region. **B**. Single subunit from MscS, showing notable domains and parts. TM1 and TM2 form the paddle. Helices are shown as *bendices*, a representation developed as part of this thesis. For details, see chapter 3.

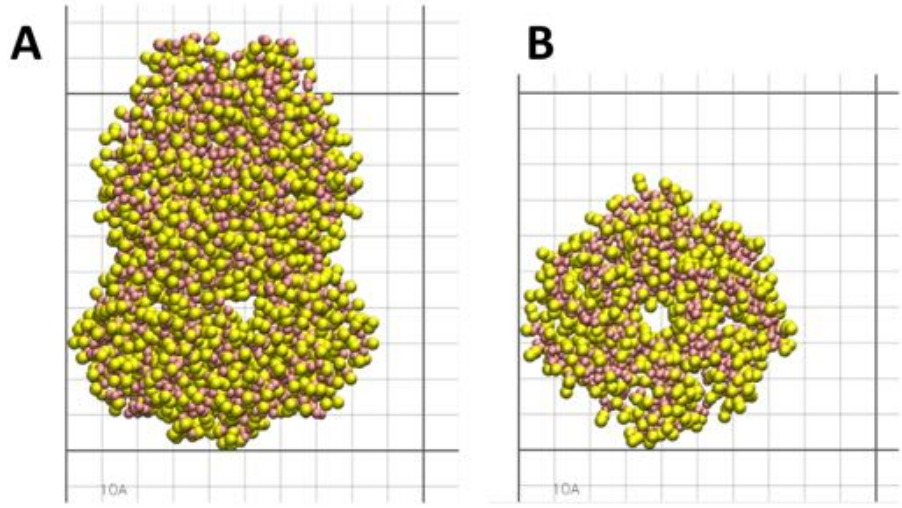

Figure 1.4.2.: The size of MscS
A. Side-view of coarse-grained MscS **B**. Top-view of coarse-grained MscS. Grid-size 10 Å. Image captured in VMD 1.9.1.

1.4.2.1. TM domain

The TM domain, as resolved in the crystal structures (discussed below), is made up of three helices, transmembrane helices 1 to 3 (TM1 to 3), that form an in-plane cross-sectional area of approximately 50 nm^2 (c.f. Fig 1.4.2). The chain N-terminus starts in the pericellular space and crosses the inner membrane three times, ending in the cytosol [Miller et al., 2003a]. DSSP annotates the constituent residues of each helix to be [Kabsch and Sander, 1983]:

1. Ala28 - Ser58
2. Ala63 - Val89
3. Ala94 - Phe127, whereof
    a) Ala94 - Gly113
    b) Ser114 - Phe127

TM1 and 2 pack together in an antiparallel fashion and are referred to as the 'paddle', named after the voltage-gated potassium channel sensor paddle.The name recognises that TM1 and 2 feature multiple positively charged residues. TM3 is subdivided into segments A and B by a circa 50° kink at residue 113 [Bass et al., 2002] which is a Glycine, a known helix breaker [Barlow and Thornton, 1988, Gunasekaran et al., 1998]. EPR measurements confirm this kink in the crystal structure [Vásquez et al., 2008a]. TM3a forms the pore and has the highest degree of conservation [Levina et al., 1999]. In its wall, conserved Gly and Ala in adjacent chains pack closely to form 'knobs and holes' [Wang et al., 2008]. Leu 105 and 109 face into the pore and form two rings that constitute the hydrophobic gate (see Figure 1.4.1). Out of the two hydrophobic residues, Leu105 is particularly conserved [Naismith and Booth, 2012].

TM3b lies parallel to the membrane and forms an amphipathic cage ceiling [Levina et al., 1999, Bass et al., 2002].

1.4.2.2. Cage domain

The cytoplasmic domain has seven 12 Å diameter portals that solutes pass through. The portals are lined by polar residues and are located between the top β-sheet and α/β domain [Bass et al., 2002]. Simulation results suggest that the cage provides a degree of cation selectivity, with potassium ions (K+) preferred over the negatively charged metabolite glutamate [Gamini et al., 2011]. However, experimental results suggest a slight anion selectivity [Martinac et al., 1987, Sukharev et al., 1993, Edwards et al., 2008].

1.4.3. Available crystal structures

Detailed knowledge of the *E.coli* MscS structure comes from two published crystal structures, one of which is presumed to depict the closed state (PDB id 1MXM, later superseded by PDB id 2OAU) and a second crystal that features the open state (PDB id 2VV5).

1.4.3.1. The 'closed' state crystal structure

The first obtained crystal structure was resolved at 3.9 Å resolution [Bass et al., 2002]. It showed that MscS has a pore constriction point formed by Leucines 105 and 109. This bottleneck diameter was originally estimated as 11 Å, but was later revised to 9.6 Å in a refined crystal structure at 3.7 Å resolution (Fig 1.4.3 **A**) [Steinbacher et al., 2007]. In comparison, the pore diameter of the *Mycobacterium Tuberculosis* MscL crystal structure is 2 Å [Chang et al., 1998] .

The state of the 'closed' crystal structure is debated The first resolved structure was originally assumed to be in the open state, but subsequent molecular dynamics (MD) studies have shown that the constriction point is too narrow and hydrophobic to be stably hydrated or permit ion flux at physiological transmembrane voltage [Sotomayor and Schulten, 2004, Anishkin and Sukharev, 2004, Spronk et al., 2006]. However, in a field that is an order of magnitude higher than is physiologically reasonable (1.2 V) the restrained, original (non-refined) closed crystal structure, became hydrated and conductive [Spronk et al., 2006]. The tentative conclusion that 2OAU is at least not completely closed may be valid as MD studies of MSCs often require extreme, non-physiological conditions to capture physiological events during the limited time that one can simulate. So it is possible that the resolved MscS structure allows at least water passage at zero surface tension.

Alternative closed state structures Experimental and computational results have also raised concerns that the true closed state of MscS could look different from the resolved crystal structure. While the location of the pore bottleneck remains unchallenged, both the location of the TM3 kink [Akitake et al., 2007] and the existence of the cavity formed between the paddle and TM3 have been questioned [Anishkin et al., 2008a, Vásquez et al., 2008b]. I discuss these concerns in detail in sections 1.6 and 1.7.1 .

1.4.3.2. The 'open' state crystal structure

A second crystal structure of MscS was resolved a few years later (PDB id 2VV5, see Fig 1.4.3 **B**) [Wang et al., 2008]. The structure was obtained at 3.45 Å resolution using sevenfold averaging, and features a mutation in the pore, Ala106Val. While the construct was a mutant, the authors argue that the structure reflects the true open state as the pore diameter is increased to 13 Å. Ala106Val disturbs the wildtype pore's close-knit 'knobs-and-holes' by introducing a large residue that widens the channel's gate. In addition, the pore-lining helix is less tilted with respect to the membrane normal and rotated 15° clockwise when viewed from the pericellular space, which rotates Leu 105 and 109 from the centre of the pore to the pore side, and this also contributes to the larger pore diameter. Taken together, the increased pore diameter is argued to break the vapour lock and is, importantly, larger than the second bottleneck to solute passage, the cage portals [Wang et al., 2008]. The pore-lining helix still kinks at Gly113, suggesting that altered helix distortion is unnecessary for gating.

The mutant 'open' structure also aligns TM helices from the same chain, while the closed crystal structure has the paddle aligned with the adjacent chain's TM3 (c.f. Fig. 1.4.3). This conformational change is relatively distal from the introduced mutation, which leads the authors to suggest that the mutation has enabled a true stable open state, rather than forcing an unnatural conformational change.

1.4.4. Experimental support for an inactive state

In addition to the closed and open states, an inactive state (also called a desensitized state) has been identified. If a pressure ramp is applied several times, fewer channels open in subsequent pressure applications compared to the first time that pressure is applied. This is indicative of use-dependent channel inactivation [Sukharev, 2002]. The open state is not accessible by tension from the inactive state, which makes it different from the closed state.

1.5. Mutations shed light on MscS structure-function relationship

The Table in Figures 1.5.1 and 1.5.2 encompasses mutation and conservation data from 13 papers. Table 1.5.1 lists the meaning of the abbreviations used for mutations and conservation, as well as the keys to references. Conservation data for the YggB MscS subfamily are taken from Pivetti et al. [2003] (only $\geq$60 % relative residue conservation is shown) while all other conservation data are from

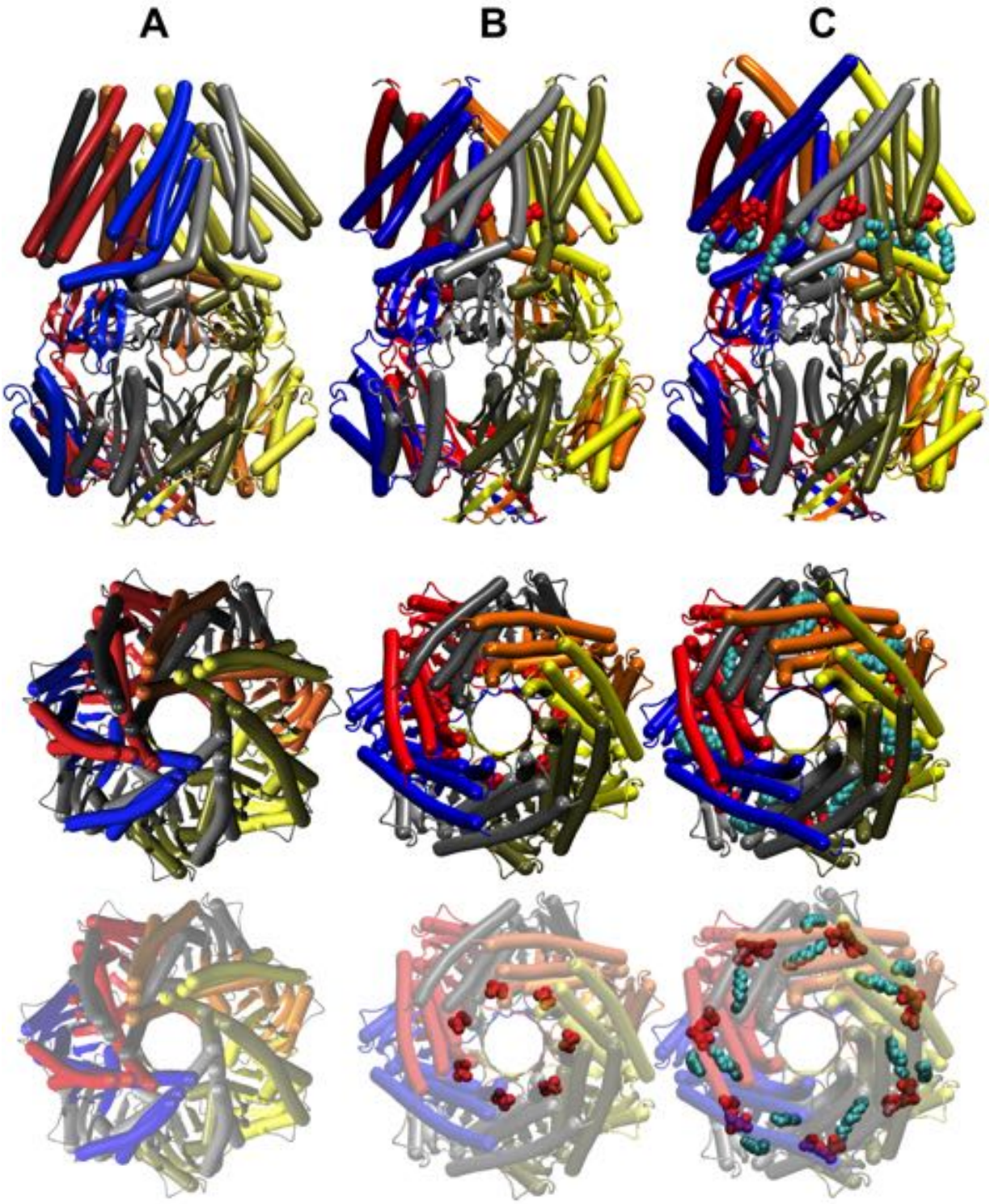

Figure 1.4.3.: Available crystal structures.
Top to bottom: Top row shows side-views of available MscS structures. The middle and bottom rows features MscS viewed down the pore from the pericellular space. The bottom row highlights mutant residues (red space-filling representations) and electron density modeled as alkyl chains (cyan space-filling representations) while the rest of the protein is transparent. **A.** The closed structure with PDB id 2OAU. **B.** The open structure with PDB id 2VV5. The mutation A106V is highlighted. **C.** A new open structure, yet unpublished. See Chapter 5 for details. The spin label at residue 67 and electron densities, modeled as alkyl chains, are highlighted. Protein rendered using bendix with uniform colour per chain (see Chapter 3).

Number	Reference
1	Vásquez et al. [2008a]
2	Nomura et al. [2006]
3	Spronk et al. [2006]
4	Miller et al. [2003a]
5	Nomura et al. [2008]
6	Belyy et al. [2010a]
7	Edwards et al. [2008]
8	Wang et al. [2008]
9	Akitake et al. [2007]
10	Schumann et al. [2004]
11	Malcolm et al. [2011]
12	Okada et al. [2002]
13	Rasmussen et al. [2007]

Abbreviation	Meaning
NEU	Neutral
POS	Positive
NEG	Negative
NHB	Non-breaking helix
ANY	Any
DEL	Deletion
B	Bacteria
A	Archae
E	Eukarya
YggB	MscS YggB subfamily
γ	Tension required for gating
L/S	MscL/MscS
I	Conductivity

Table 1.5.1.: References and abbreviations used in the table of mutations

Balleza and Gomez-Lagunas [2009]. A black box denotes that the mutation found on the lefthand side gives rise to the column phenotype. Numbers in the black box shows the quantified phenotype.

General features The tables show that TM3 is the most conserved domain of MscS. Mutations in this region produce either increased or decreased mechanosensitivity (MS), whereas paddle mutations generally decrease MS. Studies where more comprehensive sequences of mutations were made found that substitution of hydrophobic residues by residues capable of forming hydrogen-bonds lower MS if made near the interface between lipid headgroups and tails, but increase MS if made closer to the membrane middle [Okada et al., 2002, Nomura et al., 2006].

Function of the missing N-terminal While deletion of most of the 26 missing residues in the crystal structure still results in 100% downshock survival [Miller et al., 2003a], individual residue substitution suggest that it plays a role in gating. Survival experiments imply that several residues are important for MscS function [Vásquez et al., 2008a], and Trp16 mutation to Phe, Tyr or Leu render MscS less mechanosensitive [Rasmussen et al., 2007]. Combined with EPR data which suggest that the N-terminus lies at the membrane interface [Vásquez et al., 2008a], it is likely that Trp16 acts as a membrane anchorage point [Killian et al., 1996].

Saltbridge connectivity between TM and cage Mutational studies implicate Asp62 of the sensor paddle in an electrostatic bond to Arg131 on the cage domain that immediately follows TM3b. Unless the charges at position 62 and 131 are complementary (a positive and a negatively charged residue, irrespective of position), MscS requires more tension to open [Nomura et al., 2008]. Connectivity between the paddle and cage is also supported by contact analysis in MD, but between Asp62 and Arg128 [Sotomayor and Schulten, 2004].

For a more detailed discussion of mutations in the light of my thesis results, see Chapters 5 and 6.

1.6. Proposed MscS gating mechanisms

1.6.1. Conformational change necessary for gating

Solvent passage through the closed crystal structure is obstructed by Leu 105 and 109 that line the pore and form the hydrophobic gate. The cage portals constitute the second passage bottleneck,

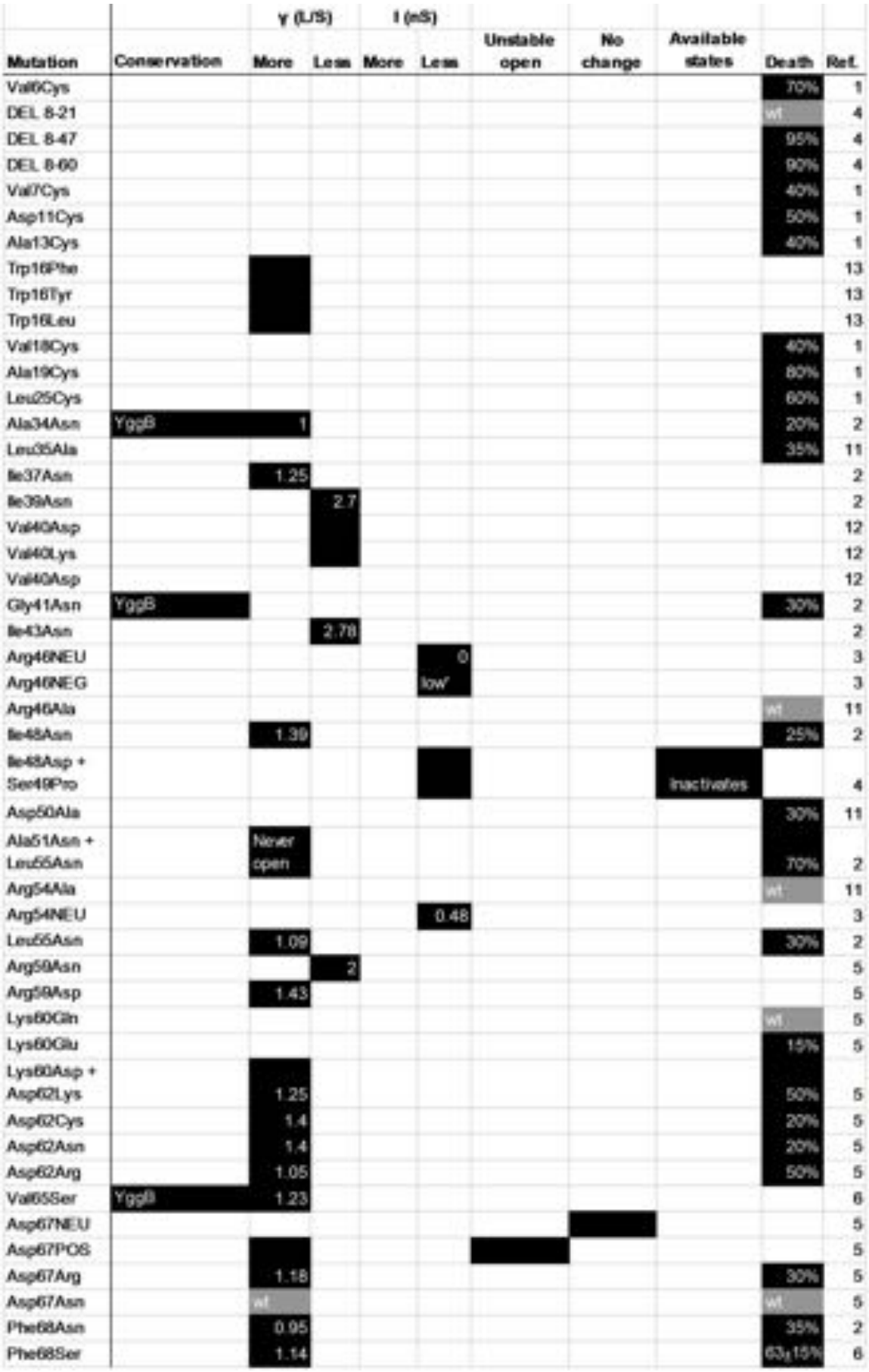

Figure 1.5.1.: Mutations performed in MscS and their phenotypes
See text and Table 1.5.1 for meanings of abbreviations and key to references.

		γ (L/S)		I (nS)						
Mutation	Conservation	More	Less	More	Less	Unstable open	No change	Available states	Death	Ref.
Arg74NEU				0.85						3
Arg74Ala					0					3
Arg74Ala									wt	11
Ile77Asn									70%	2
Leu82Ala									wt	11
Ala85Asn	YggB	0.92							30%	2
Leu86Asn		0.95							50%	2
Leu86Lys		1.12							75%	2
Leu86Asp		0.98							40%	2
Leu86Val									30%	2
Arg88NEU					0.44					3
Arg88Ser		1.39								7
Val89Asn									50%	2
Thr93Arg	YggB		1.8							4, 7
Leu100Ser	YggB	1.23								6
Gly101Asp	BAE							No inactivation		7
Ala102Pro	B (chem) +YggB		1.71		0.625			No inactivation		4, 7
Ala102Gly	B (chem)							No inactivation		7
Ala103Gly	B (chem) +YggB							No inactivation		7
Ala106Val	B									8
Val107Ser	YggB		2.56							6
Leu108Ser	B (chem) +YggB									4
Leu109Ala	B (chem) +YggB							No inactivation		7
Leu111Ser	YggB	1							87%	6
Leu111Cys	YggB									6
Gln112Gly	B (chem) +YggB									9
Gly113Arg	YggB	1.43				"Flickers"		No inactivation		7
Gly113Asp	YggB	1.31						No inactivation		7
Gly113Pro	YggB	1.33			0.75	"Flickers"		No inactivation		7
Gly113Ala	YggB	1.45						No inactivation		7
Gly113Met	YggB	1.54								7
Gly113Phe	YggB									9
Leu115Ser	B (chem)	1.1							91%	6
Gly121Phe	BAE + YggB									9
Arg128NEU										5
Arg128NEG										5
Arg131NEU										5
Arg131NEG										5
DEL Cage					0.03					7
DEL 266-286										7, 10
DEL 272-286								Less reopening		10

Figure 1.5.2.: Mutations performed in MscS and their phenotypes
See text and Table 1.5.1 for meanings of abbreviations and key to references.

but portal dimensions are assumed to be more or less static. A proposed gating mechanism must therefore find a way to increase the diameter at the hydrophobic gate, and/or move the hydrophobic side chains out of the pore. Published crystal structures, mutational data and computational modeling assist in resolving the gating mechanism of MscS, as discussed below.

1.6.2. Twist hydrophobic side chains out of the pore (twist and tilt gating)

The first hypothesis is based on the two available crystal structures and mutation data. Gating is proposed to ensue as the paddle aligns with its own chain TM3 and tilts, pulling on TM3a which moves peripherally and twist leucines out of the gate. The TM apex widens but the in-plane area remains constant. Notably, the kink at Gly113 remains throughout gating [Wang et al., 2008].

In two separate studies, mutations were made at points along the pore to increase or decrease steric hinderance to TM3a twist and tilt, and results support that they are both part of the gating mechanism of MscS [Edwards et al., 2005, Wang et al., 2008]. Moreover, paddle rotation requires movement across adjacent TM3b, and mutations there are known to cause rapid pore closure post-opening [Nomura et al., 2008].

1.6.3. TM3 curvature magnitude and location (collapsible strut gating)

The second gating hypothesis is based around mutation of helix-breaking pore residues. If Gly113 is mutated to a residue that favours helicity (Alanine), less tension gates MscS, whereas mutation to a helix-breaking residue causes MscS to require more tension to gate. The same behaviour was observed for Gly121. While Gly121Ala slows inactivation, Gly113Ala prevents inactivation. Based on this, Akitake *et al.* (2007) argue that un/kinking at Gly121 denotes open and closed states, while a kink at Gly113 (as seen in both crystal structures) denotes inactivation [Akitake et al., 2007]. The double mutant, which favours helicity throughout TM3, kills the bacterium, suggesting that it causes MscS to be permanently open. Straightening of TM3 is thus key to gating, and curvature at Gly113 would, somehow, stop force transmission between the paddle and gate. Steered MD of the 'closed' crystal structure shows that TM3 straightens, and force release causes TM3 to kink at Gly121 [Akitake et al., 2007]. Another MD study of MscS in a high voltage field supports that TM3 straightens and makes contact with the paddle as the pore widens to 13 Å [Sotomayor et al., 2007]. Less kink at Gly113 is also seen in MscS subject to steered MD where the pore reaches a near-open diameter, suggesting that de-kinking at Gly113 is part of the conformational change that MscS undergoes as it opens [Sotomayor et al., 2006]. For details on what MD entails, refer to Chapter 2.

1.6.4. Paddle function in the gating mechanism

The paddle has the weakest electron density of any MscS helix, suggesting that it may be mobile [Bass et al., 2002]. MD shows that the paddle is moved 2 to 3 Å cytosolically by a 400 mV transmembrane field, as a consequence of the positive charge on Arg46. Depolarisation would release the paddle, which could have an effect on the gate [Spronk et al., 2006]. However, the authors admit that their 10 ns simulations were too short for conclusive results. Other authors have suggested that the cytosolic ends of the sensor paddles might move peripherally, whereas the pericellular paddle end stays in place, acting as a pivot [Bass et al., 2002, Bezanilla and Perozo, 2002].

Theoretical calculations of paddle and TM3 orientation based on side-chain volume suggest that the closed state paddle ought to be proximal to its own chain TM3, leaving no void inbetween. High surface tension and steered MD of the paddle in this 'optimised' state successfully transmits tension to the pore, which moves simultaneously, gating MscS. Apolar contacts between paddle and TM3 are credited for tension transmission [Belyy et al., 2010a]. The necessity for contact between the paddle and TM3 is supported by a second study that achieved a smaller pore by high tension and steered MD of the paddle, compared to steered MD of TM3b to bring it proximal to the paddle [Sotomayor et al., 2006].

1.7. Structure controversy and new, complementary data presented in this thesis

1.7.1. Debated cavities between paddle and pore-lining helix

The existence of cavities between the paddle and pore-lining helix has been questioned since no evident support exists for their structural integrity [Akitake et al., 2007]. Moreover, experiments show that Leu72 and Val99 (on TM2 and TM3a, respectively) can be cross-linked in a setup at zero surface tension, which suggests that these residues are separated by a smaller distance than is proposed in the closed crystal [Anishkin et al., 2008a]. EPR measurements of spin-labelled MscS also suggest a more compacted structure where TM2 is more parallel to the membrane normal and tightly packed to TM3a than in the closed crystal structure [Vásquez et al., 2008a], and volume conservation studies also support compaction [Belyy et al., 2010a]. However, more recent studies show that these cavities remain intact in both detergent [Pliotas et al., 2012] and bilayer setups (Ward *et al.*, submitted), disputing that they are crystal artifacts. It has been speculated that these cavities could contain lipids [Wang et al., 2008].

1.7.2. A new structure features electron density in the cavities

My experimental collaborators generated a new crystal structure of the open state during the course of this DPhil work. To distinguish this structure from the previously published open state (PDB id 2VV5), I will refer to this latest structure as the 'new open structure'. Figure 1.4.3 **C** shows it alongside the previously published crystal structures. Notably, it features electron density in the cytosolic TM region; part of the debated cavity between the paddle and pore-lining helix. This density is proposed to be of lipid origin. If true, this places lipid at the intersection of the sensor paddle and gate, where it could play an active role in the gating mechanism of MscS.

This structure is described in more detail in Chapter 5, where I perform simulations of it in order to explore whether the density is lipid.

1.8. Work presented in this thesis

The theme of this thesis work is two-fold: to abstract and reduce the inherent complexity of MD results, and to explore MS at the molecular level, as an interplay between membrane and protein. Emphasis is put on result reproducibility, not just when planning the simulation samples used in this thesis, but also by developing accessible tools to make abstraction and statistical analysis manageable beyond this work.

Chapter 2 introduces the fundamental concepts of MD. I discuss the statistical mechanics that link macroscopic parameters to particle behaviour as well as the components of the algorithm that computes system behaviour. The system itself can have different levels of resolution; such *multiscaling* approaches are explained. In the details of practical implementation, I also explain the parameters that are chosen to mimic experimental conditions for MscS, and that are used throughout the simulations in this thesis.

Ultimately, biologically relevant MD systems are often large, on the order of 100,000 to 1,000,000 particles, and analysis of their dynamics is challenging. To that end, **Chapter 3** describes the visual analytics software *Bendix*, that abstracts protein structure and computes helix geometry. Because multiscale MD methodology is key to this thesis, I ensure that Bendix supports the most universal multiscale techniques, and that extensions exist to any particle type. Software features are demonstrated on the transporter protein Mhp1, where helix distortion is key to functionality. With the aid of Bendix, I reveal novel features of the Mhp1 transport mechanism and recognise a state that had only been proposed earlier.

While my analysis of individual simulations are greatly aided by the Bendix tool, reproducibility is imperative in MD studies. **Chapter 4** presents a second tool for reducing the complexity of MD systems; this time across both monomers and simulations. *InterQuant* is a software that tracks, categorises and orders proximity between two particle subsets in a system. InterQuant's meta-analysis generates statistics across replicate simulations, and derives any statistically significant differences between samples at the resolution of individual chains, residues or atoms.

WIth these tools at hand, I set out to investigate molecular level MTD using MD of MscS. MD allows me to introduce the available MscS crystal structures into a physiological environment where I can study their dynamics under conditions of equilibrium as well as membrane perturbation. **Chapter 5** explores lipid interaction with crystal structures at equilibrium, assuming that these represent the true closed and open states. Since membrane tension is necessary and sufficient for channel opening, identified changes to lipid interactions should control the propensity of MscS to enter either state. I identify high membrane curvature about MscS, causing MscS to largely relocate to the cytosol where charges along the MscS surface can be satisfied. However, the cytosolic region is not devoid of lipid; on the contrary, lipid penetrates between chains into the MscS gate. This lipid overlaps with electron density found in the new structure, suggesting that it is deeply buried POPE lipid. Differential lipid interaction, analysed at multiple levels of resolution, suggest key elements of a novel gating mechanism where lipid takes centre stage.

However, MscS opens physiologically under non-zero surface tension. **Chapter 6** successfully opens MscS from its closed crystal structure state, using tension. This is the first time that MscS has been observed to open by tension alone in an MD setup. Without restraints, MscS enters a state where helices are more evenly distributed, as observed in the open structure, but other key features of the open structure fail to show. Globally, membrane curvature decreases with increasing tension, effectively pulling MscS into the bulk membrane and bringing the hydrophobic gate in line with the membrane centre. As MscS enters the membrane proper, the apex of MscS is pulled radially by closely associated lipid, causing the TM domain to adopt a cylindrical shape and straightening the pore-lining helix. The speed at which tension is applied is found to correlate with the strength of association between TM domain periphery and the gate, which suggests a molecular basis for the experimentally observed *Dashpot* gating mechanism. As observed in Chapter 5, lipid interaction appears pivotal to channel opening, and this last chapter complements my equilibrium studies with a dynamic view of MscS-lipid interaction. My results add new insights both to the MscS open state and gating mechanism, both of which are subject to ongoing debate.

2. Theory and methods

A molecule's behaviour can be understood through its constituent particles' interactions and dynamics, which are described by quantum mechanics and electrodynamics. While very small biomolecules have been simulated *in silico* using software packages that incorporate such *quantum mechanical molecular mechanics* methods, the necessary calculations are too computationally intensive to allow simulation of full proteins [Leach, 2001]. In its place, classical molecular dynamics (MD), where electrons and nuclei are grouped into a point with a net charge, mass and effective potential, is used to probe larger, biologically relevant systems. To model particle interactions, classical MD employs computationally efficient equations that are appropriate near interaction equilibrium. This chapter is an overview of classical MD, which is used in simulations throughout this thesis. In the remainder of this thesis, MD refers to classical MD.

Statistical thermodynamics links MD simulation results and experimental data. System parameters are calculated using known statistical relationships between macroscopic properties such as pressure, and microscopic properties such as the movement of particles. In addition, MD naturally explores system details at atomic level resolution, which is impossible, or very hard, to interrogate by experiment, so is routinely used to investigate protein dynamics in order to complement experimental data [Karplus and McCammon, 2002]. Further, experimental interrogation necessarily involves interaction with the system to retrieve system properties. For example, the existence or position of a biomolecule is queried using attached dye, diffusion constants are retrieved using bleaching, and an AFM cantilever explores surface topology and conductance. MD removes the need to interact with the system in order to gain understanding about it, thus eliminating the potential bias that interaction implies.

In structural biochemistry, structures obtained through crystallography and nuclear magnetic resonance can suffer from the necessarily fierce experimental conditions such as low temperature, detergent, high salt concentration, close-packing and averaging. MD is used to bridge such data with a more physiological, equilibrated environment, which over time can evolve a more biologically relevant approximation of an initial input structure. Moreover, this evolution, or *trajectory*, reveals important details about protein dynamics. The physiological environment is not necessarily a constant; it can be manipulated to replicate the different conditions that the protein operates under, such as varying membrane composition, osmotic pressure, crowding, voltage and temperature.

Taken together, MD provides a highly customizable and detailed, complementary methodology to experimentally retrieved data.

2.1. The MD force field

The MD force field relates potential energy V to particle position r with respect to adjacent particles. Particles are individual atoms in classical MD, but can also represent several atoms, as is the case in united atom and coarse-grained models. All-atom, united atom and coarse-grained models require different force-fields for simulation of their respective particle types. Examples of such biomolecule force fields are GROMOS96 [van Gunsteren et al., 1996], which I use in my atomistic simulations, OPLS [Jorgensen and Tirado-Rives, 1988], CHARMM [Brooks et al., 1983, MacKerell, 1998], AMBER [Case et al., 2005] and MARTINI [Marrink et al., 2007, Monticelli et al., 2008]. United-atoms and coarse-grained models are described in more detail in Sections 2.1.4 and 2.1.4.2.

Common for all force fields is that force on a particle i is solved numerically by computing the potential energy landscape about the particle:

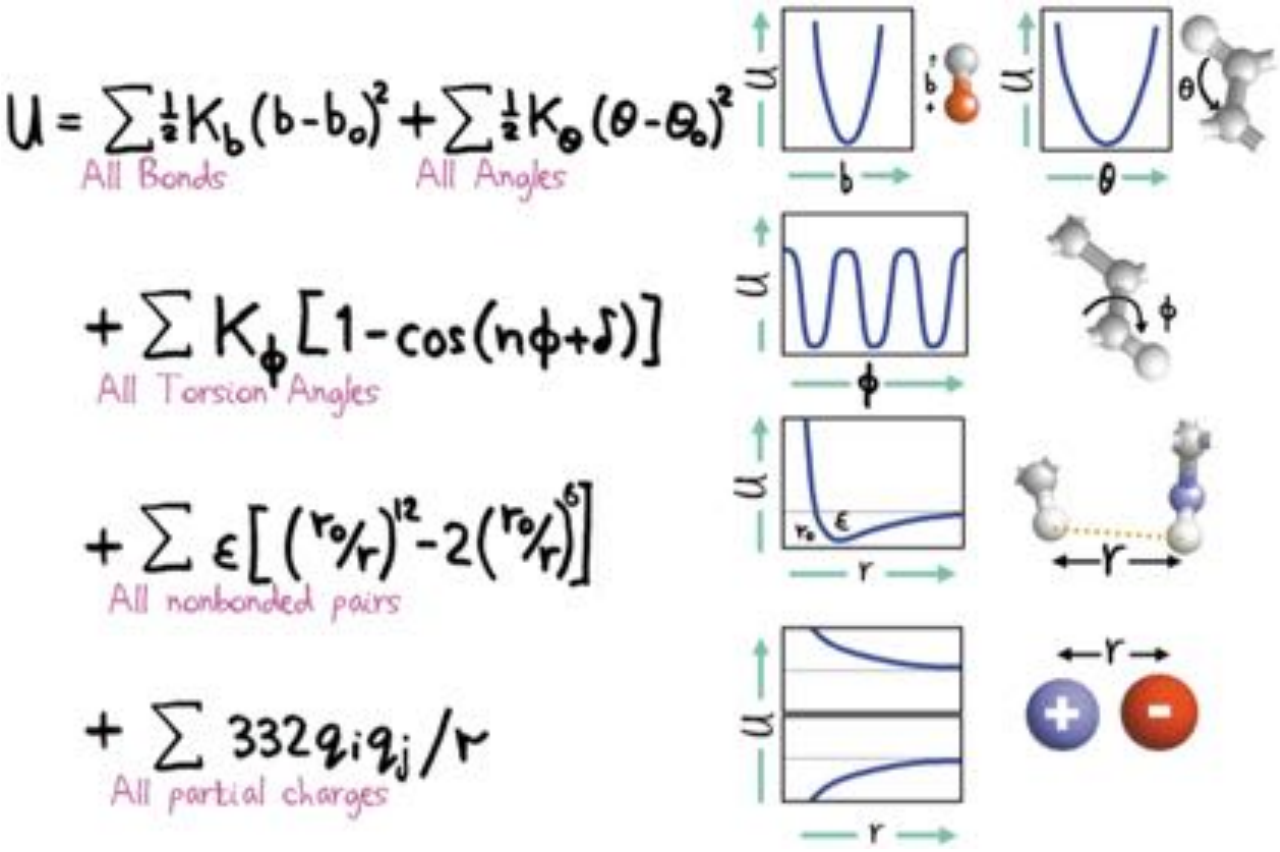

Figure 2.1.1.: Components of the MD energy landscape
Figure from Levitt [2001]. In addition to these terms, improper dihedral angles can also be used in MD. See main text for details.

$$-\frac{\partial V(\mathbf{r})}{\partial \mathbf{r_i}} = \mathbf{F_i} = m_i \mathbf{a}_i \tag{2.1.1}$$

where m_i and $\mathbf{a}_i$ are the mass and acceleration of particle i, respectively, and

$$V(\mathbf{r}) = V(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N). \tag{2.1.2}$$

Since

$$\mathbf{a_i} = \frac{\partial \mathbf{v_i}}{\partial t} = \frac{\partial^2 \mathbf{r_i}}{\partial t^2} \tag{2.1.3}$$

the future position of particle i is solved. Key to computing particle trajectories is therefore the potential energy function of particle interactions, which underlies the potential energy landscape. The contributions to the MD force field are grouped into bonded and non-bonded interactions, where bonded interactions are defined as those between covalently bound particles, while non-bonded interactions act on all particles. Therefore

$$V(\mathbf{r}) = U_{bonded}(\mathbf{r}) + U_{non-bonded}(\mathbf{r}). \tag{2.1.4}$$

An overview of the components of bonded and non-bonded energy contributions can be seen in Figure 2.1.1.

Energy is added particle pair-wise, and since the number of particle pairs grows as the square of atoms in the system, $U_{non-bonded}(\mathbf{r})$ can quickly become very computationally expensive. MD method developers have found ways to reduce this complexity, and this is discussed in Section 2.1.2. First follows an overview of common treatments of bonded and non-bonded interactions in MD force-fields.

2.1.1. Bonded interactions

Bonds between two, three and four adjacent particles are evaluated for stretching, angle and torsion, respectively. In addition, four-particle 'improper dihedral angles' are often added to retain stereochemistry. Together, they add to give bonded potential energy:

$$U_{bonded}(\mathbf{r}) = U_{bondlength}(\mathbf{r}) + U_{angle}(\mathbf{r}) + U_{dihedral}(\mathbf{r}) + U_{improper}(\mathbf{r}). \tag{2.1.5}$$

Bond lengths

The potential energy associated with particle separation is modeled as a harmonic function (as illustrated in Figure 2.1.1) or as a Morse potential [van der Spoel et al., 2010]. Bond lengths are often made constant to increase computational efficiency (see Section 2.2.6).

Bond angles

The potential energy of a bond angle is a harmonic function centered about the ideal angle between three adjacent particles.

Dihedral angles (Torsion angles)

The position between the first and last particle in a four-particle chain can give rise to steric clashes, and therefore contributes to bonded potential energy. In the example of an alkane chain, low-energy conformations such as anti and gauche are interspersed by high-energy eclipsed conformations where peripheral particles are more proximal. The resulting function is thus periodic and is commonly approximated by a cosine function, as illustrated in Figure 2.1.1.

Improper dihedrals

This other four-particle energy term is used to restrain particles so that stereochemistry is preserved in chiral structures and rings remain planar [van der Spoel et al., 2010]. Such improper dihedrals have been used to model e.g. sugars in order to ensure that their most abundant chiral conformations are featured [Guvench et al., 2011]. While dihedral angles can rotate (so are periodic), improper dihedrals have a single potential energy minimum and are commonly modeled as harmonic potentials.

2.1.2. Non-bonded interactions

The Lennard-Jones potential

The Lennard-Jones (LJ) potential describes both proximal inter-particle Pauli repulsion due to electron cloud overlap, and distal attraction due to induced-dipole effects as adjacent particles' electron clouds adapt to one another. Attraction and repulsion have been shown to be inversely proportional to the sixth and twelfth power of particle separation, respectively. While the repulsive 12th power term is empirical, the attractive 6th power comes from the the London formula for the interaction energy of uncharged, identical molecules [Evans, 1992]. The result potential energy function between particles i and j is

$$V_{LJ}(r_{ij}) = \varepsilon(\frac{\sigma^{12}}{r_{ij}^{12}} - 2\frac{\sigma^{6}}{r_{ij}^{6}}) = \frac{C_{ij}^{(12)}}{r_{ij}^{12}} - \frac{C_{ij}^{(6)}}{r_{ij}^{6}}$$

where ε is the potential energy minimum (the strength of interaction) and σ is the equilibrium centre of mass separation [Leach, 2001]. The right-most expression shows the form commonly adopted, where ε and σ are incorporated into the nominators for computational efficiency [van der Spoel et al.,

2010]. Hetero-particle interactions' $C_{ij}^{(6)}$ can be generated by averaging homo-particle interactions, as follows:

$$C_{ij}^{(6)} = \sqrt{C_{ii}^{(6)} C_{jj}^{(6)}}$$

and similarly for $C_{ij}^{(12)}$. Alternative approaches exist, such as the Lorentz-Berthelot rule, where the average equilibrium COM separation and strength of interaction is computed separately:

$$\sigma_{ij} = \frac{1}{2} \left(\sigma_{ii} + \sigma_{jj} \right) ,$$

$$\varepsilon_{ij} = \sqrt{\varepsilon_{ii} \varepsilon_{jj}} \, .$$

The Lennard-Jones cut-off

Since only attraction acts at a distance, and is inversely proportional to separation as r_{ij}^{-6}, it approaches zero even at relatively small separation. Therefore, in order to avoid computing the minuscule LJ contribution to V at larger distances, a LJ cutoff is commonly used, beyond which LJ is not evaluated. Only LJ interaction to particles that are located within particle i's cutoff contribute to i's potential energy landscape.

Pair exclusion

Same-chain, covalently bonded particles can be positioned so close that non-bonded potential functions describe their interactions poorly. Therefore, two, three and sometimes four adjacent particle interactions are modified or excluded completely from $U_{non-bonded}(\mathbf{r})$ [Leach, 2001, van der Spoel et al., 2010].

2.1.3. Electrostatic interactions

Electrostatic interactions occur between charged or polar particles, and are modeled according to Coulomb's law. Particle charge can be integer or partial; the latter to reflect polarity. However polarity does not redistribute; all charge is fixed to the particle COM [van der Spoel et al., 2010]. The electric potential and force between charges q_i and q_j is

$$V_e = \frac{q_i q_j}{4\pi\epsilon r_{ij}} = \frac{332 q_i q_j}{r_{ij}}$$

$$\mathbf{F}_e = -\frac{q_i q_j}{4\pi\epsilon r_{ij}^2} \hat{r}_{ij}$$

where ϵ is the electrical permittivity of the medium; the amount of energy that a dielectric substance requires in order to realign itself according to the field that results from the charges q_i and q_j. The signs of the charges determine whether the interaction is attractive or repulsive, as reflected in the two graphs for "All partial charges" in Figure 2.1.1.

Compared to the LJ potential, where energy of interaction falls off fast with distance and can be safely omitted beyond a cutoff, coulomb interactions fall off very slowly and thus act on large distances. Crucially, this becomes problematic when the system size is infinite, as is the case when system periodicity is implemented (see Section 2.2.2 for details). While pair-wise computation of electrostatic interactions would be highly computationally expensive (as discussed earlier), truncation has been shown to produce simulation artifacts [Norberg and Nilsson, 2000, Patra et al., 2003]. Instead, methods have been developed that compute electrostatic interactions explicitly within a given cutoff, and evaluate interactions beyond the cutoff in a more computationally economic manner.

The particle-mesh-Ewald method

The particle-mesh-Ewald (PME) method is a commonly used technique to evaluate long-range electrostatic interactions in MD [Darden et al., 1993]. PME uses Fourier transforms to model interactions that are inherently periodic across boxes when periodic boundary conditions are implemented. This allows the use of mathematical techniques from the Fourier transform 'toolbox', which simplifies and ultimately solves the energetic contribution of long-range interactions.

The electrostatic potential is written as a sum over all boxes N_{box} and particles N:

$$V_e = \frac{1}{2}\sum_{N_{box}}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{q_i q_j}{4\pi\epsilon|r_{ij}+N_{box}|}$$

where $\frac{1}{2}$ avoids double counting the particle double sum. This equation is then rewritten as two terms using the relationship

$$\frac{1}{r} = \frac{f(r)}{r} + \frac{1-f(r)}{r}$$

$f(r)$ is chosen so that it falls off to zero in both real space, as separation increases, and in reciprocal (Fourier) space. To aid convergence, an evaluation cutoff is used in real space, and limited number of modes are used in Fourier space. Further, charges are positioned onto a grid of finite resolution and $f(r)$ is formulated as though each charge was surrounded by a Gaussian distribution of neutralising charge with respect to the centre charge. This allows V_e to be reformulated as

$$V_e = V_{e1} + V_{e2} + const$$

where

$$V_{e1} = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{|N_{box}|=0}\frac{q_i q_j \xi(\alpha|r_{ij}+N_{box}|)}{4\pi\epsilon|r_{ij}+N_{box}|}$$

$$V_{e2} = \frac{1}{2}\sum_{k\neq 0}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\pi L^3}\frac{q_i q_j}{4\pi\epsilon}\frac{4\pi^2}{k^2}\exp\left(-\frac{k^2}{4\alpha^2}\right)\cos\left(k\cdot r_{ij}\right) .$$

V_{e1} is evaluated in real space while V_{e2} is evaluated in reciprocal space, using the fast Fourier transform (the Fourier transform for discretised space) for conversion between reciprocal and real space. ξ is the complementary error function and convergence ultimately depends on the choice value of α [Leach, 2001].

Charge groups

The GROMOS96 force field uses charge groups; approximately spherical volumes that contain a few spatially adjacent atoms that, when combined, have zero or integer charge. Charge groups are used to reduce the effects of truncation when using cutoffs to compute electrostatic interaction [van Gunsteren et al., 1996, van der Spoel et al., 2010]. The following algorithm is used:

1. Pick a charged atom, i.
2. Identify its charge group.
3. Identify charged atoms outside of atom i's charge group that are within cutoff.
4. If charged atoms are found,
   a) identify their charge groups.
   b) Per identified charge group,
      i. compute all pair-wise electrostatic interactions between this and atom i's charge group.

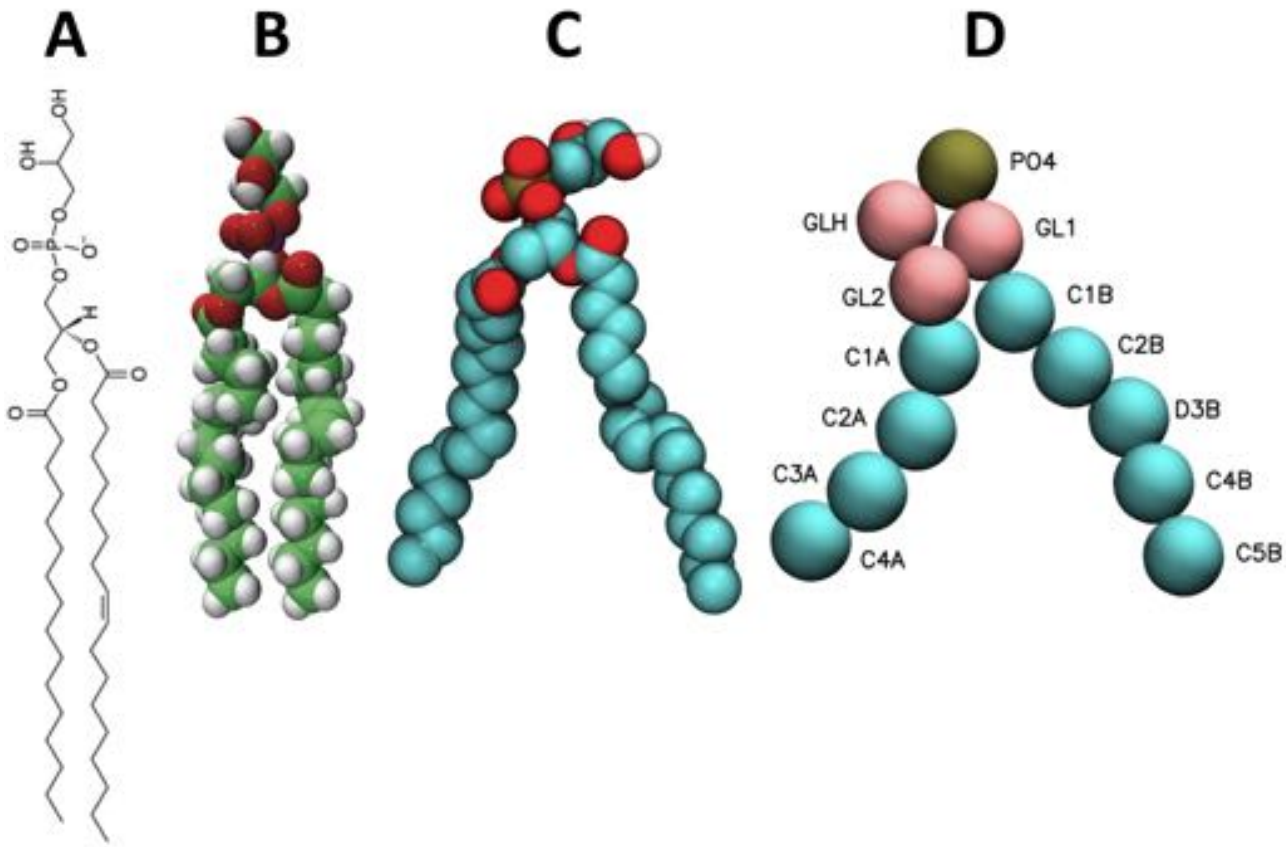

Figure 2.1.2.: POPG lipid in all-atom, united atom and coarse-grained form
A. Chemical structure drawing of POPG. Image credit: Avanti lipids. **B.** Space-filling model of all-atom atomistic POPG. Image credit: Avanti lipids. **C.** United-atoms model of POPG. **D.** Coare-grained POPG in particle-type coloured van der Waal representation and indicated particle names. The bead color shows hydrogen (white), oxygen (red), carbon (green). Brown beads show phosphorus (B, C) or phosphate (D), while turqoise beads show aliphatic compunds.

2.1.4. Particle constitution

2.1.4.1. United atoms

In aliphatic moieties CH_1, CH_2 and CH_3, GROMOS96 unites the non-polar hydrogens with their respective carbon, thus forming a larger radius particle. The rationale for this is that such hydrogens only interact with the rest of the system by LJ potentials, such hydrogens are numerous in biomolecules (in particular lipid) and that system size reduction speeds up MD trajectory computation [Leach, 2001]. For a comparison between all and united atom representation of POPG, see figure 2.1.2 **B** and **C**.

2.1.4.2. Coarse-graining methodology

Similar to united atoms, coarse-graining (CG) joins adjacent atoms into larger 'functional group particles', thereby reducing the number of particles in the system. However, the CG methodology merges not just non-polar hydrogens but all atom types into groups of three to four equivalent united atoms (see Figures 2.1.2 and 2.1.3). Such system reduction speeds up computation to the point where physiologically relevant simulation time- and length scales can be attainable [Marrink et al., 2007, Monticelli et al., 2008].

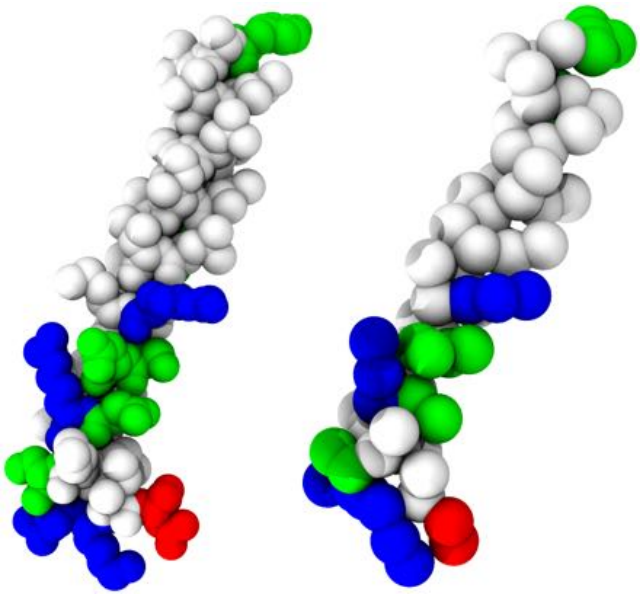

Figure 2.1.3.: Example of coarse-grained protein.
On the left hand side is TM1 of the closed state MscS (PDB id 2OAU) shown in united-atom AT representation, and on the right hand side is the same helix, converted to Martini CG. Note in particular the loss of resolution (fewer particles) and increased particle diameter. Polar residues are shown in green, positively charged in blue, negatively charged in red and hydrophobic residues in white van der Waal representation.

Martini

In this thesis, I employ the Martini 2.1 CG forcefield [Monticelli et al., 2008]. Martini CG particles are of one of four types based on their strength of LJ interaction (ϵ): apolar (C), polar (P), non-polar (N) or charged (Q). Polar particles are typically composed of atoms with partial charges, while hydrophobic atoms make up apolar particles. Non-polar particles contain a mix of polar and apolar atoms, while charged particles have full charges. Particles are further subclassified according to either hydrogen-bonding capacity (non-polar and charged particles: donor, acceptor, donor and acceptor, or neither) or degree of polarity (apolar and polar particles: 1 to 5, with 5 being the most polar). The equilibrium COM separation (σ) is set to 4.7 Å for all particle types except for special cases of aromatic and 'anti-freeze' particles. Anti-freeze particles were introduced in order to disrupt CG water particles from lattice packing and subsequently freezing, as this occurs stochastically by nucleation at 280 to 300K [Marrink et al., 2007]. Aromatic particles are relatively small, made up of two or three united atoms, and have reduced σ and ϵ. Their planarity is ensured by rigid constraints or improper dihedrals.

All particles have a mass of 72 Da, equivalent to four water molecules, and charges interact by a Coulomb potential that is smoothly shifted distally so that it equals zero at 12 Å.

MD systems made up of MARTINI particles have a higher diffusion rate in the liquid phase [Marrink et al., 2007]. This causes faster effective sampling, and some authors argue that it is appropriate to scale MD simulation time by a factor of 4 to take this into account [Ollila et al., 2011]. In this thesis, however, all simulation time reported is unscaled.

"Bond"

The "Bond" CG forcefield contains less particle types than the current Martini force field [Bond and Sansom, 2006]. "Bond" is based on an early version of Martini, where protein support was still lacking. Therefore, while individual parameters do differ, the two force fields are broadly similar; for example, the number of types of LJ interactions and treatment of Coulomb interactions are the same. I used the "Bond" CG model for preliminary studies where I studied the effect of varying elastic

network models (see below), but use the Martini model for all other simulations.

Elastic network model

Since atomic-detail interactions that are direction-sensitive, such as hydrogen bonds, are not present in a CG model, other means of preserving protein secondary and tertiary structure must be used. An elastic network model (ENM) is a network of harmonic potentials, that join the coarse-grained protein backbone to stabilise secondary to quaternary structure [Rader et al., 2006]. There are four main ways to design an ENM:

1. between all particles within a certain cutoff separation, often only applied to the protein backbone
2. in regions with specific secondary structure
3. between hydrogen-bonding particles
4. customise according to prior knowledge

In this project, all four types of ENM were implemented. See chapters 5 and 6 for their respective ENM setups, and the rationale behind them.

2.1.5. Force field parameterisation and choice

What distinguishes one force field from another are the free parameters in the functions for particle interaction, and, as described in the CG section, the level of particle abstraction. When creating a force field, free parameters are tuned according to experimental and quantum chemical data which are often selected in order to reproduce certain biomolecular features. The GROMOS family of force fields, for example, focus on e.g. large alkane properties (43a3) and reproduction of solvation enthalpies of amino acids in cyclohexane and water (53a5 and 53a6; the latter is used in this thesis) [Schuler et al., 2001, Oostenbrink et al., 2004]. Reference data come from crystal structures, NMR and various other biochemical techniques depending on the property of interest [Leach, 2001].

2.2. MD simulation

Once particle properties and interactions have been selected in the form of a force field, MD simulation computes the evolution of an initial system setup over time.

2.2.1. Equations of motion and numeric integrators

As mentioned in the introduction to MD force fields, the potential energy landscape about a particle determines the force that the particle experiences, which in turn is equal to particle mass times acceleration:

$$-\frac{\partial V(\mathbf{r})}{\partial \mathbf{r_i}} = \mathbf{F_i} = m_i \mathbf{a}_i \,. \tag{2.2.1}$$

In MD, acceleration is numerically integrated in order to arrive at the future position and velocity of the particle. This is done using a numerical integrator which uses a discrete time step Δt. Different integrators exist and are incorporated in different MD engines, but they all extrapolate position and velocity using different numbers of terms from Taylor expansion about the current particle position and velocity. The Euler method is relatively simple and uses only the two first terms

$$r(t + \Delta t) = r(t) + \frac{dr(t)}{dt}\Delta t \equiv r(t) + \mathbf{v}(t)\Delta t$$

$$v(t+\Delta t) = v(t) + \frac{dv(t)}{dt}\Delta t \equiv v(t) + \mathbf{a}(t)\Delta t$$

while the Verlet integration scheme uses the first five terms [Leach, 2001]. The *leap-frog algorithm* is another variation that is relatively computationally efficient and implemented in the GROMACS MD package [van der Spoel et al., 2010], which is used throughout this thesis. Here, particle velocities are computed for $t + \frac{1}{2}\Delta t$:

$$\mathbf{v}(t+\frac{1}{2}\Delta t) = \mathbf{v}(t-\frac{1}{2}\Delta t) + \mathbf{a}(t)\Delta t\,. \tag{2.2.2}$$

This future velocity is then used to compute the particle's position at integer time-steps

$$\mathbf{r}(t+\Delta t) = \mathbf{r}(t) + v(t+\frac{1}{2}\Delta t)\Delta t\,. \tag{2.2.3}$$

Time step used in this thesis

In AT MD simulations, a 2 fs timestep was used, while the corresponding time step in CG simulations was 20 fs.

The importance of statistics in MD: the sampling problem

The numerical integration schemes described above show that MD simulations are deterministic. Results are therefore highly sensitive to initial setup. Since MD does not sample all of state space, variations to otherwise similar initial conditions cause changes to the values of system properties [Elofsson and Nilsson, 1993, Caves et al., 1998, Hess, 2002]. This is referred to as the "sampling problem", and refers to the fact that a simulation may, by chance, show an outlier result compared to the underlying population distribution; the result is a so-called *rare event*. Longer and more simulations sample more of state space, and while convergence of simulation values is hard, if not impossible to obtain at present [Hess, 2002] (but see [Hall et al., 2014]), replicate simulations reduce the risk that the mean is biased by outliers, which promises greater reproducibility. In chapter 4, I describe a software that I developed in order to make automated statistical analysis easier across replicate simulations.

2.2.2. Periodic boundary conditions

To avoid edge effects, it is desirable to simulate an infinite rather than finite system. This is most commonly done by implementing periodic boundary conditions (PBC), whereby the box experiences copies of itself on each side (see Fig. 2.2.1). In this setup, a particle that leaves the box at one side is reinserted into the opposite side with an intact velocity vector. Thus, the box has no spatial limits, which has consequences for e.g. neighbour lists (see below), which, at the edge of the box are populated by particles located both at that edge and the opposite edge. All other interactions are similarly treated. The system size should be sufficiently large to disallow a particle to 'see itself'. In practice, replicates should be separated by a distance beyond the LJ and electrostatic cut-off. Similarly, a system needs to be large enough to avoid propagation of any internal correlations, or such correlations can be artificially amplified, and the result trajectory erroneous. If these considerations are met, a PBC system is likely to be a good model of liquid bulk phase behaviour [van der Spoel et al., 2010].

2.2.3. The neighbour list

In order to reduce the number of non-bonded interaction calculations made per trajectory step, the box is divided up into units with side greater than the applied LJ cutoff. The identities of particles

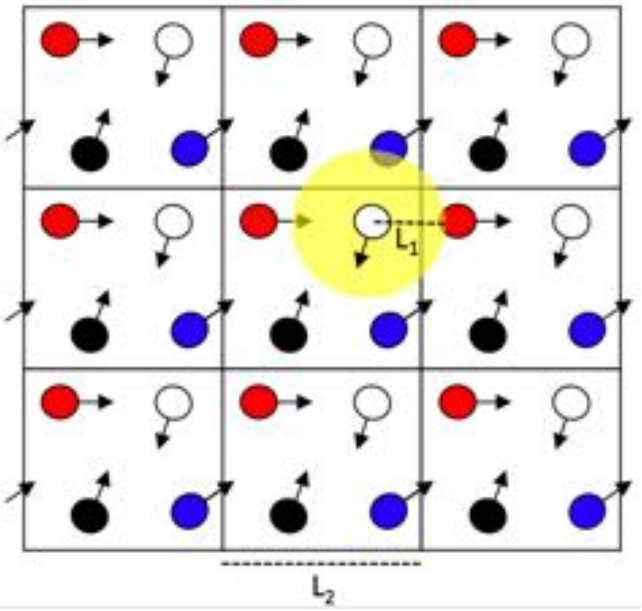

Figure 2.2.1.: Periodic boundary conditions in 2D
Showing four particles in a box with PBC. The system is effectively replicated infinitely to avoid edge effects. The box size L_2 must be greater than twice the longest interaction cutoff between contained particles, L_1 , so that a particle never interacts with itself.

located in each such system subdivision are retrieved, and so-called 'neighbour lists' are created for each substituent particle by only evaluating separation to particles from its own and immediately adjacent boxes. Thus, pair-wise separation evaluation to all system particles is effectively avoided. Neighbour lists are updated every 10 to 20 time steps, and the rationale for this is to strike a balance between computational efficiency and avoiding the scenario where particles have time to move into adjacent areas where they contribute to another particle's LJ potential [Leach, 2001, van der Spoel et al., 2010].

2.2.4. Ensemble choice

In order to compute macroscopically observable variables, such as temperature and pressure, from MD system properties, the MD system must be at statistical equilibrium and be set up to obey a statistical ensemble. All MD systems in this thesis obey the Canonical ensemble, where the number of particles, system volume and temperature are constant.

2.2.5. Temperature and pressure control

While energy should, in principle, be conserved in an isolated system such as the MD setup, integration errors can cause energy fluctuations. Further, spontaneous interconversions of potential and kinetic energy can cause temperature fluctuations in the system as a whole [Leach, 2001]. To avoid this, MD engines employ thermo- and barostat algorithms that sacrifice constant energy and volume in order to control temperature and pressure, respectively.

Thermostat

In a system where the number of particles (N), volume (V) and temperature (T) are constant, as is the case in a Canonical ensemble, the equipartition law of statistical mechanics relates kinetic energy, K, to macroscopic heat and individual system particle velocities v_i as follows

$$K_{NVT} = \frac{1}{2}\sum_{i=1}^{N} m_i v_i^2 = \frac{3}{2} N k_B T \,. \tag{2.2.4}$$

While k_B (Boltzmann's constant), N and m_i are constants, T and v_i are variables that allow change in velocity to be reflected in temperature. This equation can be rearranged to give T ,

$$T = \frac{1}{2}\sum_{i=1}^{N} \frac{2}{3} \frac{m_i {v_i}^2}{N k_B} \,.$$

To correct T to a requested value, T_{new}, the change to T required is

$$\Delta T = T_{new} - T(t).$$

Velocities are subsequently multiplied by the scaling factor λ to achieve a change to the temperature at time t, $T(t)$:

$$\triangle T = \frac{1}{2}\sum_{i=1}^{N} \frac{2}{3} \frac{m_i (\lambda v_i)^2}{N k_B} - \frac{1}{2}\sum_{i=1}^{N} \frac{2}{3} \frac{m_i {v_i}^2}{N k_B} = \lambda^2 T(t) - T(t) \,,$$

and rearranging to give λ :

$$\lambda = \sqrt{\frac{T(t) + \triangle T}{T(t)}} = \sqrt{\frac{T_{new}}{T(t)}} \,.$$

Effectively, however, a rapid shift in system temperature in a single step can de detrimental to the system. The Berendsen thermostat resolves this by scaling velocities gradually as though the system was in contact with an external bath at the requested temperature, T_{bath} [Berendsen et al., 1984]. The rate at which T_{bath} is approached is proportional to the difference in temperature between system and bath, and is controlled by the parameter τ as follows:

$$\frac{dT(t)}{dt} = \frac{1}{\tau}(T_{bath} - T(t))$$

$$\Delta T = \frac{\delta t}{\tau}(T_{bath} - T(t))$$

$$\lambda^2 = 1 + \frac{\delta t}{\tau}\left(\frac{T_{bath}}{T(t)} - 1\right) \,.$$

In the extreme case, $\tau = \delta t$, in which case the Berendsen algorithm simply implements the single step, rapid temperature change that it was implemented to avoid. GROMACS recommends that τ is at least ten times the chosen time step, or velocity scaling would have too much influence on MD dynamics [van der Spoel et al., 2010].

Stochastic fluctuations in kinetic energy are necessary to generate a proper canonical ensemble, but are suppressed by the Berendsen thermostat. In AT equilibrium simulations, I therefore employed the velocity-rescaling thermostat, which adds a stochastic term to the Berendsen formula to ensure a correct kinetic energy distribution [Bussi et al., 2007].

Alternative thermostat algorithms exist, such as the Langevin and Nosé-Hoover methods [Leach, 2001], however the Berendsen thermostat is used throughout this thesis.

Temperature coupling groups Energy does not exchange perfectly between system parts, which commonly results in water heating up. Conservation of energy deems that such water heating is done at the expense of the adjacent area, often protein, which thus cools down in response [Lingenheil et al., 2008, van der Spoel et al., 2010]. In this thesis, this is avoided by coupling protein, lipid and solvent to independent thermostat baths.

Barostat

The theorem of Clausius links particle position to system pressure using the virial W. The virial function states that

$$W(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = \sum_{i=1}^{N} \mathbf{r}_i \cdot \mathbf{F}_i$$

i.e. the virial is the sum of all forces acting on system particles in the direction of particle movement. If F is converted to it's time derivative according to Newton's laws, the time average of W across a trajectory is written as

$$\langle W \rangle = \lim_{t\to\infty} \frac{1}{t} \int_0^t d\tau \sum_{i=1}^{N} \mathbf{r}_i(\tau) \cdot m_i \mathbf{a}_i(\tau) \,.$$

If integrated by parts, this returns

$$\langle W \rangle = -\lim_{t\to\infty} \frac{1}{t} \int_0^t d\tau \sum_{i=1}^{N} m_i |\mathbf{v}_i(\tau)|^2 \,,$$

which equals the negative of twice the kinetic energy (c.f. equation 2.2.4). By extension,

$$\langle W \rangle = -3Nk_BT \,. \tag{2.2.5}$$

In the ideal gas law, where particle interactions are prohibited, temperature, container volume and pressure ($\frac{\mathbf{F}_\perp}{area}$) are related according to

$$PV = \frac{\mathbf{F}_\perp}{area} V = Nk_BT \,.$$

Inserting this into Equation 2.2.5 and also accounting for particle interactions $v(r)$ gives

$$\langle W \rangle = -3Nk_BT = -3PV + v(r) = -3PV + \sum_{i=1}^{N} \sum_{j=i+1}^{N} r_{ij} \frac{dv(r_{ij})}{dr_{ij}} \,.$$

This may be rearranged to give P as follows:

$$P = \frac{Nk_BT}{V} + \frac{v(r)}{3V} = \frac{1}{V}\left[Nk_BT + \frac{1}{3}v(r)\right]$$

and since $\mathbf{F} = -\nabla E_p$, the inter-atomic force equivalent $\mathbf{f}$ may be incorporated as follows:

$$P = \frac{1}{V}\left[Nk_BT + \frac{1}{3}\sum_{i=1}^{N}\sum_{j=i+1}^{N} r_{ij}\mathbf{f}_{ij}\right] \,.$$

Therefore, similarly to how system temperature was modulated by editing particle velocity, system pressure can be tuned by scaling particle separation (and thus system size). In a manner analogous to the temperature bath used for the Berendsen thermostat, the Berendsen barostat achieves a requested pressure in increments with respect to a connected 'pressure bath' [Berendsen et al., 1984], where the scaling factor, that particles' separation is multiplied by, equals

$$\lambda = 1 - \kappa \frac{\delta t}{\tau_P}(P - P_{bath}).$$

New particle positions are then given by

$$\mathbf{r}_i{}' = \lambda^{1/3}\mathbf{r}_i\,.$$

Here, $\mathbf{r}_i{}'$ are re-scaled particle coordinates, κ is the isothermal compressibility and τ_P is the pressure coupling time constant. τ_P is related to system volume fluctuations as a function of pressure and temperature as follows [Leach, 2001]:

$$\kappa = -\frac{1}{V}\left(\frac{\partial V}{\partial P}\right)_T\,.$$

I use an isotropic compressibility of $5 \times 10^6\, bar^{-1}$ for equilibrium CG systems, which is recommended for the Martini force field [Monticelli et al., 2008]. For AT systems the corresponding compressibility value used is $4.6 \times 10^5\, bar^{-1}$, which is correct for most liquids at 300 K [van der Spoel et al., 2010].

All system dimensions need not be coupled to the same pressure bath. This is commonly used in bilayer simulations, where dimensions in the plane of the bilayer (x and y) are re-scaled together while the membrane normal axis is scaled independently. Such *semi-isotropic* pressure coupling aids membrane self-assembly and establishment of an accurate pressure profile.

Alternative pressure coupling algorithms exist, such as the Parrinello-Rahman barostat [Parrinello and Rahman, 1981].

2.2.6. Bond length control

The fastest motions in an MD system are the vibrations about covalent bonds. These limit the magnitude of the time step. If a too high time step is utilised, a particle at high velocity has sufficient time to move into an area where it has high potential energy, causing the system to subsequently explode. Since covalent bonds do not have the capacity to break or reform in an MD setup, their exact length is of little interest to the evolution of the system. For reasons of computational speed-up, covalent bonds are therefore often constrained to their equilibrium lengths.

The GROMACS MD package implements three constraint algorithms of increasing complexity: SHAKE, LINCS and SETTLE [van der Spoel et al., 2010]. LINCS is used throughout this thesis. In the LINCS method, bonds are first updated as if unconstrained, then the bond's vector length is rescaled to within tolerance levels [Hess et al., 1997].

2.3. Software and parameter choices that approximate experimental conditions for MscS

All MD simulations in this thesis were performed in the GROMACS MD package, version 4.5.5 [Berendsen et al., 1995, Hess et al., 2008].

Pressure and temperature

E.coli has an internal hydrostatic pressure of 3 atm. However, all bacterial patch-clamp experiments are done on spheroplasts; bacteria treated with penicillin to break down their cell walls, since a patch clamp does not fit on a small bacterial cell body [Naismith and Booth, 2012]. Once the cell wall is removed, bacteria swell since they can no longer sustain the 3 atm pressure potential with respect to the outside. To produce comparable simulation and experimental results, MscS was therefore simulated at 1 atm.

In all simulations, the temperature was set to 310 K; equivalent to 36.85° C, approximate body temperature.

Membrane composition

Simulations were performed using an approximately physiological *E.coli* membrane, composed of 4:1 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE):
1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoglycerol (POPG). As described in 2.3, experimental conditions remove the outer membrane prior to electrophysiology, so only the inner bilayer was modeled.

Membrane surface tension

Even though a bacterium undergoes significant volume expansion upon loosing its cell wall during spheroplast production, experimentalists assume that the membrane returns to equilibrium surface tension, and that ensuing patch clamp pressure is the only factor affecting membrane tension. The equilibrium surface tension was therefore set to 0.

Ionic environment

The MscS MD system carries an inherent charge due to the +42 charge of the TM domain, as well as negatively charged POPG lipid. These charges were neutralised to avoid artifacts due to long-range electrostatics [Essmann et al., 1995]. In addition, 150 mM NaCl was added to approximate the physiological internal bacterial environment [Fagerbakke et al., 1999].

2.4. Data visualization

All molecular visualization was done in Visual Molecular Dynamics (VMD) 1.9.1 [Humphrey et al., 1996]. Graphs are plotted using either Microsoft Excel or VMD, or are custom drawn using graphics primitives in the OpenGL package which VMD employs for display purposes.

3. Analysis of helix distortion in membrane proteins

3.1. Introduction

3.1.1. Prevalence of helix distortion, and its origins

α-Helices have a variety of roles in proteins, including structural, recognition, gating and in DNA-binding motifs such as leucine zippers and zinc fingers. Helices are particularly common in transmembrane regions of proteins since they satisfy all backbone hydrogen-bonds internally [Picot et al., 1994, Kuhlbrandt et al., 1994, Pebay-Peyroula et al., 1997, Walz et al., 1997, Cheng et al., 1997, Bass et al., 2002]. Many such helices deviate from the linear standard Pauling-Corey model of helix structure [Pauling et al., 1951] by featuring curves or kinks [Kumar and Bansai, 1998, Riek et al., 2001]. Such "non-ideal" helices can allow the protein to avoid steric clashes under conditions of close packing, or maximisation of the number of H-bonds between solvent and the carbonyl of helix backbone amide groups in the case of solvent-exposed helices [SWISS-MODEL, 2011]. Curvature can also be an energetic optimisation response to hydrophobic mismatch with respect to the lipid bilayer environment [Yeagle et al., 2007]. Indeed, research suggests that backbone hydrogen bonds can shift partners to accommodate TM helix bends at minimal energy cost (< 1kcal/mol) [Cao and Bowie, 2012]. Sequence-generated distortion is also observed; the cyclic side-chain of a proline residue breaks two hydrogen bonds if located within a helix, which cause a 20° kink (unless the protein overcomes the energetic cost of straightening this kink) [MacArthur and Thornton, 1991]. Glycine, on the other hand, is sufficiently small to not resist helix distortion, and is seen e.g. in the pore-lining helix of MscS [Bass et al., 2002].

3.1.2. Physiological relevance of helix distortion

α-helix flexibility about this Glycine is implicated in the gating mechanism of MscS [Akitake et al., 2007]. MscS is not the only membrane protein to make use of helix structure plasticity; apart from channel gating, it is also known to enable membrane protein functionality such as receptor activation and molecule transport [Sansom and Weinstein, 2000]. Examples can be seen in the voltage-gated and inward-rectifying potassium channels and in the NaK channel, where molecular hinges within channel-lining helices elicit gating [Jiang et al., 2002, Domene et al., 2005, Shi et al., 2006, Alam and Jiang, 2009, Choe and Grabe, 2009]. In membrane proteins with hydrophilic interiors, interior-facing prolines allow helices to narrow a passage through the membrane [MacArthur and Thornton, 1991], and helix curvature can also more generally ensure structural integrity by allowing close packing [Ceruso and Weinstein, 2002].

3.1.3. Need for characterization by visual and quantitative means

Visualization As protein structure obtaining technology develops, an increasing number of structures of increasing complexity are being deposited into the Protein Data Bank (PDB) [PDB, 2014a]. The three dimensional structure of a protein is critically involved in its function, so a researchers' ability to comprehend it is crucial. Even a moderate number of particles produce a visual clutter that is hard to decipher, so necessitates visual abstraction (c.f. Figure 3.1.1). The situation is even more complicated in dynamic MD trajectories of protein structure behaviour, and the possible duration of

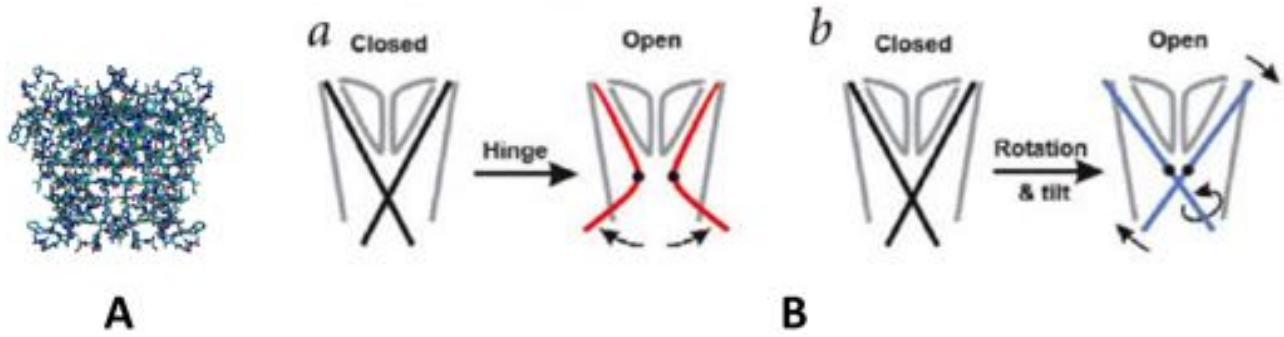

Figure 3.1.1.: Use of visual abstraction: example of KscA gating
Protein structure is commonly heavily abstracted in order to convey ideas regarding conformational change. Comparison of the original resolved protein structure of KcsA, shown in stick representation with atom-type coloration (**A**) to gating concepts developed for KscA, where a high degree of abstraction is used (**B**). **Ba** and **Bb** show alternative gating mechanisms.

such simulations (and therefore complexity) is increasing with progress in technology [Klepeis et al., 2009, Dror et al., 2012].

Good structural abstraction provides researchers with intuition for molecule behaviour by facilitating e.g. identification of the topological relationship between secondary structure elements, but only if such abstractions do not sacrifice structures that are related to function, such as helix distortions. Two helix abstractions are widely used; rigid cylinders between the first and last α-carbon in the helix, and ribbon interpolation through the helix backbone [Richardson, 1981, Carson and Bugg, 1986]. Considerations of helix flexibility are compromised by their classical representation as idealized cylinders. While one can view helix distortion using ribbon representations, this detailed view of the protein backbone entails much complexity, which can complicate structural comprehension in anything but smallproteins. For coarse-grained proteins, the situation is worse still; no current molecular graphics viewer has built-in support for CG protein secondary structure visualization or even CG particle connectivity. While command-line based scripts were developed during the course of this thesis, representations are rudimentary and offer little possibility for customization.

Analysis MD is used to explore transmembrane α-helices in an increasing number of studies [Sessions et al., 1998, Karplus and McCammon, 2002, Spronk et al., 2006, Bond et al., 2008, Choe and Grabe, 2009, Deplazes et al., 2012]. However, given the importance of helix distortions, improved methods for analysis of helix dynamics are needed. Ideally such applications should incorporate intuitive visual analytics so that helix geometry becomes a more fundamental metric that is easier to retrieve.

3.1.4. Bendix meets this demand

The prevalence and importance of kinked helices calls for improved helix abstraction and intuitive analysis tools. Furthermore, it is important to make such developments available to the wider community by incorporation into existing, publicly adopted visualization software [Goddard and Ferrin, 2007]. We have therefore developed Bendix, a program that meets these demands.

Bendix features custom helix representation and geometry indicative graphics that assist analysis of both static structures and complex molecular dynamics simulation trajectories. The software is a multi-platform and free plugin for the software Visual Molecular Dynamics (VMD). With its multiple settings, the Bendix output is easily tailored, and integrates well with the rest of VMD, making it possible to combine Bendix with existing visuals and analytics.

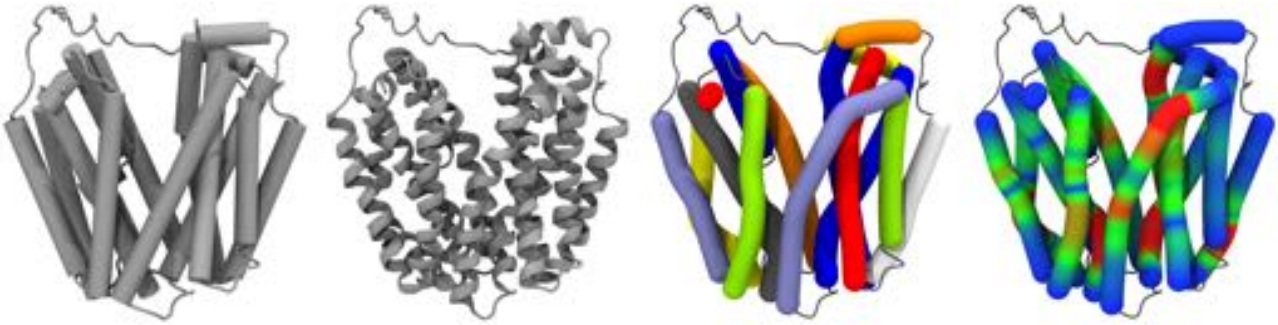

Figure 3.1.2.: Comparison of the Bendix default helix representations to classical abstractions
Left to right: The lactose transporter LacY represented as cylinders, ribbons, Bendix's uniformly coloured bendices, bendices using angle-indicative colouring. While the classical representations of helices can be coloured individually in order to increase inter-helix contrast, this is not by supported by default.

3.1.5. Chapter overview

The chapter starts with details of thetechnical implementation of Bendix, including choices made for computing the helix axis. Then I discuss the development of both the front- and backend of the program, and the usability issues that were tackled. Next, I describe the available analyses and capacity for customization, in particular support for non-AT proteins, and options for results visualization and export. Last I apply the Bendix functionality to analyse both crystal structures and MD simulation data of the transporter Mhp1.

3.2. Methods

3.2.1. Software implementation

Bendix is written in Tcl with a user interface implemented in Tcl's graphical user interface extension, the Tk toolkit [Welch and Jones, 2003]. This makes the functionality of Bendix platform-independent, the user interface adopts the native look and feel of the user's operating system, making it easier to use.

3.2.2. Use of the VMD plug-in API

Bendix is distributed as a plugin for the widely used and freely available molecular graphics software Visual Molecular Dynamics (VMD; http://www.ks.uiuc.edu/Research/vmd/) [Humphrey et al., 1996]. As a VMD plugin, Bendix uses in-built VMD features such as STRIDE's secondary structure prediction for AT proteins, rendering options and VMD's flexible syntax for selecting subsets of atoms for Bendix characterization. All structure and trajectory file formats supported by VMD can be analysed in bendix.

The VMD platform uses the OpenGL application programming interface (API) for graphics. OpenGL allows real-time, interactive 3D rotation, translation, cropping and zooming, which is important for depth perception and to gauge spatial relationships [Humphrey et al., 1996]. The API also allows custom graphics to be drawn, which is used extensively in bendix.

Bendix will ship as part of the next VMD release, version 1.9.2.

3.2.3. Helix axis computation

The Bendix helix axis generation is a Tcl implementation of the C language algorithm used by Dalton [Dalton et al., 2003]. This algorithm is also used in another program that computes helix geometry,

HELANAL [Bansal et al., 2000], and was originally presented by Sugeta and Miyazawa [1967]. This algorithm calculates a helix axis coordinate using vector algebra on four consecutive backbone alpha carbons. This window slides along the length of the helix in steps of one alpha carbon. The computed axis is robust to changes to the number of residues per helix turn.

3.2.3.1. Algorithm

See Figure 3.2.1 for an illustration of the concepts below.

Helix axis

1. Retrieve vector coordinates for four alpha carbons, $\mathbf{C}_1 - \mathbf{C}_4$.
2. Compute the three vectors between adjacent Cα, i.e. $\mathbf{B} = \mathbf{C}_{N+1} - \mathbf{C}_N$.
3. Compute the two vectors between the adjacent $\mathbf{B}$ vectors computed in step 2, i.e $\mathbf{P} = \mathbf{B}_N - \mathbf{B}_{N+1}$. These two vectors, $\mathbf{P}_1$ and $\mathbf{P}_2$, are perpendicular to the helix axis.
4. Get the average helix radius at the site of the four input alpha carbons:
   a) Compute magnitudes for $\mathbf{P}_1$ and $\mathbf{P}_2$: $|\mathbf{P}_1|$ and $|\mathbf{P}_2|$
   b) Compute the dot product between $\mathbf{P}_1$ and $\mathbf{P}_2$
   c) Get the angle between $\mathbf{P}_1$ and $\mathbf{P}_2$: $\cos\theta = \frac{\mathbf{P}_1 \bullet \mathbf{P}_2}{|\mathbf{P}_1||\mathbf{P}_2|}$
   d) Then radius, $r_{helix} = \frac{\sqrt{|\mathbf{P}_1||\mathbf{P}_2|}}{2(1-\cos\theta)}$, i.e. the average axis-perpendicular vector length divided by a normalisation factor based on the angle between adjacent residues.
      i. An angle of 60° gives a denominator of 1, while larger angles (e.g. 100° in the standard Pauling-Corey α-helix [Pauling et al., 1951] and 120° in the 3_{10}-helix) increase the denominator, thus reducing the local helix radius. This is to be expected as fewer residues per helix turn result in higher angles between adjacent residues and a smaller helix radius.
5. Retrieve a local helix axis coordinate
   a) Get the unit vector for $\mathbf{P}_1$; $\hat{\mathbf{P}}_1$.
   b) Scale $\hat{\mathbf{P}}_1$ to the size of the local helix radius
   c) Reduce $\mathbf{C}_2$ by $\hat{\mathbf{P}}_1$ to approximate the helix centre at the level of $\mathbf{C}_2$.
6. Advance the analysis window by one Cα and repeat steps 1 to 5.

Special circumstances at the ends of helices The above algorithm generates a helix axis that is three residues short of the original helix length; one at the N-terminus and two at the C-terminus. To remedy this, Bendix computes the local helix vector at the ends of the helix, and extends the helix by one residue's worth in each direction. The extension vector was optimised for different helix sizes, evaluated by the number of helix axis coordinates computed for the current helix. For the C-terminus, if the number of collected axis points is:

1: Retrieve a second local helix axis coordinate:

1. Get the unit vector for $\mathbf{P}_2$; $\hat{\mathbf{P}}_2$.
2. Scale $\hat{\mathbf{P}}_2$ to the size of the local helix radius
3. Reduce $\mathbf{C}_3$ by $\hat{\mathbf{P}}_2$ to approximate the helix centre at the level of $\mathbf{C}_3$.

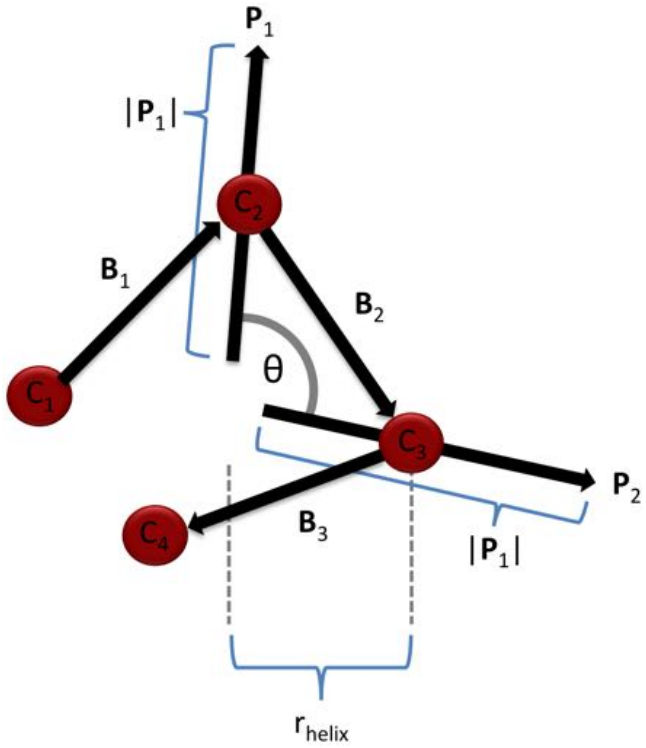

Figure 3.2.1.: Algorithm for helix axis detection
The Sugeta and Miyazawa [1967] polymer axis retrieval algorithm applied to an approximately α-helical example. Red nodes denote Cα-atoms and black arrows show vectors as described in the main text.

2: Compute the vector between the last and next-to-last axis coordinate. Add this vector to the last axis coordinate.

>2: Compute the average vector and single-residue helix rise magnitude between the last three axis coordinates. Scale the vector down to the averaged helix rise magnitude. Add this vector to the last axis coordinate.

If the helix has > 4 residues, the helix is then extended by one residue's worth of length at both the N- and C-termini. There are now ≥ 2 helix axis coordinates available, and extensions are made based on the most peripheral two or three axis coordinates, as described above.

3.2.3.2. Hermite spline implementation

A spline inputs coordinates, *control points*, and generates an averaged, smooth curve through them. The Hermite spline moves through input coordinates (the helix axis, in this case) using a four-point window, generating *spline points* that form a smooth curve between the middle points, guided by the two most peripheral points. Bendix uses Hermite spline and the Catmull-Rom definition of the gradient with the constant set to 0.5. The number of spline points generated is controlled by the user by the setting Resolution (see Figure 3.3.4).

As the curve is only generated between the middle-most coordinates in the 4-point window, spline application shortens the helix length. To counter this issue, I generate *Phantom knots*, artificial helix axis extensions, prior to spline-calculation. This way the helix axis length is preserved.

Custom control point interval By implementing helix extensions based on the local vector and phantom knots, I ensure that there is a one-to-one correspondence between helix alpha carbons and the final list of spline control points. The user chooses how many of these control points to use

when generating the final spline by setting “Use every [N] residues” (see Figure 3.3.4). An interval of 4 residues causes the result spline to follow the helix axis less closely than it would using every 2 residues, but generates a smoother, more global approximation of the axis.

Preserved helix end-points The most peripheral axis points at the helix N- and C-terminus are always preserved, irrespective of the interval chosen by the user. The exact number of spline points generated at the N-terminus is normalised to the separation to the previously chosen control point.

3.2.3.3. Helix axis angles

The angle along the helix axis is evaluated using a triangle with side length equal to the spline control point interval. The unit for angle side length is therefore number of alpha carbons.

Angle-values for helix ends are standardised Angles can not be evaluated for axis points that are located less than one angle side’s distance from a helix end. Instead, the nearest angle that it is possible to compute is evaluated, and angle measurements along helix ends are made to decrease linearly to 0 at the tip. These angle values should be ignored, and were only implemented to orient the user by denoting the real length of the helix.

3.3. Results

3.3.1. Software architecture

bendix.tcl is 6,300 lines long (circa 150 A4 pages of code) and consists of 39 functions (procedures). A flow-chart overview of the Bendix algorithm can be seen in figure 3.3.1.

3.3.1.1. Code structure

Procedures are grouped according to utility into the following sections:

- Graphical User Interface setup and maintenance
- Load and save files
- User help and acknowledgements
- Draw non-bendices
- 2D and 3D helix analysis (whereof scatter plots are generated by Multiplot)
- Trajectory and molecule ID detection
- Particle and subset validation
- Informative error messages
- Hide, erase and reset visuals
- Main bendix procedure:
  - Input verification
  - Obtain user-assigned or automated helicity (latter by STRIDE or input file text-mining)
  - Coordinate retrieval
  - Generate helix axis

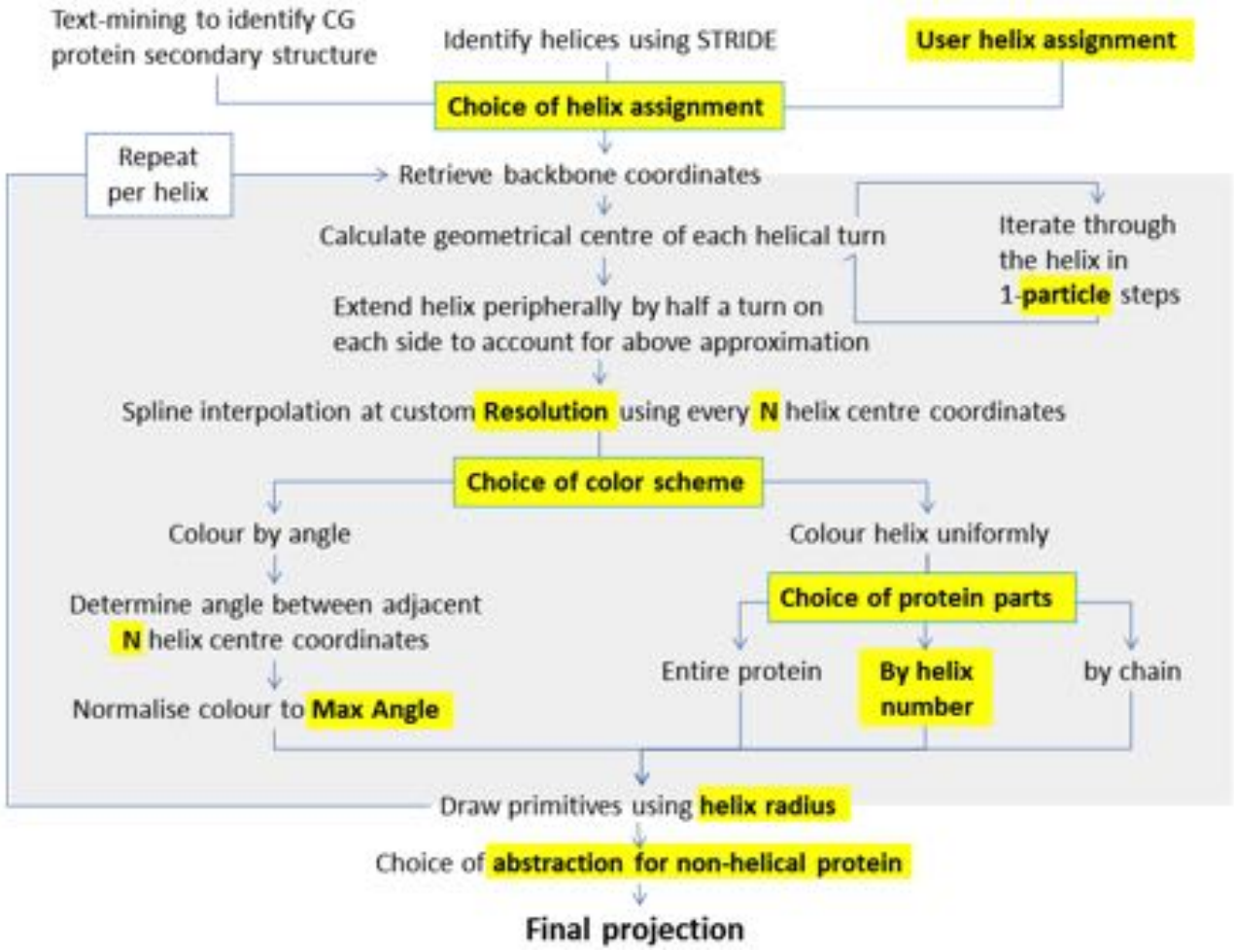

Figure 3.3.1.: Flow-chart overview of the Bendix algorithm.
Blue arrows show the direction of the algorithm. Highlighted text in bold show settings that are customizable from the Bendix GUI (see Figure 3.3.4). The gray central box shows the part of the algorithm that is run for each detected helix.

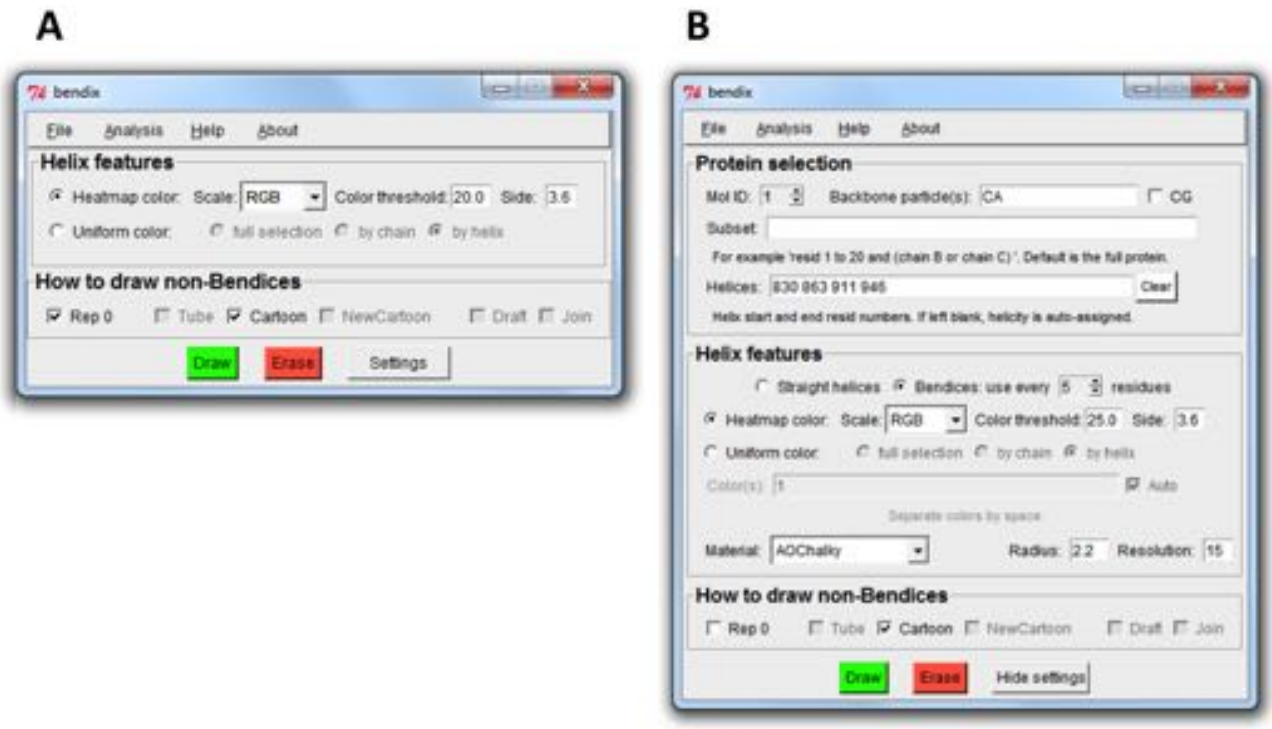

Figure 3.3.2.: Slim and expanded GUI views
A. The slim startup-view, optimised for smaller screens **B**. The expanded view with settings, available by pushing the bottom right “Settings” button in **A**.

- Extend the axis peripherally to reclaim length lost to the axis generating algorithm
- Apply Hermite spline to the axis according to user settings
- Retrieve user settings for screen output (heatmap, material, resolution et.c.)
- Draw to screen
- Calculate and store data for 2D and 3D helix analysis
- Reset variables to enable rerun

The code is modulated to speed up operation and avoid unnecessary calculations. Where possible, protein atom indices are stored for fast coordinate retrieval while drawing trajectories to the screen.

3.3.2. Graphical User Interface

Bendix comes with a user-friendly graphical user interface (GUI) which offers customizable settings, e.g. resolution, helix diameter, angle side length, helix assignment and choice of uniform or angle-indicative graphics. The main GUIs can be seen in Figure 3.3.2.

Usability was a key concern during bendix development, and therefore the GUI evolved as I received feedback from users and novel features were added. In order to create a compact interface, the front-end underwent major reconstruction to hide analysis tools, export and import functions in a menubar. The 2D and 3D graph manipulation interface was also moved to its own separate window, only made visible when relevant. The current implementation features the slim view, which allows access to the most commonly used features, with the option to extend the application window to access full settings. These two alternate views can be seen in Figure 3.3.2.

3.3.3. Available analyses

Refer to Figure 3.3.4 for an overview of how to customize the Bendix analysis.

3.3.3.1. Helix geometry

Each helix is analysed for geometry, defined as the local angle of the helix axis with respect to adjacent axis coordinates. The helix resolution; the number of helix axis points that are evaluated for geometry (see *Methods* for details), is set by the user.

Helix geometry data are visualised both qualitatively and quantitatively. Qualitatively, the angle along the helix is computed and displayed using visual analytics, directly onto the helix area in question, using heatmap colour-coding. This is facilitated by the helix abstraction of Bendix, which follows the helix axis; see Figure 3.3.3 **C** for an illustration. If quantitative data are requested, the angle along the length the helix is plotted in 2D graphs where the geometries of multiple helices can be simultaneously compared.

Additionally, 3D graphing is available for dynamic data (MD input with multiple frames). Dynamic data also benefit from a plot of maximum helix angle over time, per helix. Bendix supports analysis of helix geometry for all VMD compatible input, including MD data and static structures, such as crystal structures.

Option to measure local or global helix geometry The helix geometry algorithm uses a triangle with side length in units of residues. A large angle side measures a global angle, whereas a small angle side measures more local angles. The angle side used is the control point interval, set in the user interface under “Bendices: user every [N] residues”.

Retrieval of helix axis coordinates The user can also export the axis points (including N- and C-terminal extensions).

3.3.3.2. Customizable settings

Particle type to use for analysis and visualization Bendix implements a helix axis algorithm that works on any helical polymer. Following feedback from users who wish to use Bendix on inorganic molecules, users can supply a name or regular expression to define what particles to apply Bendix to. This makes it possible to analyse and visualize any helical VMD input.

Helicity Stride [Heinig and Frishman, 2004] is automatically called in VMD to calculate the secondary structure of loaded proteins. However, alternative resources, such as DSSP [Kabsch and Sander, 1983], exist to determine secondary structure, and their algorithms differ, with differential helicity as a result. Bendix users benefit from the ability to edit, export and import helicity. Users simply state which residues are helical by entering the start and stop residue for each helix.

3.3.4. Data visualization

Use of Multiplot for simple graphs VMD’s Multiplot plugin is used to plot maximum angle over time and single frames’ local helix axis angles. Multiplot is a convenient choice as it allows data export and displays data point values in the Terminal if the user hovers over the point with the mouse.

Custom 3D graph viewer for VMD As part of the Bendix software, I built an interactive graph viewer for use in VMD’s OpenGL window. This makes it possible to rotate the graph to get better contrast, or draw attention to parts of the graph. The helix’ curvature magnitude is expressed by a heatmap that is quantified in the Heatmap colour scale bar in the 2D version of the data, while angle data are shown as 3D surface peaks and colouration in the 3D graph (c.f. left and middle images in Figure 3.3.5).

Multiple settings exist to customize the graph, so that additional graphing software is unnecessary. Axis compression is useful for compact data display. This is beneficial to condense data for a long

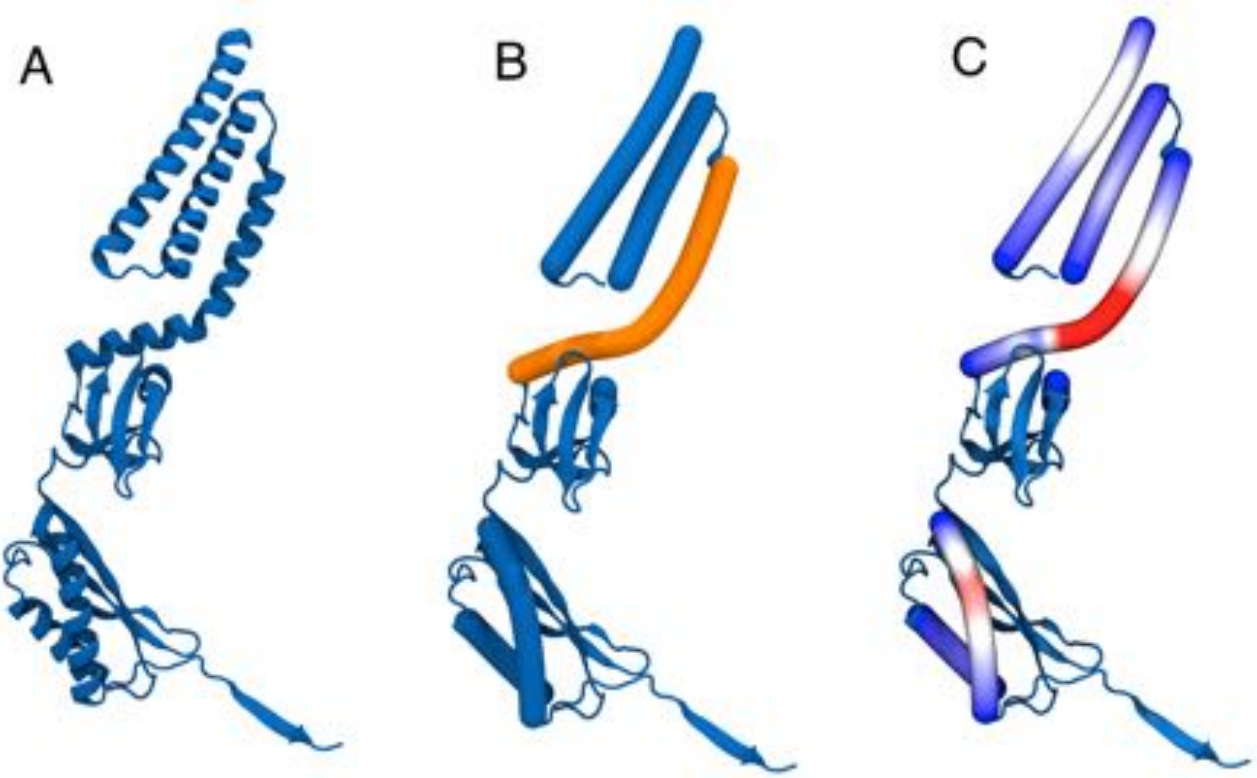

Figure 3.3.3.: Bendix rendition compared to the classical cartoon representation.
A single chain of MscS, with empahsis on the geometry of the pore-lining helix TM3. **A.** Classical cartoon representation **B.** Bendix representation using uniform coloration. TM3 is shown in orange, else in blue. **C.** Bendix representation using helix angle-indicative coloration. Helix angles are indicated using a red-white-blue colourscale, where red indicates high angle and blue indicates low angle.

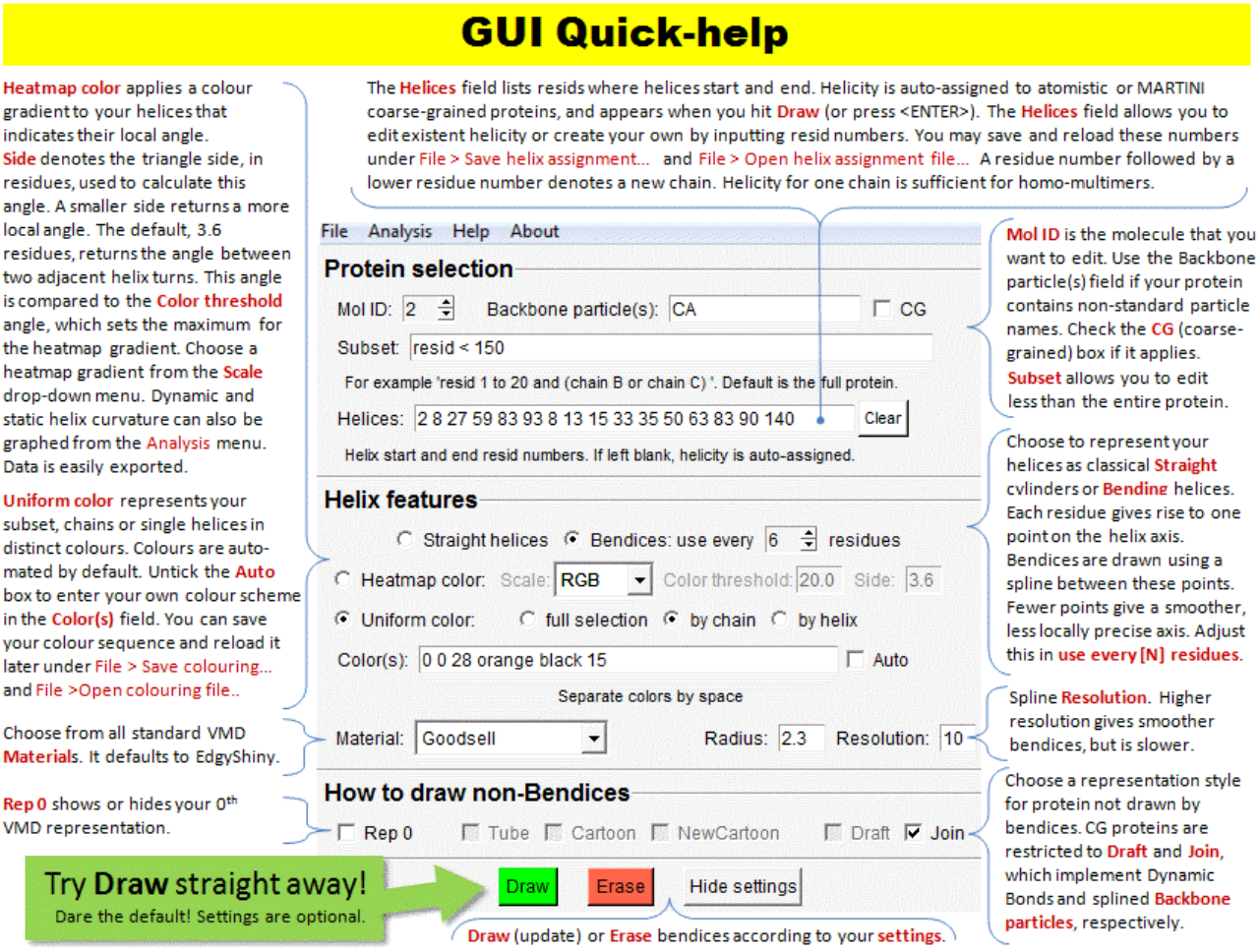

Figure 3.3.4.: Explanation of the Bendix GUI
This explanatory image is available from the Bendix Help menu.

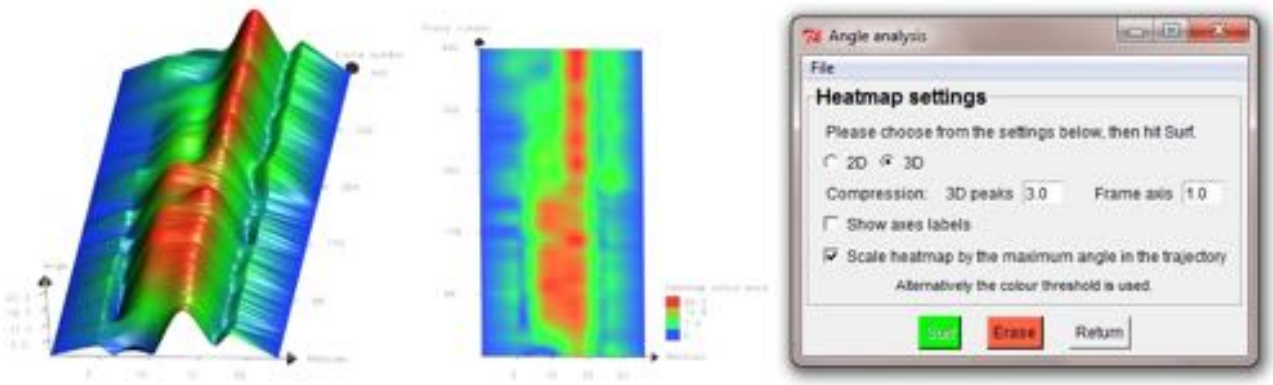

Figure 3.3.5.: The 2D and 3D data visualization unit available in Bendix
Left to right: 3D heatmap visualization, 2D heatmap visualization, the visual analytics GUI.

trajectory into a lucid graph that fits the page, or scale 3D peaks up or down to emphasizing peaks. The heatmap is by default scaled by the maximum angle computed in the trajectory, but can be customized to any value using *Colour threshold* in the Bendix main window.

3.3.5. Secondary structure visualization

3.3.5.1. Helix abstraction

Bendix abstracts protein helices as cylinders that follow the helix axis. This captures conformational information that is lost if you use classical, straight cylinders. While the structural complexity revealed by Ribbon representations do not sacrifice helix distortion, it complicates structural comprehension in anything but small proteins (see Figure 3.3.6). In such situations, bendices help clarify protein structure by reducing the number of visible edges in the image. Helices can be custom visualized using custom radius, a range of materials and uniform colors. The sequence of uniform colours to use can be saved and reloaded from the menu.

3.3.5.2. Full support for coarse-grained proteins

Software does not currently exist for prediction of secondary structure for coarse-grained proteins, and visualization is sparse, often limited to disconnected spheres (as in VMD, seen in the top row of Figure 3.3.7).

Bendix implements CG secondary structure representation. Automated secondary structure representation is available only for Martini coarse-grained protein simulations [Monticelli et al., 2008], where Bendix text-mines the input file for particle types. This only works if the particle type reveals the local secondary structure, which MARTINI coarse-grain particles do. β-sheet and backbone display is also available for Martini CG proteins (see Figure 3.3.7 **C**, **D** and **E**). For all other CG protein types, only bendices and backbone are drawn, and helicity must be submitted by the user.

3.3.5.3. Non-proteinacious input visualization

Bendix supports axis-visualization of any user-assigned particle type. This means that, whether the input structure is proteinacious or not, if part of it is helical, its axis can be visualized. Backbone display (the joining of sequential particles by a spline) is also supported.

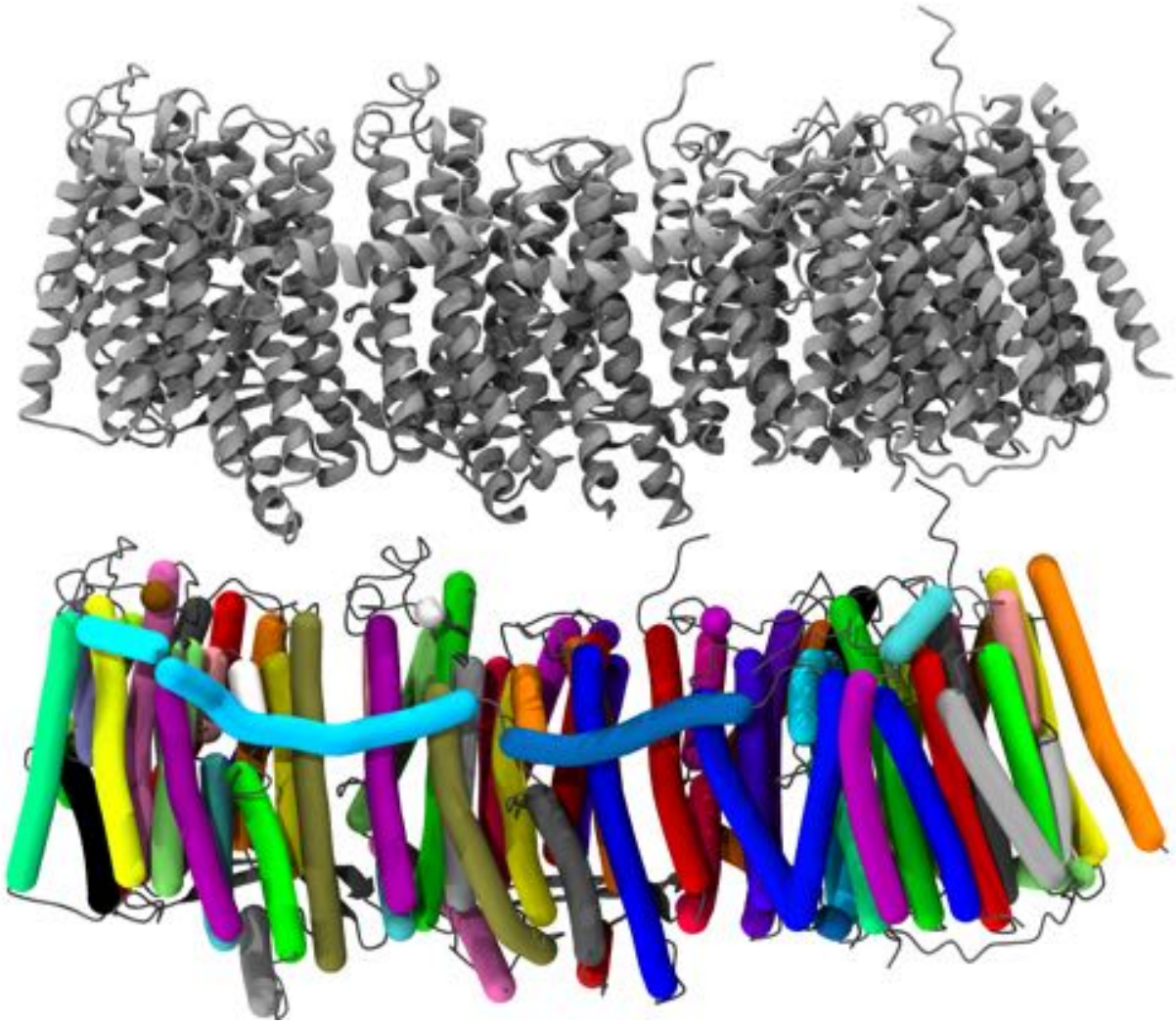

Figure 3.3.6.: Comparison between classical ribbon and default Bendix representation of Respiratory complex 1

Ribbon (top) and uniformly coloured Bendix representation (bottom) of PDB id 3RKO. The respiratory complex is the first and largest enzyme of the respiratory chain, consisting of 64 TM helices [Baradaran et al., 2013].

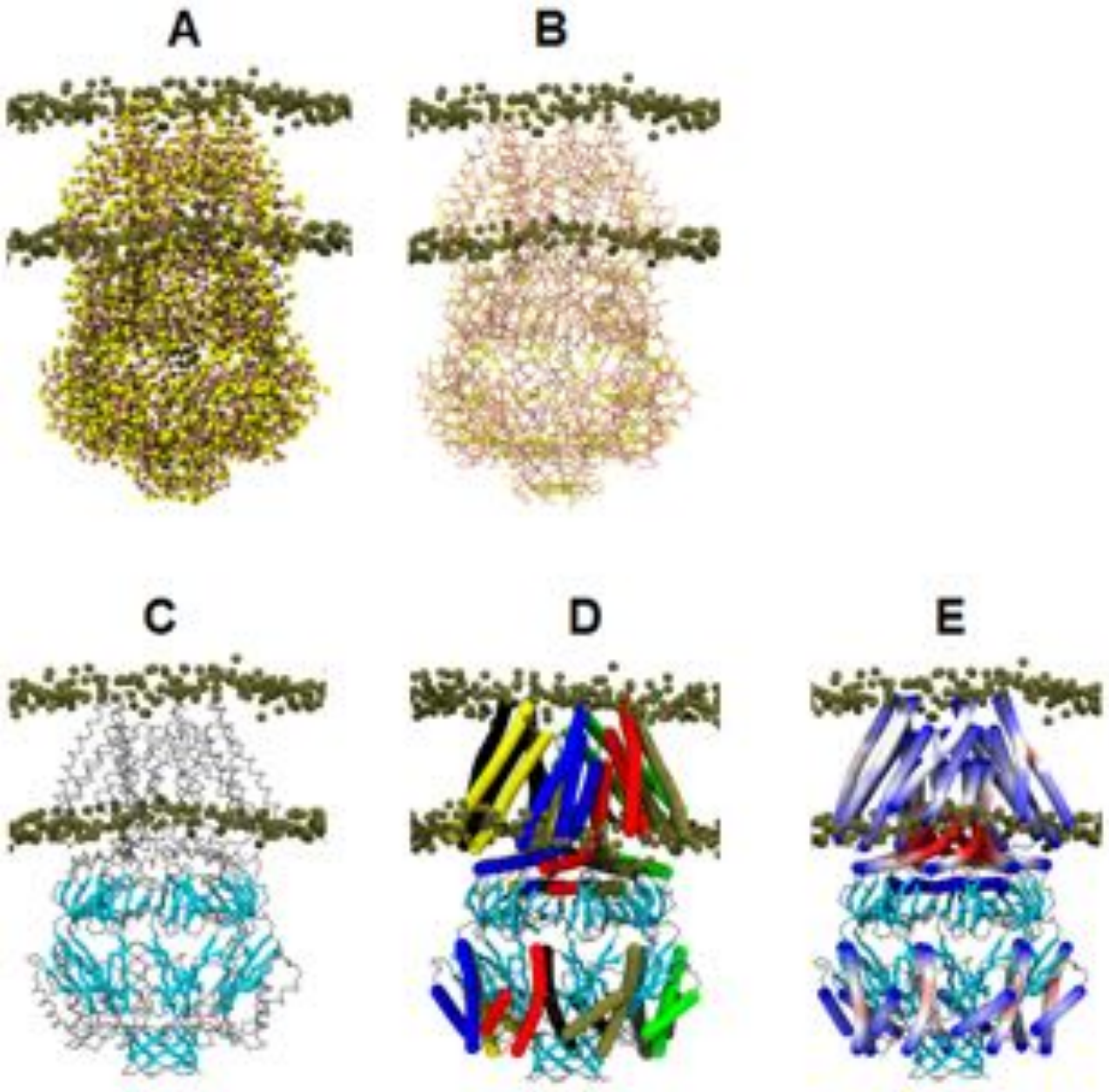

Figure 3.3.7.: Bendix visualization of CG proteins compared to existing visualization.
Top row: Existing visualizations for CG proteins in VMD 1.9.1, showing Martini coarse-grained MscS as an example. **A.** Van der Waal representation. **B.** DynamicBonds, where any atoms that are within a given cut-off are bonded. **Bottom row:** Martini coarse-grained MscS visualized using Bendix. **C.** Backbone representation. Non-beta-sheet alpha-carbons (any particle of the user's choice) are joined by a spline and beta-sheets are represented by arrows to indicate sheet direction. **D.** As in C, but helices are represented as bendices, coloured by chain. **E.** As in D, but helices are coloured according to local geometry.

3.3.6. Data export

All generated data can be exported. Output data written by Bendix is formatted to be compatible with major graph and analytical software such as MatLab, Excel and R. An explanatory note at the top of the file states what rows and columns mean. Multiplot also comes with several in-built export options.

3.3.7. Website

I built a website for Bendix that contains extensive documentation and wherefrom the source code can be downloaded. The index website address is http://sbcb.bioch.ox.ac.uk/Bendix/ and it contains another four sub-pages with the following content: a gallery with example usage of Bendix, an extensive tutorial, technical details including installation instructions, answers to frequently asked questions and contact details. At the time of writing (August 2014), the Bendix website has had 4,400 visits made by 2,800 unique visitors.

3.4. Method evaluation using the Mhp1 transporter

The membrane protein Mhp1 is a symporter which features functional helices whose degree of distortion characterises the transporter state, making it a good protein to evaluate the bendix analysis. Mhp1 moves benzylhydantoin and sodium into the cytosol by coupling the uphill movement of benzylhydantoin with the energetically favorable downhill gradient of sodium [Suzuki and Henderson, 2006].

Three crystal structures exist for Mhp1. They portray three different states in Mhp1's transport mechanism; outward, outward occluded and inward open [Weyand et al., 2008, Shimamura et al., 2010], where an occluded state refers to when substrates are trapped and thus prevented from leaving. Mhp1 belongs to the LeuT superfamily of transporters, which transports substrates by the alternating access model [Abramson and Wright, 2009, Krishnamurthy et al., 2009]. Shimamura et al. [2010] argues that Mhp1's transport mechanism is a four-state process where the bending of 'thin gates' and the rotation of the rigid 'thick gate' determines the transporter's state (see Fig. 3.4.1). TM10 and 5 make up the pericellular and cytosolic thin gates, and TMs 3, 4, 8 and 9 constitute the thick gate, or *hash*. Notably, a structure for the proposed inward occluded state (shown in gray in Fig. 3.4.1) has not been captured.

Using the published crystal structures and MD data kindly provided by Dr Oliver Beckstein, I investigate the gate dynamics of Mhp1 using the Bendix visual analytics in the last part of this chapter.

3.4.1. Methods

3.4.1.1. Data

Crystal structures The following PDB IDs were used: 2JLN (outward open), 2JLO (outward occluded) and 2X79 (inward open).

MD simulations Mhp1 trajectory data were kindly provided by Dr Oliver Beckstein. Dr Beckstein supplied five dynamic importance sampling (DIMS) MD trajectories of the transition from the outwards open state to the inward open state. In addition, he provided equilibrium simulations started from the inwards open conformation. No substrates were present in any simulations.

3.4.1.2. Analyses

Thin gate size The smallest separation between helix and adjacent surface was computed for each thin gate. I reduced COM separation by 3 Å to give the approximate van der Waal surface separation.

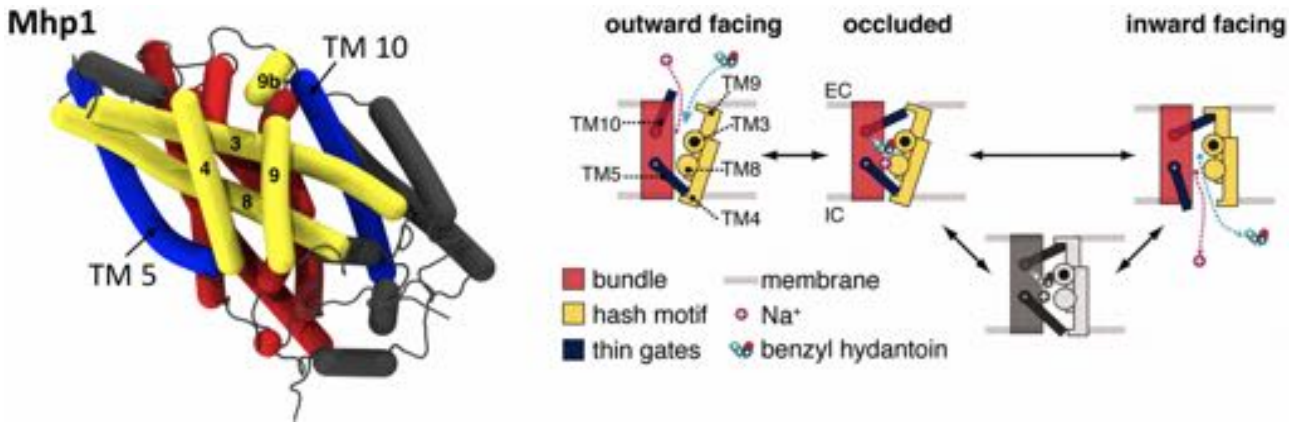

Figure 3.4.1.: Mhp1 structure and proposed gating mechanism
Left: Structure of Mhp1 in outwards open configuration (PDB id 2JLN) in bendix representation. The thin gates, hash and TM9b are labelled. Right: Abstracted transport mechanism proposed by Shimamura et al. [2010]. Colours as in the left-most image.

Bendix analysis Every four residues were used as control point interval, and subsequently the angle measured had a 4-residue side. An angle side with four residue length captures angles between sequential helix turns in a standard α-helix. This avoids measuring angles between adjacent residues (too high resolution), yet the angle is not so global that a substantial amount of helix periphery is sacrificed in order to make the measurement. Angle measurements will only be missing for six residues - three residues at either helix end.

Hash motif angle I scripted a custom VMD plugin that measures the angle between TM4 of the hash and the C-terminal half of TM7 (part of the bundle, shown in red in Fig. 3.4.1). The bendix helix axis algorithm is used to compute the unit vector in 3D for each of two selected helices. Next, the 2D screen projection used by the user to view Mhp1 is detected and the projection that each 3D helix vector makes with this plane is evaluated. The angle between the result 2D vectors is evaluated over time.

3.4.2. Investigation of the crystal structures

Thin gate distortion is visualized onto each crystal structure in Figure 3.4.2 **A**. Distortion is additionally plotted in Figure 3.4.2 **B**. The bendix results reveal that a localised kink versus broad curvature distinguishes the TM10 open and closed states, respectively. TM10's peak angle remains almost identical, irrespective of state. TM5, on the other hand, undergoes kink relocalisation to the N-terminus and almost doubling of the peak angle as the gate opens.

3.4.3. Simulation results

3.4.3.1. Gate dynamics during inwards substrate transport

In the outward open state, TM10 forms a long α-helix with TM9b and TM9 which reaches across the top of the Hash motif. The presence of Phe421 in the adjacent TM11 induces slight helix curvature around Leu361. As Mhp1 transports substrates to the cell interior, the first step involves TM10 extending across the substrate by forming a 3_{10} helix, thus closing the exterior substrate binding site (see Figure 3.4.3). Meanwhile the conserved C-terminus remains in place, possibly rigidified by salt-bridges that I identify in the area (data not shown). This leads to the outward occluded state, which is characterised by static helix angles and broad curvature along TM10. Subsequently, the hash motif rotates so that the cytosolic part of Mhp1 is poised for substrate release to the cell interior. This

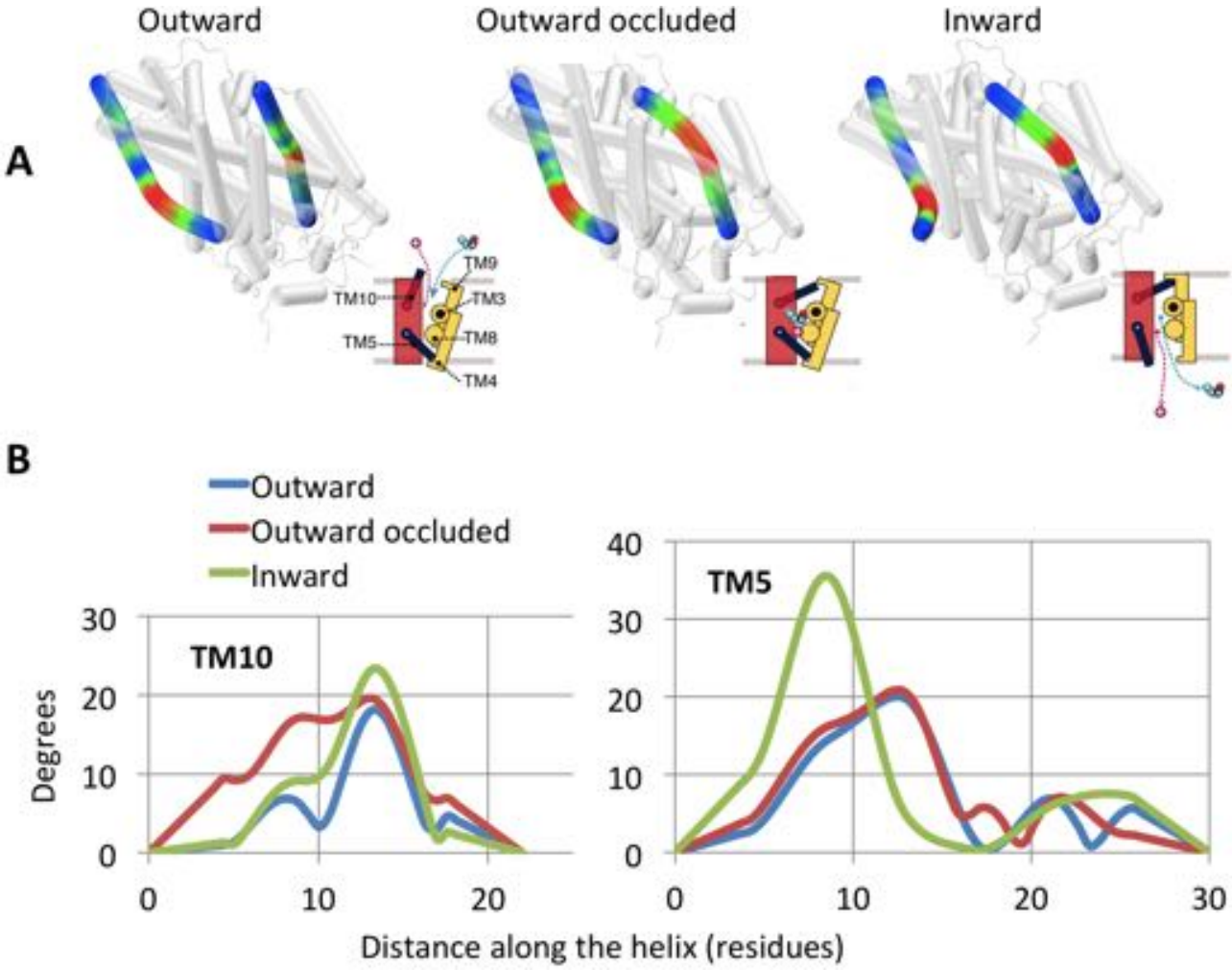

Figure 3.4.2.: Thin gate distortion in available crystal structures
A. Crystal structures visualized using bendices. Thin gates are shown as angle-indicative bendices and the rest as transparent bendices. The heatmap used is BGR; blue areas denote low angle while red areas denote high angle. Cartoons on the lower right of each structure show the suggested state of the structure (taken from Shimamura et al. [2010]). **B**. Angle along the length of each thin gate in published crystal structures. The graphs show a quantified version of the qualitative heatmaps in **A**. Left: TM10, right: TM5. See section 3.4.2 in the main text for a discussion of results.

rotation occurs simultaneously with TM5 gate widening, which ends in a high TM5 helix bend around Val160. Notably, the transporter does not pass through the proposed inwards occluded state, where the hash is angled inwards but the inner gate is closed. As the inner thin gate opens, TM10 remains closed but its initial broad high angle is lost. Phe421 induces TM10 helix distortion by steric clashes throughout inwards transport.

3.4.3.2. Gate dynamics following substrate release

Thin gate distortion hysteresis is observed upon return from the inward open state towards the outward open state. While the cytosolic thin gate passage does narrow until it reaches the closed gate size observed at the start of the Outward state, the kink around Val160 that characterises its 'inward open' state remains largely in place (c.f. Fig 3.4.3 and 3.4.4). Meanwhile TM10 remains firmly closed but releases its structural restraint on the immediately adjacent hash motif. The broader angle around Leu364, seen in the outwards occluded state, reappears briefly but a high, localised kink around Ala360 caused by Phe421 dominates its conformation in the latter half of the trajectory. The hash motif only rotates by a quarter of the necessary angle for the outwards conformation, so remains largely inwards. This closed cytosolic thin gate combined with the inwards-facing hash motif constitutes the missing state proposed by Shimamura et al. [2010], and is marked with the gray cartoon in Fig. 3.4.4.

3.4.3.3. Investigation into hash motif dynamics aided by bendix abstraction

A conserved ridge with capacity to aid substrate transport The pericellular-facing side of TM3 is lined with conserved large aromatic and/or hydrophobic residues: Leu104, Phe107, Trp108, Phe111, Trp114, Leu115 and Leu118. As the hash rotates, this ridge tilts to cover the substrate binding site, so probably plays a role in trapping substrate.

Association between hash and thin gates TM5 is closely associated with TM4 of the Hash. Together they form a complete αhelix which is only broken as TM5 moves peripherally in order to open the cystolic gate for subsequent release of substrates to the cytosol. This resembles the association between TM10 and the hash, and echoes the "inverted repeat" motif present in the LeuT superfamily of transporters.

At the start of substrate transport, TM10 closing pulls on TM9, which rotates the Hash outwards by 3°. Then, as the hash motif rotates inwards, the TM10 gate re-opens by 2.5 Å. Indeed, both thin gates go through a low curvature state as the hash transitions between its outward and inward conformation, which ought to structurally 'liberate' the hash. The hash rotation inwards occurs simultaneously with the TM5 gate's opening, which pulls the cytosolic part of the hash radially outwards.

The correlated dynamics between gates appears to be a function of substrate presence While TM5 movement occurs simultaneously with hash rotation when substrate is present, absence of substrate appears to sever the link between thin and thick gates. As TM5 closes in unrestrained simulations without substrate present, the hash only rotates by seven degrees (a quarter of the total closed angle). It is plausible that TM5 closure simply enables the hash to close eventually by releasing the strain imposed by the open thin gate conformation.

3.4.4. Discussion

I show here that Bendix can be used to help reveal the functional role of helix flexibility in a complex transport mechanism. Helix distortion distribution is visualized across residues by heatmap colour-coding according to local angle magnitude, which highlights non-linear helix behaviour. In the

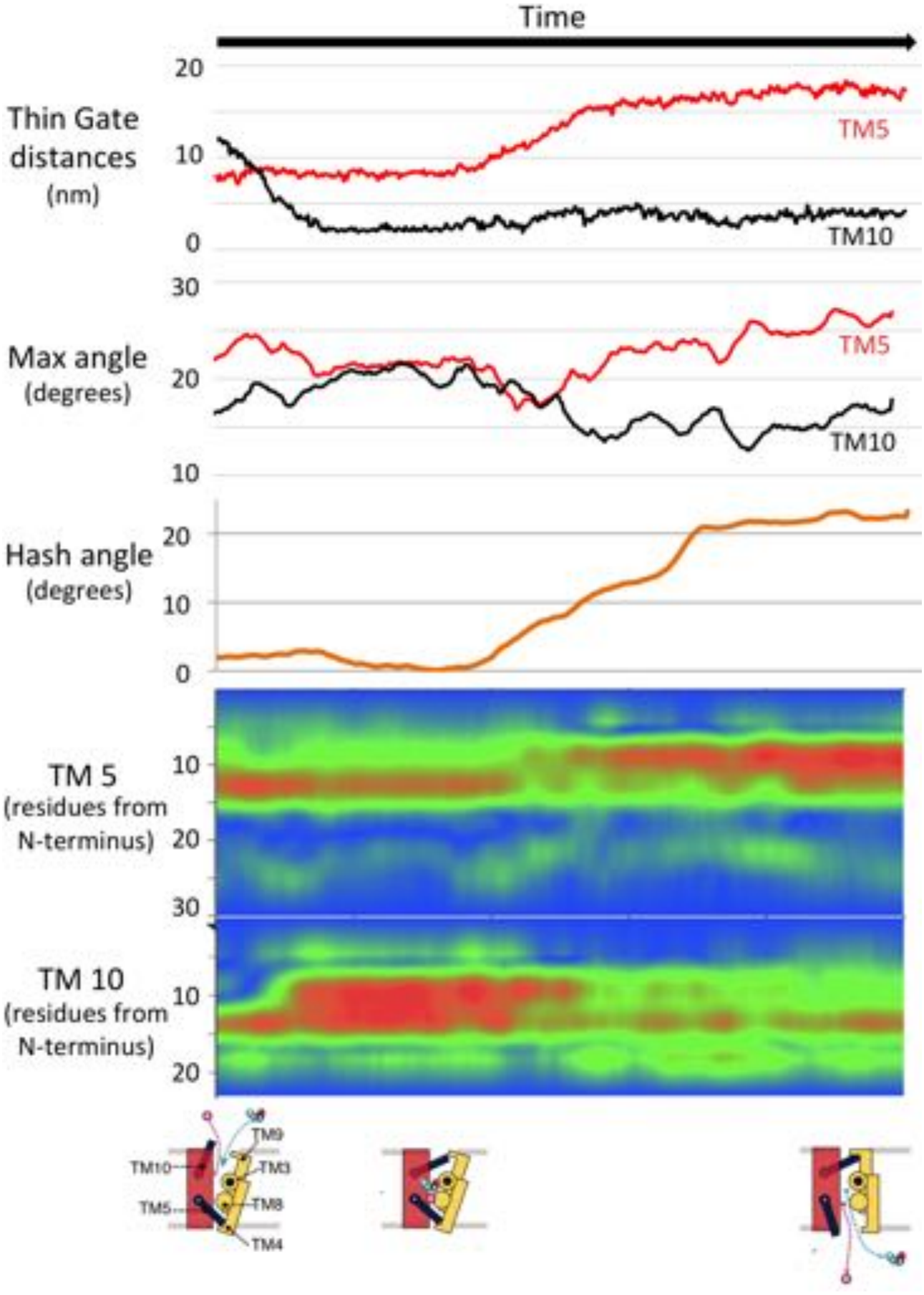

Figure 3.4.3.: Mhp1 gate dynamics during inward substrate transport.
Representative DIMS MD results for gate dynamics. Bottom: Cartoon abstraction of the transport states observed.

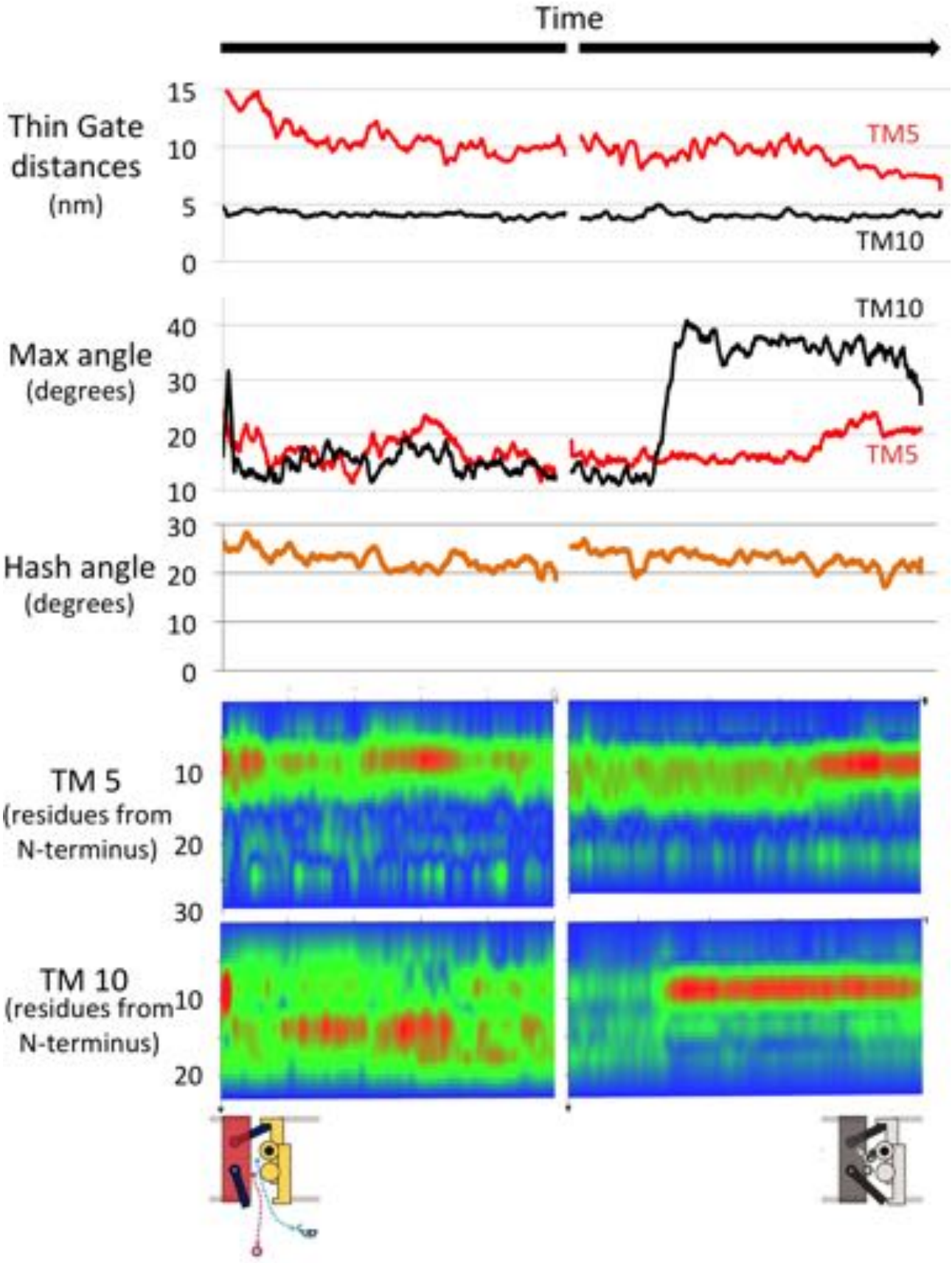

Figure 3.4.4.: Mhp1 gate dynamics during transition from inward to occluded.
Unrestrained MD results for gate dynamics. Two MD simulation were performed starting from the inward open state, thus the 'break' on the time axis. Bottom: Cartoon abstraction of the transport states observed.

example of Mhp1, this feature exposes important changes in helix TM10 as Mhp1 adopts its different functional and structural states.

I suggest that transported substrates could act as structural pivot points, resulting in rigidity that allows thin gate conformations to be relayed more effectively to the hash than when substrate is missing. Only in the absence of substrate are thick and thin gates uncoupled to the point that the inwards occluded state is possible. Inwards transport leads to 23° rotation of the hash motif relative to the bundle, close to the 30° observed by Shimamura et al. [2010]. While hash rotation permits substrate passage inwards, its cytosolic ridge of bulky aromatic and hydrophobic residues may also actively aid substrate transport.

Lastly, I identify that Phe421 influences TM10 geometry in all three crystal structures, and causes sporadic distortion in equilibrium simulations, so clashing is unlikely to be the result of a crystallisation artifact. The aromaticity of residue 421 (and thus relative bulkiness) is conserved, and the other side of the helix faces lipid, suggesting that the orientation of Phe421 could be under annular lipid influence. It remains to be investigated whether this helix distortion is functional for Mhp1's transport mechanism.

It should be noted that since the DIMS methodology biases the simulation trajectory, more simulations are required to substantiate these claims.

3.4.4.1. Proposed transport mechanism based on bendix data

Thin gates TM10 and TM5 are used as "reins" to control the rigid hash motif. In the outwards open conformation, one can assume that sodium is omnipresent in the binding pocket due to the sodium motive force. Therefore, although Mhp1 is an AND gate composed of sodium and benzylhydantoin binding sites, TM10 gate closure in the presence of benzylhydantoin is the bottleneck. This mechanically relays substrate presence to the adjacent hash, the top of which is lined by bulky conserved residues. As the hash rotates, it therefore facilitates the substrates' inward passage by both shielding it off from the outside and opening the inward passageway.

Hysteresis in terms of the gates' correlated dynamics is seen in equilibrium MD simulations started from the inward open state, where substrate is absent and both thin gates relax and compact. Upon thin gate closure, the hash, now fitted with 'slack reins', slowly returns towards the outward open conformation through the hypothesised inward occluded state. The inward occluded state is only accessible in the direction in- to outwards open, and appears to be prevented by substrate presence, which provides the structural link necessary for the thick gate to respond immediately to thin gate dynamics.

3.5. Conclusions

Curved α-helices are prevalent in membrane protein structures, and their flexibility plays a key role in biological function. Despite this demonstrated importance, methods for helix motion analysis are limited. I have developed a new tool for analysis of helix conformational change which aids structure comprehension and allows interactive analysis, visualization and manipulation of both AT and CG simulations. Up until now, CG visualization was very sparse, so the addition of Bendix to the VMD package will help the progress of multiscale MD research. In addition, it extends VMD's visualization capacity to non-standard input, such as inorganic systems. Intuitive heatmap display of helix curvature at the point of interest allows the user to more readily detect potentially relevant helix distortions, thus aiding analysis of the relationship between structure, dynamic conformational changes and biological function. Crucially, the Sugeta and Miyazawa [1967] helix axis algorithm is robust to variable number of residues per helix turn, which supports not just α-helices but also e.g. 3_{10} helices.

Bendix is highly intuitive to use, easily customizable and hosts in-built graphing with options to easily export data. It is open source, multi-platform and free, and the feedback that I have received

form users around the world suggest that Bendix is very useful. It has already contributed to fundamental research in several journal papers [Dimitrov et al., 2013, Giudice et al., 2013, Lee et al., 2013, Tikhonova et al., 2013, Stelzl et al., 2013] and theses.

3.5.1. Contrast to similar tools

There are a number of dedicated analysis software for static helix structures, e.g. HELANAL [Bansal et al., 2000], but support for analysis of helix dynamics over time is restricted to global descriptors of helix geometry [see e.g. TRAJELIX [Mezei and Filizola, 2006]. This limits applicability and lacks geometric context. Bendix resolves these issues and provides numeric results that are accompanied by axis-based helix visualization. This resembles a representation that is implemented in YASARA (http://www.yasara.org/), but is functionalised in Bendix by geometry-indicative graphics and analysis.

As part of VMD, Bendix incorporates important features of existent helix analysis software and complements them with intuitive, real-time data visualization, a clear interface and extended data type support and user control.

4. Statistical treatment of system interactions

4.1. Introduction

4.1.1. Future MD simulations are likely to be longer, larger and feature more replicates

Grid computing is a rapidly developing computer science area that is increasingly being adopted by computational biophysicists. MD simulation is a well suited application of this technology, as the distributed system architecture makes creation and management of replicate MD runs easier [Rhee et al., 2004, Klepeis et al., 2009]. The researchers behind the successful initiative Folding@home used this framework to their advantage [Beberg et al., 2009], and frameworks for automated setup of replicate MD simulations have recently been developed [Bertini et al., 2010, Hall et al., 2014]. Therefore, replicate simulation availability is likely to become more prevalent in future MD studies.

Moreover, as computer power increases, so has the feasible size of single MD systems, today reaching millions of atoms [Lindahl and Sansom, 2008, Zhao et al., 2013, Ngo et al., 2012], and such massive systems often incorporate multiple instances of the same protein species, which lends an individual system to statistical investigation. Recent examples of such massive MD systems are models of bacterial photosynthetic membrane vesicles [Sener et al., 2007], satellite tobacco mosaic virus [Freddolino et al., 2006], rabbit hemorrhagic disease virus [Wang et al., 2013], southern bean mosaic virus [Zink and Grubmuller, 2009] and the HIV [Ayton and Voth, 2010, Zhao et al., 2013] and human influenza virion [Chavent et al., 2014]. Given sufficient spatial separation, such single-system protein replicates could qualify as pseudo-independent observations. Very large MD systems therefore offer a growing resource of data for statistical analysis.

Lastly, progress in both force-field and hardware development have recently enabled experimental results (obtained by NMR and crystallography) to be approximated by very long MD simulations [Shaw et al., 2010, Dror et al., 2012]. When long trajectories, rather than replicate-based ensembles, are used to retrieve experimental parameters, the trajectory likely contain multiple instances of the event of interest (e.g. folding and unfolding). Observations could therefore be pooled from within a trajectory, rather than from multiple trajectories, to retrieve event parameters.

Whether individual observations are captured from replicate, large or long MD simulations, it is clear that increasing numbers of observations are becoming a more mainstream MD phenomenon. However, tools for automated analysis and statistical treatment of MD data are lacking [Likic et al., 2005]. While tools that perform statistical analyses exist (for example the R software and MatLab), these tools would require that MD data are manually dissected and ported prior to analysis. Molecular dynamics visualization packages such as VMD do not contain dedicated statistics engines. I wrote InterQuant to meet this demand, and to make statistical tests easy and accessible to perform.

4.1.2. InterQuant: automated proximity analyses and statistics

In this chapter, I present InterQuant, a plugin for the Visual Molecular Dynamics (VMD) package [Humphrey et al., 1996] that probes system interactions using statistical inference methods. Specifically, InterQuant measures a large set of proximity-based metrics for individual simulations, and performs statistical analyses on results of multiple simulations, presenting the user with intelligible, interactive conclusions. It supports all VMD-compatible input, including but not limited to static, dynamic (MD) atomistic and coarse grained data. As VMD supports most MD and static structure file-formats, InterQuant is almost universally applicable to structural bioinformatics data. The program

naturally separates into two parts: the proximity analysis part, called InterQuant (IQ) as it quantifies distances between system parts, and the statistical inference engine which inputs IQ results from many simulations, called Inter-InterQuant (IIQ). From here on, 'InterQuant' will refer to the overall program, while the acronyms will refer to each respective sub-component.

4.1.2.1. IQ

IQ computes the geometric relationships between two selections of atoms in a loaded system. Specification of selections is easy since VMD comes with in-built regular expressions for generating custom queries of system substituents, and IQ supports these expressions fully. In addition, IQ supports averaging of results across system components, which is useful for e.g. protein homo-multimers. The resolution at which proximity is analysed, e.g. between whole protein chains or individual substituent atoms, is chosen by the user.

A range of analyses are available that output e.g. ordinal, categorical and/or time series data. Where possible, results are intuitively displayed directly onto the selections using a heatmap. Because of the types of results on offer, IQ's analyses of system interactions lends itself well to explorative, first port of call analysis. While analysis of MD systems can be daunting because of their size and four-dimensional nature, IQ handles large datasets easily and rapidly filters, categorizes and quantifies contacts, making further system investigation easier.

4.1.2.2. IIQ

IIQ allows users to easily import and post-process previously obtained proximity results to retrieve any statistically significant results across repeat simulations, and to contrast interactions obtained using different systems, like those obtained by change to content (e.g. wildtype versus mutated protein) or starting configuration. For comparison between samples, IIQ reveals what proximity metrics are significantly different between samples, using a custom level of significance. This software module was named Inter-InterQuant to underscore that its function is to compare between IQ results.

By building InterQuant, I wanted to implement statistical methods that accommodate all sample sizes, but particularly small ones ($N < 30$) as MD research often does not involve more than a few replicate simulations. For this reason, InterQuant implements the *t-test*, which was especially developed to deal with small sample sizes (but is equally applicable to larger data sets) [Boslaugh, 2013]. The t-test assumes that underlying populations are normally distributed. Near equilibrium, particle interaction energies are approximated by a quadratic function in MD, and it is therefore it is valid to assume that interaction-based population distributions should approximate normal curves. However, the t-test is also relatively robust to non-normal population distributions [Student, 1908, Lehmann, 1999]. Finally, remains the most commonly used method for hypothesis testing about means [Zar, 2010]. This makes the t-test a good choice for statistical treatment of MD data. A more detailed description of the t-test and related statistics concepts is available in Appendix B.

4.1.2.3. Usability and compatibility

Throughout, InterQuant is very user-friendly. It verifies user selections against the loaded system, shows error messages that suggest how to resolve problems and displays explanatory notes on both how to operate the tool and the analysis algorithms used. Results are visualized as graphs and heatmaps as far as is possible. Notably, statistics results are heatmapped onto the structure, which makes it easy to pinpoint the location of e.g. statistically significant differences as a result of mutation. Results can be reported as both graphs and in a text file. Lastly, InterQuant is available in both terminal and graphical interface mode, and is compatible with most operating systems, including Windows, Mac and Linux.

4.1.3. The importance of proximity and contact analyses

Proximity analyses, including contacts, is often amongst the first analyses performed on an MD trajectory, and most MD studies include some aspect of proximity study. Proximity analysis is fundamental to biomolecular study since interactions; long- and short-range, unspecific and specific, are responsible for biomolecular processes [Hakes et al., 2007, Karplus and Kuriyan, 2005]. A protein's non-bonded interactions contribute to its local and long-range electrostatic energy, which influence overall conformation and has downstream effects on, for example, ligand docking. Therefore the identification and quantification of interactions is a key concern for structural biologists and computational biophysicists. The dynamics of non-bonded interactions suggest the conformational state space that the protein explores as it performs its function [Karplus and Kuriyan, 2005], and is therefore important to investigate in order to understand the protein's structure-function relationship. Besides, from a technical MD viewpoint, the evolution of an MD system is a function of its substituent particles' proximity. Therefore, by analysing proximity, I can reason about the evolution of the MD trajectory.

4.1.4. Example applications of InterQuant

Applications for InterQuant include:

- Ligand binding pocket characterization; key residue IDs and residue type preference per ligand atom given the present system
- Statistical comparison between MD and crystallographic positioning of co-crystallised ligand
- Membrane binding domain localization in soluble proteins
- Protein-protein interactions ranked according to favourability
- Quantified differences between force-fields
- Track system equilibration progress by monitoring how interactions changes over time
- Identify and quantify conserved docking site contacts across multiple drugs that are known to bind the same general area
- Heatmap-colour molecules according to results obtained from InterQuant or elsewhere

Because of the way that selections are treated, InterQuant lends itself naturally to investigation of solvent-solute interactions, where solvent is used in the broadest sense of the word, to include e.g. water, ions, lipid and drugs. InterQuant assays a solute's solvent accessibility and preference, normalised to system content. While both ion and water have been found to be important mediators in protein-protein interactions [Rhee et al., 2004, Samsonov et al., 2008, Dawe et al., 2013], lipid interaction with protein is of particular interest to this thesis, and is also a growing research area in its own right [Yoshimura and Sokabe, 2010].

4.1.4.1. The physiological relevance of lipid-protein interactions

Lipid is the natural environment of membrane proteins, which make up 60% of drug targets [Overington et al., 2006]. It is known to modulate many membrane proteins' structure and function, e.g. in the nicotinic acetylcholine receptor [daCosta et al, 2013], bacterial mechanosensitive channels [Perozo and Rees, 2003, Powl et al., 2005] (see also later chapters) and K2P channels [Honoré, 2007], but lipid has also been found to regulate soluble and viral proteins [Li et al., 2014]. Many proteins even contain explicit lipid-binding sites, in particular for ionic lipid [Suh and Hille, 2008, Hansen et al., 2011, Li et al., 2014]. It should come as no surprise, then, that aberrant lipid-protein interactions are implicated in disease [Fabelo, 2011].

In recent years, MD research has produced fine-tuned lipid force-fields with increasing amount of physiologically relevant complexity to accurately predict experimental values using both atomistic [Klauda et al., 2010] and coarse-grained MD setups [Lopez et al., 2013]. It is therefore a good time to present tools for facilitated analysis of lipid-protein interactions.

4.1.4.2. Membrane-protein interaction applications of InterQuant

- Contacts made by different kinds of lipids to a membrane protein; absolute amount of contacts and relative to system content
- See how the dynamics of lipid interaction to each part of the protein changes over time
- Compare the binding profile of different lipid composition setups using the distribution of number of contacting lipid types versus continuous and binned time in contact
- Locate a lipid binding site, and investigate residue or atom contributions to lipid binding by obtaining their number of frames in contact with lipid
- Find out if lipid binding sites are conserved across homo-monomers of a multimeric protein
- Retrieve population parameters for lipid-protein interactions across repeat simulations for the last part of each trajectory, where interactions are the most equilibrated
- Compare lipid contacts to wildtype and mutant protein; retrieve statistically significant differences and a confidence interval for this difference per protein residue
- Compare lipid contacts to the same protein under different circumstances, such as absence or presence of a transmembrane voltage or membrane stress
- Find the lipid type preference per protein part

4.1.5. Organisation of this chapter

The chapter starts with an overview of how InterQuant was implemented, the available user interfaces and documentation. This is followed by a presentation of the proximity and contact analysis suite available through IQ, as well as settings. The IQ program algorithm and the various output options is also described, including results visualization. Then I present the inference tools that are implemented in IIQ, the IIQ program algorithm and its output options. The chapter ends with conclusions and ideas for future feature implementations. For an overview of statistical concepts that are relevant to InterQuant, including rules for when one statistical test is preferable to another, please refer to Appendix B.

4.2. Results

4.2.1. Software implementation

InterQuant is implemented as a plugin to Visual Molecular Dynamics (VMD) [Humphrey et al., 1996] and is written in *Tcl*, a cross-platform compatible programming language that comes with a graphical user interface (GUI) toolkit called *Tk*. Tk is used to generate the GUI front-end, which exists alongside the functionally equivalent terminal front-end. InterQuant's back-end is split between the trajectory analysis suite, *InterQuant* (IQ), and the statistical inference engine *InterInterQuant* (IIQ). IQ analyses any VMD compatible input for proximity and contacts between selections, while IIQ inputs IQ results in order to infer their parameters. It has no dependencies beyond VMD, so is easy to run.

The code is circa 14,000 lines long (approximately 230 pages) and consists of 39 procedures (Tcl functions) that handle the following tasks:

- Core software management
  - InterQuant loading in VMD GUI environment or Terminal
  - InterQuant namespace variables and package requirements
  - GUI setup and maintenance
  - File handling: load, save and export data
  - Informative error messages
  - User help and acknowledgements
  - Reset variables and erase OpenGL visuals
  - Reset and quit the program
- Input pre-processing and analysis
  - VMD system calls to detect VMD frame change
  - Pre-processing and selection validation prior to input analysis
  - IQ analysis: compute metrics for requested analyses
  - Pre-processing of separation data for within-cutoff display only (optional)
  - IIQ input pre-processing: variance homogeneity, IQ analyses that are common to all input files
  - IIQ inference per analysis: Standard t-test for one or two samples, Welch's t-test, dependent samples' t-test
  - Normal, T and F distributions for p-value and critical value retrieval
- Results presentation
  - 2D and 3D surface graph generation that employs OpenGL graphics
  - Heatmapping of VMD-loaded structures according to results
  - Graph output customised per analysis: to text file, using Multiplot, gnuplot or as Terminal messages

All code is original and written by me.

InterQuant computes efficiently and at high precision The code is modulated throughout by encapsulation into procedures and conditionals to speed up operation and avoid unnecessary calculations.

All mathematical expressions are computed as Tcl floating points. Rounding is only done when values are written to file, at which point values are reported to three decimal places.

4.2.1.1. User interfaces

InterQuant has two front-ends to accommodate both graphical and terminal user input.

Terminal user interface The terminal front-end was built to accommodate analysis without the overhead cost of the graphical display, and to allow script automation.

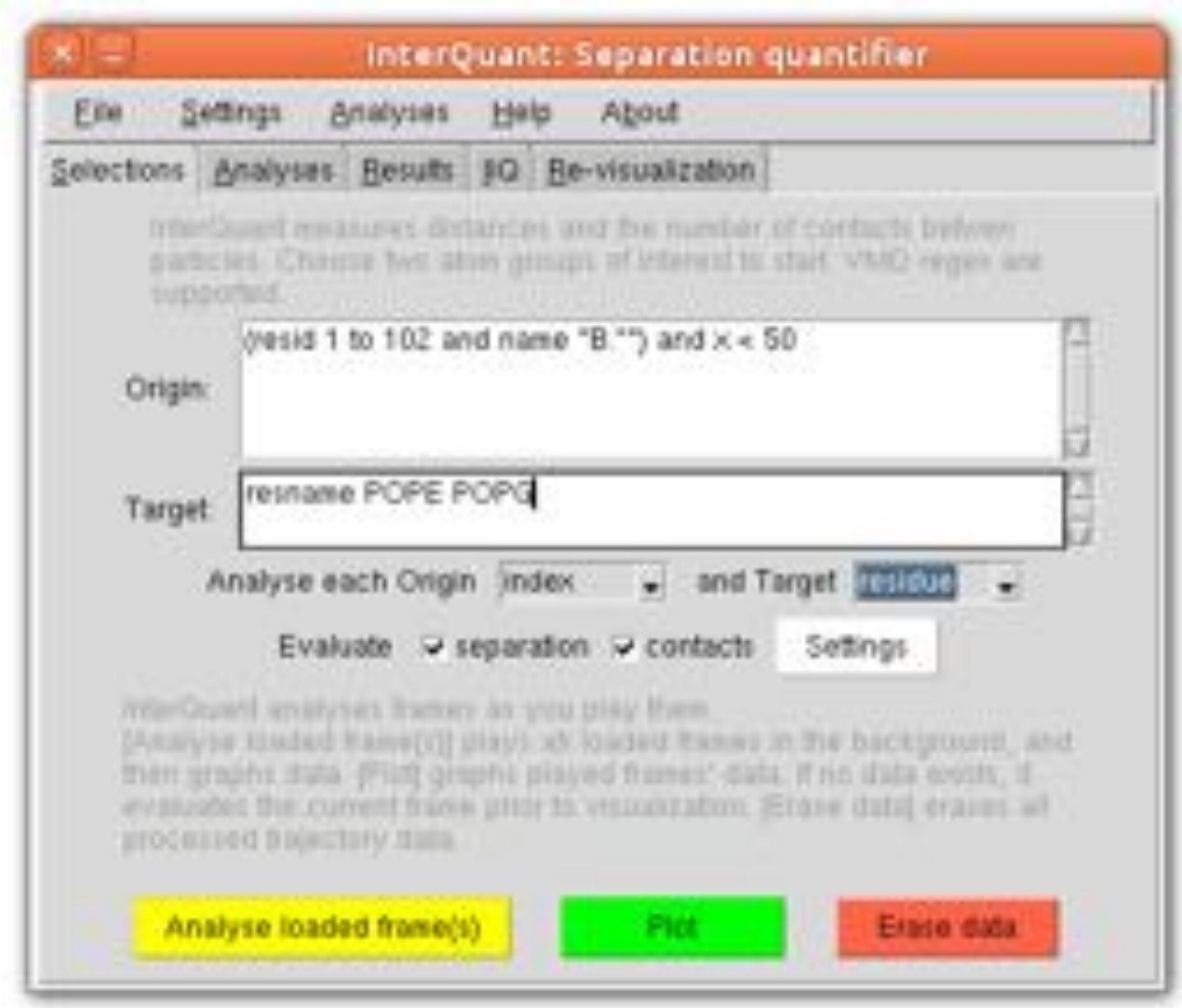

Figure 4.2.1.: InterQuant's graphical user interface
Screenshot from a PC running Ubuntu Linux. The GUI adapts to the operating system that it runs on, so will look different on Mac OX and Windows. However features remain the same.

Graphical user interface The InterQuant GUI is fully compatible with the VMD GUI interface and OpenGL display. When loaded, it incorporates into VMD's Analysis suite that is found under the Extensions menu. I built the front-end GUI with emphasis on compactness, using a tab layout to switch between user input, analyses settings and result presentation modes (c.f. Figure 4.2.1). The GUI window expands when additional space is necessary, to accommodate multiple samples in IIQ. A compact GUI layout is necessary from a usability point-of-view, to allow InterQuant's full functionality on smaller screens. Likewise, GUI field display is managed so that only relevant fields are enabled and visible at any time. An input- and results-driven interface modification help users navigate InterQuant more efficiently and optimises use of available space.

4.2.1.2. Documentation

A website contains a general introduction to InterQuant, lists available analyses and settings, shows screenshots of InterQuant in both GUI and terminal mode, and gives detailed instructions for how to use the terminal mode both for single file input and for automation of InterQuant's IQ and IIQ analyses. Also included are step-by-step installation instructions and a Troubleshooting section that details both VMD plugin installation and InterQuant behaviour and error messages. Because InterQuant is yet

unpublished, code and documentation are kept offline. For access, please contact me.

4.2.2. Proximity and contact analyses implemented in InterQuant

InterQuant (IQ) analyses system interactions for static or dynamic VMD input. The user loads a system and specifies two selections of atoms from the system, *Origin* and *Target*, that IQ will consider in analyses. Only Origin and Target selections are necessary IQ inputs; defaults exist for all other settings.

4.2.2.1. Selections, selection settings and output

- Origin selection

  All VMD regular expressions are supported, as well as the symbol "|||" (three vertical bars), which I implemented specifically for InterQuant. This is used to partition the Origin selection into equally large subsets of particles. Analyses will then average results per Origin particle across these selections. This can be beneficial if the Origin is a homo-multimer and the user seeks to average the protein's interactions across chains. The vertical bars were chosen to not interfere with any existing, VMD-accepted regular expression notation and for its similarity to the conditional OR statement that is used in most programming languages.

- Origin particle type (OP)

  This is the resolution at which IQ analysis is performed for the Origin selection. Choose between *resid*, *residue* or *index*. The default is resid

- Target selection

  All VMD regular expressions are supported.

- Target particle type

  This is the resolution at which IQ analysis is performed for the Target selection. Choose between *resid*, *residue* or *index*. The default is resid.

- Output file name

4.2.2.2. Available analyses and analyses settings

Below, each IQ analysis is described in words alongside its mathematical expression. Table 4.2.1 describes mathematical notations that are commonly used in IQ equations. Pseudo-code for each analysis is also available as comments in the code.

IQ employs the standard deviation (SD) equation that assumes that the full population data are available, since the full trajectory data *are* all population observations with respect to that trajectory (so no parameter inference is necessary, c.f. Section B.2). IIQ, on the other hand, infers the SD parameter. For more details, see Appendix B.

The accompanying InterQuant GUI for selecting IQ analyses and settings is shown in Figure 4.2.2.

Proximity

Analyses of the exact separation between Origin and Target.

The proximity algorithm identifies the smallest distance between Origin and Target's geometric centres. It avoids exhaustive computation of the Origin's separation to all system Targets. Rather, I implemented a set-decomposition algorithm that seeks Targets within a user-defined radius of the Origin, and if no Targets are found, the search radius r_c is incrementally increased by r_{step} to r'_c until either one or more contacting Targets are found, or the radius equals a user-defined search cutoff:

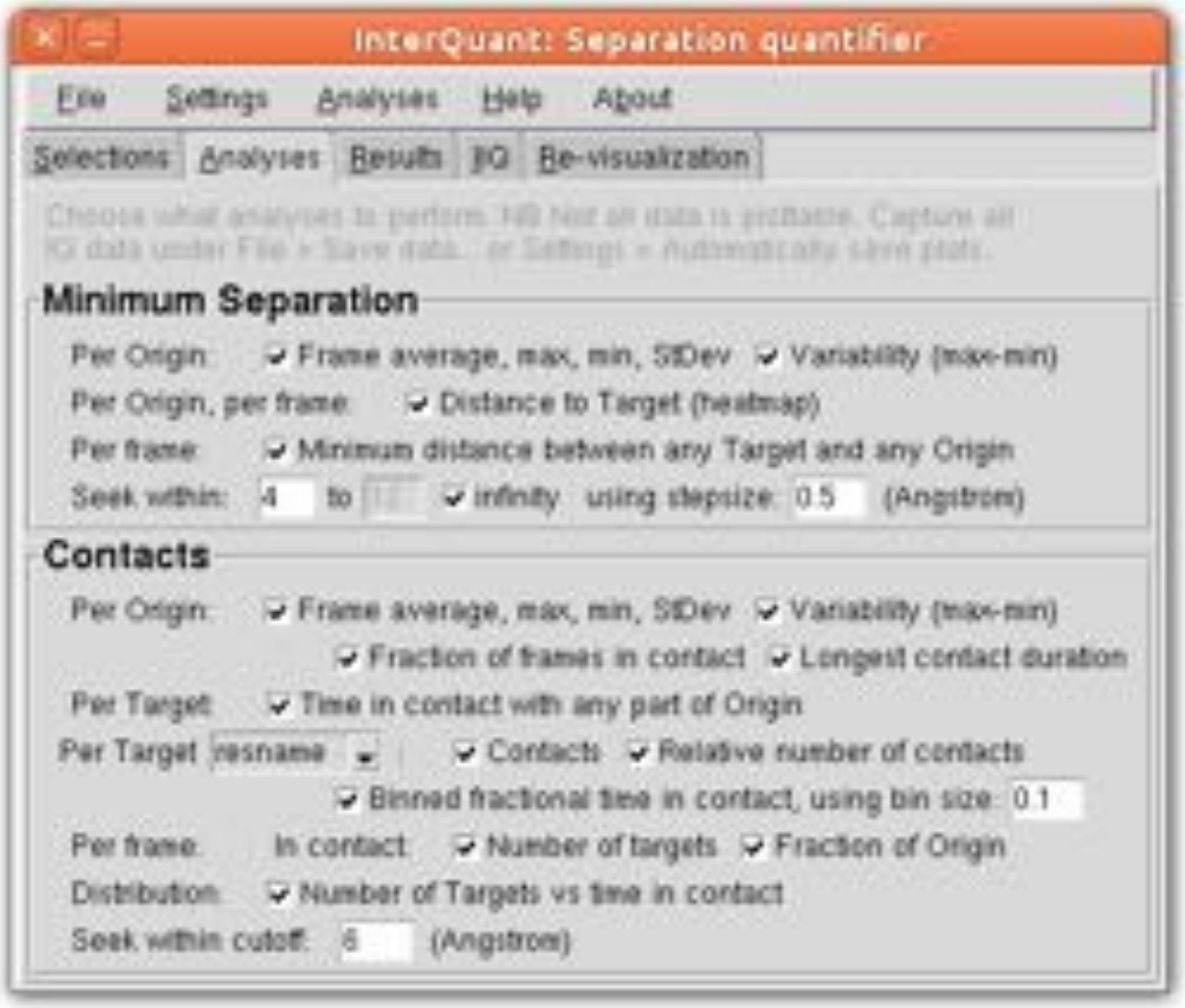

Figure 4.2.2.: Available IQ analyses and their respective settings, viewed in the GUI.

Symbol	Description	Relevant equation
r	separation	
OP_i	Origin particle i	$OP_i : \forall i \in \{1, 2, \ldots, N_{OP}\}$
T_j	Target particle j	$T_j : \forall j \in \{1, 2, \ldots, N_T\}$
T_{type}	Target type, e.g. resname	$\forall k \in \left\{1, 2, \ldots, N_{T_{type}}\right\}$
k	Target type of T_j	$k \in T_{type}$
r_{ij}	separation between OP_i and T_j	$\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}$
r_i	separation between OP_i and any T	
r_j	separation between T_j and any OP	
t	time	
N_t	Number of frames	
N_{OP}	Number of Origin particles	
N_T	Number of Targets	
c	Number of contacts	
r_c	Cutoff radius that defines a contact	
c_{ij}	Contact between OP_i and T_j	$c_{ij} = \begin{cases} 1 & r_{ij} \leq r_c, \\ 0 & r_{ij} > r_c. \end{cases}$
c_i	Number of T within r_c of OP_i	$c_i = \sum^{j} c_{ij}$
c_j	Number of OP within r_c of T_j	$c_j = \sum^{i} c_{ij}$

Table 4.2.1.: IQ analysis notation

$$r_c' = \begin{cases} r_c + r_{step} & c_i\left(r_c\right) = 0 \ \wedge \ r_c \neq r_{max}, \\ r_c & c_i\left(r_c\right) \neq 0 \ \vee \ r_c = r_{max}. \end{cases}$$

The increment with which the search radius is increased, r_{step} , is also user-defined. Once $c_i\left(r_c\right) \neq 0$, substituent atoms of both the Origin particle and all contacting Targets are identified. For example, if the Origin particle resolution is equivalent to an amino acid residue, it consists of several substituent atoms. Then, exhaustive pair-wise separation is computed in order to find the smallest separation between any of the Origin particle's atoms and any contacting Target atom. If neither Origin nor Target are composed of sub-particles, and only a single Target is within radius ($c_i\left(r_c\right) = 1$), the exact separation between the Origin and Target particle is computed immediately.

This algorithm performs a minimal amount of computations between the Origin and neighbouring Targets only, excluding distant particles. It is therefore efficient and scales well for larger systems, which should be advantageous considering the increasing complexity of MD systems.

- Per Origin particle: average and standard deviation across all frames, and the highest and lowest distance to any Target encountered in any frame

  $r_{i,\mu} = \frac{1}{N_t} \sum_{t=0}^{N_t} r_i(t)$

  $r_{i,\sigma} = \sqrt{\frac{1}{N_t} \sum_{t=0}^{N_t} \left[\left(r_i(t) - r_{i,\mu}\right)^2\right]}$

  $r_{i,max} = max\left(r_i\left(t\right)\right), \forall t \in \{1, 2, \ldots, N_t\}$

  $r_{i,min} = min\left(r_i\left(t\right)\right), \forall t \in \{1, 2, \ldots, N_t\}$

- Per Origin particle: Most extreme variation in distance to any Target

The smallest distance to Target detected in any one frame, subtracted from the largest distance detected.

$r_{i,var} = r_{i,max} - r_{i,min}$

- Per Origin particle, per frame: Distance to any Target

  $r_i(t)$

- Minimum per frame: Minimum distance between any Target and any Origin

  $r_{min}(t)$

Proximity settings

- Seek Targets within this radius: inner bound (Ångstrom)

  The radius away from each Origin particle's geometric centre to start screening for Targets. Defaults to 4 Å.

  $r_{c,min}$

- Seek Targets within this radius: outer bound (Ångstrom)

  Beyond this radius of each Origin particle's geometric centre, Targets are not sought. If no Targets are encountered within this outer cutoff, distance to any Target is defined as Outer cutoff radius + 1 A, to discriminate these results from real separation values.

  $r_{c,max}$

- Seek Targets within this radius: infinite outer bound (overrules any set outer bound)

  Targets are sought using incrementally increasing radii until they are found. This may be a time-consuming loop, but it is not an infinite loop, as both Target and Origin selections' existence in the loaded system are verified during IQ pre-processing. Therefore a Target will always eventually be found. This is the default cutoff setting.

  $r_{c,max} = \infty$

- Seek Targets with this incremental step size in radius (Ångstrom)

  The size of the increment to use when increasing the radius from an origin particle's geometric centre when no Target is found within the current radius. Defaults to 0.5 Å.

  r_{step}

Contacts

Analyses of number and type of Targets within and including a cutoff radius of the Origin. Irrespective of the exact separation between Origin and Target particles, if separated by a distance that is less than or equal to the user-defined cutoff radius, they are said to be in contact. Regardless of the extent of contact, a contact is simply reported as an integer boolean value; 0 or 1.

- Per Origin particle: average and standard deviation across all frames, and the highest and lowest number of unique Targets within cutoff encountered in any frame

  $c_{i,\mu} = \frac{1}{N_t}\sum_{t=0}^{N_t} c_i(t)$

  $c_{i,\sigma} = \sqrt{\frac{1}{N_t}\sum_{t=0}^{N_t}\left[\left(c_i(t) - c_{i,\mu}\right)^2\right]}$

  $c_{i,max} = max\left(c_i\left(t\right)\right), \forall t \in \{1, 2, \ldots, N_t\}$

  $c_{i,min} = min\left(c_i\left(t\right)\right), \forall t \in \{1, 2, \ldots, N_t\}$

- Per Origin particle: Most extreme variation in numbers of unique Targets within cutoff

  The smallest number of Targets detected within cutoff in any one frame, subtracted from the largest number detected.

  $c_{i,var} = c_{i,max} - c_{i,min}$

- Per Origin particle: Fraction of frames in contact with any Target

  The fraction of all frames that at least one Target is within cutoff.

  $c_{i,frac} = \frac{1}{N_t} \sum_{t=0}^{N_t} c_{i,bool}(t)$

  where $c_{i,bool}\left(t\right) = \begin{cases} 1 & c_i\left(t\right) \neq 0, \\ 0 & c_i\left(t\right) = 0. \end{cases}$

- Per Origin particle: Longest contact duration to a unique Target

  The Target that is within cutoff the most number of frames (not necessarily continuously) is identified. The fraction of frames that this Target is in contact, as well as Target type identifying strings such as resid, chain and residue name are provided so that a user can easily follow up on this analysis with more detailed interaction analyses. If no Target is ever in contact, this is also reported.

  $c_{i,long} = max\left(N_{t,T_j}\right) * \frac{1}{N_t}, \forall j \in \{1, 2, \ldots, N_T\}$

  $N_{t,T_j} = \frac{1}{N_t} \sum_{t=1}^{N_t} c_{ij}(t)$

- Per Target particle: Fraction of frames in contact with any Origin

  All Targets that are ever within cutoff of any part of Origin are identified and sorted according to contact length. The number of frames in contact (not necessarily continuously), as well as Target type identifying strings such as resid, chain and residue name are provided so that a user can easily follow up on this analysis with more detailed interaction analyses.

  $N_{t,T_j,frac} = \frac{1}{N_t} * N_{t,T_j}$

  $N_{t,T_j} = \sum_{t=1}^{N_t} c_{j,bool}(t)$

  where $c_{j,bool}\left(t\right) = \begin{cases} 1 & c_j\left(t\right) \neq 0, \\ 0 & c_j\left(t\right) = 0. \end{cases}$

- Per Target type: Number of contacts

  Each Target $T_j \in T$ has a type $k \in T_{type}$ associated with it. This type is an identifying string chosen by the user; e.g. the Target's residue name or atom type. Targets that contact any part of Origin are identified per frame, and their type is retrieved. At the end of the trajectory, contacts per Target type are summed.

  $c_k = \sum_{j=1}^{N_T} \varphi\left(T_j\right)$

  where $\varphi\left(T_j\right) = \begin{cases} N_{t,T_j} & T_{type}\left(T_j\right) = k, \\ 0 & T_{type}\left(T_j\right) \neq k. \end{cases}$

- Per Target type: Number of contacts, relative to system content

  As above, but results are presented as a fraction of the system content of each particular kind of Target (again chosen by the user). Results that have been normalised in this way reveal Origin attraction to Target, unbiased by system content.

$$c_{k,rel} = \begin{cases} c_k \left[\left(\frac{N_k}{N_T} \right) \sum_{k=1}^{N_{Type}} c_k \right]^{-1} & c_k > 0, \\ 0 & c_k = 0. \end{cases}$$

where the total number of different Target types is the system is $N_{T_{type}}$ and the total number of Target particles of type k is denoted N_k .

- Per Target type: Binned duration of contacts, as fractions

  Results for *Per Target particle: Fraction of frames in contact with any Origin* are binned using a bin size chosen by the user. Results are presented per Target type (e.g. residue name or atom type), which is also chosen by the user.

  When $N_{bin} = \frac{1}{c_{bin}}, \forall bin \in \{1, 2, \ldots, N_{bin}\}$:

  $$c_{k,bin} = \sum_{j=1}^{N_T} \phi(T_j)$$

  where $\phi(T_j) = \begin{cases} N_{t,T_j} & \left(N_{t,T_j,frac} \leq c_{bin} * bin\right) \wedge \left(T_{type}\left(T_j\right) = k\right), \\ 0 & \left(N_{t,T_j,frac} > c_{bin} * bin\right) \vee \left(T_{type}\left(T_j\right) \neq k\right). \end{cases}$

- Per frame: Fraction of Origin in contact with any Target

  Each Origin particle is evaluated for contacting Targets, and irrespective of the number of Targets in contact, if one or more contacts exist, that Origin particle is defined as having contacts. The fraction of all Origin particles with contacts is evaluated per frame and presented as a fraction of the total number of frames.

  $$t_{c,fracOP} = \frac{1}{N_{OP}} \sum_{i=1}^{N_{OP}} c_{i,bool}(t)$$

- Per frame: Number of unique Targets in contact with any part of Origin

  $$t_{c.N_T} = \sum_{j=1}^{N_T} c_{j,bool}(t)$$

- Per duration of time: Number of unique Targets in contact with any part of Origin

  The number of frames that each unique Target contacts any part of Origin is collected per Target. The number of unique Targets that contact Origin for a given number of frames is then counted, and plotted against time in contact. The result distribution reveals the macroscopic profile of Target attraction to any part of Origin, and includes the number of Targets that remain bound throughout the trajectory.

  $$t_{c,duration} = \sum_{j=1}^{N_T} \psi_{bool}(T_j, t)$$

  where $\psi_{bool}(T_j, t) = \begin{cases} 1 & N_{t,T_j} = t, \\ 0 & N_{t,T_j} \neq t. \end{cases}$

Contact settings

- Cutoff radius within (and including) which to seek Targets for contacts, in Ångstrom. The default is 6 Å.

  r_c

- Target type level; the string to retrieve for contacting Targets, to reveal its identity. Users choose between resname, name, type or chain. The default is resname.

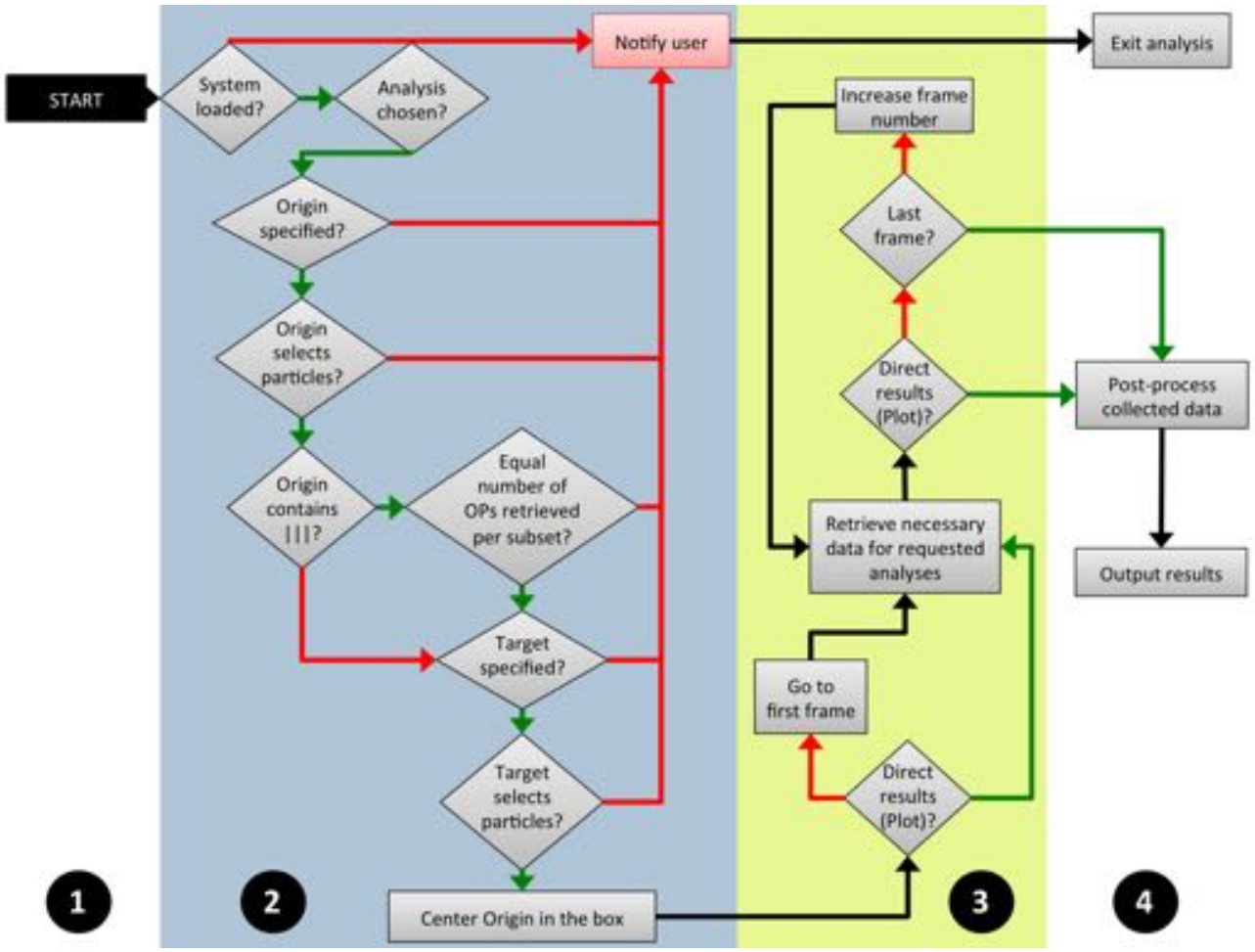

Figure 4.2.3.: Flowchart of the IQ algorithm at runtime.
The flowchart is broken down into the following sections (distinguished by colour): **1.** Input and input settings submitted by user **2.** Input pre-processing **3.** Analysis **4.** Post-processing and results output. At conditional points in the flowchart, arrow colour denote decisions: Green = yes, Red = no. Black arrows show unconditional algorithm pathways.

- Bin size, as a fraction, to use when binning Target duration of contact. The default bin size is 0.1.

  c_{bin}

4.2.2.3. The IQ algorithm at runtime

Figure 4.2.3 shows a flowchart view of the algorithm that is used whenever a user clicks either 'Analyse loaded frames' or 'Plot' in GUI mode (see Figure 4.2.1), or uses IQ in Terminal mode. In addition, the user can choose to pause analysis at any time, in which case IQ stops at the nearest frame. In that case, analysis data are stored in the event that the user wishes to continue analysis. In Terminal mode, on the other hand, all input frames are automatically analysed; a 'Plot' or 'Pause' equivalent does not exist.

IQ features significant input pre-processing to reduce the risk of error, and specifies sources of error using Terminal or pop-up messages (Fig. 4.2.3, section 2). Only if no input errors were encountered does analysis proceed.

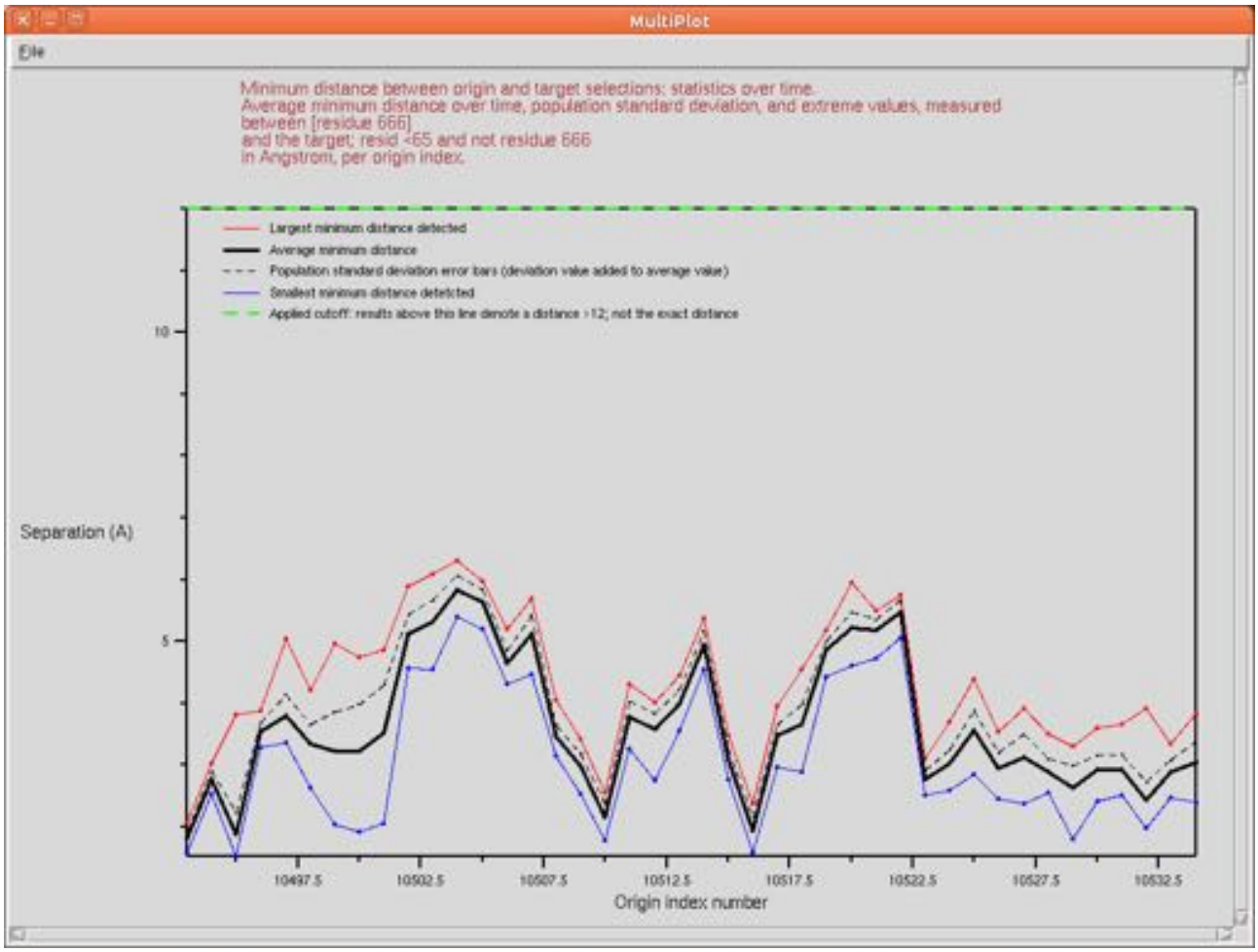

Figure 4.2.4.: IQ results for separation to Target, plotted in Multiplot.
Trajectory statistics on separation to Target per OP, where Origin and Target are a peptide transporter and antibiotic, respectively. Traces show maximal detected variation (red and blue), average (bold black) and SD (blue dash). The green dashed horizontal line denotes $r_{c,max}$, the cutoff distance beyond which Targets would no longer be sought. No Targets were ever outside this boundary (c.f. green dash and red lines). This graph encompasses the heatmap data shown in Figure 4.2.5 **A**.

4.2.2.4. Result output options

When relevant loaded frames have been analysed and post-processed, results are outputted according to user settings.

Visualization in GUI mode

All analyses graphs display automatically as individual windows using MultiPlot (a built-in VMD graph module). An example of an IQ Multiplot graph is shown in Figure 4.2.4.

If the user chose to analyse the evolution of proximity per OP, the VMD display is switched to the interactive 3D heatmap graphing mode. Alternatively, IQ detects what analyses were performed and enables the relevant radiobuttons which allows the user to explore IQ results visually, directly onto the loaded system. Both these GUI modes are shown in Figure 4.2.5. When data exist for both visualization modes, they are easily interchanged in the GUI.

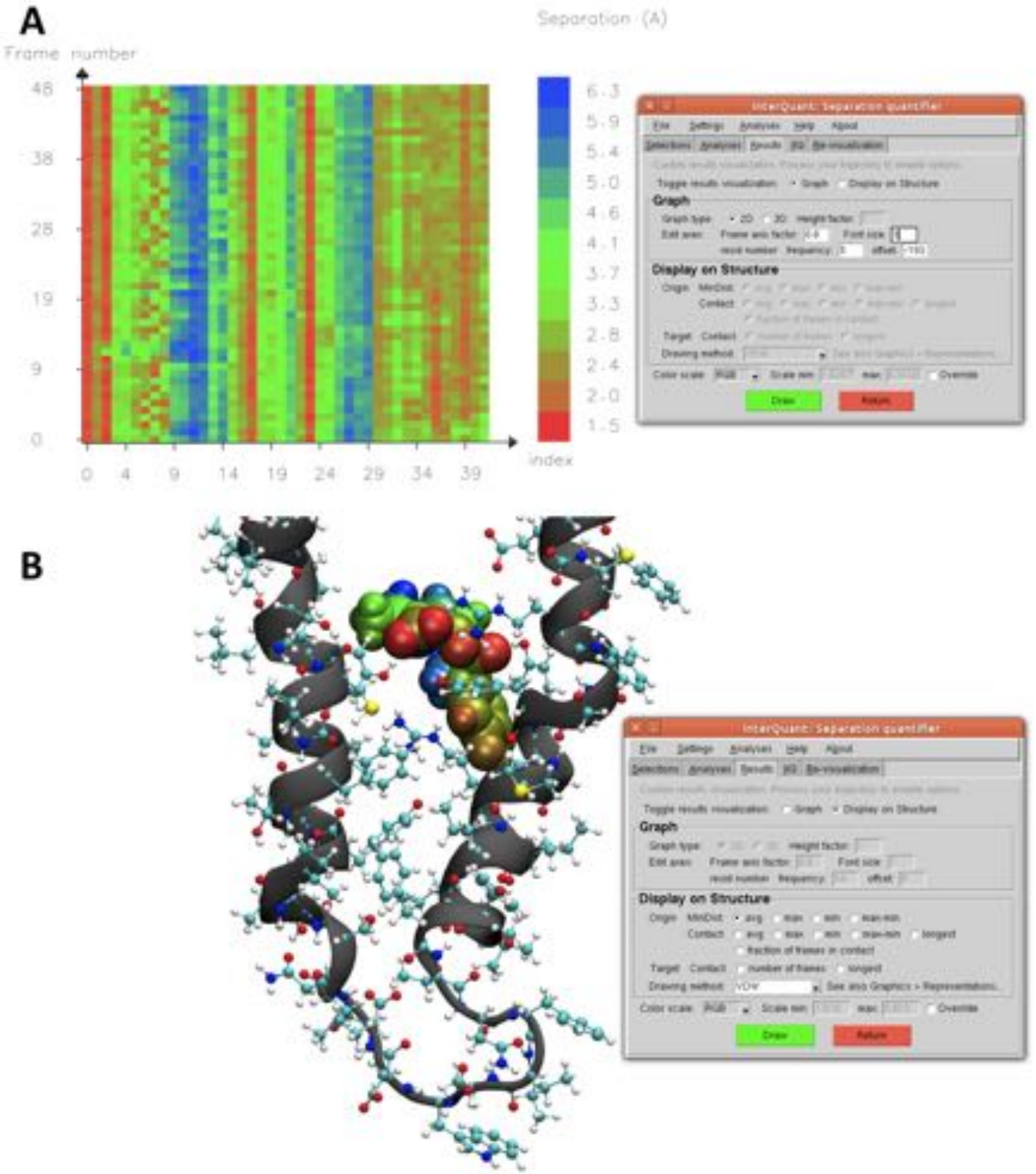

Figure 4.2.5.: Visualization of IQ results in the VMD display.
A. 2D heatmap surface of results for minimum separation to Target per OP. The GUI features several settings to customize the graph. **B.** Results for the time-average minimum separation between a human peptide transporter, PepT1, and individual atoms of the Target, a beta-lactam antibiotic called cefadroxil, is heatmapped onto the antibiotic. Results for for drug-protein interactions are intuitively displayed in its structural context. These results are also displayed quantitatively in Figure 4.2.4. MD data courtesy of Firdaus Samsudin.

Saving data

The user can choose to save IQ data at any point (in the terminal version, IQ data are saved automatically), or pre-set IQ to save results automatically upon completion of analysis. In the latter case, the user supplies a directory and an output file name. If no IQ data exist upon saving, the current frame is analysed and results are written to file. The output text file lists the input file, chosen settings and file content at the top of the file, followed by a clearly distinguishable section per analysis. Each analysis has a header that states the relevant settings that were for that analysis, and a brief explanation of the algorithm used to generate the data.

Heatmap data are saved both in text form and in Gnuplot format, wherefrom PNGs can be easily created, and the user is informed of what Terminal command to run to do this. Other analyses are available as .plot graph output, which can be viewed in e.g. XMGrace. The option to automatically save graphs of results is available from the IQ GUI menu and as a flag in Terminal mode.

4.2.3. IIQ: inference techniques for previously IQ-analysed data

Inter-InterQuant (IIQ) is the statistics engine of the InterQuant plugin. While IQ provides simple statistics for a single trajectory, IIQ allows the user to load previously collected IQ data to estimate population parameters, and evaluate whether two samples' differences are statistically significant for any collected metric. InterQuant's IIQ section of the GUI is shown in Figure 4.2.6.

T and F distributions

I translated the student's t-distribution and the F distribution from the Perl module Statistics::Distributions version 1.02 [Kospach and Trautner Kromann, 2003] to the Tcl programming language. In addition, critical t-statistic values are tabulated in IIQ for 1 to 40 df for rapid computation. InterQuant therefore supports analysis of any number of loaded IQ files.

4.2.3.1. Compatibility across time-series data with dissimilar length

In order to compare dynamics across files, data from trajectories with identical lengths can be compared on a frame-by-frame basis. Alternatively, the user can choose to pre-process time-series data. Time is then made into fractions of the total trajectory duration using a user-defined bin size. Results are averaged per bin, which allows comparison of time-series data across multiple trajectories that do not necessarily have the same length.

To bin frame data, the number of frames in each trajectory, N_{frames} is compared against the minimum number of frames that is required to populate each bin by a minimum of a single frame, $N_{bins} = \frac{1}{bin\,size}$. If all input files contain the necessary number of frames, the analysis goes ahead, otherwise the user is informed which file has insufficient length and is advised to increase the bin size to accommodate this. If the analysis goes ahead, the following algorithm is applied per trajectory:

If there is no remainder for $N_{frames}\%N_{bins}$, frame results are simply averaged per bin. If a remainder exists, a non-integer number of frames belong in each bin. In this case, the number of frames per bin is increased in ulterior bins to accommodate the remainder number of frames. In practice, $floor(\frac{N_{frames}}{N_{bins}})$ frames are averaged for the first few bins until the bin is reached where $floor(\frac{N_{frames}}{N_{bins}}) + 1$ frames is required to account for all frames. Thereafter, $floor(\frac{N_{frames}}{N_{bins}}) + 1$ frames are averaged for each remaining bin. Statistical analysis of pre-processed data then goes ahead as usual.

4.2.3.2. Hard-coded settings

Throughout, IIQ uses $\alpha = 0.05$ to test for variance homogeneity. The $H0$ is that there is no difference between samples. The two-tailed parts of the distributions are used in all cases, as it is in the interest

Figure 4.2.6.: The IIQ tab in the GUI.
In this example, IQ files are loaded into both samples and the user acknowledges that the samples should be treated as paired. A range of frames to use for the analysis 'separation to Target per origin particle over time' is set per file, and the user has specified that only proximal separations ($\leq$ 13 Angstrom) are of interest in that analysis. Each analysis will contain the 95% confidence interval for the sample difference, and only highly statistically significant differences between the samples are graphed ($p \leq 0.005$).

of the user to know which sample has the highest (or lowest) value for a given statistic.

4.2.3.3. Algorithm

The number of samples and their size (number of input files) are counted, and depending on this, IIQ proceeds down one of four routes (described below). Common to routes 1 to 3 are that input files are mined for analysis content, and statistical analysis is only performed for analyses that are common to all input files. For non-categorical IQ data, the amount of data (e.g. the number of Origin particles that observations exist for) must be identical in all files. Otherwise statistical analysis is aborted for that metric, and an error message is written to the terminal as well as stored, so that it can be written out if the user chooses to write results to file. If data for separation per Origin particle per frame exist in all input files, these data are averaged over the time specified per file by the user. If no start and stop frame numbers were supplied, or either is missing, the full number of frames is used from the input file. An overview of the IIQ algorithm is shown in the flowchart in Figure 4.2.7.

1. One sample of size larger than one: single population parameter estimation

The following results are obtained per analysis for the full sample:

- Mean per particle
  - Mean for all particles
- Standard deviation per particle
  - Mean standard deviation
- CI for the population mean (upper and lower bound for a given confidence level)

2. Two samples, whereof one sample has a sample size of one: One-sample t-test

The reference is automatically taken to be the sample with a single observation, irrespective of which field it is loaded into (c.f. Figure 4.2.6). The same types of results are obtained for the one-sample t-test as for the two-sample t-test. See below.

3. Two samples with sample sizes larger than one: Two-sample t-test

Unless the user has indicated that the samples are dependent, Hartley's test for homogeneity of variances is performed across observations. Data with insignificantly heterogenous variance ($p > 0.05$) are tested for significant difference between sample means using the standard t-test, while data with significantly heterogenous variance ($p \leq 0.05$) are tested using Welch's t-test. Dependent samples are tested using the dependent samples' t-test.

Output results per analysis are:

- Mean per particle in sample 1
  - Mean of sample 1
- Mean per particle in sample 2
  - Mean of sample 2
- Difference of sample means per particle: sample 1 - sample 2
  - Mean difference between sample 1 and sample 2
- Standard deviation of the difference between sample means per particle
  - Mean standard error of difference between total sample means

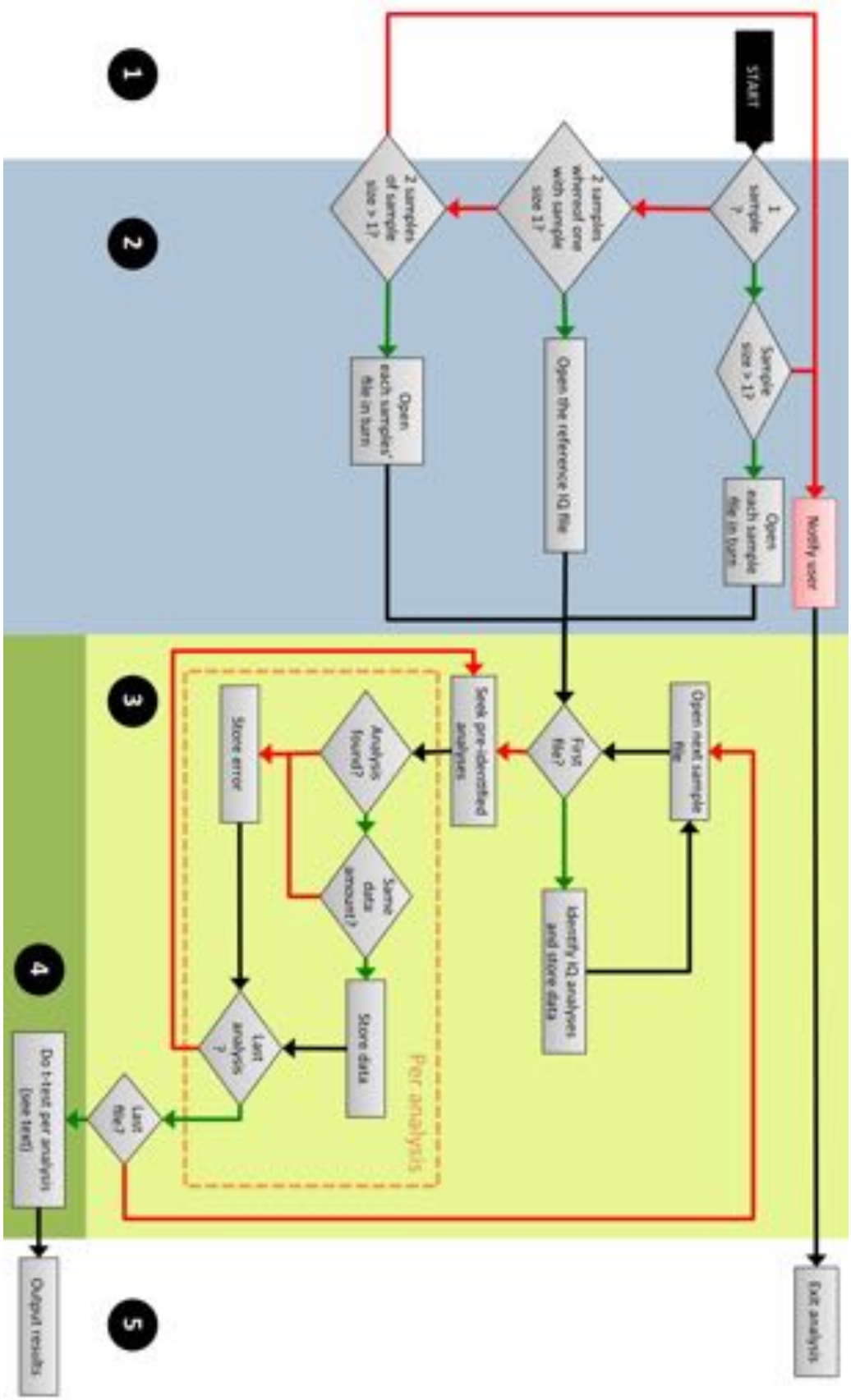

Figure 4.2.7.: Overview of the IIQ algorithm at runtime.
The flowchart is broken down into the following sections (distinguished by blocks of colour): **1.** Input and input settings submitted by user **2.** Input pre-processing **3.** Input file data mining **4.** Analysis **5.** Post-processing and results output. At conditional points in the flowchart, arrow colour denote decisions: Green = yes, Red = no. Black arrows show unconditional algorithm pathways.

- 95% CI for difference between population means: upper and lower bound per particle
  - Mean 95% CI for difference between the upper and lower bound of population means
- Raw t values of sample mean difference per particle. This can be used to retrieve exact significance if required.
  - Mean raw t value of sample mean difference per particle
- Highest probability of there being no difference between population means per particle, i.e. the highest significance as a function of t value and degrees of freedom, per particle. Significance is rounded up to the nearest one of the following values: 0.1, 0.05, 0.02, 0.01, 0.005, 0.002 or 0.001.
  - Highest significance for the mean raw t value and the degrees of freedom of the sets
- Data that display significant differences between samples are listed in descending order of significance alongside the computed difference of means and significance (this feature is only reported to text file)
- If a custom significance cutoff was applied: a list of qualifying significantly different values between sample 1 and sample 2, and the particles that they belong to, are provided in order of descending significance (this feature is only reported to text file)

4. None of the above

IIQ does not handle a single sample of sample size one, or two samples where both samples have a sample size of one. If this is loaded and IIQ is started, the user is notified via pop-up in the GUI, or by a message printed to the terminal if IIQ is run in terminal mode.

4.2.3.4. Output options

Visualization in GUI mode

Results are graphed alongside their statistical significance, and automatically presented using Multiplot. If a significance cutoff level was chosen, only differences between means at or beyond this level of significance are shown, while insignificant results are zeroed. Similarly, for separation to Target per frame data, the user can also apply a separation cutoff, above which data are not t-tested. If a molecule is loaded in VMD, IIQ tests whether IQ files' Origin selection retrieves atoms from the current system. If it does, radiobuttons that are relevant to successfully IIQ-tested analyses are enabled. Results for differences in sample means can then be heatmapped directly onto the loaded molecule from the Results tab, exactly as in IQ mode. Additionally, the user may interactively choose to only show results that qualify for a certain significance cutoff. An example of IIQ data visualization is shown in Figure 4.2.8.

Save to text file

In terminal mode, results are automatically written to file at the end of analysis. In GUI mode, results can be saved to file from InterQuant's main menu under *File > Save IIQ data...* . IIQ results do not need to be graphed before being saved to file. The text file layout is the same as for IQ text files. The type of t-test performed is stated, and if any errors occurred during analysis, such as a failed requested analysis, the error and the reason for the error is written out at the end of the file. For example, IIQ requires each input file to contribute observations towards statistical analysis. If one file does not contain observations for a particular analysis, IIQ of that analysis is stopped, and the name of the failed analysis and the file name that did not contain the required observations are printed to file, along with an explanatory note.

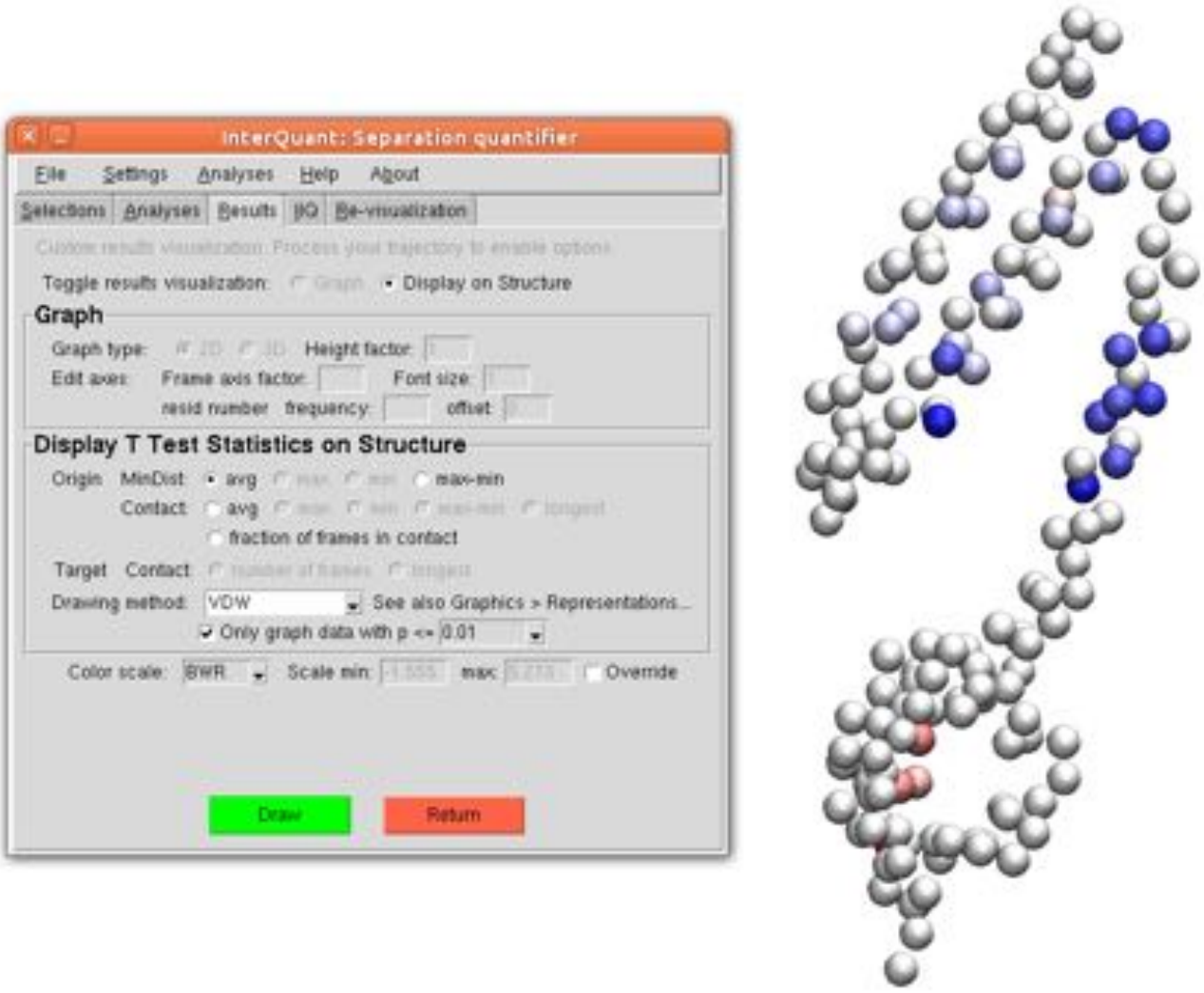

Figure 4.2.8.: IIQ data visualization
Note the change to the GUI heading and the added drop-down menu for the statistical significance cutoff, compared to the Results tab for IQ results (c.f. Figure 4.2.5). Only significant differences between samples ($p \leq 0.01$) are heatmapped unto the molecule. For clarity, only the Cα backbone is shown of the protein.

Save to graph

Both terminal and GUI modes support automatic output of gnuplot (.plot) graphs that can be opened by xmgrace.

4.2.4. Additional InterQuant feature: Visualization of user-defined results

In GUI mode, InterQuant also features a tab where users can input their own results for a selection of particles in a loaded system. InterQuant parses both the particle selection and the results, and heatmaps the selection according to results. All of InterQuant's custom heatmap editing tools are available, including choice of heatmap and heatmap scaling according to a custom range.

4.3. Conclusions

InterQuant is a tool that facilitates statistically sound quantitative analysis of geometric system interactions across any number of simulations. It features population parameter inference for one or two, dependent or independent samples, and tests for variance heterogeneity in order to use the correct statistical method. Statistically significant differences between samples can be pin-pointed to system areas at custom levels of resolution and visualized directly onto the analysed system, making analysis highly intuitive. It can also compare MD ensemble data to crystallographic data containing unidentified electron density, to spot statistically insignificant differences, thus clarifying the source of the electron density.

Statistical analysis is performed on a range of proximity-based metrics that are evaluated per input system. These metrics rapidly filter, categorize and quantify contacts and exact separation between custom selections, which makes further system investigation easier. This is necessary as progress in the MD forcefield as well as both software and hardware development means that today's MD systems are on the verge of accessing (or already doing so) physiologically relevant complexity, time and length scales [Lindahl and Sansom, 2008], thus posing a huge logistic problem for analysis.

InterQuant is a complex software composed of 17,000 lines of code. This was necessary in order to make it the intelligent, user-friendly and customizable tool that it is. It integrates with VMD, where it supports all VMD input file formats and regular expressions, has no dependencies and runs on most operating systems.

5. Study of the Mechanosensitive channel of small conductance at zero membrane tension

5.1. Introduction

The bacterial mechanosensitive channels of small and large conductance, MscS and MscL, are the best studied MSCs [Kung, 2005, Haswell et al., 2011, Naismith and Booth, 2012]. They gate to release osmolytes in response to high membrane tension under hypo-osmotic conditions, which saves the cell from rupture [Levina et al., 1999, Blount and Moe, 1999]. Membrane tension alone is sufficient to gate both channels; neither accessory proteins, ligands nor a cytoskeletal network is necessary [Sukharev et al., 1993, Hase et al., 1995] , which greatly facilitates their study both experimentally and in an MD setup. Whereas only non-conducting states of MscL have been crystallised [Chang et al., 1998, Liu et al., 2009], the *E.coli* MscS has been been crystallised in both an open and closed conformation, which makes it particularly useful for exploring the MSC gating mechanism [Bass et al., 2002, Wang et al., 2008, Lai et al., 2013].

Previous MD studies of MscS have explored conductivity [Anishkin and Sukharev, 2004, Sotomayor and Schulten, 2004, Sotomayor et al., 2006, Spronk et al., 2006, Sotomayor et al., 2007, Vora et al., 2006], conformational change as a result of unrestrained simulation [Spronk et al., 2006, Vásquez et al., 2008a], voltage [Spronk et al., 2006, Sotomayor et al., 2007], surface tension and steered MD [Sotomayor et al., 2006, Anishkin et al., 2008b, Belyy et al., 2010a], and the role of the cytoplasmic domain [Sotomayor et al., 2006, 2007].

Lipid interaction with MscS has been investigated to an extent. Using MD methodology, Malcolm et al. [2011] identified key lipid-interacting residues by their interaction energy minima using MD, but the system does not include the cage or POPG lipid, and data rely on a 10 ns simulation of the closed state only. Moreover, MscS is erroneously positioned with respect to the membrane when energy minima are computed, with unknown consequences for results. Sotomayor and Schulten [2004] mention annular lipid positioning in passing but do not elucidate beyond qualitative snapshots of the MD system. Using experimental techniques, Nomura et al. [2006] replaced hydrophobic residues by polar ones and measured pressure sensitivity and downshock survival. This mutagenesis study was followed by another where the importance of charged residues was investigated [Nomura et al., 2008] (for an overview of results, see Table 1.5.1 and 1.5.2 in the Introduction). While these data in some cases point to important sites of lipid-protein interaction (and sometimes protein-protein interaction), they do not offer molecular detail.

This chapter aims to present a detailed account of lipid behaviour about MscS using MD methodology. A comparison of lipid interaction to both closed and open states of MscS should clarify how MscS is able to gate by membrane tension alone, and has never previously been done. The data presented here are the result of the longest simulations of MscS ever performed, over 6 μs in total, to allow lipid-protein contacts to properly equilibrate prior to analysis. These data were used as input to InterQuant to deduce statistically significant differences in lipid positioning between states.

5.1.1. New open structure with resolved electron density that is hypothesised to be lipid

While an open structure has been published, this study uses an unpublished structure that was recently resolved by my experimental collaborators. It agrees very closely with the published open

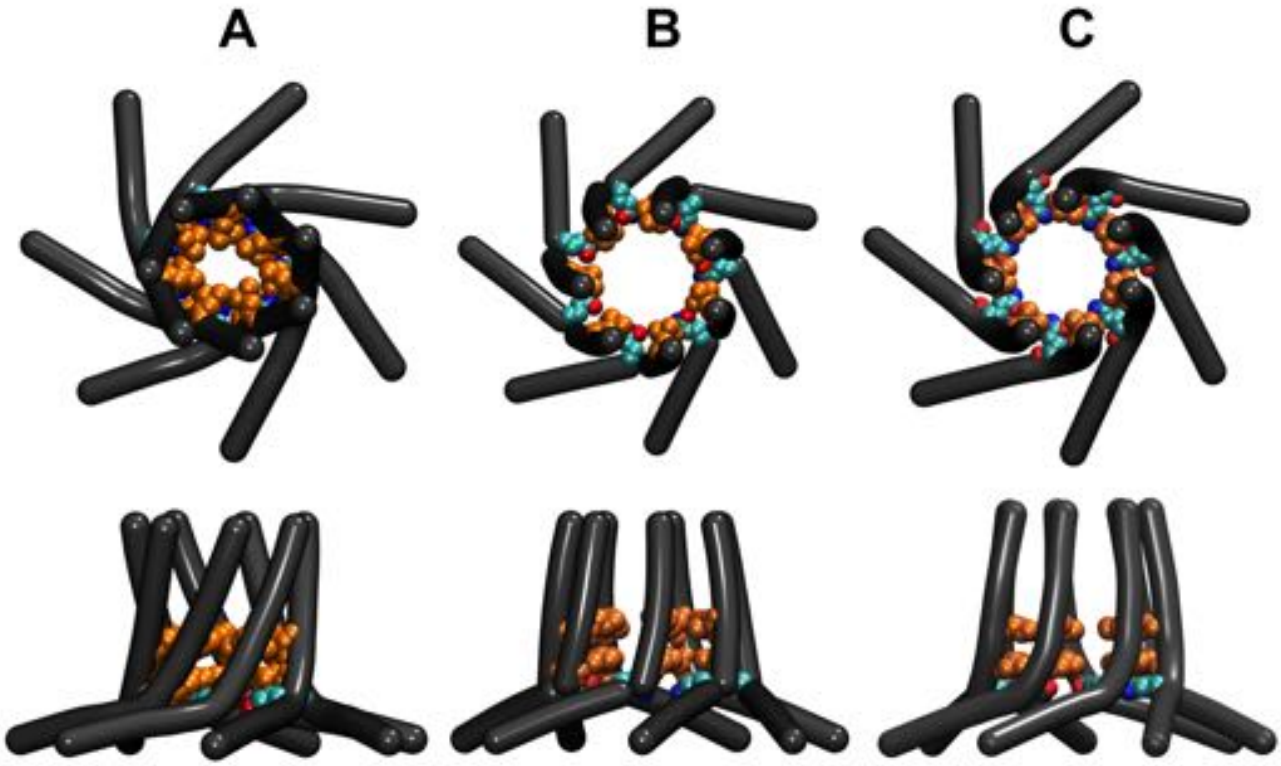

Figure 5.1.1.: Differences in the gate region across structures.
Top row: view down the pore from the pericellular space. **Bottom row**: side-view. **A.** The closed structure. **B.** The published open structure **C.** The new open structure. Pore-lining helices are shown as black bendices. The hydrophobic gate, made up of two rows of Leucines 105 and 109, and Glutamic acid 112 is shown as space-filling representations. Leucines are shown in orange while Glutamic acid is shown in turqoise, red and blue, denoting hydrocarbon and the carboxyl group, respectively. Glutamine 112 protrudes into the pore of the new open structure (**C**), narrowing the pore diameter at this residue by approximately 0.5 Å compared to PDB id 2VV5 **(B).**

structure in terms of conformation, but features electron density that could be lipid. This study will therefore also examine whether MD results supports that captured electron density is lipid.

5.1.1.1. The location of electron density supports the lipid origin hypothesis

The hypothesis that density could be hydrocarbons is consistent with the presence of alkyl chains in the crystal setup, and the electron density's location adjacent to the hydrophobic 'ceiling' of the cytosolic cage, formed by the amphipathic helix TM3b. With respect to the central axis through the channel, electron density that is not of MscS origin is positioned peripheral to the hydrophobic gate (see Fig. 5.1.2). Electron density that is located deeper in MscS is resolved in more chains than peripheral density. The innermost density (with respect to the pore axis) is present in all chains, followed by ID 500 which is present in all but one chain. Five out of seven chains have either ID OCT or D10 present most peripherally. This suggests that their interaction with MscS is less stable than that of the other densities. Results from this chapter will help resolve this matter.

The lower TM3 region contacts all electron density InterQuant analysis reveals that residues 106 to 122 are within 6 Å of all alkyl-modeled electron density that is found in the new open structure. These residues line the cytosolic part of the cavity formed between the sensor paddle and TM3, as shown in Figure 5.1.3, and encompass most of the structure that comes together to form the

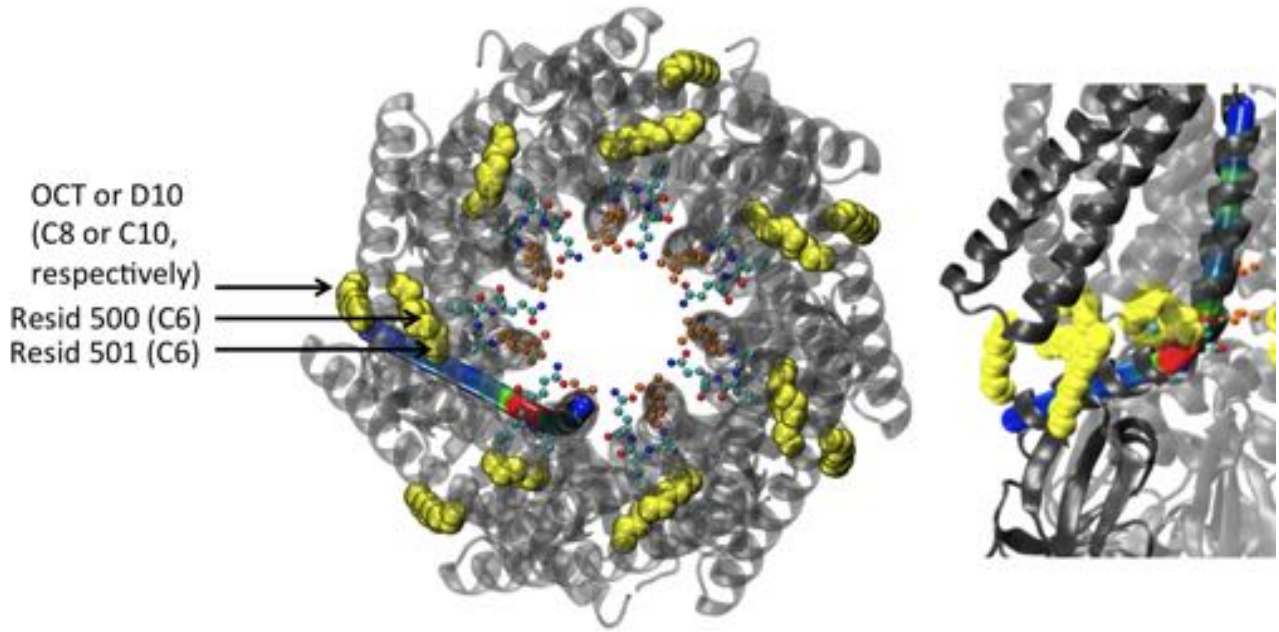

Figure 5.1.2.: Location of electron densities, modeled as alkyl chains, in the new open structure TM. View down the pore from the pericellular space (centre image) and from the side (right-most image). Alkyl chains are shown as yellow space-filling representations, with arrows indicating their ID and alkyl length in parenthesis. The hydrophobic gate, formed by Leucines 105 and 109, is shown in orange and polar Gln112 in atom-type indicative colour, as employed in Figure 5.1.1. Chain A's pore-lining helix is shown in angle-indicative bendix representation to indiate electron densities' proximity to the gate. The rest of the TM is shown in transparent gray cartoon.

hydrophobic gate. I use this subset in contact analysis and will hereafter refer to it as the *lower TM3 region*.

5.1.2. Assumptions

This chapter operates on the assumption that the states captured by crystallisation are the closed and open states of MscS. Differences in conformation and interaction with lipid between the two structures - the latter is particularly crucial for a lipid-gated protein such as MscS - should therefore reveal insights into the gating mechanism of MscS. Notably, all three crystal structures share the same TM3 kink magnitude and location (see Appendix D.1), so assuming that the crystals portray the true open and closed state, the conformational change associated with gating must occur elsewhere.

5.1.3. Previous CG models of MscS

This chapter adopts a multiscale approach to modeling MscS, that involves both AT and CG representations. Two studies have employed CG methodology to study MscS previously. Sotomayor et al. [2006] used BioMOCA to study electrostatic and conduction properties of MscS for 100 to 200 ns, where protein atoms were immobilised point charges and membrane is modeled as a dielectric where substituent atoms are not resolved [Van Der Straaten et al., 2005]. Protein and membrane dynamics were therefore not treated, which makes the study incomparable to this one.

The aim of the second study by Li et al. [2009] was not to study MscS, but rather to engineer an in silico toy model that was based on MscS and widened its gate by tension application, but was very low resolution. Only three particle types were employed to model the protein, the cage was excluded, and rigid rows of particles of varying length represented helices. The kink which subdivides TM3 is missing.

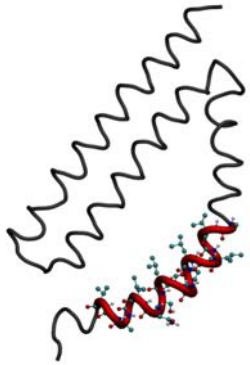

Figure 5.1.3.: Location of lower TM3 region
A single chain from the closed structure is shown in ribbon representation. Lower TM3; residues 106 to 122, is coloured in red and its constituent residues are shown in atom type coloured CPK.

5.1.4. Organisation of this chapter

This chapter starts by a qualitative overview of membrane behaviour about MscS, and distinguishing features between the two states. These differences are then explored quantitatively in order to relate lipid positioning to a gating mechanism. Metrics that show significant differences between states at the level of individual residues are presented and compared, followed by contact analysis in the form of cumulative radial contact distributions. Lasting contacts between protein and lipid are distinguished from temporary contacts, which leads to the identification of a persisting population of lipids in MscS, and their relation to resolved lipid in the new crystal. The molecular origin of both such persisting and temporary contacts is explored in the light of new, unpublished experimental data on lipid selectivity. Last, I explore my collaborators crystallographic and biophysical data in the light of my MD results. This is used to construct a new gating mechanism that is discussed in the light of published mutation data.

5.2. Methods

5.2.1. Simulated Protein

PyMol's *Mutagenesis Wizard* was used to mutate the new open structure's nitroxide spinlabel at residue 67 back to wildtype aspartic acid in the new, open crystal structure. The spin label was removed, and because it was large compared to the wild-type aspartic acid, no steric clashes resulted as a result of restoring the wild-type residue. Residues 27 to 280 were used for simulation of both the closed 2OAU structure and the new open structure.

5.2.2. CG–MD Simulations

Protein structures were coarse-grained (CG) using the MARTINI version 2.1 force field [Monticelli et al., 2008]. An elastic network model was applied to the protein Cα particles, using a distance cutoff of 7 Å and a force constant of $10\,kJ\,mol^{-1}\mathring{A}^{-2}$ [Scott et al., 2008]. MD simulations were performed using Gromacs (version 4.5.5) (www.gromacs.org) [Van Der Spoel et al., 2005].

A physiological bilayer

A bilayer self-assembly protocol [Scott et al., 2008] was used to produce five 4:1 POPE:POPG bilayers. 600 POPE, 150 POPG, 150 Na^+ and water molecules were randomly positioned in a 160 Å × 160 Å × 160 Å periodic box. Each simulation was run for 50 ns, with the bilayer generally forming within 10 ns. The final membrane had dimensions 152 Å x 152 Å.

Merging protein and bilayer

To assess the interactions of lipid with the MscS channel, preformed 4:1 PE:PG bilayers were aligned with the TM region of the protein, and lipids that overlapped the protein were removed. The resultant positioning of MscS relative to the membrane was in good agreement with the proposal by Bass *et al*. [Bass et al., 2002].

Solvation

The system was solvated by approximately 18,000 water particles, charge neutralised and a further 0.15 M of Na+ and Cl- were added for correspondence to the *E.coli* cytosolic environment.

CG equilibration and simulation duration

Each system was equilibrated for 10 ns during which the protein was position restrained, followed by simulation for 1 μs, with a total of five simulations performed for each MscS state.

5.2.3. CG to AT conversion and simulation

The final coordinates of the CG systems were converted to atomistic models using a CG to atomistic (AT) protocol [Stansfeld and Sansom, 2011]. This was done successfully for all CG simulations of the new open state, and for four out of five CG simulations of the closed state. The last simulation of the open state could not be converted because of clashes within the system, that gave rise to high energy that could not be circumvented by energy minimisation. Each resulting AT system was composed of approximately 275,000 particles.

5.2.3.1. Simulation setup

Atomistic MD simulations were performed with Gromacs using the GROMOS96 [Scott et al., 1999] 53a6 force field and the SPC water model [Hermans et al., 1984]. Simulations were performed using semi-isotropic pressure coupling with the Parrinello–Rahman barostat [Parrinello and Rahman, 1981]. The temperature of the system was coupled to an external bath held at 310 K, using the Berendsen thermostat [Berendsen et al., 1984]. The LINCS algorithm was used to constrain bond lengths [Hess et al., 1997], and harmonic positional restraints were placed on non-hydrogen atoms using a force constant of $10\,kJ\,mol^{-1}\mathring{A}^{-2}$. Long-range electrostatics, up to 10 Å, were modeled using the particle mesh Ewald (PME) method [Darden et al., 1993], and a cutoff of 10 Å was used for van der Waals interactions. Gromacs was used to add hydrogen atoms to the models.

5.2.3.2. Equilibration and simulation under partial restraints

Following 1 ns equilibration where the protein was positionally restrained, each system was then simulated for 100 ns, during which coordinates were saved every 10 ps for analysis. Analysis of the simulations was performed using InterQuant (see Chapter 4).

In order to equilibrate protein–lipid interactions prior to analysis, 0.5 to 1us of CG simulations and 50-100 ns of subsequent atomistic simulations underwent analysis. This is conservative, as the relaxation time for the fraction of the TM domain in contact with lipid is reached within 25-50 ns of simulation (see Fig. 5.2.2).

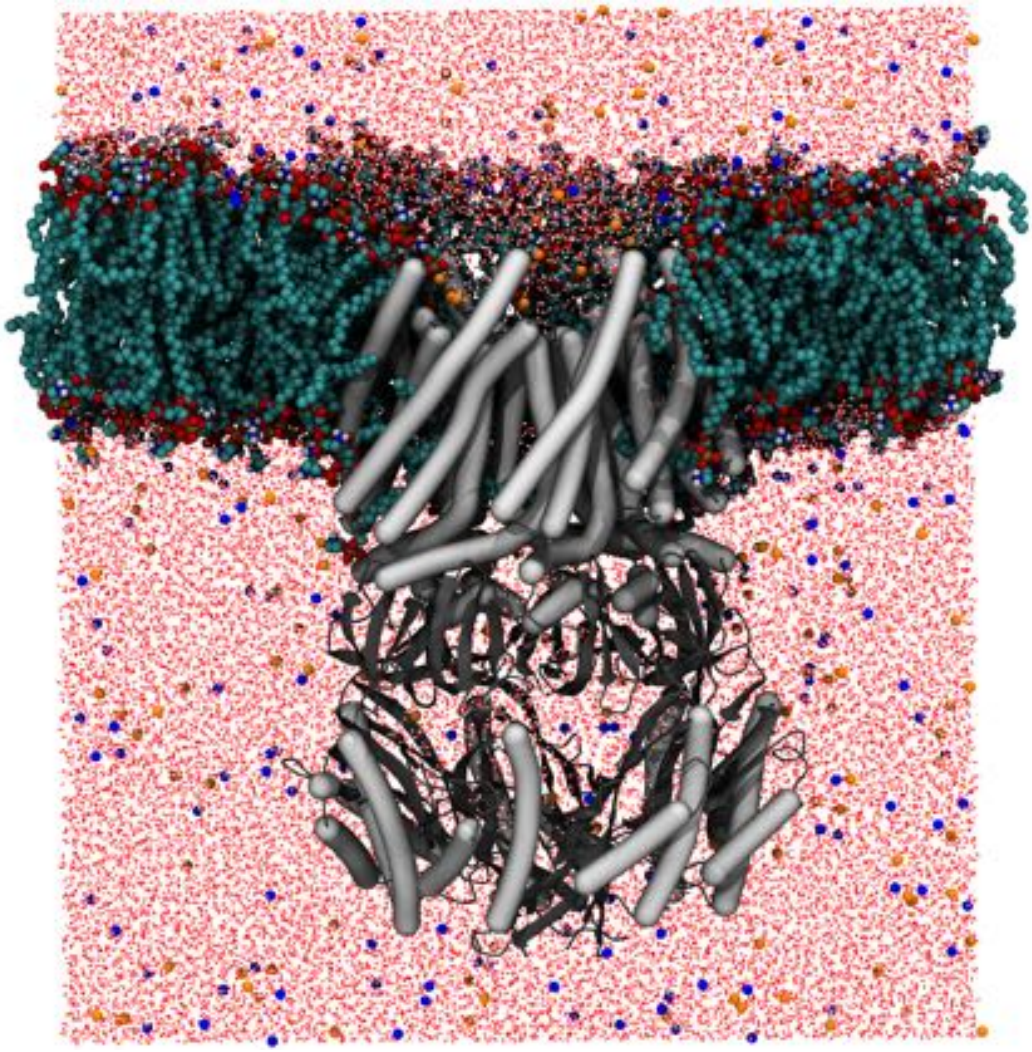

Figure 5.2.1.: New open MscS AT system after conversion from CG
MscS in gray bendix representation, lipid in particle-type coloured space-filling representation, solvent waters as red and white sticks, Na^+ in blue and Cl^- as orange spheres. Membrane and solvent was partially removed to visually expose the protein.

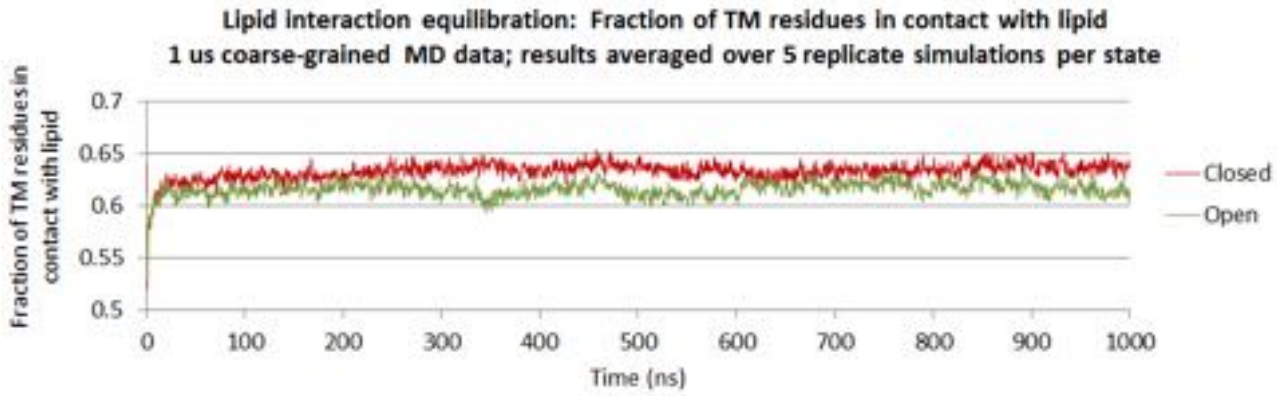

Figure 5.2.2.: Lipid-protein interactions are well equilibrated in time for analysis
Graph of the fraction of TM residues that are in contact with lipid during 1 μs CG MD data. Results were averaged over five replicate simulations per state.

5.2.4. Analysis

5.2.4.1. Proximity and contact analyses on MD data

Sampling rate for proximity studies

Proximity to lipid, as well as number and duration of lipid contacts to MscS were analysed using a sampling rate of 4/ns for atomistic data and 0.4/ns for coarse-grained data, and averaged across simulations. This sampling frequency generated 200 data points per simulation, which was deemed sufficient for the purpose of the analysis.

Contact cutoffs used

A distance cutoff of 6 Å, as used in previous CG studies [Hall and Sansom, 2009], was used to define contacts made between lipid and the MscS channels during the CG simulations. The corresponding value for atomistic simulations was 5 Å. This lower contact cutoff was chosen to acknowledge the smaller radii of AT particles compared to the corresponding CG system.

For comparison between the number of lipids that interact with closed and open MscS crystal structures, unique lipids within 12 Å of residue 106 to 122 on all chains were identified after each system had undergone 1μs CG MD. This was followed by 100 ns atomistic MD for investigation of the atomistic structure's interaction to lipid.

To quantify specifically and unspecifically bound lipid, proximal (6 Å) and distal (12 Å) contacts were analysed for the full lower TM3 region (residue 106 to 122) and the full TM (residue 27 to 128). Proximal contact number, duration and exact separation to lipid were also computed per TM residue.

Lipid headgroup selectivity per TM residue

The last 50-100 ns were used from each 100 ns AT trajectory. Contacts within 4 Å to lipid headgroups were sought per residue, as this distance permits both hydrogen bond and saltbridge formation [Barlow and Thornton, 1983]. If at least one sought lipid headgroup part was within (and including) cutoff, the residue was deemed to be in contact in that trajectory frame. Each residue was evaluated across all seven monomers, and their average contact value computed. The fraction of time that each residue was in contact with the lipid headgroup (not necessarily continuously) is evaluated at the end of the trajectory. 'Strong' contacts were defined as contacts that lasted throughout at least two thirds of the simulation. A 'significant' contact difference was defined as a difference with $p \leq 0.01$, and is reported alongside its standard error of sample means.

Statistics settings

Where analysis was performed per residue, results were averaged across chains per trajectory, and data were collected per state (closed and new open). Significant differences per residue between open and closed conformations were evaluated in InterQuant using a two-tailed Student's T test, as described in chapter 4. A cutoff of p $\leq$ 0.01 was used, and significant differences were mapped unto the closed crystal structure.

5.2.4.2. Comparison between MD and experimental data

Proximity comparison between MD and crystal structure data

For comparison between MD data on lipid interaction and the crystal structure's interaction to electron density modeled as alkyl chains, residues 27 to 128 of the new open crystal structure were aligned to atomistic MTS67Asp that had undergone 1 μs CG MD followed by 100 ns atomistic equilibration.

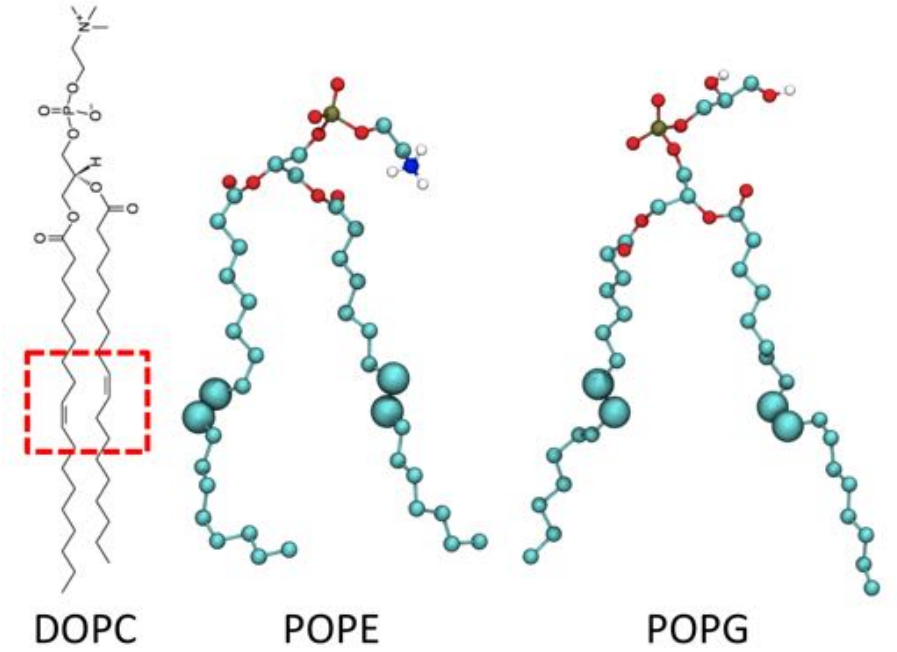

Figure 5.2.3.: Comparison of the DOPC bromination site and the carbons used in POPE and POPG For quenching experiments, DOPC is brominated at the unsaturated alkyl bonds marked in red. DOPC diagram courtesy of Avanti Lipids. POPE and POPG are shown as CPK representations, with particle-specific colouring as follows: alkyl in turquoise, oxygen in red, phosphorus in brown, nitrogen in blue and hydrogen in white. The site that best correspond to the experimental bromination sites are enlarged. These were subsequently chosen for separation and contact analysis in AT MD.

Details of experimental quenching data, and approximations used for comparison with MD data

Dr Tim Rasmussen performed brominated DOPC quenching experiments on MscS tryptophan mutants to identify the location of alkyl chains with respect to protein. Quenching data were obtained from wildtype MscS (apart from the Tryptophan mutation), so the protein is likely to have been in the closed state during experiments. These experimental results were compared to InterQuant data from Bond coarse-grained MD simulation results of the closed state in 4:1 POPE:POPG bilayer. These lipids are not brominated, nor do they have two oleyl chains, so the approximate site where the bromine group would have been attached was used in analysis. This corresponded to the site of the unsaturated bond in the oleyl chain, and the middle particle in the palmitoyl chain (this corresponds to CG particles C3A and D3B in Fig. 2.1.2). Per residue, the separation to the site of lipid bromination and the number of unique bromine groups within cutoff 6 Å were retrieved and averaged across chains and simulation time.

For AT MD analysis, separation was analysed to the two alkyl carbons that approximate those that participate in the double bond in DOPC, see Fig. 5.2.3.

Derivation of quenching data from MD proximity data

Degree of quenching is a function of separation between fluorophore and quencher, and has been modeled using a hard-sphere approximation where no quenching occurs beyond a critical separation R_c [Birks, 1970]. At a distance $r \leq R_c$, fluorescence adheres to

$$\frac{F}{F_0} = e^{-\pi R_c^2 C + \pi r^2 C}$$

where $\frac{F}{F_0}$ is the fluorescence in the presence of quencher as a fraction of possible fluorescence without quencher present, and C is the area concentration of lipid with quencher attached [Chattopadhyay and London, 1987]. This equation is accurate for $r \leq 0.9R_c$ but underestimates fluorescence between $0.9R_c < r < R_c$. Experimental data for monobromobimane and Tryptophan support this equation, where quenching is approximately extinguished at $r = 11\mathring{A}$, suggesting $R_c = 11\mathring{A}$ [Mansoor et al., 2010]. This cutoff is supported by results for other small probes, including fluorophores [London and Feigenson, 1981, Chattopadhyay and London, 1987, Mansoor et al., 2010]. However, it should be noted that in these resports, either quencher or fluorophore are attached to its 'host molecule' by means of a flexible linker, which provides the functional group with some degree of flexibility. Parameters are therefore naturally imprecise. Reference data for the exact setup that I attempt to model; brominated lipid interacting with Trp in protein, are also lacking.

$C = \frac{0.5}{R_c^2}$ is used in Chattopadhyay and London [1987] and corresponds to only a quarter of the brominated lipid concentration in my MD setup, where all lipids are treated as though they were brominated. However, if I use MD concentration settings in the equation, this results in near 100% quenching regardless of separation. This is unrealistic and inconsistent with my collaborators' data, which show that 100% brominated lipids result in a range of values for quenching (c.f. Fig. 5.3.21). In the light of this, I use the suggested value for C; $C = \frac{0.5}{R_C^2}$.

Experimental data were reported in terms of quenching, $Q = 1 - F$. Using the parameters justified above,

$$\frac{Q}{Q_0} = 1 - e^{-0.5\pi + \pi r^2 0.00413}. \tag{5.2.1}$$

5.3. Results

5.3.1. Global protein-lipid interactions revealed by multiscale MD

5.3.1.1. Only half of the TM region is integral to the membrane

It was found that only half of the MscS TM domain actually resides in the membrane. The cytosolic half of the TM domain resides in the cytosol (c.f. Figs. 5.3.1 and 5.3.2), and this arrangement persists in both atomistic and coarse-grained system setups.

5.3.1.2. High local membrane curvature about MscS

The closed structure induces high local membrane curvature, particularly in the pericellular leaflet (Fig. 5.3.2 and 5.3.3). This curvature is also present in POPC membrane, and presents after both system self-assembly and protein insertion into pre-assembled bilayer (data not shown). The new open structure induces less curvature, and only in the pericellular leaflet (Fig. 5.3.2). This supports the proposal that the local membrane environment around MscS is highly distorted [Phillips et al., 2009]. Previous simulations suggested a smaller degree of curvature [Sotomayor and Schulten, 2004], but those simulations were shorter (< 10 ns) than the current simulations and used an unphysiological lipid bilayer (POPC).

5.3.1.3. The TM domain contains sequestered lipid and most lipid locate to the cytosolic TM region

Lipids are found as far cytosolically as between TM3b and the TM1-2 linker. Visual inspection of all simulations reveal that polar lipid headgroups remain peripheral, close to the MscS perimeter, and do not enter the hydrophobic area between chains. Lipid tails, on the other hand, commonly penetrate between chains. Lipid tails in the TM3b region often enter this space parallel to TM3b and, where space permits, curve into a normal orientation with respect to the membrane to fill inter-helical cavity

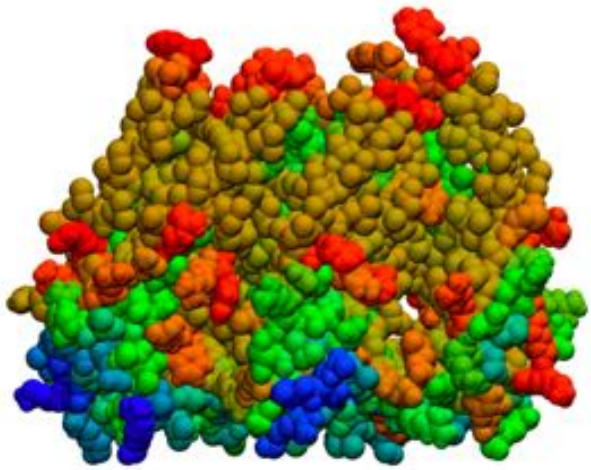

Figure 5.3.1.: Lipid phosphate group proximity to protein reveals the protein's position in the membrane.
Snapshot from the end of 100 ns AT MD of the closed structure in 4:1 POPE:POPG membrane. Residues are heatmapped to show proximity to lipid phosphate groups using an RGB colour scheme. The minimal distance to phosphate is 1.9 Å (shown in red) and the maximum distance is 10.8 Å (shown in blue). Averaged values across time, but not chains, so each residue shows its individual proximity to phosphate. The band of red around the top and middle of the closed structure reveals the location of the membrane. Analysis in InterQuant.

space, following the TM3a helix. A second lipid fits between chains, adjacent to the lipid that runs immediately parallel to TM3.

Membrane-sequestered lipids fill the inter-helical voids of the cytosolic TM region in both open and closed structures, and lipid density is highest inside the cytosolic TM. Visual inspection suggests that more lipids locate in between chains of the closed structure than the open, more compact structure (Fig. 5.3.4).

5.3.2. Investigation into the gating mechanism of MscS by comparison between closed and open states

5.3.2.1. The closed and open states' interaction with membrane, by residue

The three metrics proximity to lipid, contact duration and number of contacts, all collected per residue using InterQuant, show significant differences across the two states of MscS . Number of contacting lipids, proximity to lipid and time spent in lipid contact all decrease as one moves from protein periphery (TM1) to interior. This decrease is more pronounced for the open state. Out of all helices, TM1 has the closest contacts to lipid, with average separation at its lowest at circa 3 Å for residues Tyr27, Arg46, Asn50, Arg54, Ser58 and Arg59. The last two residues form the most C-terminal part of the helix, and makes closer contact to lipid in the open state compared to the closed state.

A drop in lipid presence can be seen at linker-regions between helices. A single lipid lines the TM3b region (c.f. Fig. 5.3.5, residues 113 onwards) and it becomes more dislodged towards the the C-terminus since separation increases and lipid spend less time in contact with residues here (this is more pronounced in the open state, c.f. **B** and **C**).

The fluctuating traces reflect the helical secondary structure of the TM domains, where half of the helix is exposed to lipid, and the other half is buried in the protein interior. The middle of TM2 is an exception to this. The graph shows that lipid can access this helix from all sides, particularly in the closed state (c.f. residues 72 to 80 in **A**, **B** and **C**). This is probably the result of TM2 lining the internal cavity formed by the sensor paddle and TM3. In the closed state, TM2 is angled diagonally across to the adjacent chain's TM3, exposing itself to contained lipid. In the open state, however, TM2

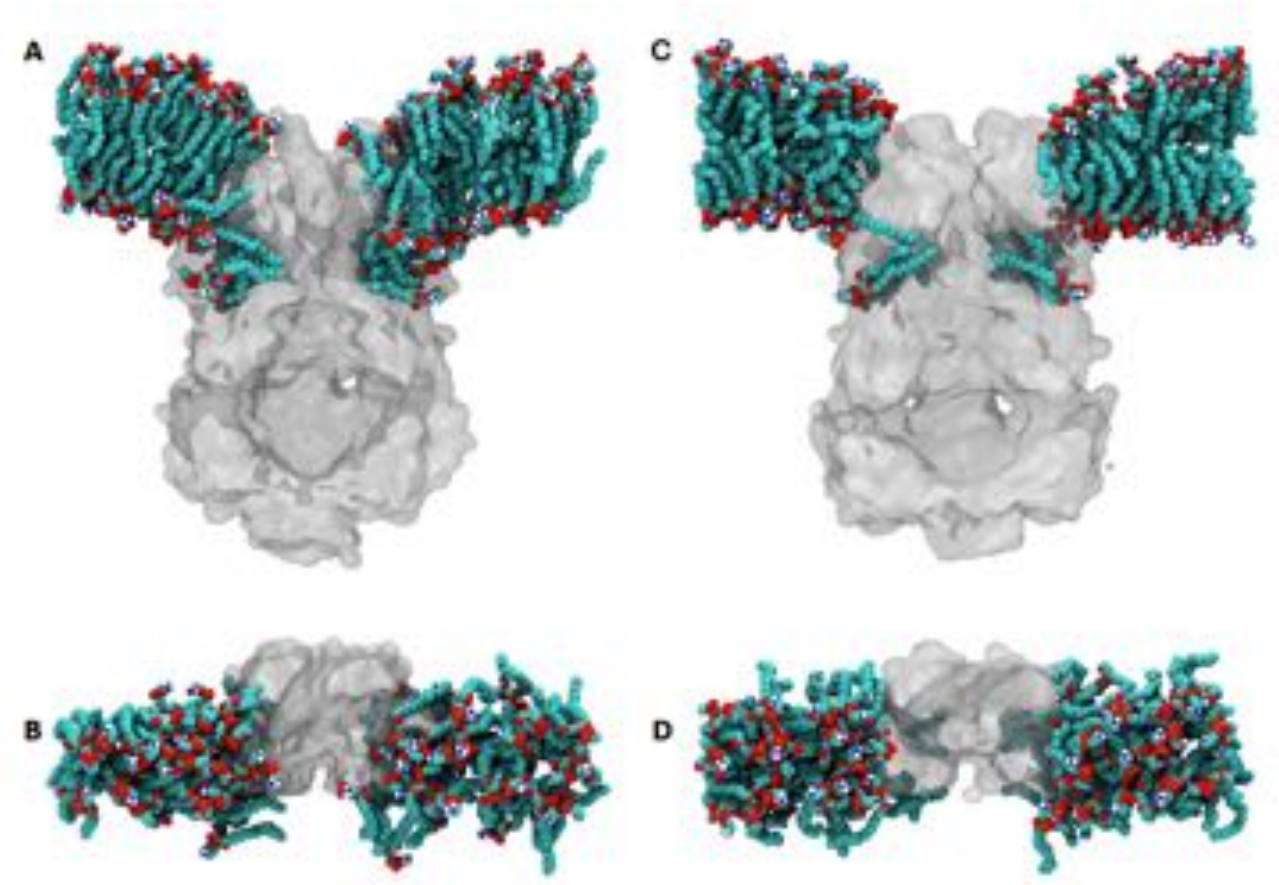

Figure 5.3.2.: High curvature about MscS in closed and open states.
Snapshots (at 100 ns) of atomistic simulations of the closed (A, B) and new open (C, D) conformation of MscS in POPE:POPG (4:1) lipid bilayers. **A.** Side view of closed MscS where half of the protein is removed to display penetrating lipids. Curvature is present about the protein in both leaflets. **B.** Top view of A from the pericellular space to show the extent of protein and lipid removed for visual purposes. **C.** Side view of open MscS where half of the protein is removed to display penetrating lipids. Curvature is only present about the protein in the pericellular membrane leaflet. **D.** Top view of C from the pericellular space to show the extent of protein and lipid removed for visual purposes.

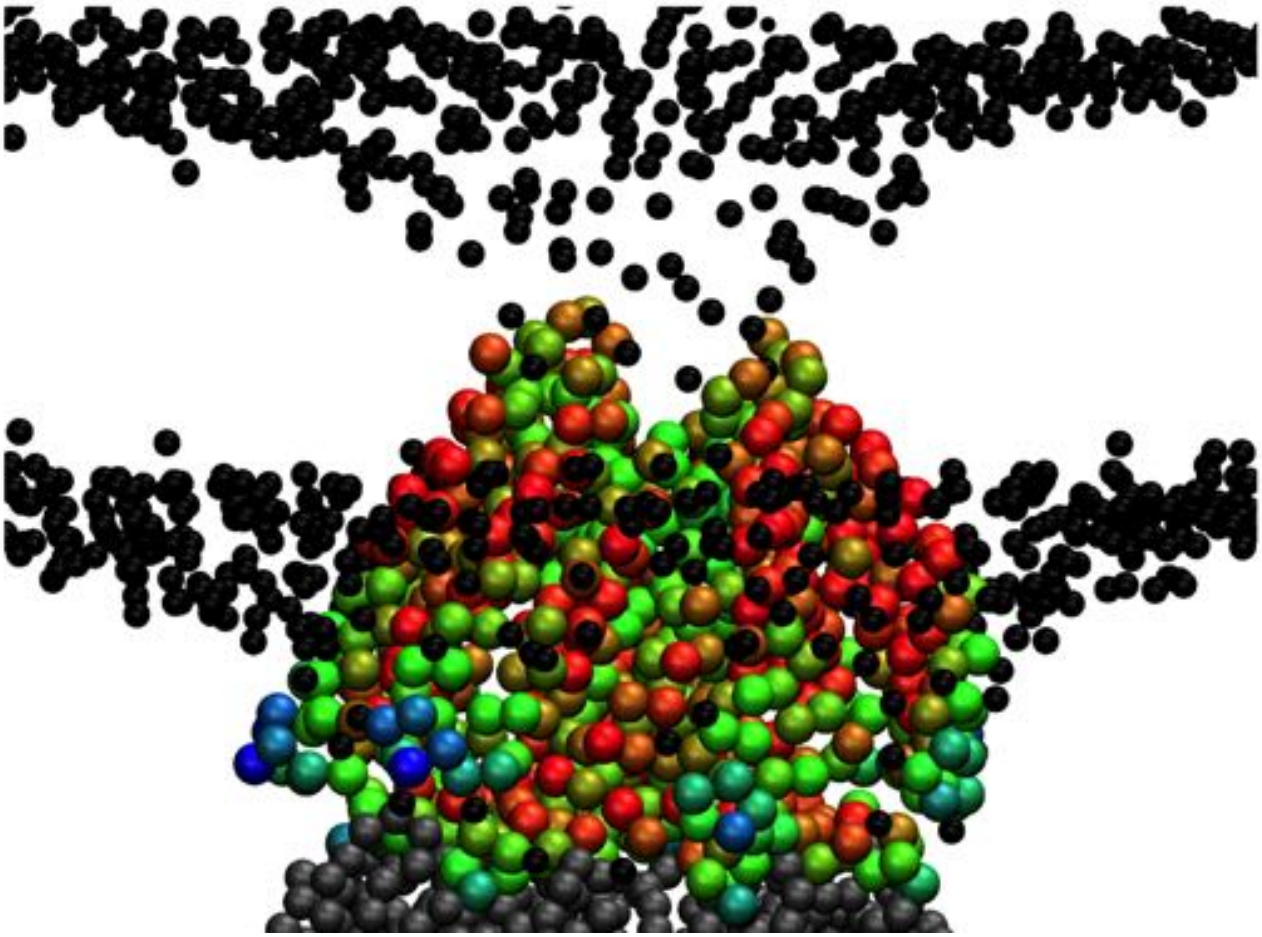

Figure 5.3.3.: Phosphate groups reveal extreme membrane deformation about MscS.
Snapshot at the end of 1 μs CG MD of closed MscS. Lipid phosphate groups, rendered as black spheres, show the extent of membrane curvature about MscS compared to the surrounding bilayer. The C-alphas of the MscS TM are shown, and these are RGB heatmap coloured according to proximity to sites of lipid bromination; minimum 5.4 Å in red and maximum 18.5 Å in blue (see Section 5.3.5.3 for more details).

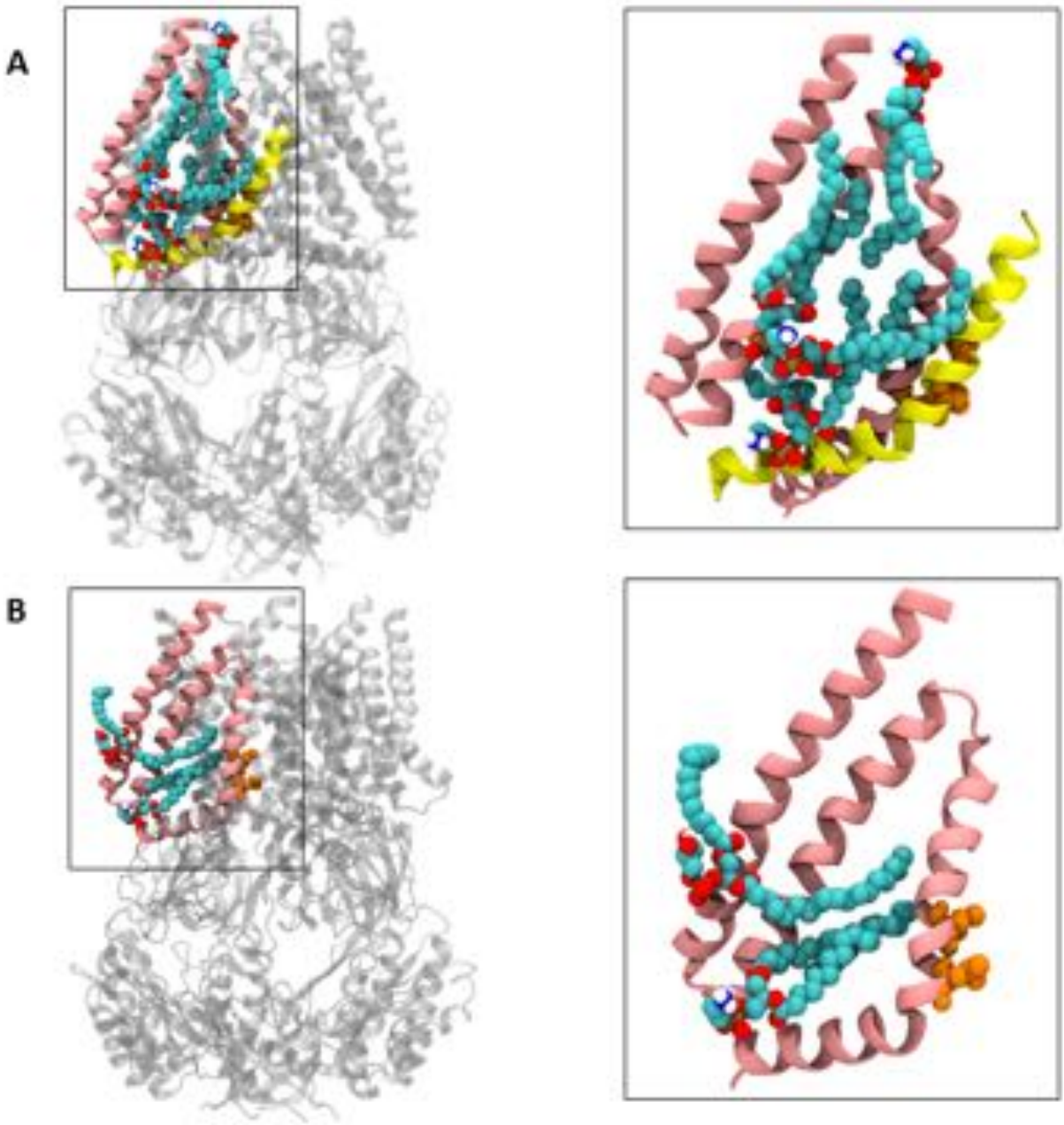

Figure 5.3.4.: Differences in lipid penetration between closed and open states.
Snapshot from the end of 100 ns AT MD of the closed (**A**) and open (**B**) states. Insets to the right show close-ups of the TM area of interest. **A.** Full view of the closed MscS structure in transparent cartoon. A section of the TM domain and contained lipid is coloured in and enlarged in the inset to show a single chain's TM domain in salmon cartoon representation, and the adjacent chain's TM3 helix in yellow, with hydrophobic gate residues in orange space-filling representation. Lipid shown in atom type coloured space-filling representation. **B:** Full view of the open MscS structure in transparent cartoon. A section of the TM domain and contained lipid is coloured in and enlarged in the inset to show a single chain's TM domain in salmon cartoon representation, with hydrophobic gate residues in orange space-filling representation. Lipid shown in atom type coloured, space-filling representation. TM3 of the adjacent chain is not shown, as same-chain TM helices are aligned in the open conformation.

is aligned with its own chain's TM3, which is probably the source of the decrease in lipid contacts. Moreover, TM2 has very high lipid proximity and fractional time in contact with lipid at residues Arg74 and Tyr75, which suggests lipid binding sites (this is explored in Section 5.3.4.2).

The reason for the pore-lining TM3a helix' exposure to lipid on all sides, only in the open state, is that lipid alkyl tails penetrate into the widened pore. This is explored in more detail in Section 5.3.3.1 .

5.3.2.2. Significant differences between states at residue level

Each residue was tested for statistically significant differences in results between states, using $p \leq 0.01$ as cutoff for significance. Significant results are shown in Fig. 5.3.6. Separation (c.f. Fig. 5.3.5 **B**) was converted to proximity, the negative of separation, for better correlation with other metrics. N.B. that fractional time in contact with lipid has a maximum magnitude of 1, so peaks appear small relative to difference in proximity and number of contacts. For another version of this data, where each metric's data set is scaled to fit the same red-white-blue heatmap, see Fig. 5.3.7. That figure should be viewed alongside Fig. 5.3.6 as it visualizes significant results with respect to the protein 3D structure. Open state data were subtracted from closed state data, so a positive value indicates that the closed state has a significantly larger value than the open state at that residue, whereas a negative value signifies the opposite. Difference with $p > 0.1$ was zeroed.

There are discernible correlations across all three metrics in terms of value and site of significant difference. This is particularly evident in Fig. 5.3.7, where colour and and intensity displays little variation across **A**, **B** and **C**. Results indicate four areas that undergo significant difference to lipid interaction as MscS transitions between states. I will consider each in turn:

- **The C-terminal TM1 and the TM1-2 linker region** comes into closer, more numerous and longer contact with lipid upon gating. This is interesting as preliminary data from my collaborators show that mutation of positive charge at the TM1 C-terminus to an apolar residue causes MscS to dysfunction.
- The change to **mid-TM2**'s proximity and fractional time in contact is low, indicating that this area remains in contact with lipid throughout gating. However, at least one lipid is lost in this region as MscS gates. This is discussed in more detail in Section 5.3.2.5.
- The fluctuations seen in **TM3a** originates in lipids' access to the pore in the open state. Therefore, significant difference does not indicate lipid gain in the open state, but rather reorientation of existing bound lipid. Interestingly, the fluctuations about zero seen in the number of contacts is indicative of the helix twist that occurs to TM3a as hydrophobic residues are turned away from the central pore axis during gating.
- **TM3b** experiences a distal loss in lipid proximity by gating. The lipid located here may be less tightly coordinated in the open state.

5.3.2.3. Cumulative radial distribution of lipids: A measure of space available in MscS

The number of unique lipids that can be fit within a given proximal radius of the protein suggests the amount of surface area that the protein exposes to lipid. Here, I investigate the number of unique lipids that are within a range of radii of MscS in different states. N.B. this method integrates the standard radial distribution function, which quantifies the number of annular molecules at a given radius.

Lower TM3
The closed state contacts significantly more lipid regardless of cutoff, but the difference is particularly significant proximally, peaking at 6 Å separation.

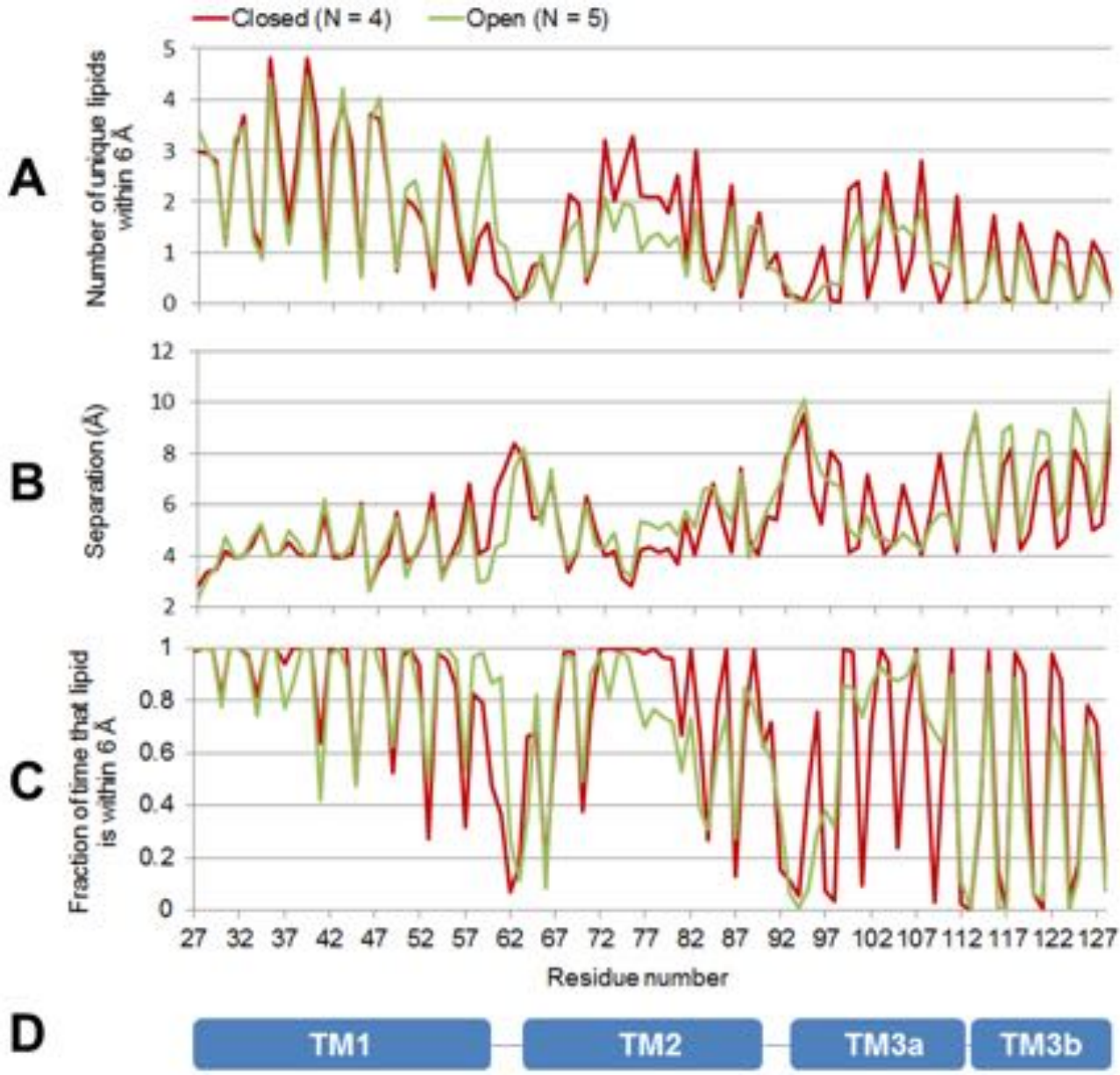

Figure 5.3.5.: Metrics that display significant differences across states.
Data from AT MD, averaged across chains, simulations and time. In all graphs, the red trace shows data from simulations of the closed state, and the green trace shows data from simulations of the open state. The x-axis is only shown in **C**, but is shared for graphs **A** and **B**. Error bars are omitted for clarity; for significant differences between states, see Fig. 5.3.6 . **A**. Average number of lipids that are within 6 Å at any point in time. **B**. Average proximity to lipid. **C**. Average fraction of time that a residue is within 6 Å of lipid. **D**. The TM domains that the residue numbers shown in **C** belong to.

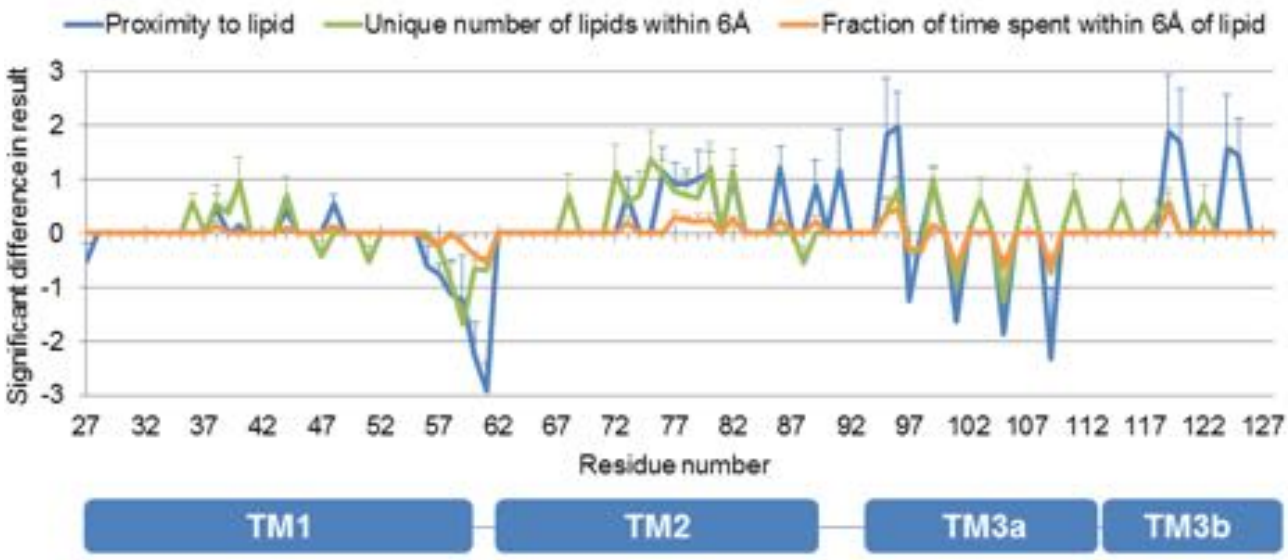

Figure 5.3.6.: Comparison across metrics with significant differences between closed and open states of MscS.

Per state, values per residue were averaged across chains, simulations and time (50-100 ns AT MD). The open structure value was subtracted from the closed structure value for each residue, and significant difference plotted (t test, $p \leq 0.01$, insignificant values were set to zero). A positive difference signifies that the closed state has a significantly larger value than the open state at that residue. Similarly, a negative value signifies that the open state has a significantly larger value than the closed state at that residue. Error bars denote the 95% upper limit confidence interval of the difference between closed and open states. Analysis performed in InterQuant. For individual values per state, rather than the value difference presented here, see Fig. 5.3.5. Refer to section 5.3.2.2 in the main text for a discussion on these results.

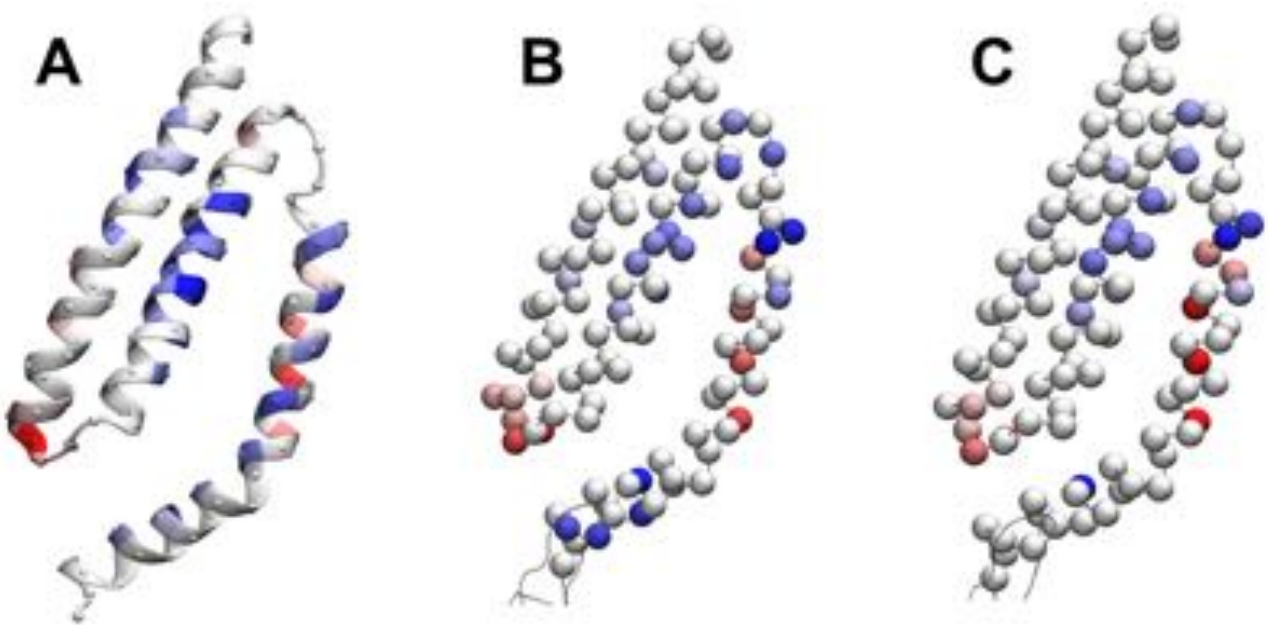

Figure 5.3.7.: Significant differences in lipid proximity and contact duration between open and closed states.

Significant differential lipid interaction between the closed and open structures investigated per residue in AT-MD ($p \leq 0.01$). Data collected from 50 to 100 ns AT MD, averaged across 4 replicates of the closed state and 5 replicates of the new open state. Results are shown as heatmapped residues onto a single chain's TM domain from the closed structure.

A. Differential numbers of contacting lipids (cutoff 6 Å). A blue residue indicates that it has more contacts to the closed state (maximum 1.7 lipids more) whereas red indicates that the open state has more lipid contacts to this residue (maximum 1.4 lipids more). **B.** Differential proximity. A blue residue indicates that it is more proximal to lipid in the closed state (maximum 2.0 Å closer) whereas red indicates that the open state is closer to lipid at this residue (maximum 2.9 Å closer). A blue residue indicates that it is more proximal to lipid in the closed state (maximum 2.0 Å closer) whereas red indicates that the open state is closer to lipid at this residue (maximum 2.9 Å closer). **C.** Fractional differential duration in contact with lipid. Values range from -0.65 (red; longer duration contacts in the open structure) to 0.49 (blue; longer duration contacts in the closed structure). Visualization by InterQuant in VMD 1.9.1.

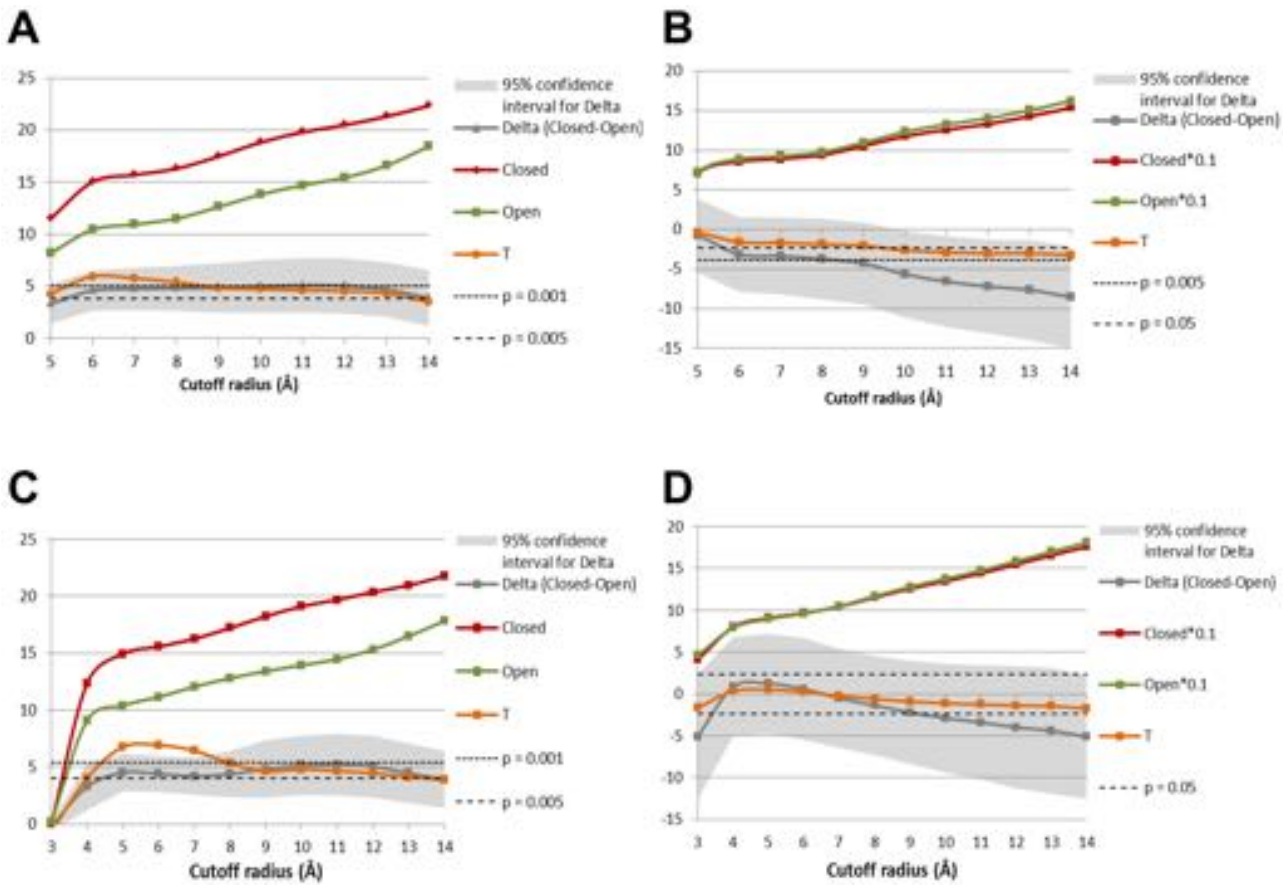

Figure 5.3.8.: Distribution of lipid around MscS
The number of unique lipids that remain within cutoff throughout the latter half of simulation; throughout 500 ns of CG data (**A**: lower TM3 region, **B**: full TM domain) and throughout 50 ns of AT data (**C**: lower TM3 region, **D**: full TM domain). Both the full TM domain (**B**, **D**) and the lower TM3 region only (**A**, **C**) were analysed. Note that for data for the full TM domain, the number of lipid contacts have been factorized by 0.1 to fit ons the graph. Both the open and closed state's results are scaled by the same amount, so are comparable. Delta (in gray), the number of lipids in contact with the open state subtracted from the number of lipids in contact with the closed state, is not scaled, so refers to the actual difference. In all graphs, T (in orange) refers to the t-value obtained from the difference between open and closed samples. Relevant significance cutoffs are shown as dashed lines; where T surpasses a cutoff, the difference is significant to that amount. Analysis made in InterQuant.

Interestingly, the difference in lipid contacts is approximately constant; 5 lipids more in the closed state, regardless of cutoff radius. Also, the increase in the number of lipids that make contact to lower TM3 is linear; circa 1 lipid per 1 Å increase in radius. AT and CG results are similar.

Full TM

Atomistic system differences between closed and open states are insignificant, regardless of the cutoff. In the CG system, the open state only contacts significantly more lipid distally, above 1nm away from the protein. This lead increases with larger radius cutoff, peaking at 8 lipids more in contact with the open state than the closed state within 1.4 nm radius.

The increase in the number of lipids that make contact to the full TM domain is linear, similar to the lower TM3 region, but with a ten times higher gradient; circa 10 lipid per 1 Å increase in radius. This can be attributed to the larger surface area of the full TM compared to the lower TM3 region.

5.3.2.4. Cumulative radial distribution of tightly bound lipids

AT data are not relevant for this analysis since lipid does not have the time to diffuse significantly in 50 ns (so analysis would show lipid as being strongly bound when, in reality, it just has not had the time to diffuse away). AT analysis data are only included to show its correlation to the number of lipids in contact (not persistent contacts), to confirm that persistent lipid analysis is not justified for AT systems (c.f. similarity in **C** and **D** between figures 5.3.8 and 5.3.9). Therefore, below analysis only refers to the CG data (**A** and **B** in 5.3.9).

Lower TM3
Proximally, circa one lipid is strongly bound per CG chain, regardless of state. Distally, the closed CG state tends to retain slightly more lipid than the open CG state does ($p \geq 0.05$). The closed state gains approximately 1 strongly bound lipid per 1 Å increase in radius between 6 and 13 Å radius, whereas the open state only binds 2 additional lipids strongly in total throughout the same radius increase. The data suggest that both states host one strongly bound lipid proximally, but that only the closed state accommodates an additional strongly bound lipid per chain distally.

Full TM domain
The closed state TM region retains significantly more lipid - approximately 1 more lipid per chain - than the open state does, irrespective of radius. Almost no difference in the number of strongly bound lipids is seen as the cutoff radius is increased, irrespective of state, suggesting that strongly bound lipids are kept in close proximity of the protein. This makes sense from an electrostatic energy perspective, as the strength of attraction between protein and lipid decreases with increased separation.

5.3.2.5. Conclusions for lipid radial distribution studies

This contact analysis reveals that the closed state's lower TM3 region contains more lipid than the same region in the open structure, irrespective of multiscaling and cutoff radius. Considering only the lower TM3 region, both states bind one lipid per chain tightly proximally, but only the closed state has the capacity to bind another lipid tightly distally. A lipid that is tightly associated distally with respect to lower TM3 in the closed state is therefore lost upon gating. This lipid may be lost as a result of movement of the sensor paddle, which is angled diagonally across to the adjacent chain's TM3 in the closed state, but is aligned with its own TM3 in the open state, where it presents less of a block to diffusion. A second reason could be the loss in cavity volume that results from gating.

Results from the full TM suggests the exact location of this lipid in the closed state; the TM tightly binds one more lipid per chain than the open state TM does, and this lipid is bound proximally. This must be the same lipid that is bound circa 1.4 nm away from the closed state's lower TM3 region. Considering that one more lipid is bound closer and for longer contacts per residue along the closed structure's middle TM2 region than in the open structure (c.f. Fig. 5.3.6), this is probably the site where lipid is bound and subsequently lost upon gating - the volume between the sensor paddle and TM3a.

5.3.3. A persisting population of lipids in the cytosolic TM region connects the sensor paddle to the gate

An unexpected outcome of the radial distribution analyses of lipid binding was that approximately seven lipids remain tightly associated with the lower TM3 region, irrespective of state. This permanently bound population of lipid appears in striking contrast to other protein associated lipids, which is clear in Fig. 5.3.10. In both the closed and open state, the majority of lipids only associate briefly with MscS. The graph tails off to the right as fewer lipids remain bound for longer periods of time, but a distinct lipid population remains bound to the protein throughout the full simulation. This population decreases in size as the closed state transitions to open, leaving lipids primarily bound to the lower TM3 region. This section explores these lipids further.

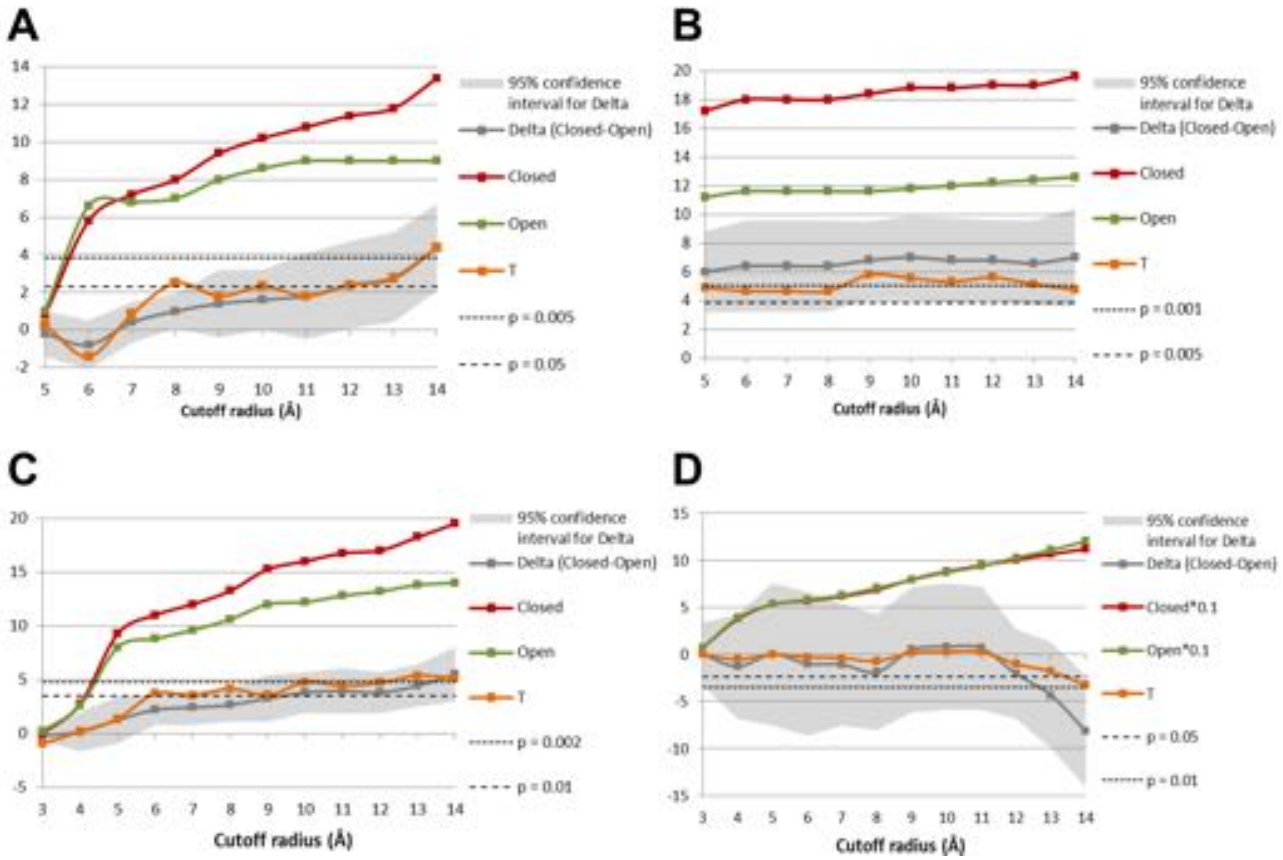

Figure 5.3.9.: Distribution of tightly associated lipid about MscS
The number of unique lipids that remain within cutoff throughout the latter half of simulation; throughout 500 ns for CG data (**A**: lower TM3 region, **B**: full TM domain) and throughout 50 ns for AT data (**C**: lower TM3 region, **D**: full TM domain). Both the full TM domain (**B**, **D**) and the lower TM3 region only (**A**, **C**) were analysed. Note that for data for the full AT TM domain, the number of lipid contacts have been factorized by 0.1in order to fit on the graph. Both the open and closed state's results are scaled by the same amount, so are comparable. Delta (in gray), the number of lipids in contact with the open state subtracted from the number of lipids in contact with the closed state, is not scaled, so refers to the actual difference. In all graphs, T (in orange) refers to the t-value obtained from the difference between open and closed samples. Relevant significance cutoffs are shown as dashed lines; where T surpasses a cutoff, the difference is significant to that amount. Analysis made in InterQuant.

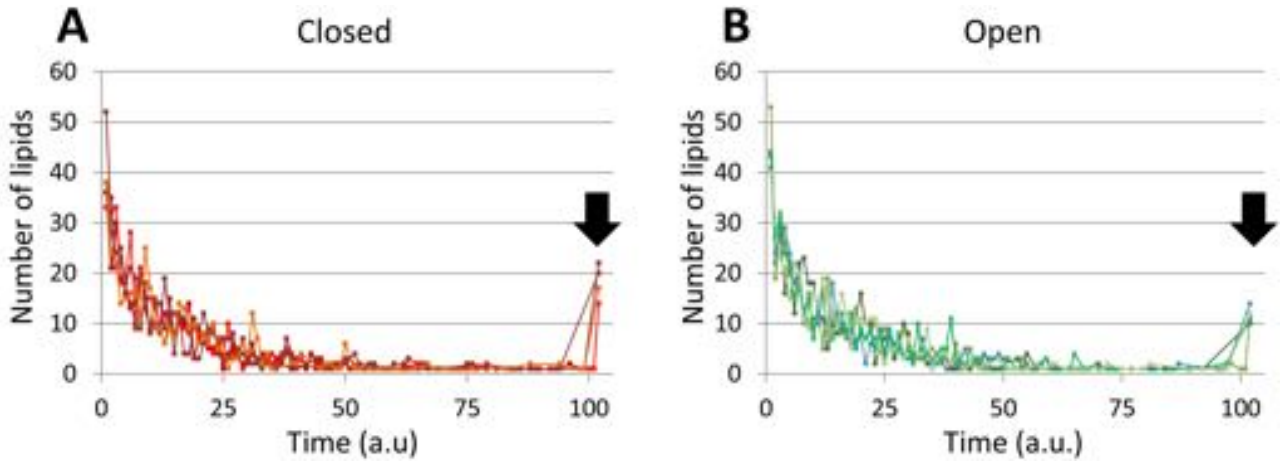

Figure 5.3.10.: Tightly and loosely associated lipid populations in MscS.
Contact duration probability graphed for POPE and POPG interaction with closed (**A**) and open (**B**) MscS during 0.5 μs CG MD (five replicate MD simulations shown per state). The contact cutoff was set to 6 Å. A distinct lipid population remains bound to the protein throughout the full simulation, and this population is larger in the closed state than in the open state (black arrows). The shape of the graph in **A** is also representative of closed MscS in POPC (data not shown).

5.3.3.1. Lipid access to the pore is only possible in the open state

Lipid that remain within 6 Å of protein throughout the last 50 ns of AT simulation was analysed for access to the pore region. I found that lipid access to the hydrophobic gate is permitted only in the open structure (Fig. 5.3.11), where distance between pore-lining TM3 helices is larger than in the closed structure. In the open structure, lipid alkyl tails contact the Leucines that form the hydrophobic gate [Bass et al., 2002] . The effect of the presence of an alkyl chain could be to stabilize the dilated open pore, but may also increase pore hydrophobicity, which would limit the access of polar solvents to the pore.

5.3.3.2. POPE lipids are overrepresented in the open state's lower TM3 region

An investigation into the type of these lower TM3 bound lipids reveals that primarily POPE binds the open state, whereas the closed state adheres to the 4:1 POPE:POPG membrane composition (see Fig. 5.3.12).

5.3.3.3. Permanent lipids are electrostatically coordinated to sensor paddle residues

POPE lipids that bind the lower TM3 region throughout the full 50-100 ns AT MD were analysed for binding sites to their headgroups to elucidate the origin of POPE specificity. POPE amine contacts are primarily made to the sensor paddle, where Aspartic acid 67 was found to be the key residue. Many contacts were also made to Arginine 59 and Lysine 60 on the cytosolic end of the sensor paddle (see Fig. 5.3.13). At this site, the open structure has more, longer and closer contacts to lipid compared to the open structure (c.f. Fig. 5.3.6).

5.3.3.4. The new open structure's electron density shows good overlap with permanently bound lipids

The separation profile for simulated open MscS and lipid bound to the lower TM3 region shows very good overlap with the separation profile for the crystal mutant and unidentified electron density (Fig. 5.3.14). This fit to crystal data is an improvement on the previous comparison to MD data, where

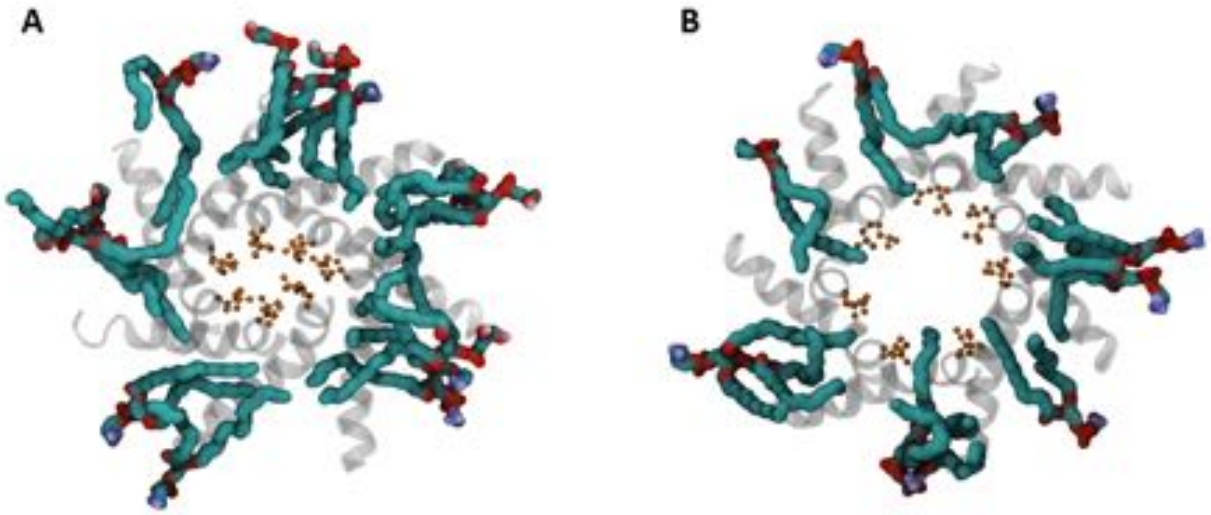

Figure 5.3.11.: Alkyl chain positioning with respect to the pore in closed and open states.
Stably bound lipids' configuration with respect to the lower TM3 region in the closed (**A**) and new open (**B**) state. Lipids that make contact (6 Å cutoff) to the lower TM3 region throughout 50 to 100 ns AT-MD are displayed in snapshots from the end of two 100 ns simulations. Bound lipids in the open structure reach the hydrophobic gate (**B**), whereas bound lipids in the closed structure is prevented from reaching the gate (**A**). Lipid is shown in atom-type coloured surface representation, lower TM3 in transparent gray ribbon and leucines that form the gate are shown in orange CPK representation.

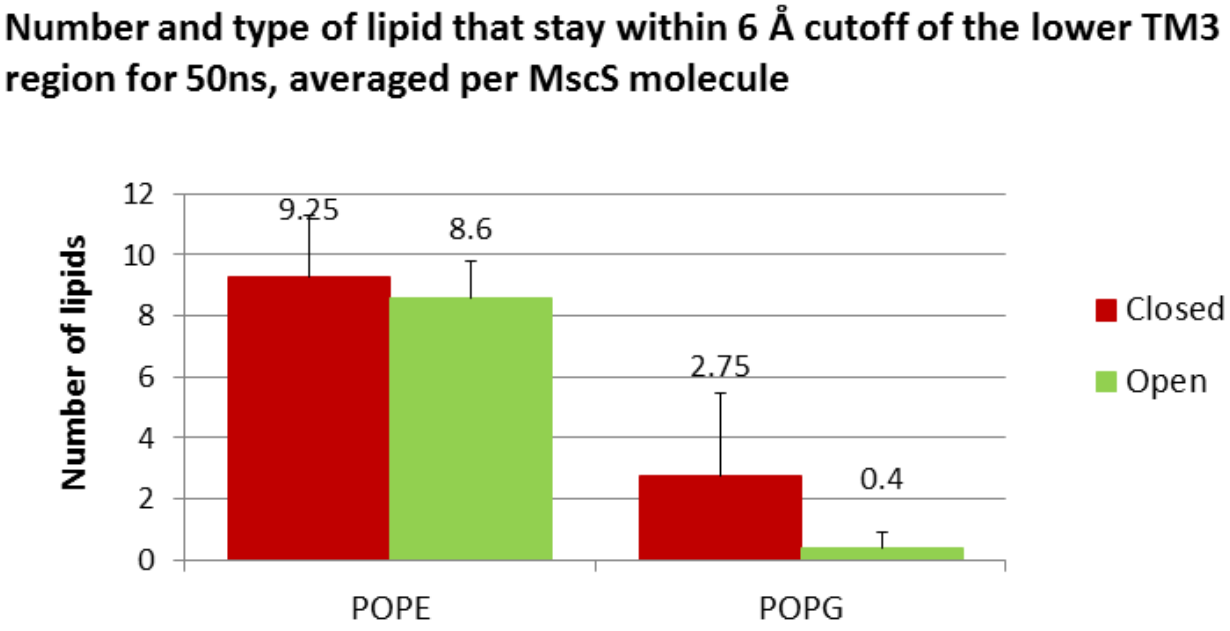

Figure 5.3.12.: Lower TM3 lipid preference
Investigation into the type of lipid that make contact to the lower TM3 region in closed and open MscS structures using AT-MD. Data collected between 50 and 100 ns and averaged over five simulations of the open structure and four of the closed structure.

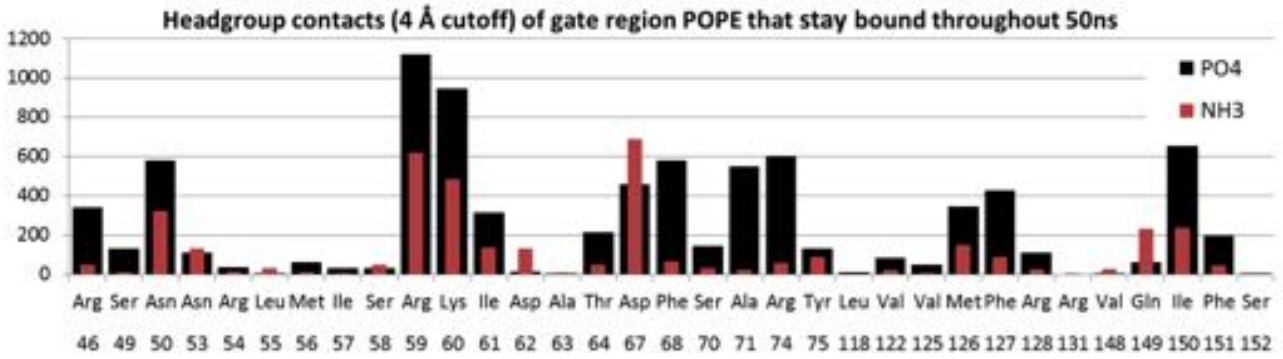

Figure 5.3.13.: Permanently bound POPE lipids are coordinated by charges along open MscS' sensor paddle.
POPE lipids that bind the lower TM3 region throughout the full 50-100 ns AT MD were analysed for binding sites to their headgroups to elucidate the origin of POPE specificity.
Results are scaled approximately linearly when a larger cutoff (5 Å) is used (data not shown).

all associated lipids were taken into account, not just lower TM3 (Fig. 5.3.18). Assuming that the unidentified electron density in the crystal is lipid, only fragments of lipid are captured, which would explain the remaining discrepancy observed between the separation profiles.

Based on the propensity for POPE to bind in this area, the separation profile suggests that electron densities are deeply buried POPE lipids (Fig. 5.3.14). Proximity to these lipids remains essentially unvaried through 50 ns of atomistic simulation (Fig. 5.3.15), indicating that binding of these lower TM3 region associated lipids is favourable and persistent, making them likely contributors to the high PE content observed in the open mutant MscS using Mass Spectrometry.

5.3.3.5. The electron density TM3b overhang is not an artifact, but is not an alkyl tail

Snapshots from AT MD show good qualitative overlap between lipid positioning and electron density in the new crystal structure (Fig. 5.3.16). The snapshot is not representative for all lipids' placements or placement across all time frames, but it supports that the electron density 'overhang' may not be a crystal artifact, as initially thought. It is likely to be a lipid headgroup, rather than the alkyl chain that it was modeled as in the crystal.

5.3.3.6. Conclusions

Lipid exits the volume between the sensor paddle and TM3 as MscS gates, but leaves lipid that lines the lower TM3 region intact. These permanently bound lower TM3 lipids make stable contacts to MscS and show very good spatial overlap with unidentified electron density in the new open crystal, suggesting that it is lipid-originated. MD data also suggest that the origin of the most peripheral electron density is lipid headgroup, not alkyl tail as previously thought, which suggests that this electron density is not an experimental artifact after all. Alkyl tails of lipids bound to the lower TM3 region enter the open pore at the level of the hydrophobic gate, but do not block the pore. Their headgroups are almost exclusively PE, which corresponds well with experimental findings that open MscS is PE-selective. Those PE headgroups primarily bind the sensor paddle, and the interaction between the amine group of POPE and Asp67 could be the origin of PE selectivity. However, PE selectivity was established experimentally for the MTS spin labeled mutant, and the spin label replaced Asp67. So Asp67 can therefore not be the origin of PE-selectivity.

Lastly, my data suggest that lipid is permanently bound to the lower TM3 region, and assuming that this is true, lipid would not be exchanged during gating. But open and closed states bind different types of lipid. Either there is no selectivity, and the significant increase in PE seen in the open state occurred by chance, or MscS is selective for PE, and the near 4:1 PE:PG composition of the lower

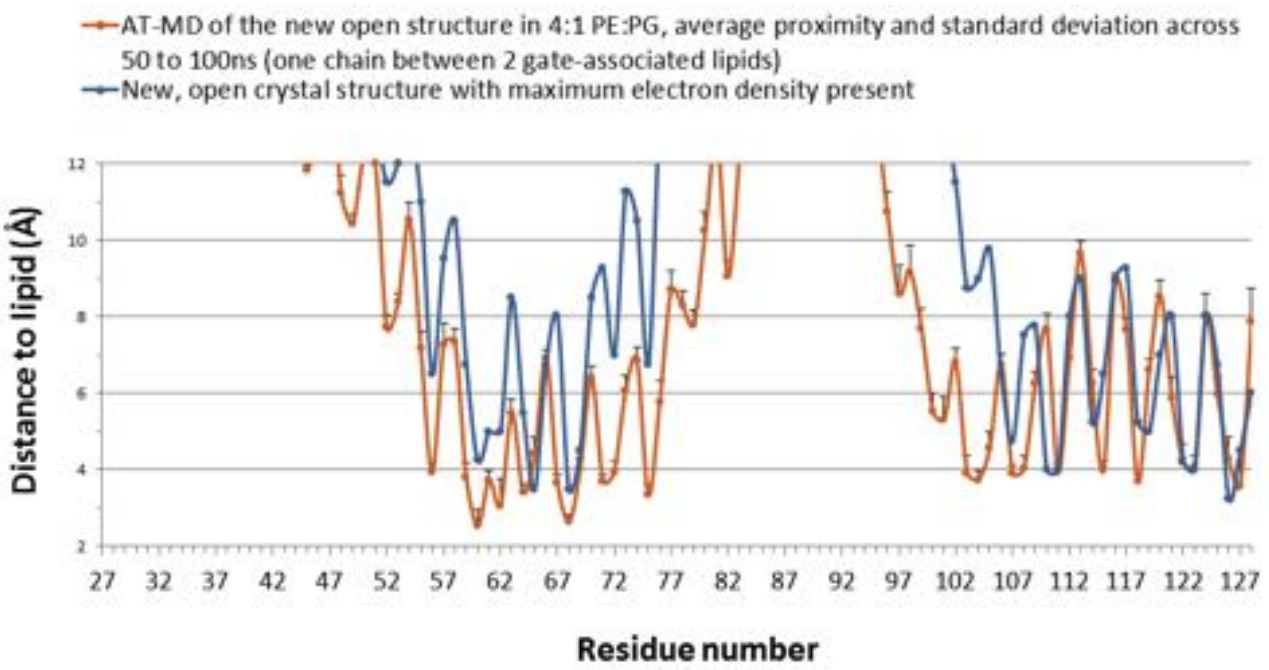

Figure 5.3.14.: Comparison between lipid placement in MD, and electron density in the new crystal. Analysis of distance between open MscS and POPE lipid bound to lower TM3 in AT-MD, compared to the separation between protein and electron density, modelled as alkyl tail, in the new crystal structure.

TM3 region in the closed state was a chance event. Experimental results suggest that MscS is PE selective, but does not specify the location of where PE should bind. MD data and crystal data both suggest a strong binding site for lipid in the lower TM3 region, but another site can not be ruled out. Therefore, I analysed the full protein for lipid selectivity.

5.3.4. Investigating the origin of lipid preference

5.3.4.1. Neither the open or closed state TM shows PE selectivity

The five repeat CG simulations per state were analysed for lipid that was ever within 6 Å of any part of the TM. Lipid type (POPE or POPG) and time spent within cutoff, not necessarily consecutive time, was stored, binned over time and plotted in Figure 5.3.17.

When lipid type preference is compared to the background membrane composition (Fig. 5.3.17, **B** and **D**), neither closed MscS (**B**) nor open MscS (**D**) show significant preference for either lipid type amongst lipid that bind the strongest to MscS ($0.9 < t \leq 1$ fraction of the time). On the contrary, considering its low content in the membrane, PG is significantly preferred at time-scales at or below 0.9 fractions of the simulation.

5.3.4.2. Lipid headgroup interactions to individual residues

Following on from the above analysis of the full TM's lipid selectivity, each residue's interactions with lipid headgroup was analysed. Headgroups were broken down into the following four categories:

- the phosphate of either POPE or POPG
- the amine of POPE
- the hydroxyls of POPG

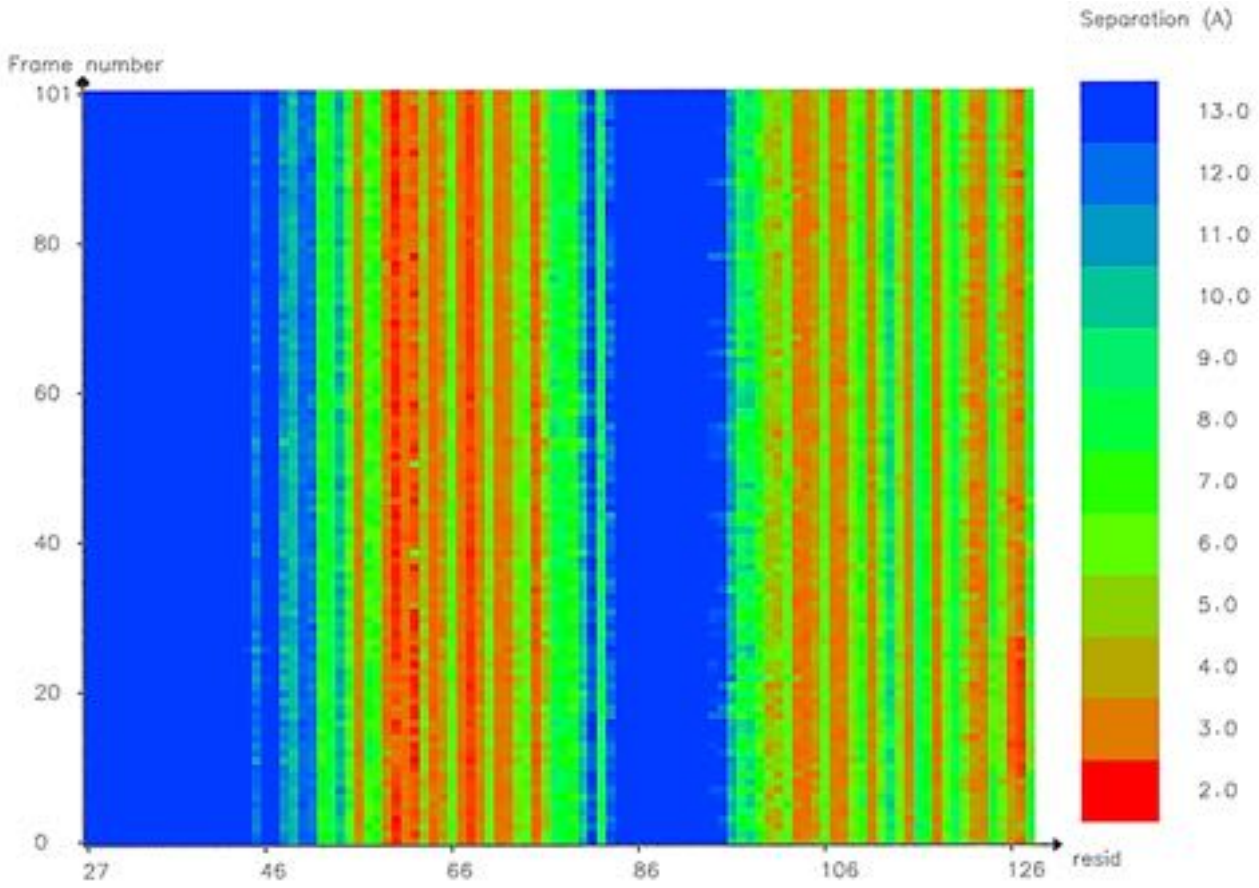

Figure 5.3.15.: Largely time-invariant proximity to permanently bound lipid in the new open structure InterQuant 2D heatmap of a single chain's TM domain's proximity to bound lipid over 50 to 100 ns atomistic simulation of the open structure. The data displayed here was used to compute the average values and error bars seen in Fig. 5.3.14.

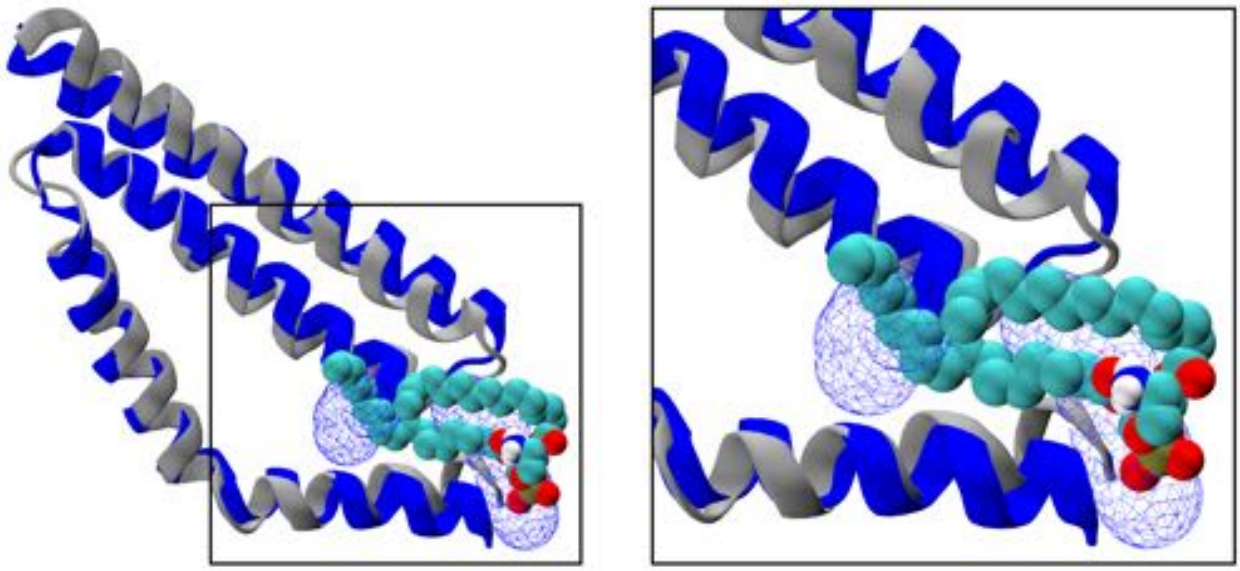

Figure 5.3.16.: Overlap between lipid placement in MD and electron density in the new crystal. Qualitative comparison of the new open crystal structure and its electron density (the TM region of chain A shown in blue cartoon and mesh, respectively), and aligned MD data of the new open structure in a lipid bilayer (chain A's TM regions shown in gray cartoon with a bound lipid as a space-filling representation in atom-indicative colours). Snapshot taken at the end of 100 ns AT simulation of the new open structure. Inset is an enlargement of the site of interest, showing overlap between electron density mesh and lipid.

- the amine of POPE or the hydroxyls of POPG

N.B. graphs and a more extensive discussion of the states' common and differential lipid contacts is available in Appendix D.2.

Lipid phosphates are responsible for persistent lipid binding to MscS Lipid phosphates make the most persistent bonds to MscS. As phosphate groups are present on both POPE and POPG, this does not constitute basis for selectivity. Even when interaction to either lipid amine or hydroxyl is accounted for, the least weak interactions, Tyr27 and Arginines 46, 54 and 59, never bind longer than 0.4 of the trajectory; that is half the duration that those same residues bind lipid phosphate groups. The occasional binding between residue and the part of the lipid headgroup that is not phosphate could therefore be due to its close proximity to strongly bound phosphate. However, it can not be ruled out that the additional attraction offered by amino or hydroxyl groups to protein may aid lipid coordination in subtle ways.

Charges on the C-terminus of TM1 and mid-TM2 show significant differential binding between the closed and open state of MscS Ser58, Arg59 and Lys60 show significant preference for lipid phosphate groups in the open state compared to the closed state. Out of these residues, Arginine 59 binds phosphate the strongest; on average throughout three quarters of the open structure simulation, compared to only half the time in simulations of the closed structure. The closed structure has on average stronger contacts to phosphate at residues along TM2; Phe68, Ala71 and Arg74, but the increase in contacts is not significant. However, Arg74 binds preferentially to POPG hydroxyl groups ($p \leq 0.005$), and only in the closed state. Arg74 is located at the middle of TM2, wherefrom lipid is lost during gating, suggesting that lost lipid is PG.

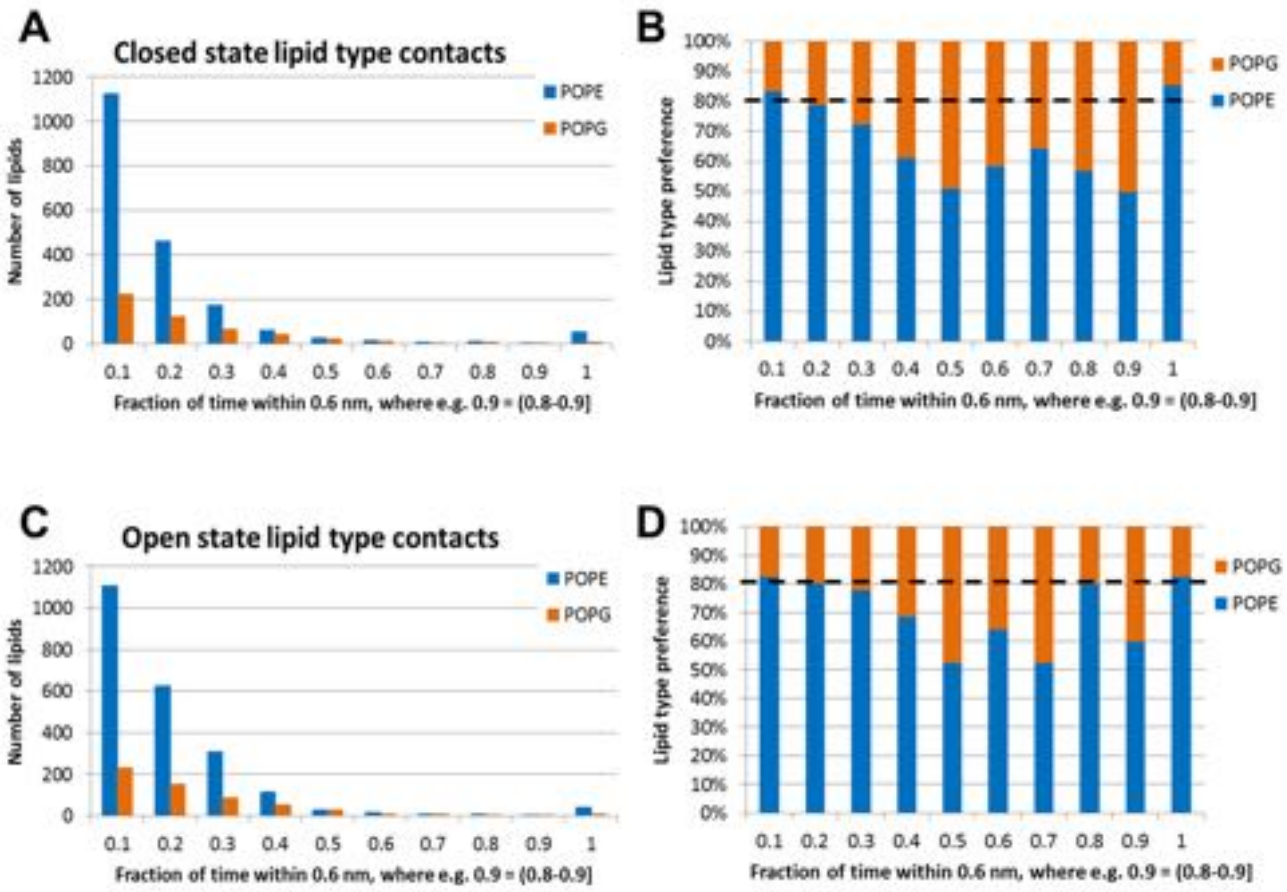

Figure 5.3.17.: Lipid type preference of the closed and open TM
Lipids within 6 Å of any part of the TM was checked for type (POPE or POPG) and time spent within cutoff in five 1 μs CG simulations per MscS state. Data are binned over time and presented for the closed (**top graphs**) and open (**bottom graphs**) state. For example, bin 0.5 refers to $0.4 < t \leq 0.5$, where t is fraction of trajectory time **A** and **C** show the absolute number of lipids bound, added over the five simulations, while **B** and **D** show type preference as a percentage per bin. The dashed black line in **B** and **D** indicates the 4:1 POPE:POPG lipid ratio used as membrane composition.

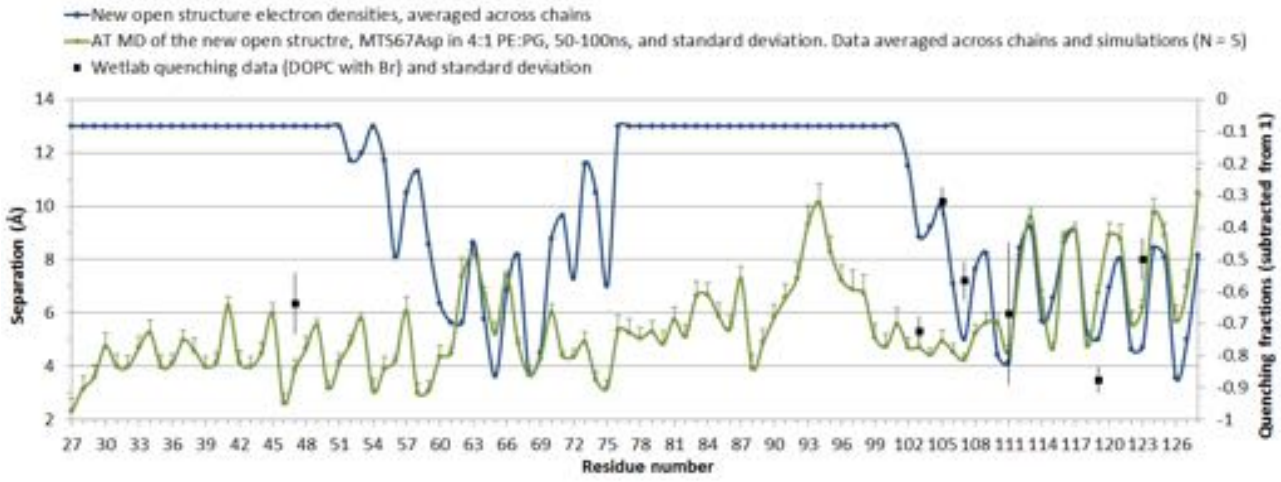

Figure 5.3.18.: The proximity of MscS to lipid and electron density: comparison across experimental and MD data.
Error bars show +1 SD. Electron density is modelled as alkyl tails. Black squares show quenching data. N.B. quenching data are not linearly correlated with separation; these data points should be taken as an approximation only.

MscS shows weak preference for POPG Both amines and hydroxyls contact predominantly polar residues with comparable strength, which - considering the 4:1 PE:PG membrane composition - suggests that MscS is slightly PG selective. This makes sense from an electrostatic perspective, as the MscS TM region is heavily positively charged.

A potential drawback of this MD study is that the system size may not be large enough to allow uneven distribution of negatively charged POPG. In vivo, POPG charge could be neutralised and re-distributed away from MscS, whereas the MD box arrangement necessitates relatively high proximity between individual POPGs, which might bias lipid positioning into an even distribution. However, in defence of the MD methodology employed, 0.15M of Na^+ was present to screen POPG repulsion.

5.3.5. Comparison between MD and experimental data

5.3.5.1. MD data suggest that the origin of the electron density is lipid

I used InterQuant to probe the new open crystal structure's proximity to electron density (modeled as Carbon), and compared this to its proximity to lipid in my AT MD data. Proximity is correlated along the cytosolic TM3 region (residues 107-128) and similarly in the linker-region between TM1-2 and N-terminal TM2 (residues 60-69), see Fig. 5.3.18. Assuming that the crystal electron densities originate from lipid, the discrepancies between crystal and MD data can be explained by the densities only featuring parts of lipids. Moreover, because the crystallization process does not feature a planar bilayer, the electron density does not feature lipids from the surrounding bilayer which is present physiologically. Therefore, MD analysis of lipid access will suggest proximity that fails to feature in the crystal (c.f. residues 27-59 and 70-105). Lipids that do remain bound to the protein throughout crystallisation suggest strong binding sites, which can be investigated by MD.

For a graph of protein proximity to lipid under conditions that better approximate the crystal setup (where only lipid that overlaps electron density is present; not the full lipid bilayer), see. Fig. 5.3.14 .

5.3.5.2. Quenching data support lipid presence away from the electron density

Where lipid is not resolved in the crystal, quenching data, albeit sparse, lend support to MD results. Quenching data support lipid presence at residues 103 and 47, where crystal data fail to resolve lipid but MD data suggest it (see Fig 5.3.18, black squares).

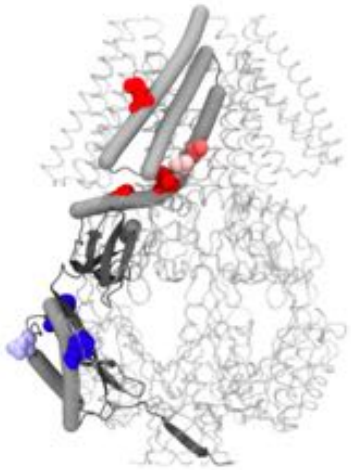

Figure 5.3.19.: Tryptophan mutant quenching results
Average DOPC quenching values heatmapped onto the wildtype residues of chain A of closed state MscS. The blue-white-red heatmap displays values from 0% (blue) to 87% quenching (red). Chain A is shown as gray bendices and cartoon, where wildtype residues, that were replaced by Trp in the experiment, are shown as space-filling representations. The rest of the protein is shown as a ribbon, for visual reference.

N.B. Quenching data do not easily convert to separation between protein and lipid, so should only be viewed as an indication of lipid presence or absence. Second, MD data in Fig. 5.3.18 refer to separation to any part of the lipid, whereas quenching data are correlated with separation to a bromine group that is attached mid-way down the lipid alkyl tail (c.f. Fig. 5.2.3).

5.3.5.3. Correlation between MD and experimental data on alkyl positioning with respect to MscS

My experimental collaborators' quenching results are heatmapped unto the closed state MscS in Fig. 5.3.19 and, focussing on the TM region, in Fig. 5.3.20 **A**.

CG MD results for the average separation to the site of lipid bromination and the number of unique bromine groups within 6 Å are heatmapped onto the closed state in Fig. 5.3.20 **B** and **C**. Proximity analysis shows a good degree of correlation with experimental data, even though the MD data are coarse-grained and linear separation is an oversimplified estimate of quenching (for a more accurate comparison to MD data, see the section below). Contact analysis is included under the assumption that the number of quenchers in contact should impact the amount of quenching, probably in a multiplicative manner with respect to separation. In hindsight, the contact cutoff should have been set to 11 Å; the distance at which quenching is extinguished [London and Feigenson, 1981, Chattopadhyay and London, 1987, Mansoor et al., 2010]. This would have captured the quenching capacity of all surrounding bromine groups.

Comparison of experimental and AT MD quenching data
InterQuant results of AT MD were converted to quenching values using equation 5.2.1. MD and experimental quenching results are shown in Fig. 5.3.21. The relationship between experimental and MD results show a high degree of similarity. The largest discrepancies are seen for residues Ala119 and 213. Ala119, located at the centre of TM3b, should display a high degree of quenching given that the bromine is located half-way down the lipid, which lies parallel to this helix in MD (see Figure 5.3.16). Equation 5.2.1 disallows quenching $> 0.8\frac{Q}{Q_0}$, which could explain the difference in result. Ala213 is located in the cytosolic cage domain, on the 'underside' of the cage. It is strange that experiment shows any quenching at all in this region (albeit low), and my collaborators wanted to discard it as insignificant quenching. MD data support this conclusion, showing an average separation

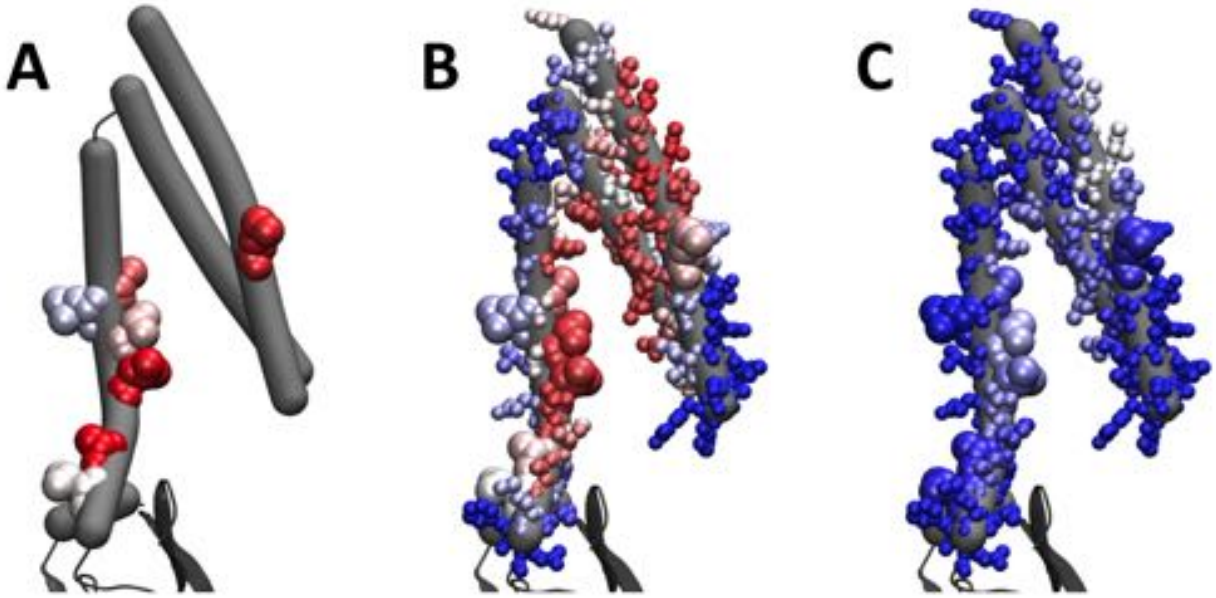

Figure 5.3.20.: Comparison of quenching and CG MD data
A. Average DOPC quenching values heatmapped unto the wildtype residues of the TM region of chain A of closed state MscS. Close-up of Fig. 5.3.19. **B.** MD results for separation to bromination sites. The red-white-blue heatmap displays values from 5.5 Å (red) to 16 Å separation (blue). **C.** Number of unique bromination sites in contact. A 6 Å cutoff was applied and the number of bromination sites within cutoff was evaluated. The blue-white heatmap runs from zero contacts (blue) to 1.5 contacts (white). Enlarged residues indicate residues for which there exists quenching data.

of 45 Å (± 8.91 SD). However, disregarding the relatively low quenching value for Ala213 also calls into question the low quenching value for Leu105, where MD does support lipid presence.

I hypothesize that if both MD and quenching data are in close agreement for a residue, it suggests that this data are valid.

5.4. Discussion

5.4.1. Proposed gating mechanism based on MD results

5.4.1.1. Insufficient lipids to allow local membrane curvature results is widening of the TM apex via associated lipid

At equilibrium membrane tension, membrane curves extensively about MscS' conical TM shape, and half of the TM locates to the bacterial cytosol. On the contrary, when lipid density decreases, such as in a membrane under high tension, curvature is not favoured (c.f. Fig. 5.3.2), and lipid pulls on the N-terminal TM1, where strong links to lipid headgroups exists (c.f. Fig. D.2.2). This results in a less cone-shaped TM with increased apex radius and hydration.

5.4.1.2. The sensor paddle moves, possibly aided by lipid, to optimise in plane expansion

Some of the effective protein length that is available for extension in the membrane plane is lost to the diagonal arrangement of the sensor paddle in the closed state. As tension increases, the sensor paddle is therefore relocated to be in line with its own chain's TM3 domain once there is enough tension to break the restraining saltbridge that links the sensor paddle to TM3b [Sotomayor and Schulten, 2004]. Lipid headgroups exist in the TM1-2 linker area (c.f. Figs D.2.2, D.2.5), proximal

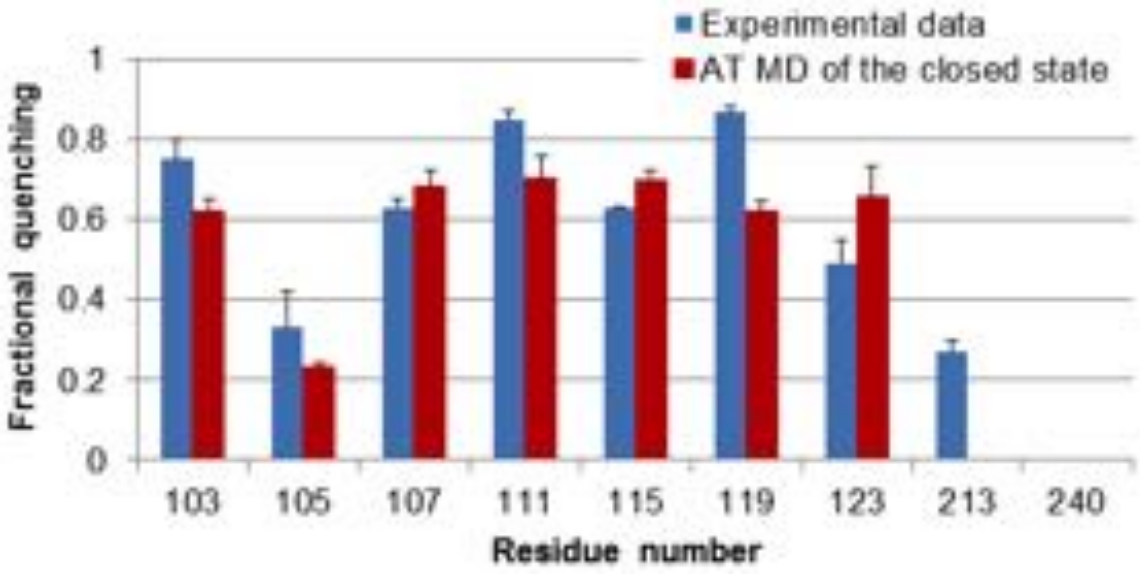

Figure 5.3.21.: Experimental quenching data compared to derived quenching for AT MD data
Error bars show one standard deviation; positive side only.

to Asp62 (one of the residues that participate in the saltbridge), and may therefore provide it with alternate coordination sites to lower the bond breaking energy.

Paddle movement could also be aided by the lipid tails that the paddle confines, that may reduce friction between the sensor paddle and TM3 by "greasing up" the passageway due to its non-specific binding, and by hosting thermal fluctuations that compromise the saltbridge. Alternatively, contained lipid may prohibit paddle movement, in which case the paddle only moves across to its own chain once bulk lipid is removed (although it is possible that lipid exit and paddle movement happens in a concerted motion). In previous mutational studies, polar residues were introduced in the place of hydrophobic ones, which made MscS require more tension to gate [Nomura et al., 2006, Belyy et al., 2010a]. Polar residues could provide binding sites for lipid headgroups, making it harder for them to leave the otherwise hydrophobic space, or to be the easily shifted grease that could facilitate sensor paddle movement.

Once the paddle shifts position to its open state conformation, it is subjected to an increase in lipid contacts. The open state shows increased lipid proximity to the TM1-2 linker region (c.f. Fig 5.3.6), where they may substitute intra-protein saltbridges between paddle and cage that could otherwise form and lock the protein in its open conformation, which would be devastating to the host bacterium.

N.B. Above, I write exclusively about sensor paddle movement. However, to align the sensor paddle with its own-chain TM3b helix, the sensor paddle must move *relative to the cage*. This means that either cage or sensor paddle could move to achieve the final conformation.

5.4.1.3. Bulk lipid leaves and TM3 moves peripherally to accommodate the loss

Tension in the plane of the membrane decreases lipid density and lipid lacking specific binding to MscS is lost to the bulk. This leaves a permanent population of lipid, probably POPE, fastened by their headgroups to the sensor paddles, and with alkyl tails pointing towards the gate (c.f. Figs 5.3.10, 5.3.12 and 5.3.13). MscS thus undergoes a transition from a state where lipid contacts are liberally distributed about the TM, to a state where lipid contacts are few and focused at the gate. TM3 moves out to accommodate the void left by bulk lipid, increasing the distance between pore-lining helices. In such a situation, vertical cage compression may aid the TM3b domain's movement outward, providing necessary extra protein length parallel to the membrane since TM3b and the cage top are closely associated.

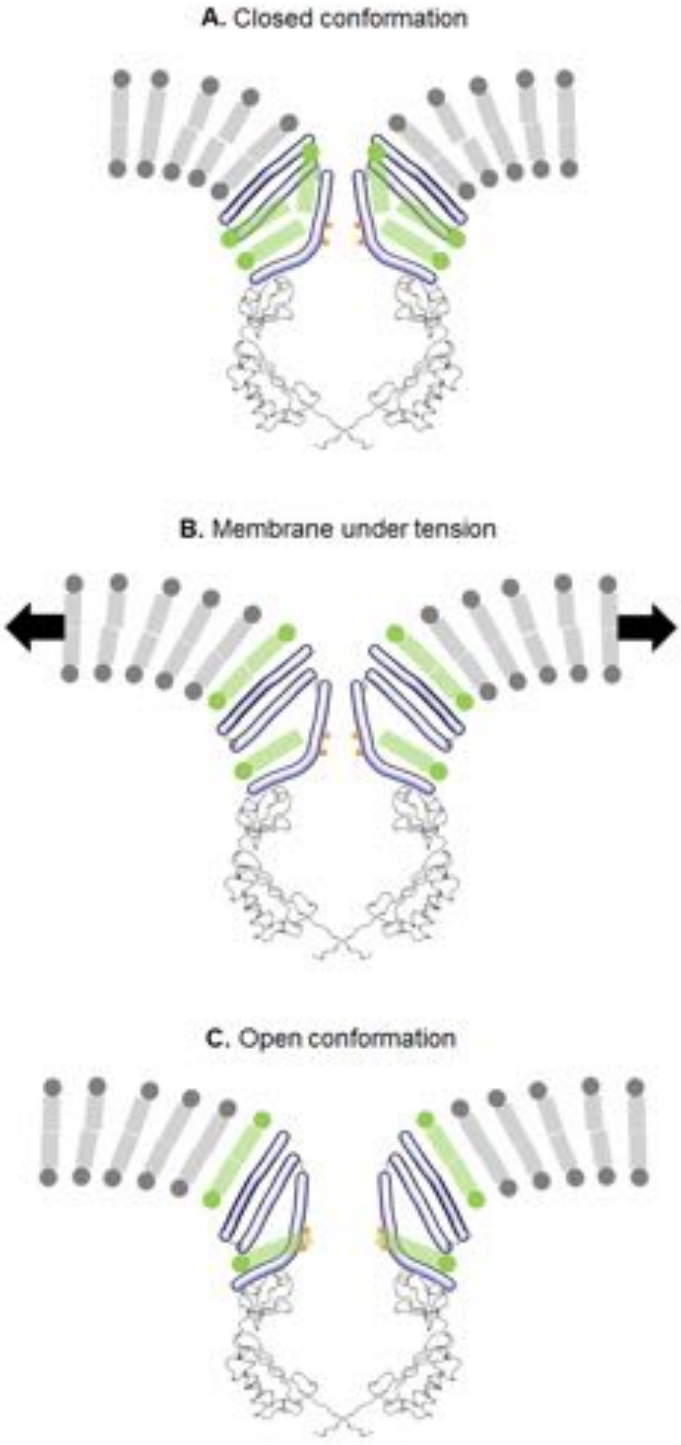

Figure 5.4.1.: Proposed gating mechanism from restrained MD simulations of the two MscS states. **A.** In its closed, relaxed conformation, MscS induces curvature in the surrounding membrane. Lipids associate with the charges on the sensor paddle but are free to partition between the cavity formed by the sensor paddle and TM3 and bulk membrane. **B.** As pressure is applied, the membrane is stretched and lipids that are not permanently bound to the lower TM3 region repartition from protein to bulk bilayer. **C.** With less lipid occupancy, the closed structure becomes less stable and the protein structure transitions to an open conformation that needs less lipid to be stable. Permanently bound lipid accesses the open hydrophobic seal (shown in orange) where it may stabilize the widened pore structure. TM helices are shown as bendices, the rest of the protein in ribbon representation.

5.4.1.4. Hydrophobicity of permanent lipids proximal to the gate help twist Leucines out of the gate

The linker between the sensor paddle and TM3 is short, and it is possible that movement of the sensor paddle results in strain along TM3. TM3 contains the hydrophobic gate, so this strain could affect the energy barrier associated with twisting Leucine out of the gate. We now have a situation where pore lining TM3s already have some degree of inter-helical gaps due to both the widened TM apex under membrane tension and structural accommodation of lipid loss from internal cavities, and water access is more proximal to the hydrophobic gate than in the closed state. The Leucine twist that is associated with the open state may therefore be further assisted by the presence of the hydrophobic alkyl chains of lower TM3 lipids, that are located immediately peripherally to the gate (c.f Fig. 5.3.11). The hydrophobic effect may lower the energy barrier associated with the twist, which in turn permits pore hydration and osmolyte flux.

5.4.2. MD results in the light of published mutation and conservation data

5.4.2.1. Hydrophobic interactions

Two published studies focus on the effect of substituting hydrophobic residues for polar residues on pressure threshold and downshock survival. Nomura et al. [2006] study membrane-facing residues along the sensor paddle, and Belyy et al. [2010a] investigate residues that line the cavity between TM2 and TM3a. Taken together, the two studies examine most of the hydrophobic residues in the TM region. Notably, both studies verify that loss of hydrophobicity is key to the result phenotype (as opposed to other amino acid features, like volume) by producing exaggerated results when replacing the polar mutant residue by a charged one, or partially rescuing the wildtype phenotype by re-introducing hydrophobicity.

Nomura *et al.* find that polarity introduced to the ends of the sensor paddle causes increase to the pressure threshold, and that a higher gating tension is correlated with higher chance of death as a result of downshock. Polarity in the middle paddle region, on the other hand, made MscS gate at lower tension. In the majority of cases, loss of hydrophobicity in the paddle domain leads to decreased MS. This makes sense also for our model, where lipid must be free to enter and exit the area about the paddle, so introduced polar residues should impede lipid exchange between protein and membrane. If polar residues are introduced in the middle of the paddle, lipid may be prevented from stably entering the MscS cavity, and in the absence of lipid, the closed state is destabilised. Additional potential binding sites for lipid headgroups, on the other hand (at either end of the paddle), could plausibly impede lipid exit, causing lipid to linger and stabilise the closed, lipid-dense state.

Second, Nomura *et al.* show that polar substitutions to the ends of the sensor paddle only have additive detrimental effect if both mutations are made to the same end of the paddle, in which case they have approximately the same effect as a full charge. The authors interpret the data as an introduced imbalance to focal points where headgroups attach and pull peripherally upon tension, disrupting gating. Compared to my results, an extra charge would probably amplify headgroup binding, and, at the cytosolic end, might dislodge the permanent lipids seen in MD.

In the second study, Belyy et al. [2010a] conclude that only polar substitutions to the cytosolic region of the paddle-TM3a cavity increase the pressure threshold of MscS. In particular, polar substitutions of three residues, Phe68, Leu111 and Leu115, cause increased pressure threshold, more death by downshock, fast inactivation and inactivation by sustained sub-threshold pressure. Combinations of these mutations cause more extreme phenotypes, and Belyy *et al.* argue that these residues, which are over 1.5 nm apart in the closed crystal structure, are proximal in a stable state where the paddle is aligned with TM3 in the same chain. The authors support this claim by their result that volume between hypothesised contacts in the TM-aligned state is conserved, in particular between the aforementioned three residues.

My MD data show that Phe68, Leu111 and Leu115 make contact to the alkyl tails of the permanent

lipid population, so decreased hydrophobicity in this region probably does not favour stable alkyl positioning. Leu111 and Leu115 are proximal to the gate, so introduced polarity will upset whatever purpose the permanent lipid population serves; whether stabilizing the open pore or lowering the energy barrier for Leucines to twist out of the pore.

Belyy *et al.* fail to explain the lowered pressure threshold that results from Ala103Ser and Val107Ser. These residues are located in the middle of the cavity and could therefore interfere with the lipid that exits the cavity as MscS adopts the open conformation. The situation is similar to that where polar residues were introduced to the middle of the sensor paddle; firm lipid docking is disallowed, rendering the lipid prone to exit at lower tension. However, the remainding residues that line the cavity interior show wildtype pressure sensitivity upon polar substitution, so the support for my hypothesis is inconclusive.

It would be interesting to see the results of introducing a charged residue into the cavity interior. This may prohibit lipid docking altogether, which could either favour the open or closed state. If the latter, it suggests that lipid exit and gate opening happens in concert, and that lipid presence in the cavity is necessary for the transition to take place.

While the added volume of Phe68, Leu111 and Leu115 is well conserved, so is that of Gly76, Leu100, Ala103, Val107 and Ala110 and their contacts. All these residues form close 4 Å contacts to lipid in my MD simulations (apart from Ala110, which forms a 6 Å contact). So whereas Belyy *et al.* assume that the cavity compacts for hydrophobic contacts to be made between the paddle and TM3a, lipid fills this void in at least one stable state, so my theory does not require TM compaction. Conserved cavity space will decidedly aid lipid fitting, and so will the fact that the hydrophobic residue character is conserved across bacteria (c.f. Fig. arrows in 5.4.2) [Balleza and Gomez-Lagunas, 2009]. This suggests that our lipid-centric gating mechanism holds true not just in *E.coli*, but across the 86 MscS homologs used in the study of Belyy et al. [2010a].

5.4.2.2. Electrostatic interactions

In a more recent study, *Nomura et al. [2008]* investigate the electrostatic linkage between the cage and cytosolic end of the sensor paddle, as proposed by Sotomayor and Schulten [2004]. They find that there exists a state where Asp62 and Arg131 are more proximal than in the published crystal structures and linked electrostatically. Substitution causes MscS to gate at higher tension and inactivate faster, suggesting that the cage influences gating. The linkage between cage and TM is intact in my MD simulations; lipid does not make contact with either Asp62, Arg128 or Arg131 (Arg128 is implicated in a saltbridge in the closed structure) (Fig. D.2.2).

Nomura *et al.* also examine adjacent charges on the paddle by substituting the charge on Arg59, Lys60 and Asp67 for neutral or oppositely charged residues. While these mutants at times produce significantly different phenotypes compared to wildtype, the changes are small in all cases except for charge reversal at Asp67, which increases the pressure threshold, causes 30% cell death by downshock and increases the rate of inactivation. Asp67 is the residue that shows the highest lipid selectivity (PE) in open state MscS (Fig. D.2.6) and is the strongest attachment site for amine groups of permanently bound lipids (Fig 5.3.13). While the reason for the almost exclusive PE nature of permanently bound lipids still eludes both myself and my collaborators, it is interesting that charge reversal to the residue that is most closely associated with the PE amine group impedes *E.coli* MS.

PE lipid 3D orientation suggests that they may connect neighbouring paddles electrostatically (Fig. 5.4.3). In the open structure, these lipids are oriented so that their amine groups contact Asp67 while their phosphate groups contact Arg59 and/or Lys60 of the adjacent chain. These are the most persistent contacts that PE headgroups make (Fig. 5.3.13), which supports a functional role for them. It is therefore surprising that charge substitution of either Arg59 or Lys60 produce small changes to phenotype, as the data of Nomura *et al.* show. However, there is the possibility that permanent lipids bond to the paddle is too important to be left to a single positive residue. The close proximity between Arg59 and Lys60 could enable one residue to take over the role as electrostatic partner to lipid phosphate groups in the event of loss of the other residue. It would be interesting to see the

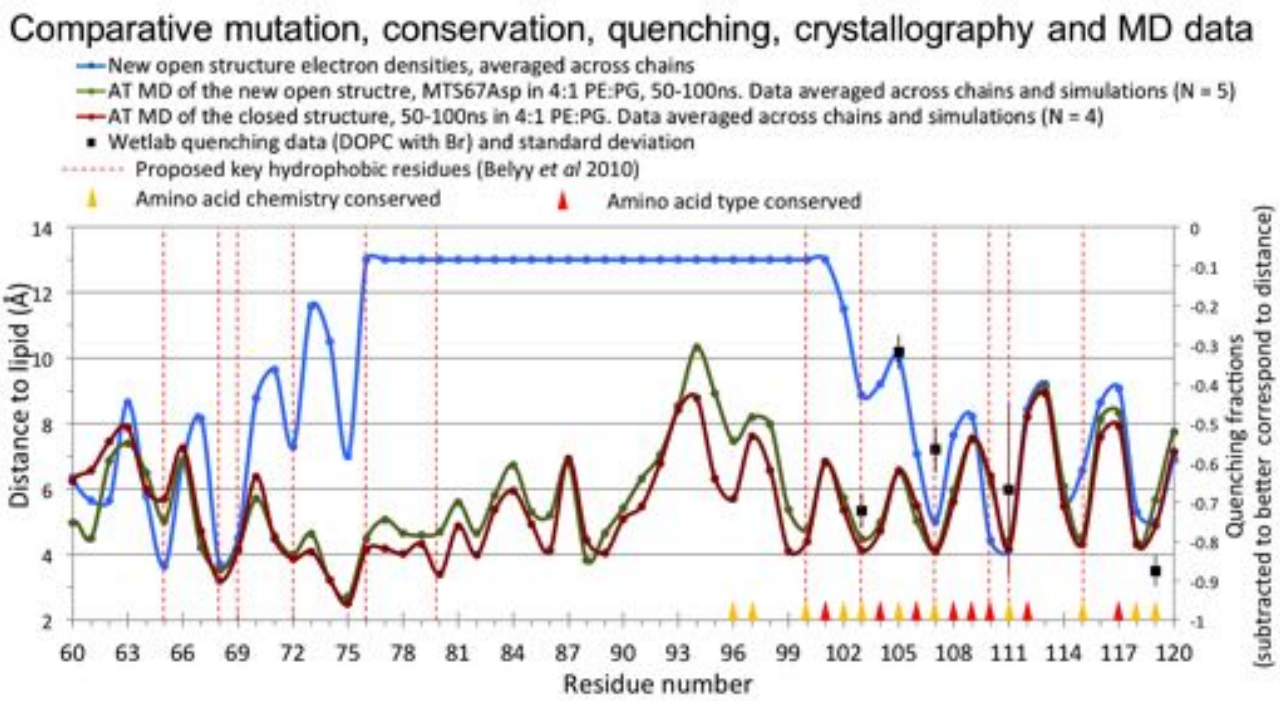

Figure 5.4.2.: Lipid localization proposed by MD and experimental data coincide with conserved hydrophobic regions where mutation is detrimental to function.

Lipid proximity profile per residue shown for the new open crystal structure (blue) and AT MD of both closed (red) and new open (green) structures. Superimposed are quenching data (black data points), residues that are conserved in terms of volume in the region between the paddle and TM3a [Belyy et al., 2010a] and residue type and chemical property conservation data [Balleza and Gomez-Lagunas, 2009].

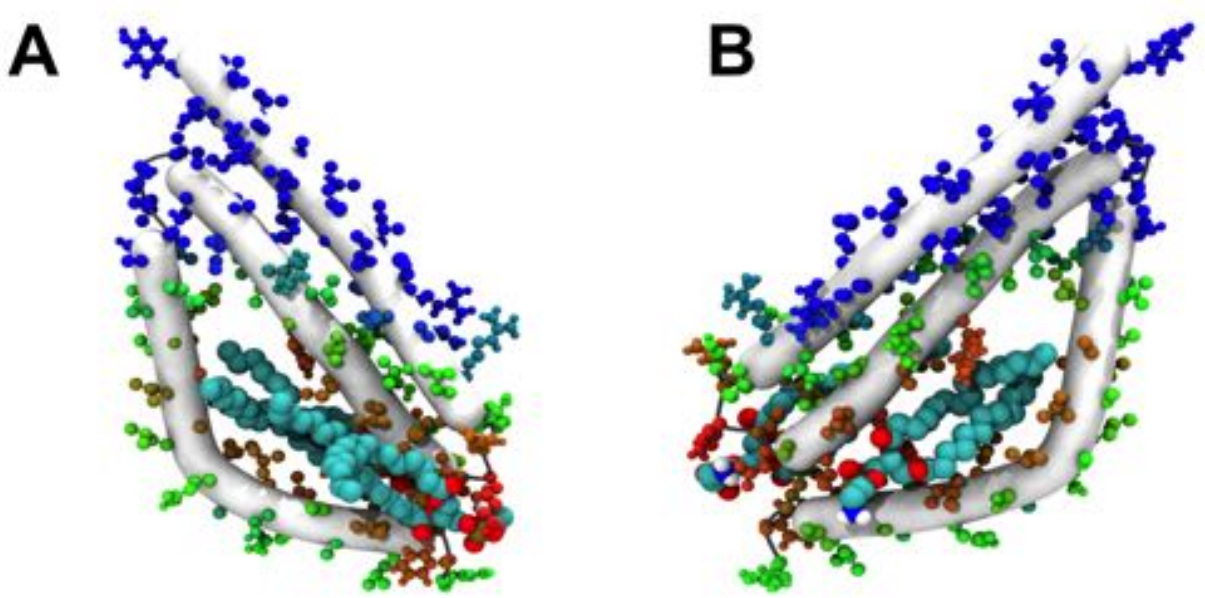

Figure 5.4.3.: Example of interaction between permanently bound lipids and the lower TM3 region. Snapshot of the TM region of a single chain in new open MscS after 100 ns AT MD. **A.** Residue side chains are heatmapped according to proximity to lipid (RGB heatmap with values ranging from 2 Å in red to 13 Å in blue). Lipid in space-filling representation, coloured according to atom type. **B.** as in A, but the chain was turned 180 ° to show residues on the opposite side.

phenotype of an Arg59/Lys60 double mutant.

No experimental mutational data exist for the other charges on the sensor paddle, which is unfortunate, since my MD data show that lipid headgroup phosphates bind strongly to most of these charges. However, charge neutralisation was done in silico for Arginines 46, 54, 74 and 88, and Asp67 in AT MD of the unrestrained closed state at 1.1V transmembrane potential [Spronk et al., 2006]. Substitution at position 46 or 74 eliminates conductance, which was attributed to subtle alterations to the electrostatic environment.

Although physiological lipid (POPE) was present in the setup, no details of protein-lipid interactions are reported, making it hard to compare these results to my data. Nevertheless, Arg74 is the region from where lipid is lost upon gating (Fig. 5.3.6), and this residue exhibits moderately strong contacts to lipid headgroup phosphates and shows PG-selectivity (Figs D.2.2 and D.2.5). Lack of POPG in the setup of Spronk *et al.* may have affected their results.

5.5. Summary and conclusions

This chapter focused on the detailed interaction of MscS and the membrane at zero membrane tension. Using multiscale MD and custom analyses, I explore both the published closed state and a novel open state crystal structure. The timescales accessible by CG MD allows sufficient lipid diffusion for analysis of protein lipid selectivity, while the subsequent AT treatment allows interactions to be analysed with high precision.

MD replicates and supports experimental data

There is extensive overlap between proximity to electron density and proximity to lipid, suggesting that the electron density is lipid. Further, MD results suggests that the most peripheral density is lipid headgroup rather than alkyl tail, which it was modeled as in the crystal. Back-calculation of quenching values from MD data also show a good fit to experimental results.

Taken together, crystal, biophysical and MD data all suggest that lipid penetrate deep into the MscS structure, proximally to the gate. Such lipid could act as structural support for voids that have been criticized as implausible in a natural membrane environment.

Further exploration of lower TM3 lipids by MD

A persistent lipid population is located in the lower TM3 region, and are found in both the closed and open state, so remain bound to MscS as it gates. These lipids form part of the electron density in the new open crystal structure. While the electron density in the crystal only resolves parts of these lipids, MD results reveal that headgroups of permanent lipids, primarily the phosphate group, make strong contacts to charges on the cytosolic end of the sensor paddle, notably Arginines 59 and 60. Lipid alkyl chains run alongside the hydrophobic TM3b top and become a physical part of the open gate of MscS, where their function could be to stabilize the distended pore. The pore of the closed state, on the other hand, prevents alkyl access, leaving tails immediately adjacent to the gate. In both closed and open states, contained lipid form physical links between the gate and the surface-exposed paddle.

Global membrane behaviour

The membrane curves noticeably about MscS, particularly about the closed state, thus submerging the TM into the cytosol. While the cytosolic half of the TM is located outside of bulk membrane, lipid still infiltrates this area with their headgroups facing the cytosol. Therefore, membrane curvature optimises the number of paddle charges that are satisfied by lipid headgroups. By comparison, in a planar membrane, paddle charges would risk contacting the hydrophobic membrane interior. Membrane curvature also minimises hydrophobic mismatch about a cone-shaped TM such as closed MscS, and has been theoretically predicted [Phillips et al., 2009].

Differential lipid interaction between states

One lipid per chain is tightly bound proximally to the cytosolic TM, irrespective of state. Beyond this, the closed state fits approximately two lipids while the open state fits only one. Another lipid is bound tightly to the middle of the TM2 helix, but is lost upon gating, possibly as a reaction to paddle movement. During gating, the cytosolic end of the paddle also gets closer and more contacts per residue, TM3a twists and contacts made by persistent lipids to C-terminal TM3b are loosened slightly.

PE selectivity is observed but its origin remains unknown

Experiments show that open MscS purifies with a large population of PE, and my simulations of the open structure show PE selectivity in the lower TM3 region. However, irrespective of whether the full TM or individual residues are analysed for lipid type preference, MD results can not explain this selectivity. My data suggest that lipid binding is not type selective, or slightly PG selective, which is reasonable since the TM has a net positive charge.

When the lower TM3 region is analysed, the data are yet more perplexing, as the residue with the most significant preference for POPE is Asp67, the residue that was mutated to MTS in the new open crystal structure. So while Asp67 might explain POPE preference in MD data, it does not explain PE preference in experimental data.

Consequences of MD results for MscS' gating mechanism

Based on these results, I propose a gating mechanism whereby the availability of lipid controls which MscS state is stable. At equilibrium, lipid partitions between the cavities of MscS and the bilayer, whereas applied tension drives lipid exit from MscS. Either the sensor paddle or the pore-lining helix

can move to fill the spot of the removed lipid. The positively charged paddle is probably horizontally restrained (or even subject to net outwards movement) by lipid that is under radial tension outwards, leaving the pore-lining helices to move peripherally to fill the cavity, which widens the pore and gates the channel. Lipid exit therefore effectively destabilises the closed state. The sequence of events during gating is unlikely to occur step-wise; rather, lipid exit and TM3 movement probably happens in concert.

The function of the resident lipids in the lower TM3 region remains unclear. They are located in a region that is conserved both in terms of chemical character and volume, which suggests functional importance. I hypothesise that alkyl tails provide a flexible hydrophobic surface that facilitates the considerable movement that the sensor paddle does relative to the TM3 domain during gating. The advantage of using a lipid mediator over direct protein-protein contacts may be to avoid undesirable bond formation, which could cause MscS to get stuck in between states. The hydrophobic alkyl chains of resident lipids could serve as “molecular ball bearings” that preserve stable van der Waals interactions throughout the gating transition.

Tension is the missing factor in these simulations

Ultimately, more data are needed to ascertain the role of lipid in the cavities within MscS. However, to properly model the situation where lipid is lost to the bulk, the membrane should not be at equilibrium, as is the case in simulations throughout this chapter. Rather, the membrane should be under tension. In the next chapter, I apply membrane tension to verify whether I can induce the lipid exit proposed here.

6. Study of the Mechanosensitive channel of small conductance under membrane tension

6.1. Introduction

The last chapter explored the differential lipid interaction between the closed crystal structure and the new open crystal structure at equilibrium. My results identify a role for transiently bound lipid as 'state inducers' by stabilising the closed or open MscS state based on lipid's propensity to leave for bulk membrane. I also identify a second, more permanently bound population of lipid, which I assign a putative role as tension relays from the membrane-exposed, charged protein surface to the gate. Included in that hypothesis is the suggestion that the paddle conformation in the open state could be different than that observed in the new open crystal structure. An alternate paddle conformation could be the result of the presence of a bilayer under tension; neither of which was present in the crystal setup. While the closed state structure was not resolved in a bilayer either, at least the condition of zero tension is physiologically relevant to the state. In fact, results from biophysical experiments presented in the previous chapter support the lipid positioning that I observe about MscS in the closed state in MD.

In this chapter, I start simulations from the closed crystal structure and go on to apply increasing amounts of surface tension. By using the natural channel stimulus, I aim to capture the conformational changes involved in the gating mechanism of MscS and, ideally, reach the open state. It should be interesting to see whether MD results can support the existence of the open crystal structure. The route that lipid takes between the closed and open channel state is also of interest, particularly in such a membrane-driven protein as MscS.

MD researchers have made previous attempts to open MscS by tension alone and failed. The common denominator for all these MD attempts is that they used AT force fields. AT MD, while more precise than CG, presents obstacles in terms of achievable timescales. Further investigations are needed to address the gating mechanism of MscS [Corry and Martinac, 2008], and researchers have used AT MD methods without success. Interestingly, the larger *E.coli* MSC , MscL, has been successfully opened by surface tension when modeled using the Martini CG force field. This has never been done for MscS, and this study will therefore adopt a CG MD methodology in an attempt to gate MscS.

6.1.1. Overview of parameter values used in earlier studies

An attempt to capture the gating mechanism of MscS requires a notion of what an open state of MscS might look like. This, however, is controversial, and this section investigates metrics that have been identified as important to MscS gating in earlier studies. I also review the key tension parameters used on bacterial MSCs . Where possible, MscS parameters are put into context alongside MscL parameters as the two channels are commonly compared in literature.

6.1.1.1. The size of the open gate

Studies of model hydrophobic pores The diameter necessary for molecule permeation depends on the chemistry and geometry of both the solvent and the pore. For example, an 8 Å long perfectly hydrophobic pore with a 36 Å mouth diameter requires a 9 Å diameter for hydration [Beckstein et al., 2001]. Given that water has a diameter of approximately 3.5 Å, the 9 Å diameter pore is sufficient

to fit three water molecules, which allows the number of intra-water hydrogen bonds in the pore to approximate the number in bulk water, thus stabilising the presence of water in the pore. For current conductance, a wider diameter of 13 Å is required to allow a hydrated ion to pass through. This assumes that there are no residues that coordinate ion dehydration, as opposed to the situation in a potassium channel. [Beckstein and Sansom, 2004].

The pore of MscS Experimental data estimate the conductance of MscS to 1 to 1.2 nS [Sukharev, 2002, Akitake et al., 2005, Edwards et al., 2005]. Theoretical evaluation estimated that this necessitates a 14-16 Å diameter pore [Sukharev, 2002]. By comparison, previous AT MD show that MscS pore widening to 16 Å allowed 1.4nS [Anishkin et al., 2008a].

Physiologically, MscS is likely to also allow passage of small metabolites such as amino acids, under conditions of zero voltage [Britten and McClure, 1962]. The lack of voltage occurs because more stretch-sensitive MSCs with smaller pore diameters open prior to MscS and eliminate the membrane potential prior to MscS opening. However, ion passage in a voltage field is the typical metric that is used in the MscS research field to indicate gate size.

The proposed open structure, with PDB id 2VV5, has a 13 Å pore diameter [Wang et al., 2008]. This is the minimum diameter required for model hydrophobic pore conductance, and just below the estimated pore diameter necessary for experimentally determined conductivity. It also fulfills the criterium that its pore diameter should be smaller than that of the open MscL.

Prior to entering the hydrophobic pore, solvents need to pass through the cage domain which contains 12 Å diameter portals lined by polar residues [Bass et al., 2002]. This makes it less restrictive to passage of polar substances than the hydrophobic gate.

6.1.1.2. Tension magnitude

Experimental values for midpoint surface tension The total number of channels in a patch is found by applying saturating tension and dividing the recorded current by the known single channel conductivity (in the case of MscS circa 1 nS). The *midpoint surface tension* is the tension at which half of that channel population is open. The midpoint surface tension is commonly referred to as the channel's *pressure threshold*. The midpoint tension of MscS is 5.5 ± 0.1 dyn/cm (55 bar*nm) [Cui and Adler, 1996, Sukharev, 2002, Akitake et al., 2005] (although the slightly higher value of 78 bar*nm has also been observed [Belyy et al., 2010b]). For comparison, MscL has a midpoint of 120 to 140 bar*nm [Sukharev et al., 1999, Belyy et al., 2010b]. This makes it approximately half as tension-sensitive as MscS. For reference, 1.6 to 2.3 times MscL's pressure threshold lyses an *E.coli* spheroplast [Nomura et al., 2006].

MD studies of MscS and MscL MD studies so far have simulated single MscS channels, rather than the population required for direct comparison to experimental midpoint tension values. Gating is evaluated by measuring the minimum diameter along the pore. Experimental surface tensions are in stark contrast to the high values used in attempts to gate MscS in MD setups. By surface tension alone, AT MD maximally achieves 7.5 Å pore diameter [Sotomayor and Schulten, 2004], which, as explained earlier, is insufficient for the experimentally determined channel conductance. Only when high surface tension is paired with steered MD does MscS gate [Sotomayor et al., 2006, Anishkin et al., 2008b, Belyy et al., 2010a].

While surface tension has never been applied to a CG MD setup of MscS, it has been used successfully to gate MscL [Colombo et al., 2003].

See the Appendix for a summary of results from all previous MD studies of MscS (Table A.1.1) and MscL (Table A.2.1).

Lytic tension At the extreme end of tension application, 700-900 bar*nm leads to rupture of bilayers in AT and CG setups [Tieleman et al., 2003, Yefimov et al., 2008, Louhivuori et al., 2010]. The high

resistance to lysis that is observed in MD could be due to the high speed at which tension is applied to accommodate the limited time-scale available to MD. It has been shown experimentally that faster application of tension cause membranes to rupture at higher tension. DOPC membrane stretched at a rate of 0.7 bar*nm/s breaks at 60 bar*nm, while it does not lyse until 140 bar*nm if pulled at the higher speed of 260 bar*nm/s [Evans and Heinrich, 2003].

6.1.1.3. Speed of surface tension application

As mentioned earlier, tension rate affects the lipid lytic tension. Tension rate thus affects lipid behaviour, which in turn affects MSC gating. This has been observed in previous CG MD of MscL in a small DOPC liposome, where a tension rate of 6,600 bar*nm/μs resulted in membrane lysis prior to MscL gating, whereas gating and lysis happened simultaneously at the lower rate of 1,400 bar*nm/μs [Louhivuori et al., 2010]. In vivo, on the other hand, MscL has evolved to gate in response to tension smaller than lytic tension. This suggests that rate of tension is important to MSC gating; unphysiologically high rates of tension will return erroneous results.

On the other extreme, slow pressure application inactivates most MscS in a patch [Akitake et al., 2005, Belyy et al., 2010b]. Akitake *et al*. hypothesise that the gate is connected to the element that transmits the tension via a velocity-sensitive "dashpot", and suggests that MscS enters an inactive state en route to the open state at slow tension application.

6.1.2. Chapter aims

In this chapter, I attempt to capture the gating mechanism of MscS using CG MD, which has never previously been done. A large focus will be on lipid-MscS dynamics during gating. This is because lipid confers membrane tension to MscS, which gates the channel, so lipid interaction with MscS is pivotal to understanding the gating mechanism of MscS.

6.1.3. Chapter overview

The chapter results begin by retrieving data that will facilitate the rest of the study; the ENM configuration necessary for the gate to be responsive to tension, and the relationship between lipid area and surface tension that is required for tension quantification. With these parameters and tools at hand, I start my investigation into the events that follow the onset of surface tension. The subsequent results naturally subdivide into two sections: protein and membrane deformation. The protein deformation section includes a systematic retrieval of the amount of tension that is required for the gate of MscS to open, as well as results for global and local changes to conformation. In the study of membrane deformation, I categorize lipid interaction according to the associated simulation outcome (e.g. whether MscS opens or not), which allows me to probe the significantly different lipid interactions that are associated with particular protein states. The chapter closes with a discussion of results in the light of published data, and the consequences of my results for the new open crystal structure in particular and the gating mechanism of MscS in general.

6.2. Methods

All simulations were set up to start from the crystal structure of the closed state of MscS (PDB id 2OAU).

6.2.1. Choice of gate and tension parameters based on published data

The open state gate diameter In the light of published experimental and computational data (as reviewed in the chapter introduction), I analyse MscS simulations up until and including when the pore

diameter is 15 Å in diameter at its minimum, classifying this as an open pore. This encompasses the 13 Å pore diameter new open crystal structure, allowing me to see whether the gating mechanism passes through this proposed state.

Surface tension I use twice the midpoint tension, 110 bar*nm. Simultaneously, I set up linear deformation simulations that capture tensions up until and including lytic tension. By monitoring the evolution of both the pore diameter and surface tension, I can estimate the surface tension at which MscS opens, and use this tension in a set of separate simulations run at instant constant tension.

Speed of surface tension I deform systems linearly to the point of membrane lysis in 1μs, and one simulation at half this velocity. Second, I have the set of simulations run at 'constant' tension, where the system is rapidly brought to a target tension and retained there. Such prolongued constant stimuli might allow MscS to adopt a conformation that is unreachable under linear deformation conditions.

I will refer to the two sets of simulations as *linear deformation* and *instant constant tension*, respectively.

6.2.2. MD implementation

6.2.2.1. Simulations

All simulations were run using Gromacs 4.5.5 and the Martini1.2 force field [Monticelli et al., 2008]. Initial tests of MscS under different ENM restraint regimes used the "Bond" force field [Bond and Sansom, 2006, Bond et al., 2008].

6.2.2.2. Derivation of an ENM that allows tension sensing

A functional model of MscS must have the ability to respond to membrane stress. The ENM imposed om the CG model may prevent MS, so must be tuned in order to be responsive. Assuming that linear deformation up until and including membrane rupture should widen the closed crystal structure pore diameter, a series of simulations were set up where the restraints imposed on the protein were systematically altered both in terms of number and location (see Table 6.3.1 for ENM regimes alongside their respective results). All ENMs used a spring constant of 2,500 kJ mol^{-1}nm^{-2} apart from one Martini CG model of MscS, which employed 1,000 kJ mol^{-1}nm^{-2}. ENM was implemented between backbone particles. The ENM implemented by the "Pore support" regime is visualized in Figure 6.2.1.

ENM-free model by force-field modification In addition to ENM variations, I investigated the possibility of ENM-free MscS, by implementing increasingly energetically favourable backbone-interactions by modifications to the Bond force field, in an attempt to mimic hydrogen bonding.

Simulations set up 100 ns equilibration and 100 ns production simulations were followed by 200 ns simulation under increasing surface tension until membrane lysis.

6.2.2.3. Membrane strain

Simulations performed of MscS are shown in Figure 6.2.2. The simulations listed under *Tension application* refer to simulations used for this chapter. The majority of them originate in the five 1μs Martini CG simulations run at equilibrium that were used in Chapter 5.

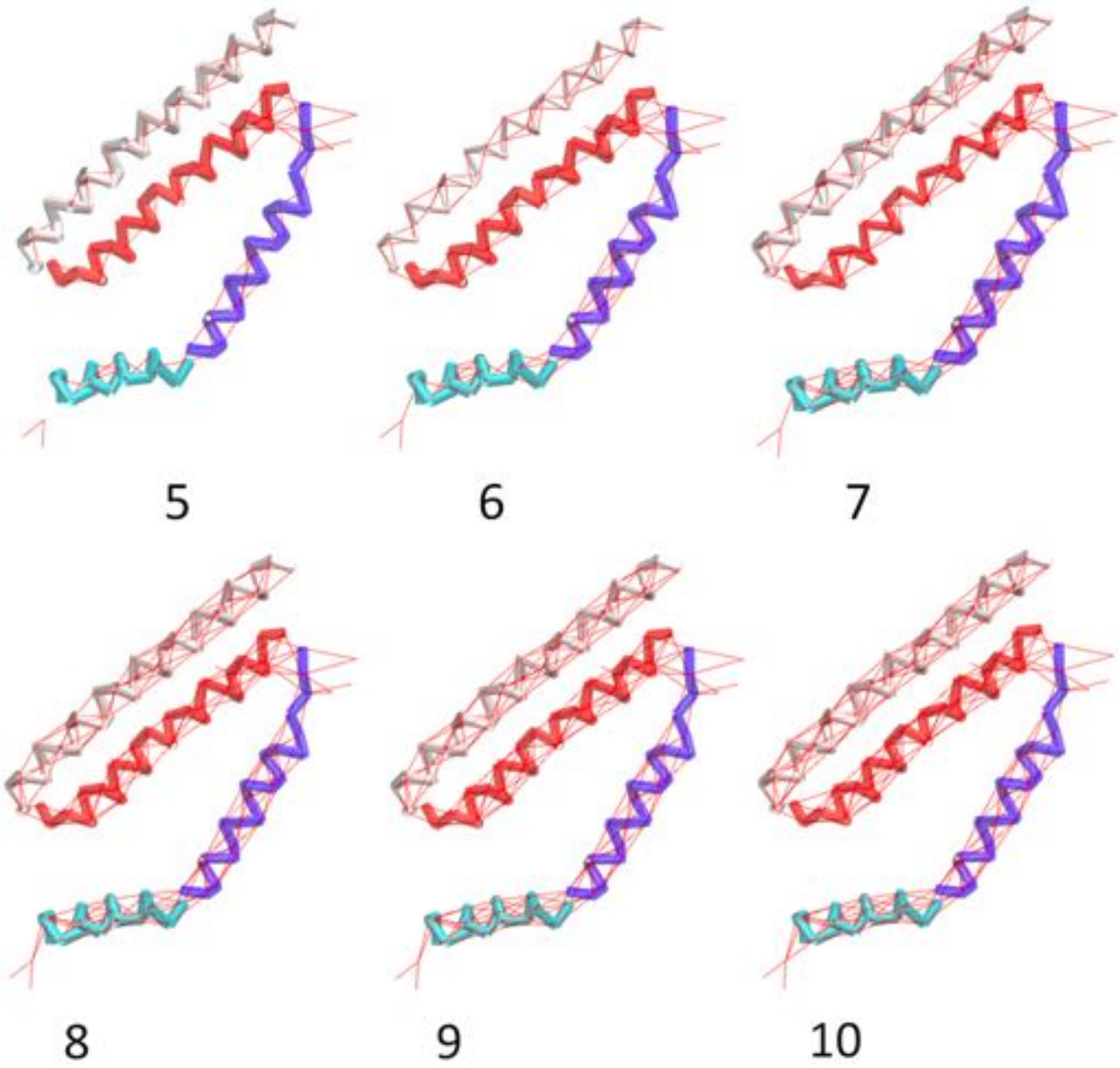

Figure 6.2.1.: ENM implemented in the 'Pore support' regime
The ENM is represented as thin red lines and TM1 is shown in white, TM2 in red, TM3a in blue and TM3b in turquoise trace of backbone particles. The numbers denote how many particles were allowed between springs. Note the constant springs between residues 90 to 92 at the top of TM3a, which were imposed in an effort to prevent pore collapse.

ENM In all simulations at non-zero surface tension, MscS β-sheets are subject to ENM. On top of this basic ENM, simulations with and without ENM between any Cα that are within 7 Å were run. These two ENMs are referred to as *Full* and *None*, and were implemented in order to clarify what lipid movement is possible under different structure constraints. A third ENM, *Cage*, refers to ENM on the cage domain only (again on top of the basic β-sheet ENM). The rationale for the Cage ENM is explained in the Results section.

Speed of tension application The time taken until the membrane surface tension reaches 600 bar*nm, noted under 'Time γ_{600}' is indicative of the speed of surface tension application. As instant constant tension simulations are instantaneous, tension speed data are not provided there.

6.2.3. Analyses

Effects of surface tension on lipid area 1 μs MARTINI CG simulations of 4:1 POPE:POPG membrane were run at instant constant surface tension using $\gamma = 0$, 55, 100, 200, 300, 400, 500, 600, 700 and 725 bar*nm. No protein was present in the system. The system energy was plotted, and where the system energy reached a plateau, the system was deemed to be equilibrated. All systems had equilibrated at 0.5 μs , so system x and y dimensions (parallel with the membrane) were averaged for the latter 0.5 to 1 μs of the simulation. The equilibrated system's area per lipid (APL) was then calculated by

$$A_{lipid} = \frac{2L_xL_y}{N_{lipids}}$$

where A is area, L is the box length and N_{lipids} is the number of lipids present. The factor 2 reflects that the membrane is made up of two leaflets, each with a surface area of L_xL_y.

Algorithm to compute surface tension from area per lipid The polynomial relationship between APL and membrane tension was computed and used to derive tension over time for systems subject to linear deformation that did include protein. In these simulations, the area taken up by MscS was approximated by a circle and this area was then subtracted from the system area. Membrane immediately adjacent to the protein, where there was high membrane curvature, was also discounted.

Time evolution of protein in-plane area Per timestep, protein in-plane area was evaluated by approximating the TM domain of MscS to a circle with maximum and minimum coordinates retrieved for x and y axes. The largest protein dimension was used as diameter and to compute protein area.

Membrane thickness as a function of surface tension GROMACS g_density analysis was used to compute the density of lipid phosphate headgroups across the simulation box for simulations run without protein at instant constant surface tension. The distance between peaks is the membrane thickness.

Helix TM1 length and tilt with respect to the XY plane The vector formed by residues 28 to 58, the helical residues of TM1 as assayed by DSSP [PDB, 2014b], was calculated using a custom version of Bendix. TM1's angle with respect to the XY plane is then given by

$$\theta = 90^\circ - \arccos\left(\frac{\hat{\mathbf{v}}_z \cdot \hat{\mathbf{v}}_h}{|\hat{\mathbf{v}}_z| * |\hat{\mathbf{v}}_h|}\right)$$

where $\hat{\mathbf{v}}_z$ is the normal unit vector of the XY plane, and $\hat{\mathbf{v}}_h$ is the unit vector of TM1.

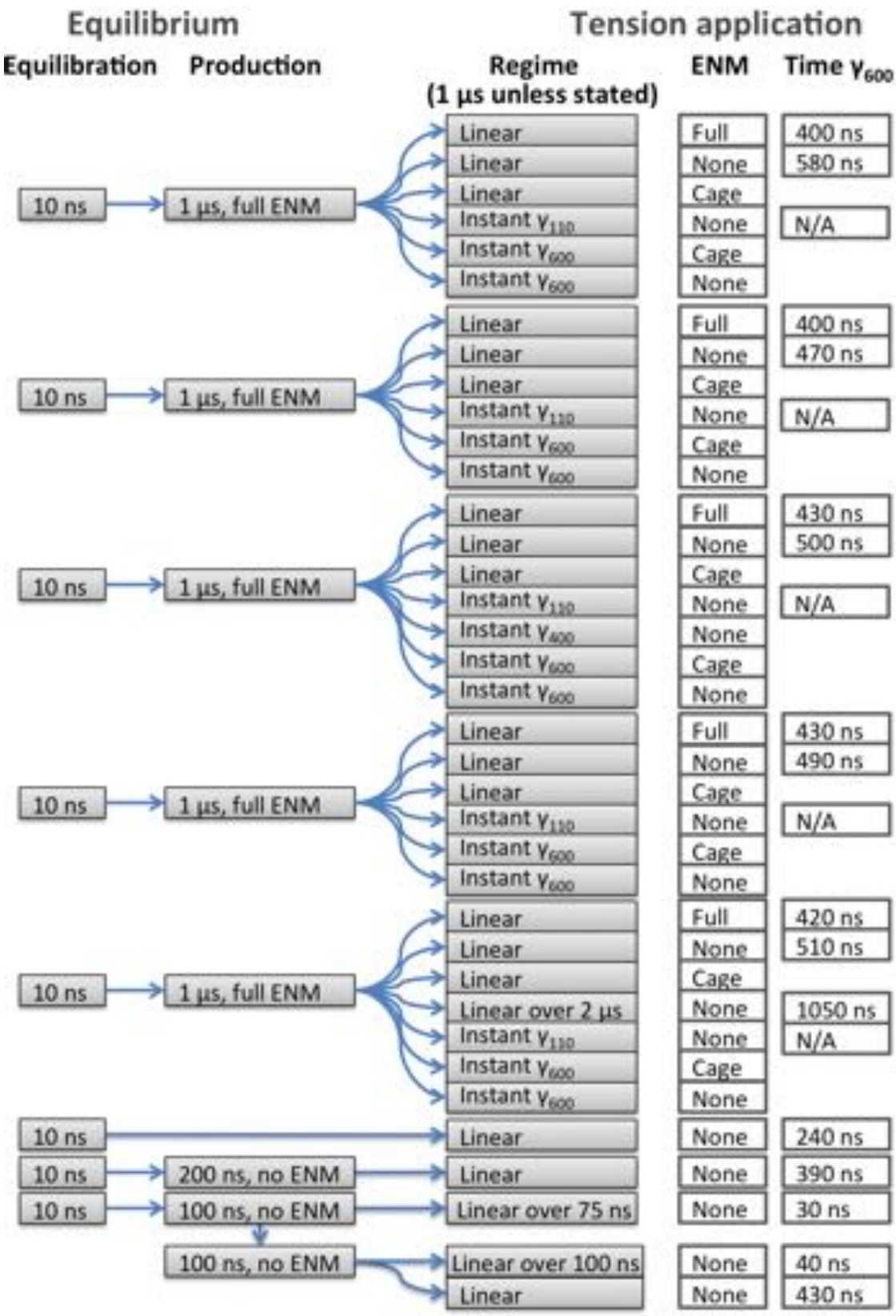

Figure 6.2.2.: Simulations run in this chapter
On the right-hand-side are equilibrium simulations used in Chapter 5. These were extended to give the set of simulations under non-zero surface tension that are used in this chapter, listed under *Tension application*. *Linear* = Linear deformation simulations until membrane lysis. *Instant* = Instant constant surface tension simulations at the stated tension (units of bar*nm). For details see the main text.

Minimum diameter of the hydrophobic gate The coordinates of Leucine 105 were identified for each chain, and the geometrical centre was retrieved. The minimal distance from this geometrical centre to Leucine 105 was used as the radius of a circle that defined the gate geometry. The same algorithm was also used to retrieve the minimal diameter at the level of Leucine 109. The MARTINI particle type AC1, which is used for the Leucine side-chain (irrespective of what the interacting particle type is) has a van der Waal atom radius is 2.125 Å [Marrink et al., 2007]. The pore diameter was therefore reduced by 4.25 Å to make my data comparable to other studies [Sotomayor and Schulten, 2004, Sotomayor et al., 2006] .

TM3 helix curvature (Bendix analysis) Helical residues Ala94 to Phe127 were analysed. A triangle with side 7 residues was employed to measure angles. In all trajectories where MscS opens, the angle per residue for each monomer and highest angle were collected over time and averaged.

TM3 helix twist Each residue in the TM domain was classified as facing bulk lipid, cleft (between chains) or protein interior. Results for simulations under tension were compared against the 'closed' and 'open' crystal structures' residue orientations.

Lipid-protein interactions as a function of state (InterQuant analysis) An increase in membrane surface tension is sufficient to gate MscS. Therefore the configuration of lipid about MscS, and its dynamics under tension, is key to understanding the gating mechanism of MscS. Residues 27 to 182, which incorporate the TM domain as well as the cytosolic beta-sheet domain, were analysed per chain. Proximities to any part of lipid, as well as proximities to only phosphate groups were analysed per chain as well as averaged across chains. 6 Å was used as the contact cutoff and separation was measured from 4 Å until infinity. In t-tests of states, only highly significant differences between samples ($p \leq 0.05$) were recorded.

6.3. Results

6.3.1. MD system configuration

This section retrieves data that were necessary for further simulations and analysis: a stretch-responsive ENM and a way to calculate surface tension.

6.3.1.1. Derivation of an ENM that allows tension sensing

Production simulations Results for the change in MscS pore diameter during production at zero membrane tension are shown in Table 6.3.1. The extra harmonic potentials implemented in the "pore support" regime to prevent pore collapse did not fulfill their function. While "Custom H-bonds" did stop the pore collapsing, they also prevent pore widening by membrane stretch, so is not applicable.

The force field modifications implemented to avoid ENM by increasing backbone attraction resulted in overall protein compaction so that the pore was hardly distinguishable (data not shown). These models were therefore deemed unfit for purpose.

Strain simulations All but three models of MscS were unresponsive to saturating membrane stretch. The pore of the 'open' crystal structure with $\leq$ 7 Å ENM widened to 1.1 nm by 1.5 nm diameter at membrane rupture tension. While I seek a stretch-sensitive ENM for the closed crystal structure, it is interesting to note that the open structure is more mechanosensitive than the closed structure under the same ENM conditions. Simulation of the 'closed' MscS structure with hydrogen-bond guided ENM widened from total pore collapse to 3 Å by stretch, which is insufficient for water flux. The third and most stretch-responsive system was the 'closed' MscS structure with β-sheets ENM, which

Structure	ENM regime	(cp)	Δ pore diameter Production	Stretch
2OAU	within 5 Å		+	0
	within 7 Å		+	0
	within 7 Å in POPC membrane		0	0
2OAU TM	within 7 Å		0	0
2VV5	within 7 Å		++	+
2OAU	"Pore support"	≤ 5	+	0
	TM: within 7 Å if between particles that are within 5, 6, 7, 8, 9 or 10 consecutive particles	≤ 6	+	0
	TM: within 7 Å for residues 90 to 92 to prevent pore-collapse	≤ 7	+	0
		≤ 8	+	0
	Cage / TM: No ENM between cage and TM	≤ 9	+	0
	Cage: ENM within 7 Å	≤ 10	+	0
2OAU TM	As above, modelling the TM only	≤ 5	+	0
		≤ 7	+	0
		≤ 10	+	0
2OAU	Hydrogen-bond guided ENM		+	+
2OAU	"Custom H-bonds"	≤ 5	0	0
	TM: Hydrogen-bonding guided ENM if between particles that are within 5, 6, 7, 8, 9 or 10 consecutive particles	≤ 6	0	0
		≤ 7	+	0
		≤ 8	0	N/A
	Cage / TM: No ENM between cage and TM	≤ 9	0	0
	Cage: ENM within 7 Å	≤ 10	0	0
2OAU	within 7 Å between β-sheet particles		+	++

Table 6.3.1.: ENM regimes tested and their results
The "Bond" CG model was used for all structures but the bottom, β-sheets ENM, which employed Martini CG. Key to pore diameter results: **0**, **+** and **++** refers to < 2.5 Å, 2.5 to 7.5 Å and > 7.5 Å change to pore diameter, respectively. **N/A** = simulation failed. (**cp**) = consecutive particles, where applicable.

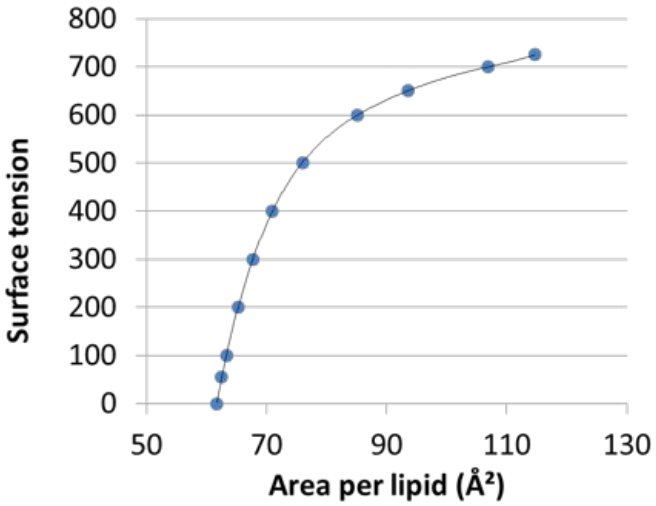

Figure 6.3.1.: The relationship between area per lipid and surface tension
Data from eleven equilibrium simulations at different surface tensions (blue dots). Polynomial fitted to the data (black curve). Surface tension units are bar*nm.

reached a 13 Å diameter gate by stretch. This model is used throughout the rest of this thesis. When other ENMs are used, such as 'full' or cage', they are applied in addition to this basal β-sheets ENM (these latter ENMs are explained later in the text).

6.3.1.2. Relationship between area per lipid and surface tension

Figure 6.3.1 shows the polynomial relationship between APL and surface tension (γ). Notably, while APL increases linearly with system deformation, surface tension does not.

Accuracy The derived relationship between APL and surface tension was tested on a set of five simulations run at a constant target surface tension of 600 bar*nm. The average surface tension results across the five simulations was 615.52$\pm$2.88 SD. My algorithm therefore performs well, and over-estimates surface tension only slightly.

6.3.2. Protein adaptation to increasing surface tension

Having arrived at an ENM that permits MscS to respond to membrane tension, and a function that allows me to quantify surface tension, I focus on the protein's response to membrane stretch. In this section, I first monitor the gate diameter to arrive at the surface tension necessary for gate opening. Subsequently I analyse the tension-response of metrics that have been associated with the gating mechanism of MscS in previous studies, namely the TM domain in-plane area, paddle tilt, cage compression and the geometry of the pore-lining helix. Guided by these data, the full conformational change undergone by MscS as it opens is treated later in section 6.3.4.

6.3.2.1. Determining the surface tension that opens MscS

Binary results for whether MscS opens or not are shown in Figure 6.3.4.

Linear deformation simulations

The minimum pore diameter at Leu 105 and Leu 109, which make up the hydrophobic gate, is shown as a function of surface tension in Figures 6.3.2 and 6.3.3. The results show that just over half of

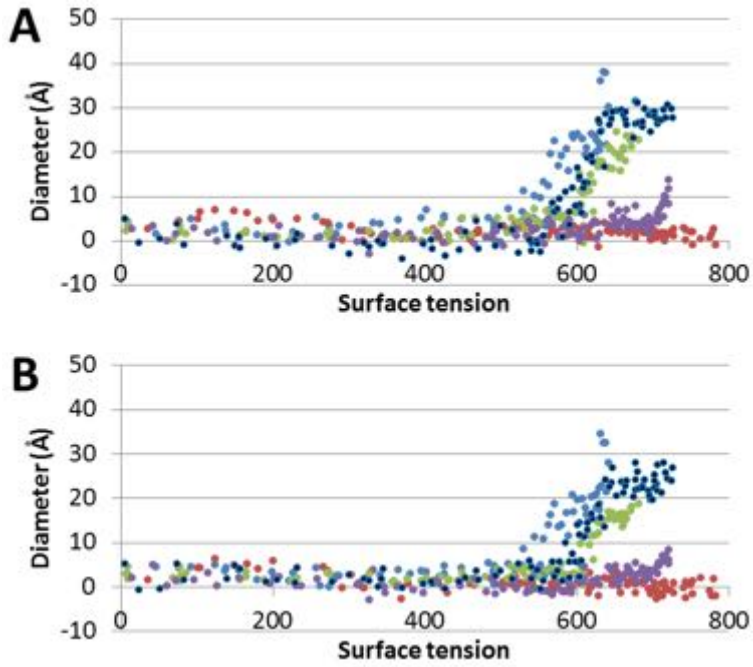

Figure 6.3.2.: Hydrophobic pore diameter as a function of surface tension:
1μs linear deformation without ENM
A. Minimum diameter at Leu 105. **B**. Minimum diameter at Leu 109. The different data point colours refer to individual simulations. Surface tension units are bar*nm.

simulations of MscS without ENM result in channel gating, despite saturating tension.

Surface tension necessary for gating Figures 6.3.2 and 6.3.3 show that surface tension below 500 bar*nm is ineffective at gating MscS. Between 500 and 600 bar*nm, most systems that do gate reach an open pore diameter. Only simulations where the production simulation, that preceeded linear deformation, were run without ENM require up to 700 bar*nm to open. Therefore, it seems that the protein might relax into a less tension-sensitive conformation from the PDB id 2OAU starting structure if given the chance to equilibrate prior to tension application.

Linear deformation of MscS with full ENM does not result in pore widening (see Figure 6.3.4).

Comparison to published data If each simulation is interpreted as representing a single channel in a patch, the effective midpoint tension is approximately 600 to 650 bar*nm. This is an order of magnitude higher than the experimental midpoint for MscS; 55 bar*nm. However, it is just above the surface tension that has been unsuccessfully used to gate MscS in AT MD setups, and lower than the tension that opens CG MscL [Colombo et al., 2003, Sotomayor and Schulten, 2004]. Therefore, this midpoint is likely to be correct, given the computational methodology.

Instant constant surface tension simulations

Guided by results from linear deformation simulations, I set up a series of instant constant surface tension simulations to explore the effects on MscS of very rapid tension onset and prolongued high, constant tension.

Results for tension below 600 bar*nm Neither twice or seven times the experimental midpoint surface tension, 110 and 400 bar*nm respectively, opens the gate (data not shown). 400 bar*nm

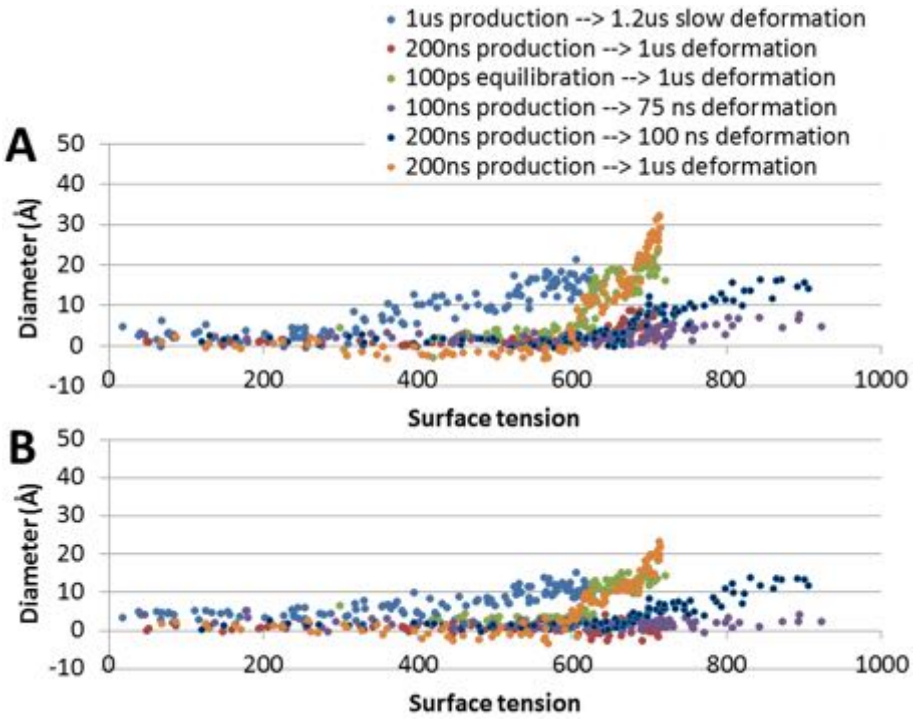

Figure 6.3.3.: Hydrophobic pore diameter as a function of surface tension:
Various linear deformation simulations without ENM. The different data point colours refer to individual simulations.
A. Minimum diameter at Leu 105. **B**. Minimum diameter at Leu 109. Surface tension units are bar*nm.

widens the gate until it reaches an apparent steady-state around 10 Å diameter, which is insufficient for gating.

MD midpoint tension MscS invariably opens by 600 bar*nm (Figure 6.3.5), which supports that the gating tension determined by the previous linear deformation simulations is correct. At this surface tension, the membrane always ruptures within 1 μs. The fact that MscS opens prior to membrane lysis, as it should under physiological conditions, supports this midpoint tension value.

The gate diameter plateaus around 30Å. This matches the diameter achieved for MscL using tension and steered MD [Deplazes et al., 2012]. Although this is a very large diameter, the MscS pore could conceivably open this wide; it would mean that the hydrophobic pore ceases to be the solute bottleneck and that the cage portals limit passage instead. However, such a large pore could also be indicative of the protein tertiary structure unravelling at near-lytic tension. For this reason, only frames where the pore diameter is $\leq$15Å are analysed when exploring the gating mechanism of MscS in future thesis sections.

The two Leucine rings do not open simultaneously

Leu 105 is invariably faster at responding to tension than Leu 109. This holds whether tension is applied rapidly or slowly.

6.3.2.2. MscS in-plane area expansion as a function of tension

The closed MscS TM domain without ENM expands by tension, whether MscS opens as a result or not, and irrespective of the speed of tension (see Figure 6.3.4). However, fast tension application generally widens the TM domain further than a slow linear deformation does.

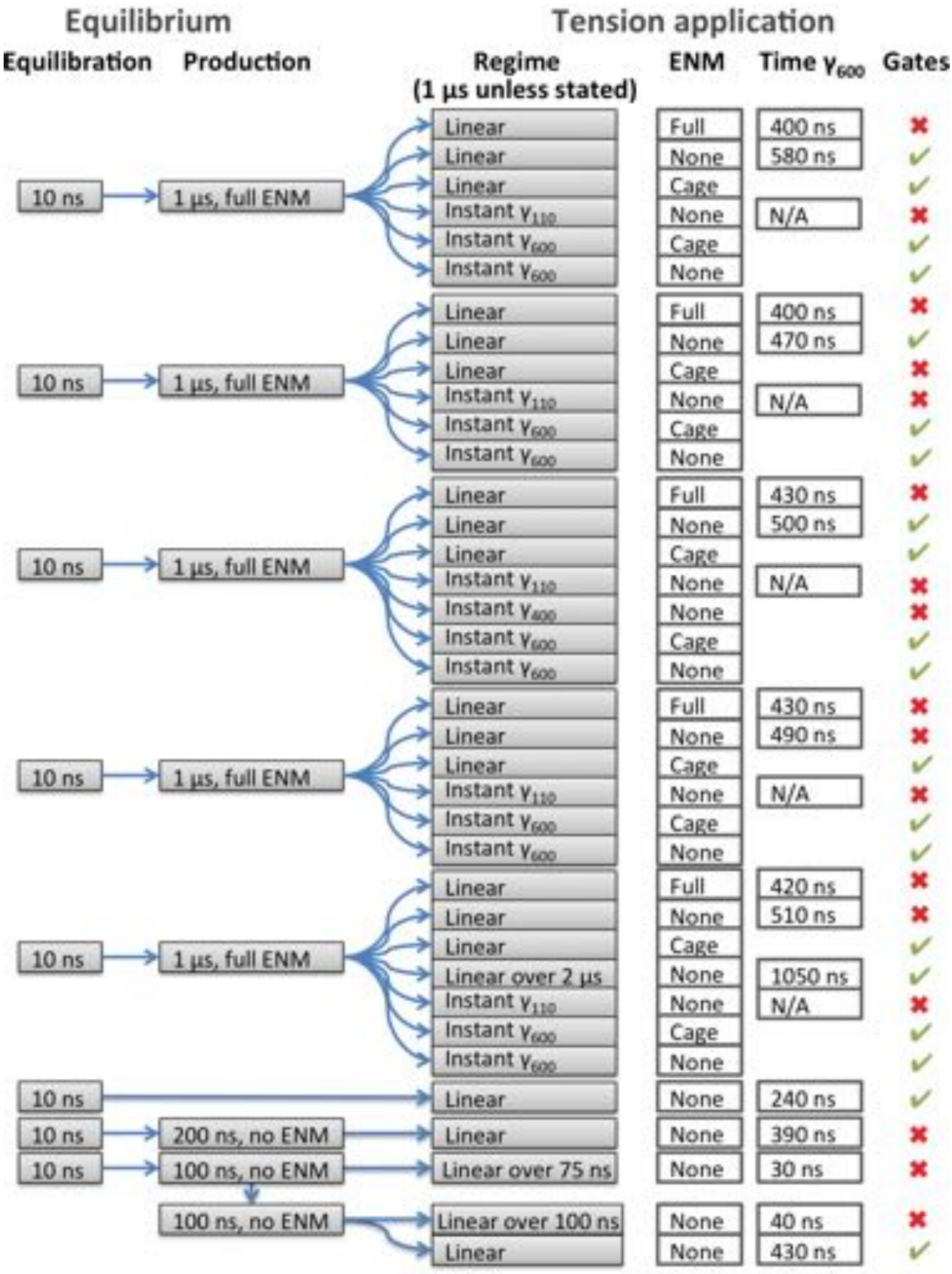

Figure 6.3.4.: Results for simulations at non-zero surface tension

"Gates" refers to whether the simulation achieves an open, 15 Å bottleneck diameter. "Time γ600" refers to simulation duration necessary to achieve 600 bar*nm surface tension. That is an indicator of the speed of tension application.

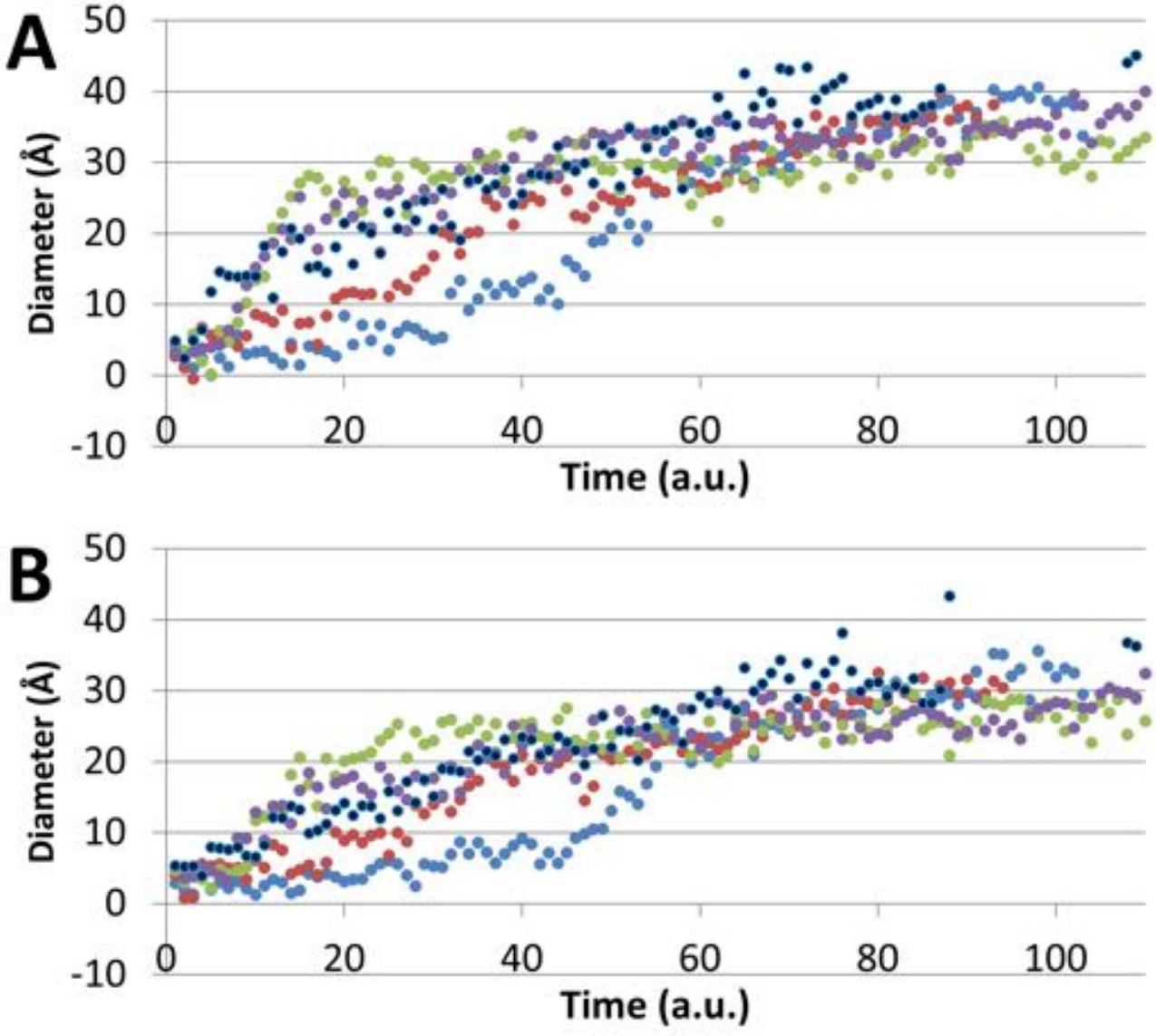

Figure 6.3.5.: Pore diameter resulting from 600 bar*nm surface tension
Frames where the membrane had ruptured were removed and the remaining time was fitted to approximately 100 frames. Time is shown in arbitrary units. **A**. Leu 105 minimal diameter **B**. Leu 109 minimal diameter

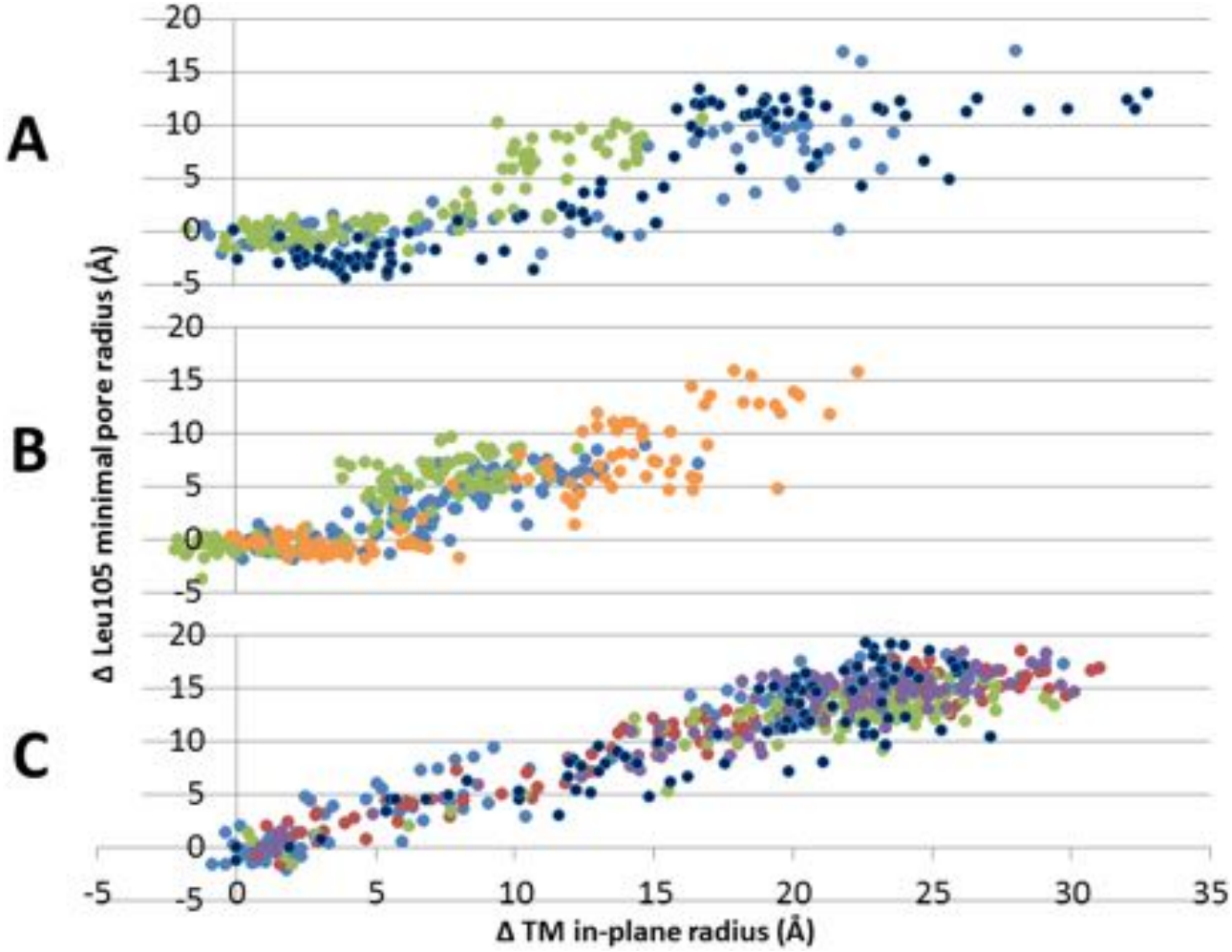

Figure 6.3.6.: TM domain versus pore expansion in simulations where MscS gates.
A. 1μs linear deformation without ENM (c.f pore diameters in Figure 6.3.2). **B**. Various linear deformation simulations without ENM (c.f. pore diameters in Figure 6.3.3). **C**. 1μs simulation at 600 bar*nm (c.f. pore diameters in Figure 6.3.3). Data point colour scheme is the same as in pore diameter graphs, for easy data identification and comparison.

Change to pore radius is plotted against change to TM domain radius for all simulations where MscS gates in Fig. 6.3.6. All changes to values are quoted with respect to the closed MscS structure dimensions. Results reveal two modes of opening, depending on whether tension is applied slowly or rapidly. In simulations were tension is slowly applied, there is lag between TM domain expansion and pore widening; the TM domain needs to expand its radius by 5-10 Å before the pore widens (see Figure 6.3.6 **A** and **B**). Rapid tension application (600 bar*nm instant constant tension), on the other hand, causes immediate gate response to TM domain expansion; their radii relationship is near linear (c.f. Figure 6.3.6 **C**). The in-plane expansion that MscS has undergone at the time of pore opening is therefore also affected by tension rate. Rapid tension application requires only 2.5 Å TM domain radius increase (equivalent to 7 nm^2 area increase) for gating, while slow tension application requires 6 to circa 10 Å TM domain radius increase (circa 15 to 30 nm^2 area expansion) before the gate is open. This is comparable to theoretical values of 8.4 and 15 nm^2 in-plane area expansion necessary for gating [Sukharev, 2002, Belyy et al., 2010b].

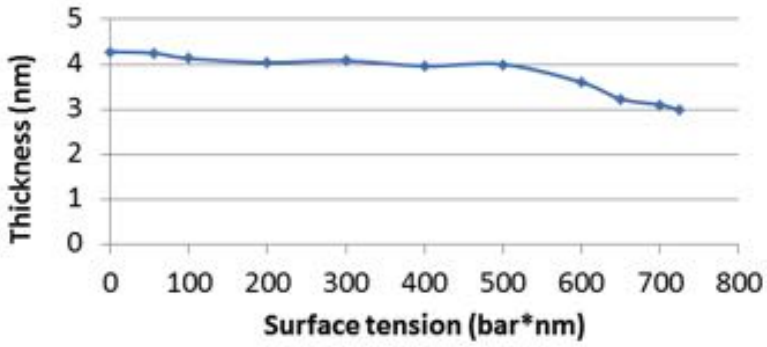

Figure 6.3.7.: Membrane thickness as a function of surface tension.
The blue line shows the distance between the peak phosphate group density of each membrane.

6.3.2.3. Membrane and protein compression by tension

Membrane thickness as a function of tension

Membrane thickness as a function of tension is plotted in Figure 6.3.7. Initially, as tension increases, bilayer thickness decreases only slightly, plateauing at 4 nm up until and including 500 bar*nm, after which thickness decreases steeply to 3 nm minimum prior to lysis. This closely resembles values obtained in previous studies of CG membrane [Colombo et al., 2003]).

TM helix tilt angle is correlated with membrane thickness

TM1, the most peripheral TM helix which makes up half of the sensor paddle, was analysed for its length and angle with respect to the membrane normal under increasing tension. TM1's length is largely invariant; 45.44 Å ($\pm$0.72 Å SD). At zero tension, TM1 is tilted at an angle of approximately 60° (see Figures 6.3.8). During linear deformation, TM1's tilt angle then follows the membrane thickness closely until 500 bar*nm, at which point TM1's angle decreases at a higher rate. Near lytic tension, TM1 would only have to be at an angle of 41.3° to remain within the 3 nm bilayer, yet it invariably tilts at more acute angles of < 30°.

Interestingly, at $\geq$ 600 bar*nm, tilt angle is indicative of the propensity of MscS to open. Simulations where MscS opens achieve a more membrane-parallel TM1 conformation at high tension than do simulations where MscS does not gate. A sensor paddle that is near-parallel to the membrane appears to be conducive to MscS gating, but only if paddle tilts fast in response to tension. Moreover, all instant constant surface tension simulations at 600 bar*nm (all of which open) also locate to the left-hand-side, supporting this hypothesis (data not shown).

6.3.2.4. Cage distortion without full ENM

In simulations without full ENM, where only β-sheet ENM is present, not only the TM domain, but also the cytosolic cage compacts in the membrane normal axis as tension increases. This leads to portal closure and relocation of the pore constriction point to the cytosolic side, as the β-barrel compacts about the cytosolic TM pore exit (see 6.3.9). Interference between cytosolic and TM domains has also been observed for MscL under tension in both AT [Gullingsrud et al., 2001] and CG MD simulations [Yefimov et al., 2008, Deplazes et al., 2012] . MscL's cytosolic domain compacts and restricts or completely blocks the TM passageway.

Experimental results reveal that the cage domain does rearrange during gating [Kaprowski and Kubalski, 2003, Miller et al., 2003b]. Indeed, the MscS β-barrel is known to have regulatory function; compaction towards the TM domain is associated with channel inactivation [Koprowski et al., 2011]. However, the large deformation of the entire cage that is seen in my simulations suggest that, rather

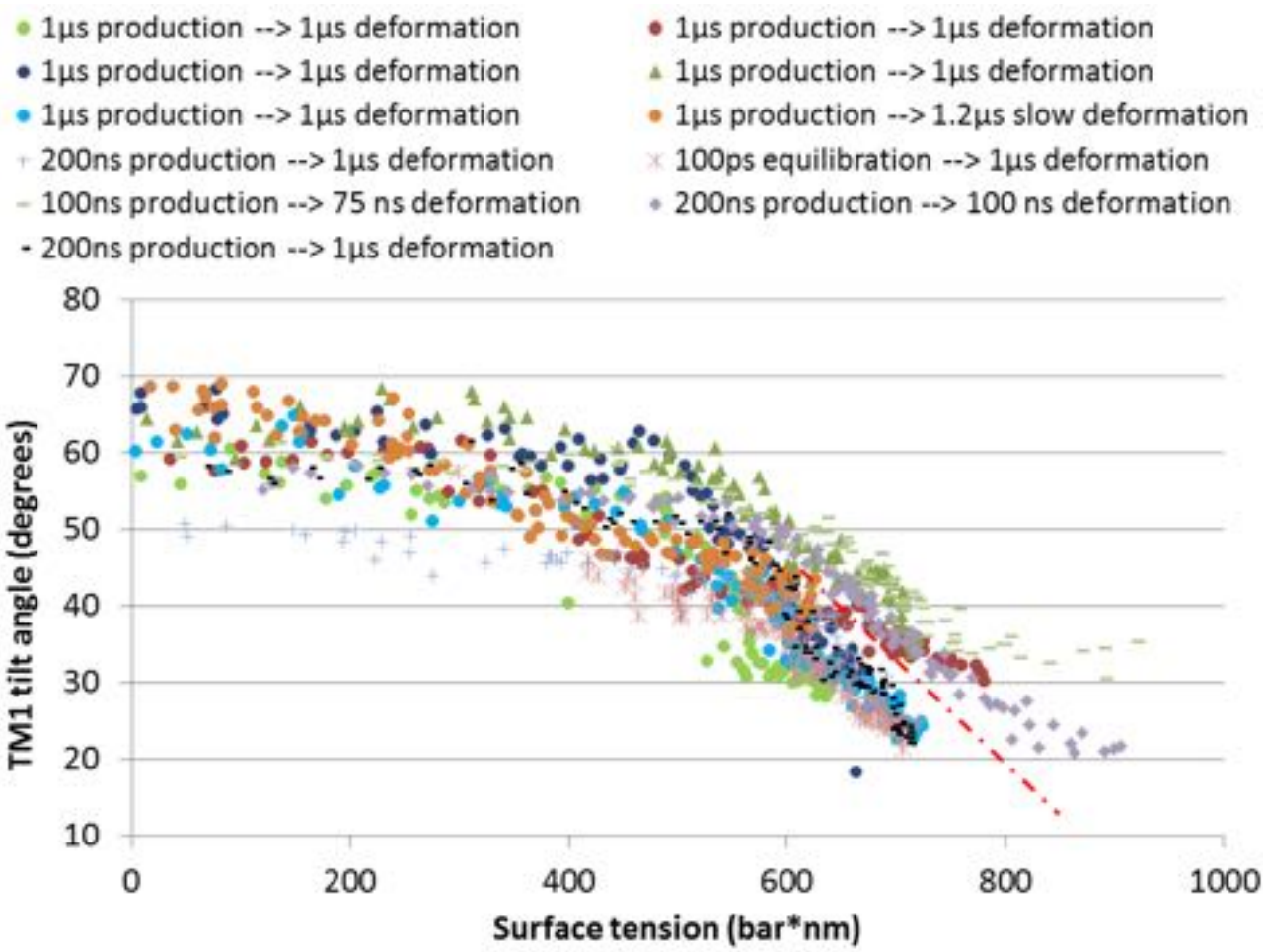

Figure 6.3.8.: TM1 angle with respect to the membrane: linear deformation simulations
Results for tilt angle with respect to the membrane, plotted as a function of surface tension, for all linear deformation simulations. The red dash-dot line delineates simulations where the gate remains closed (right-hand side) from simulations where all but one MscS opens (left-hand side).

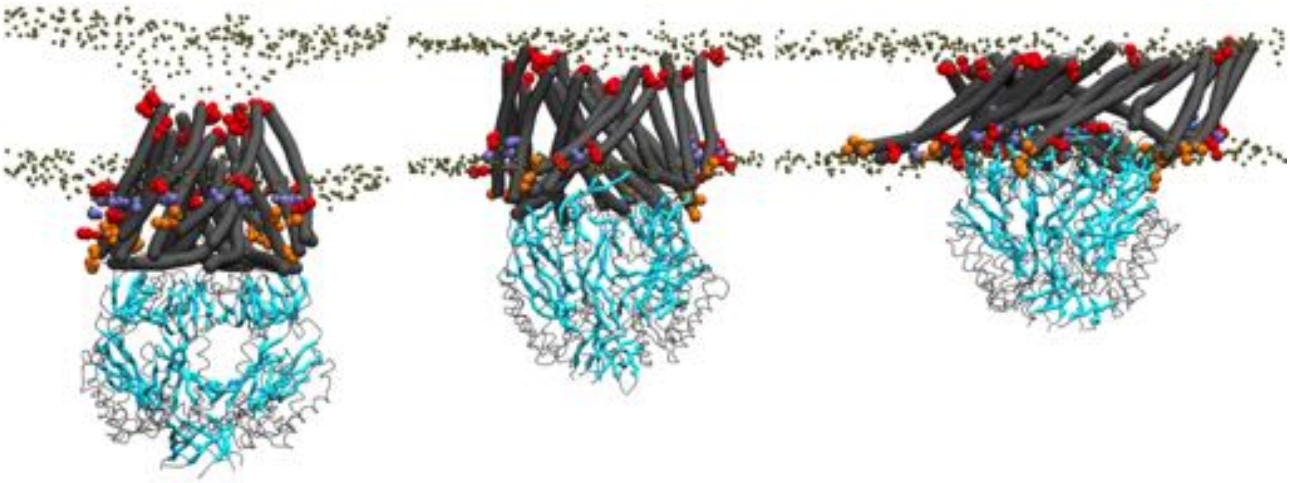

Figure 6.3.9.: Cage distortion in ENM-free simulations.
The MscS structure at simulation start (left), middle and end (right) of a linear deformation simulation where MscS gates. The cage moves into the TM domain and blocks passage. The TM domain is drawn in gray bendices and the cage is drawn using the Bendix Join representation. More details regarding residues shown in coloured van der Waal representations is provided in the section on MscS' gating mechanism. The image is representative for all linear deformation simulations where MscS is not treated with ENM. For more screenshots from the gating transition, see Appendix D.4.1.

than showing the inactivated state, that the CG MscS cage is unstable without ENM. For this reason, I performed simulations of MscS where the ENM was limited to the cage.

The effect of ENM on the cage An ENM was created for solely the cage, avoiding residues 27 to 127. ENM between the TM and cage was also rejected. This model is from now on called 'cage-ENM'. Cage-ENM was used in five 1us linear deformation simulations and five instant constant tension simulations run at 600 bar*nm. These simulations were set up as continuations of the five 1us production runs (where the full MscS structure was treated with ENM) that were previously used (c.f. Figure 6.3.4). Results are seen in Figures 6.3.9 to 6.3.12.

Linear deformation Cage-ENM results in slightly higher propensity to gate compared to MscS with only basal β-sheet ENM. MscS gates in 4 out of 5 simulations with cage-ENM, compared to 3 out of 5 with only β-sheet ENM (see figures 6.3.10 and 6.3.2). Otherwise, pore dynamics is comparable; Leu109 is the bottleneck, 600 bar*nm gates MscS on average and MscS with cage-ENM has the same capacity to widen as MscS lacking ENM. In-plane expansion behaviour is also similar; expansion requires higher tension in non-gating than gating MscS. However, gating or not, all cage-ENM simulations attain $\geq$50 nm^2 expansion, whereas expansion is visibly reduced in non-gating MscS lacking ENM (see Figure 6.3.12 **A**; expansion data for β-sheet ENM simulations undergoing linear deformation are not shown).

Instant constant surface tension While all instances of MscS lacking ENM reach $\geq$20 Å pore diameter half-way to lysis, MscS with cage-ENM is restricted and gating is a graded response to time. Near-lysis is necessary to gate all cage-ENM simulations (c.f. Figures 6.3.5 vs 6.3.11). This graded response is also seen with respect to in-plane expansion (see Figure 6.3.12 **B**).

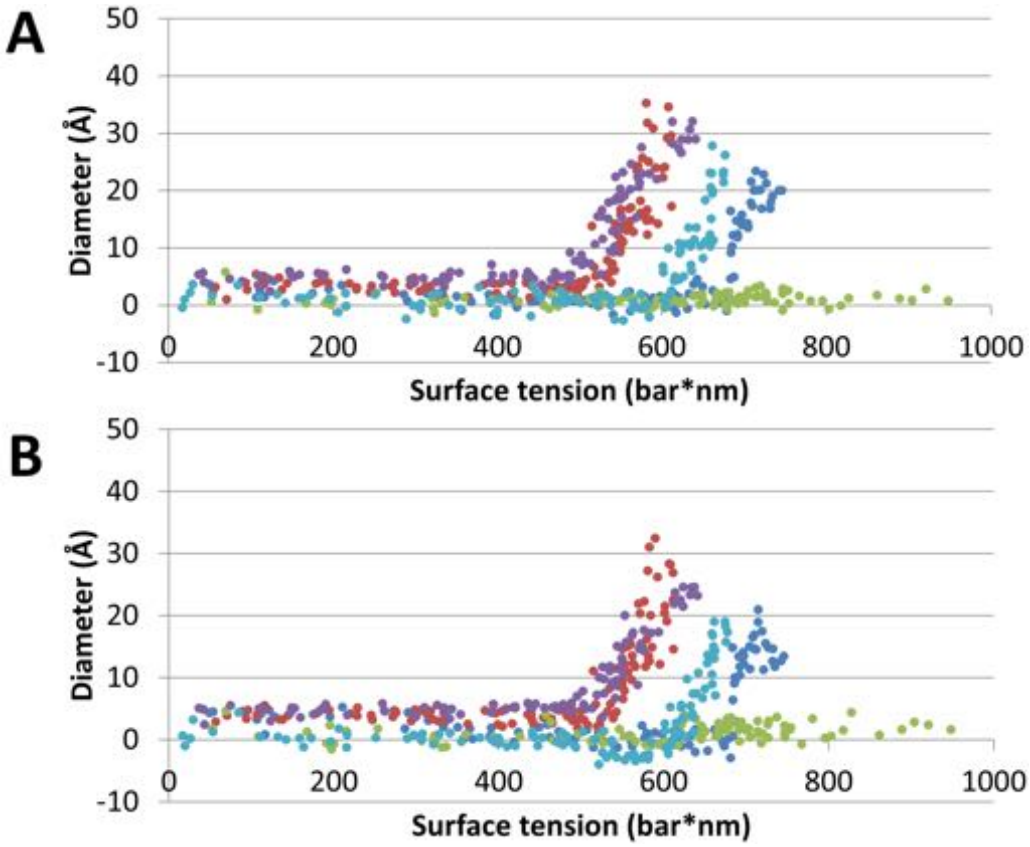

Figure 6.3.10.: Pore dimensions during linear deformation simulations of MscS with cage-ENM. MscS pore diameter at Leu 105 (**A**) and 109 (**B**). Each colour denotes a separate 1μs linear deformation simulation.

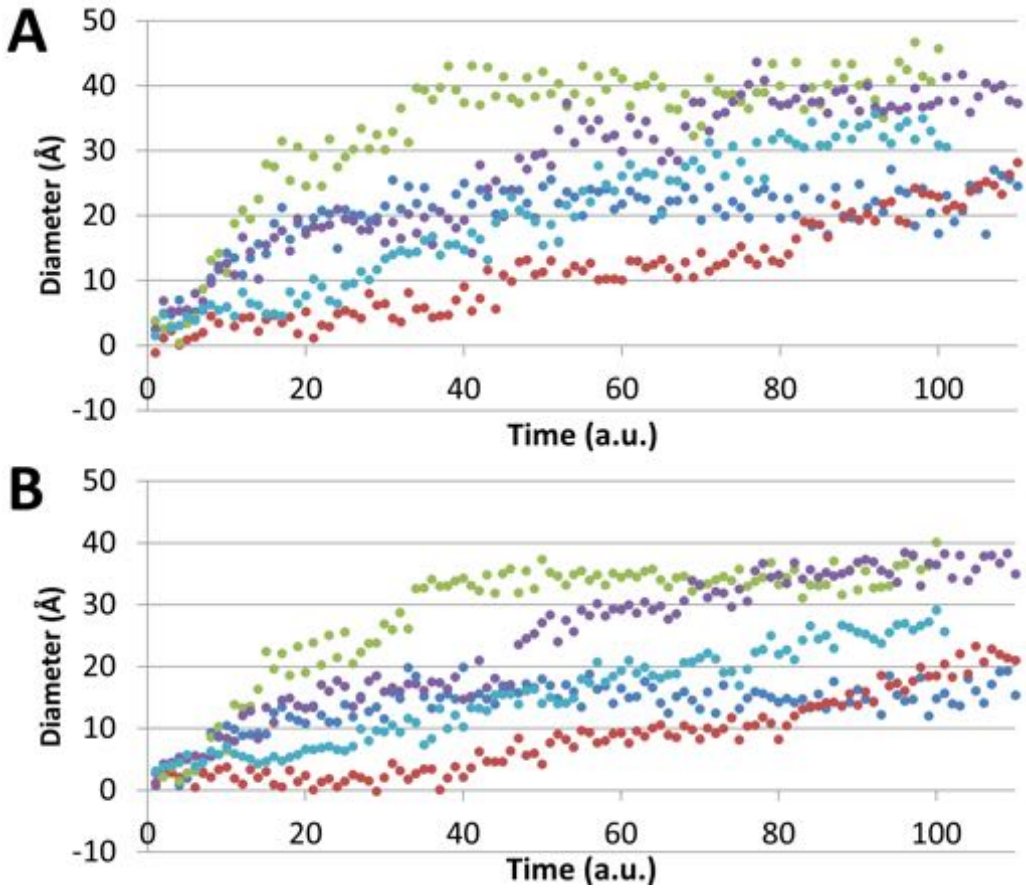

Figure 6.3.11.: Pore dimensions during instant constant tension simulations of MscS with added cage-ENM.
MscS pore diameter at Leu 105 (**A**) and 109 (**B**). Each colour denotes a separate 1μs instant constant tension simulation at 600 bar*nm.

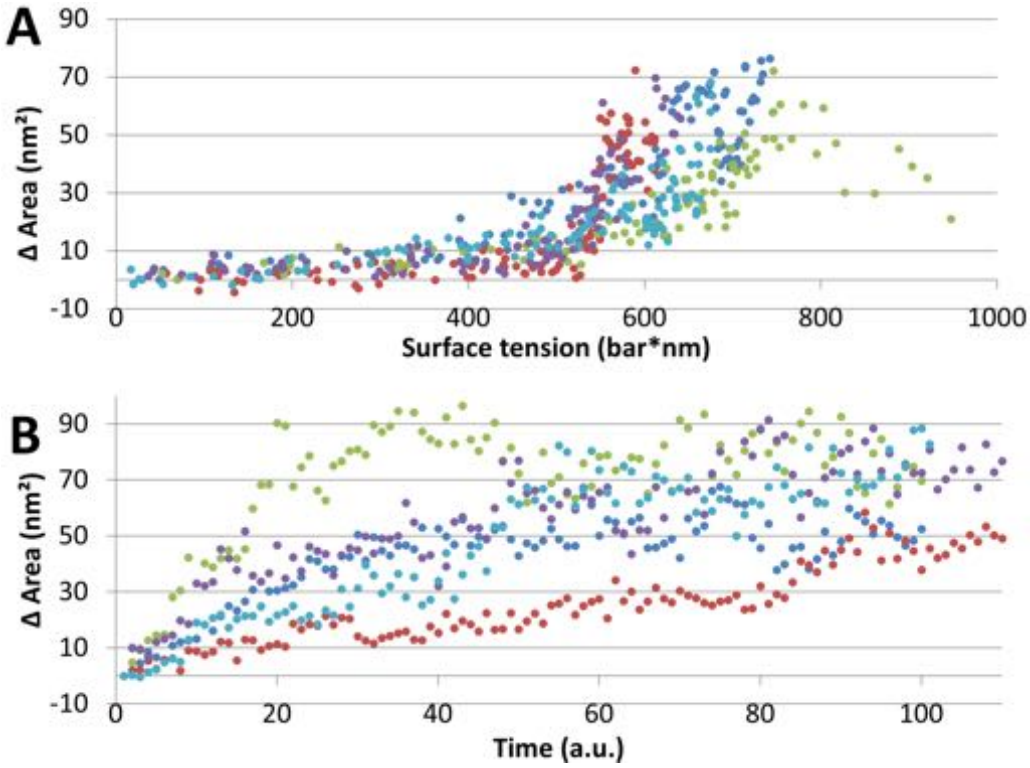

Figure 6.3.12.: In-plane expansion in simulations of MscS with cage-ENM.
MscS in-plane expansion during linear deformation (**A**) and instant constant tension simulations at 600 bar*nm (**B**). Each colour denotes a separate 1μs simulation, and the same colour is used per dataset as in Figures 6.3.10 and 6.3.11.

6.3.2.5. TM3 helix geometry

Previous studies suggest that the geometry of the pore-lining helix TM3 is critically involved in MscS gating (see the Introduction for details). It is therefore analysed in this section.

Curvature The evolution of distortion of TM3, was evaluated using Bendix for all simulations where MscS gates. Results for 77 TM3 helices were averaged to give the maximum angle evolution from zero to gating surface tension in Figure 6.3.13. The maximal helix distortion decreases from 30° to 18° as MscS gates, suggesting that the helix never straightens completely. Figure 6.3.14 reveals that the location of the maximal bend in TM3 is in the area around residues 112 to 115 when MscS is still closed, and that the angle foci moves to Leu115 one third through gating. Throughout the latter half of the gating mechanism of MscS, TM3 reaches a steady-state of 12-15°; half the initial maximal angle, which is not localized but rather spans the length of the helix (see Figure 6.3.14). TM3 thus remains gently curved as MscS opens; nothing remains of the acute angle at Gly113 that partitions TM3 into TM3a and TM3b in the closed crystal structure.

Results for TM3 distortion in individual chains is seen in Figure 6.3.15. Bendix results reveal a degree of asymmetry across chains. Chain A is distinctly more distorted than any other chain at the onset of tension, and the adjacent two chains feature bends that are 5 to 10° smaller in comparison. It is plausible that steric hinderance could be preventing adjacent chains from adopting the same high helix distortion. Note that these unequal initial states of helix distortion likely originate in the asymmetric crystal structure; the starting structure for these simulations. Also noticeable is the gradient of bends across chains C to G; while the curvature of each of these chains decrease as MscS gates, chain C has the highest initial curvature and the other chains follow suit in approximately the order of the chains.

There is also a certain asymmetry in helix curvature as MscS gates. Of all chains, Chain C comes

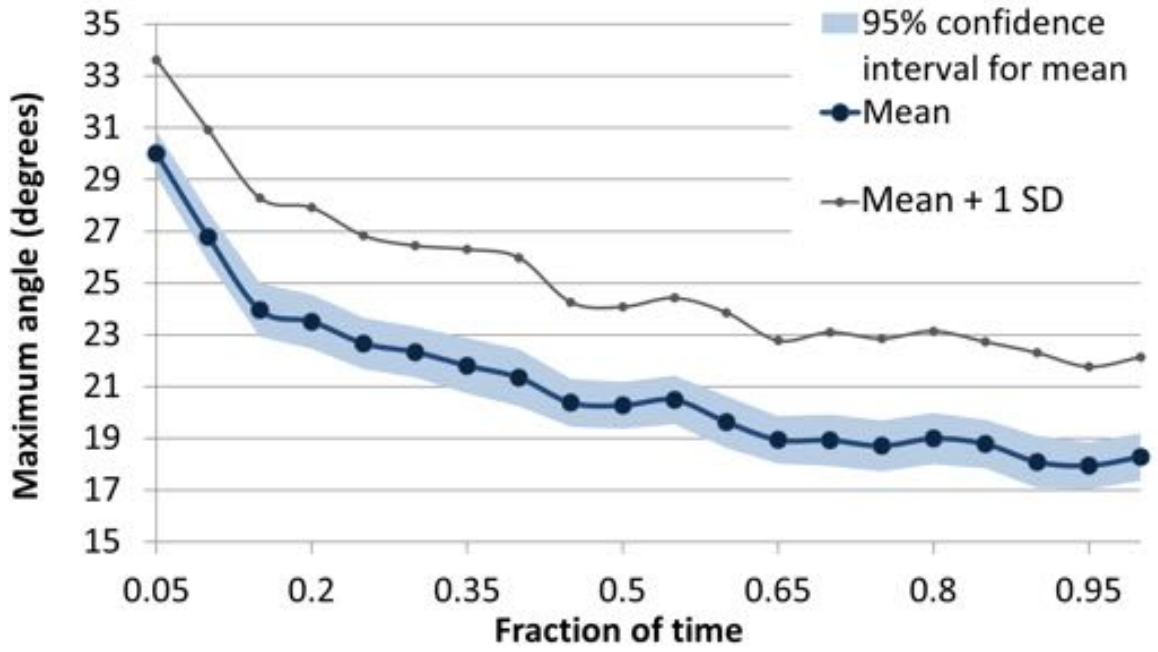

Figure 6.3.13.: Evolution of maximum TM3 angle as MscS gates.
Mean maximum angle values for 77 TM3s, 1 SD and the 95% confidence interval from zero tension to when the pore reaches 15 Å diameter. Results are binned using bin size 0.05 with respect to time: e.g. 0.2 contains results from (excluding) 0.15 to (including) 0.2 fraction of trajectory time.

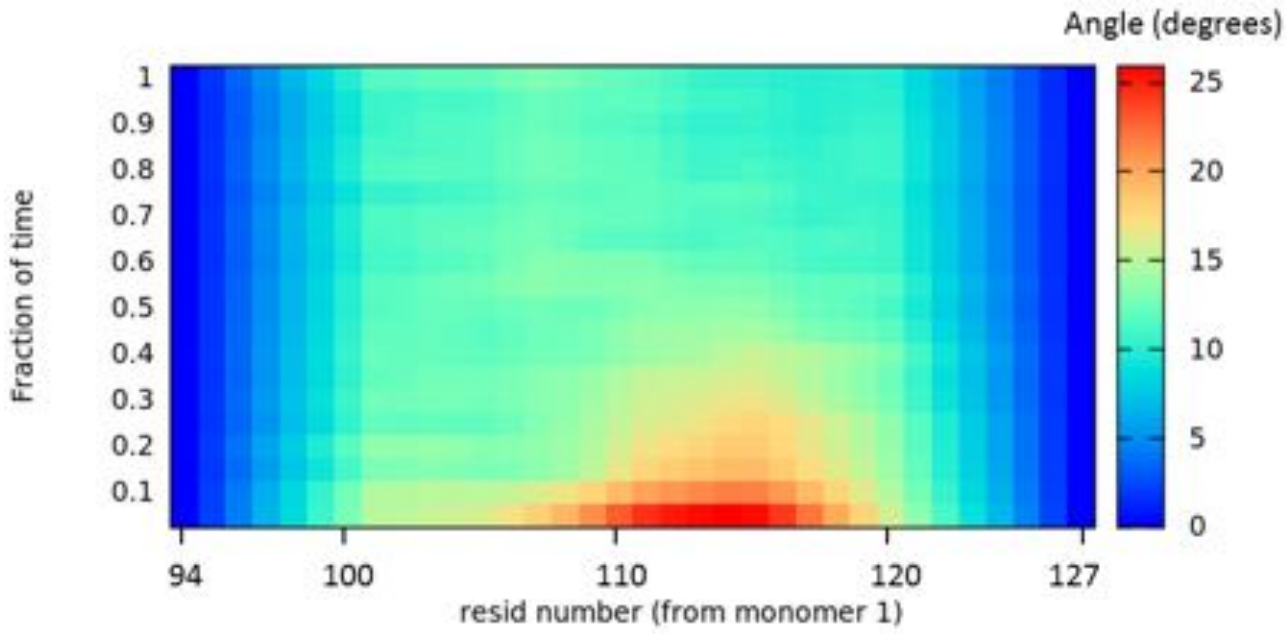

Figure 6.3.14.: Evolution of TM3 helix angles as MscS gates

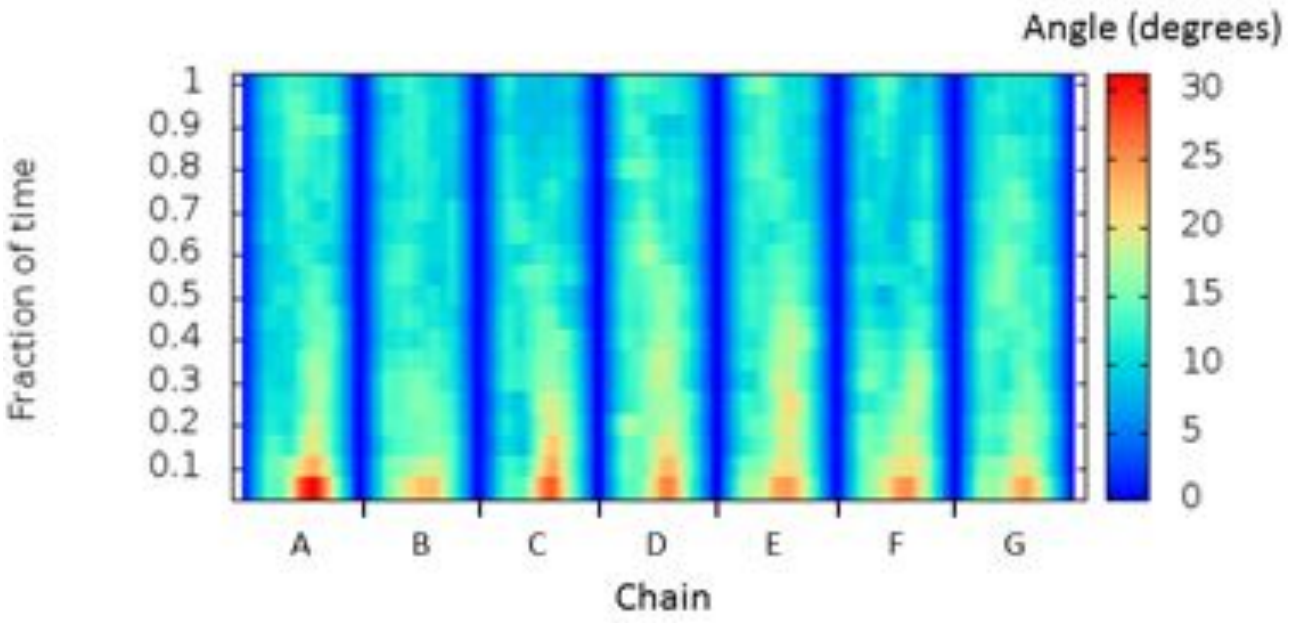

Figure 6.3.15.: Evolution of each chain's TM3 helix angles as MscS gates

closest to straightening completely, while chain D and E retain a 15° angle. On the way to MscS opening, each chain reach an approximate steady state half-way through gating, as was also observed in Figure 6.3.14.

Twist At gating tension, TM3a has twisted anti-clockwise (viewed from the pericellular space) by approximately 45 to 90°. The evolution of the twist of TM3 can be seen in Figure 6.3.16. This twist is in the opposite direction to the twist observed in the 'open' crystal structure 2VV5 (see Figure 6.3.17). It is interesting to note that tilt appears to be related to twist; higher degree of tilt is associated with higher anti-clockwise twist with respect to residue orientation in PDB id 2VV5. While TM3 does not display the kink at Gly113, so is not geometrically broken up into TM3a and TM3b, TM3b does not display twist. Other TM helices are also unaffected by twist.

ENM and gating speed appears to have an inhibiting effect on twist; TM3a twist is less in MscS with ENM on the cage and MscS that undergoes rapid gating, than in slowly gating MscS without ENM (c.f. Figures 6.3.39, 6.3.42 and Appendix D.4.2).

6.3.3. Membrane adaptation to increasing surface tension: Lipid-protein interaction dynamics

6.3.3.1. Considerations for what data to use

Open state sample Out of the 11 simulations where MscS gates, one simulation had only undergone 100 ps of equilibration prior to linear deformation. To avoid including potentially un-equilibrated lipid-protein interactions, this simulation was excluded from the sample of gating MscS. The remaining ten simulations include five simulations at instant constant surface tension and five linear deformation simulations where tension was applied using two different speeds. The 'open state' refers to frames from these ten simulations where the MscS pore diameter is approximately 15Å.

Closed state sample The closed state sample is composed of the ten production simulations at zero surface tension that preceded each simulation where MscS gates under tension. In order to reduce sampling of unequilibrated lipid-protein interactions, only the latter half of the production runs

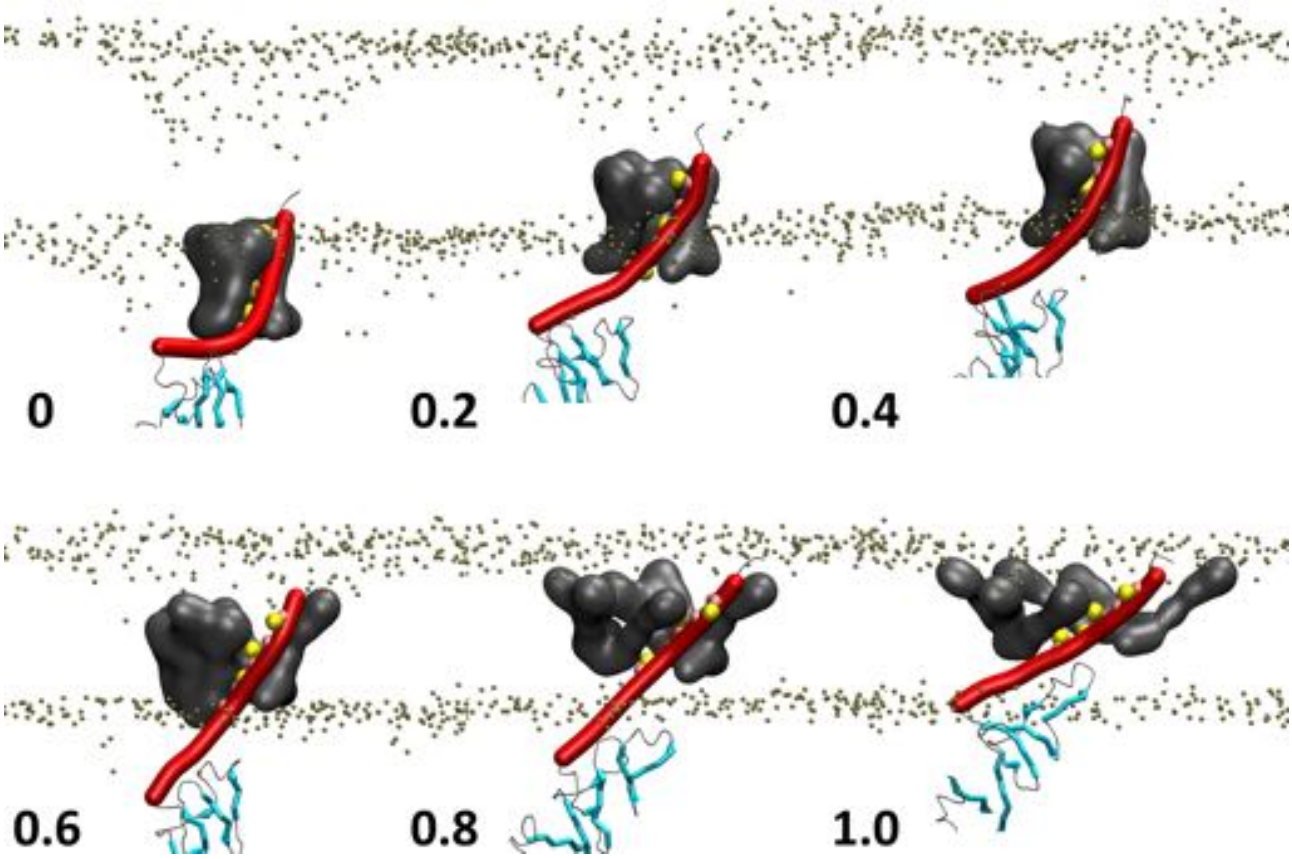

Figure 6.3.16.: Evolution of TM3 twist during linear deformation simulation
The numbers refers to fractional time of the total simulation run until gating. TM3a of chain A, shown in red bendix with pore-lining residues as pink and yellow spheres, twists anti-clockwise as MscS gates. Pore-lining residues of TM3a of other chains are shown as gray surface and phosphates of the lipid bilayer as tan spheres.

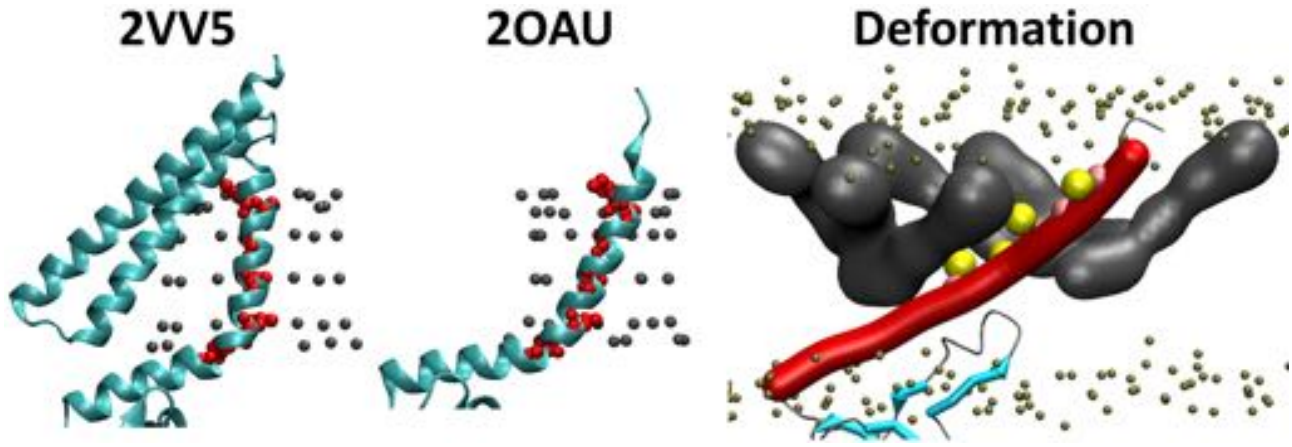

Figure 6.3.17.: TM3 twist: crystal structures in comparison to simulation
Structures ordered according to helix twist and tilt. Viewing PDB id 2VV5 as the frame of reference, TM3a in 2OAU twists by 15° as the helix tilts, and side chains are twisted by a further circa 45 to 90° by gating tension. (c.f. Figure 6.3.16). In the crystal structures, chain A is shown in turquoise cartoon representation, pore-lining residues are shown in red and Cα of the pore-lining TM3a residues of other chains are drawn as tan spheres to show the pore. Representation of the linear deformation simulation at gating tension as in Figure 6.3.16.

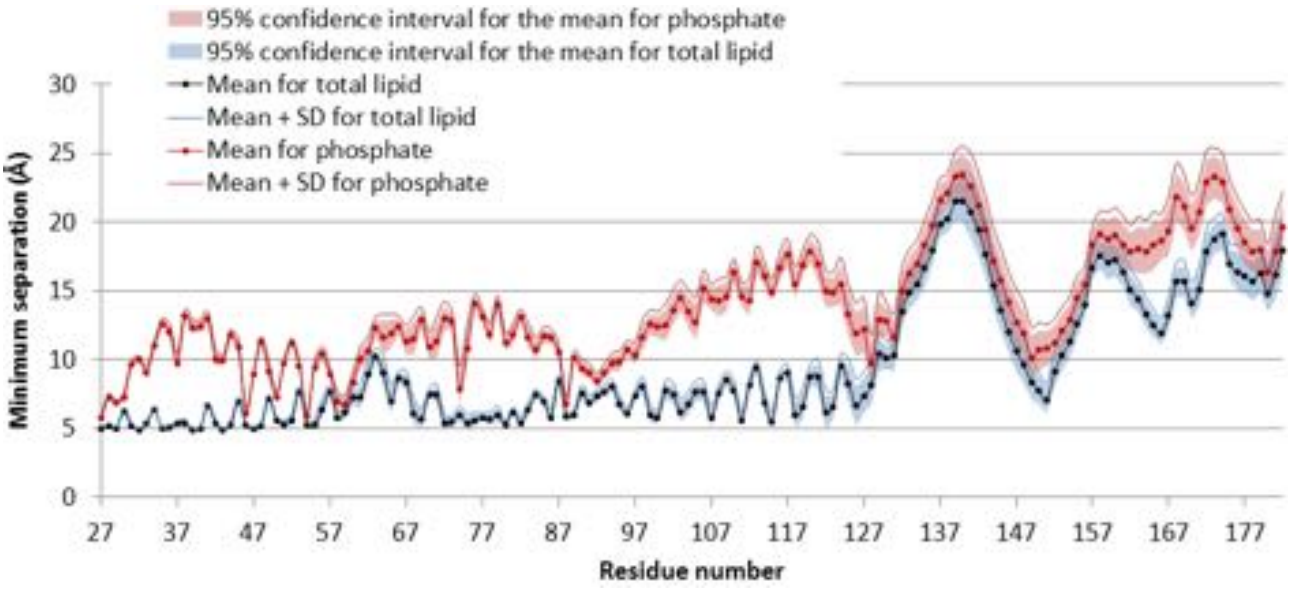

Figure 6.3.18.: The minimum separation to lipid in the MscS open state
See section 6.3.3.2 in the main text for discussion.

were analysed. As a result, nine 500 ns simulations and one 100 ns simulation were analysed in total.

6.3.3.2. The lipid interface of the open state

Persistent lipid-protein contacts across the different tension regimes included in the open state sample should suggest important interactions.

Open MscS makes more contacts to POPG than POPE, relative to system content (1.391 ± 0.262 SD vs 0.902 ± 0.064 SD). This bias increases slightly when only the lipid phosphate groups are considered; POPG's phosphate groups bind on average 1.645 ± 0.428 SD whereas POPE's phosphate groups bind 0.838 ± 0.105 SD.

Separation to lipid is lowest in the transmembrane region; residues 27 to 127 (see Figure 6.3.18). However, additional proximal contacts are seen around residue 150, which is part of the β-barrel, suggesting that it approaches the membrane during opening. Phe151 is the most lipid-proximal residue this region. Overlap between red and blue lines in Figure 6.3.18 indicates where lipid phosphate groups are principally responsible for the interaction between lipid and MscS. Proximity to lipid phosphate is particularly apparent to charges along the sensor paddle, where the most proximal interactions are made to residues Tyr27, Asn 30 and 50, and Arginines 46, 54, 59, 74 and 88. On average one phosphate group makes contact to each Tyrosine and Arginine, while Asparagines make weaker contacts (c.f. Figure 6.3.19). Curiously, while a residue can make contact with up to three different lipids at any one time, the residue only makes contact to a single phosphate group at a time. Phosphate proximity is not high around residue 150, so that interaction with the β-barrel during opening is probably not mediated by phosphate groups.

6.3.3.3. Changes to the lipid interface as MscS opens

Differences between closed and open state Paired t-tests were performed between closed and open states. Closed sample data were subtracted from open sample data.

As revealed in Figure 6.3.20, proximity to lipid changes dramatically by gating. Most residues beyond the TM region get closer to lipid, but only in the region around residue 150 does this result in actual contacts (c.f. Figure 6.3.21).

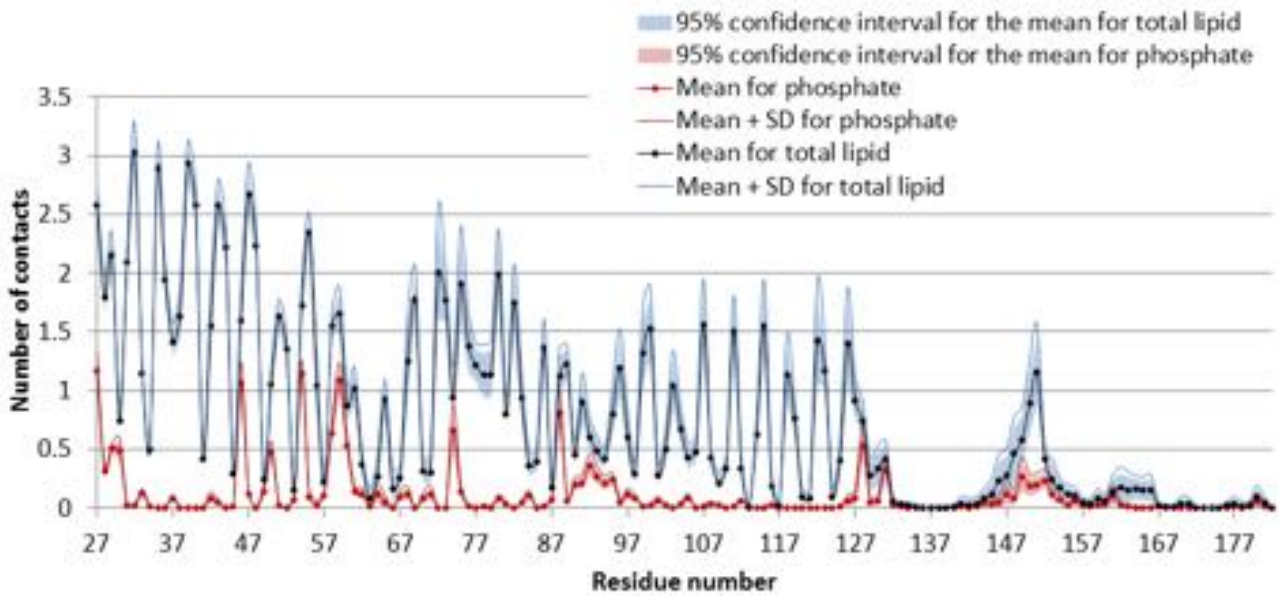

Figure 6.3.19.: Number of lipids in contact with the MscS open state
See section 6.3.3.2 in the main text for discussion.

In the TM domain, change to proximity is most prominent for lipid phosphate groups. The gap between the paddle and TM3a gains phosphate interactions, while TM3b and C-terminal TM2 are subject to a small loss. However, most of these differences between states occur distally from MscS; contacts are only lost at Arg 74 and gained at Arg 128 (see Figure 6.3.21). The distal gain in phosphate contacts to TM3a can be explained by its relocation from a shielded position within the closed protein interior to being in line with the pericellular leaflet in the MscS open state. In the open state, seven TM3a helices form a widened pore in the planar membrane under tension and, as a result, TM3a is located closer to lipid headgroups.

Contacts analysis reveals more change to protein-lipid tail interaction than protein-phosphate interaction (Figure 6.3.21). Loss of contact is prominent along TM1 and the space between the paddle and TM3a. The zig-zag pattern caused by contact loss immediately adjacent to contact gain along TM3a is reminiscent of the twisting of the pore-lining helix as MscS opens, which is seen in comparisons between the two crystal structures, and supports that this feature of PDB id 2VV5 is a state on the way to gating. The peaks and troughs of the zig-zag are asymmetric, suggesting lipid loss from the space between the paddle and TM3a, in particular at residues Ala 103 and Val107, which are located immediately adjacent to the hydrophobic gate, but face the cavity formed by the paddle and TM3.

Contacts along the cytosolic end of the paddle are altered according to location. Cytosolically located charged residues retain or gain contacts, while Arginines 46 and 54, and polar Asn50, loose more contact to phosphate the further away they are from the cytosol. Notably, the lipid contact made by Arg 59 remains insignificantly changed by gating, suggesting a persistent bond.

Dynamics of lipid interaction as MscS gates In the ten simulations where MscS opens, frames encompassing $\leq$15 Å pore diameter were used to investigate the dynamics of lipid interaction to MscS as MscS gates.

Figure 6.3.22 suggests that separation to lipid is relatively persistent during gating. While the general tendency is a small decrease in proximity to lipid, individual residues in the area between the sensor paddle and TM3 sees a slight gain in proximity at the expense of adjacent residues, reflecting the helix twist in the area. Separation to lipid phosphate groups (bottom image) is more dynamic, with the notable exception of Arginine 54 (and to a lesser extent, Arg 46). The close contact between phosphate and Arg 54 is essentially unchanged by gating, while several other interactions to

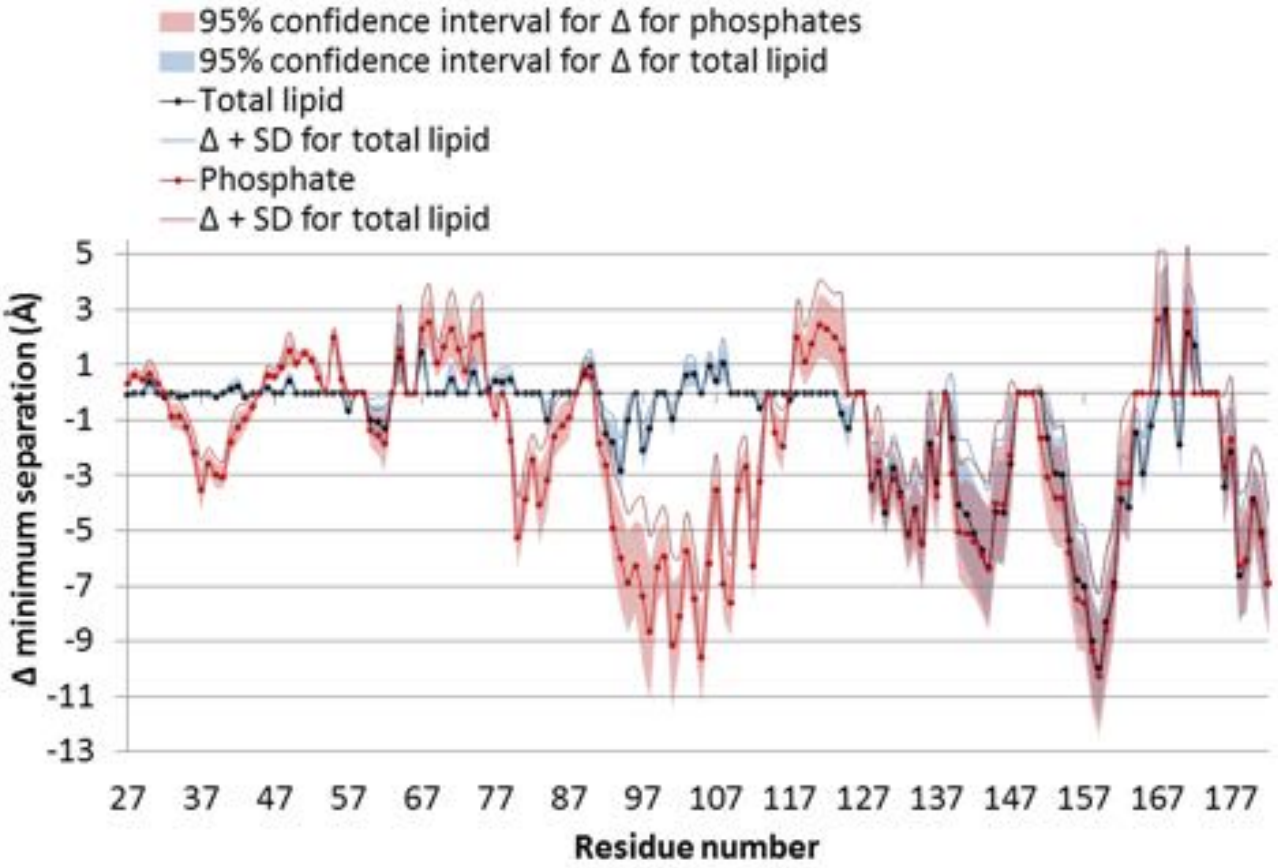

Figure 6.3.20.: Significant change to lipid-protein separation as MscS opens.
Positive values indicate that lipid is further away in the simulation under tension than at no tension, while negative values indicate that lipid gets closer by tension.

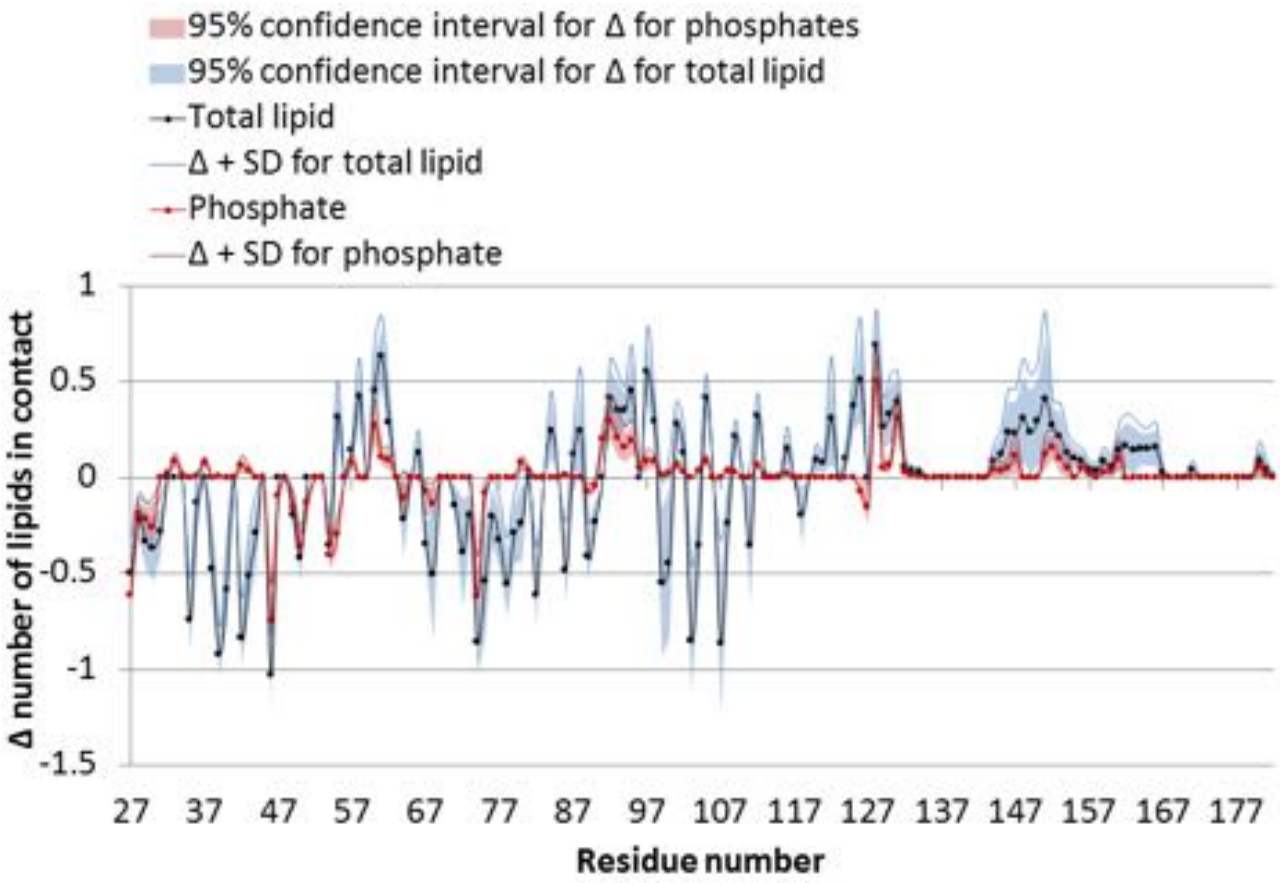

Figure 6.3.21.: Significant change to number of lipids in contact as MscS opens.
Positive values indicate that lipid makes more contact in the simulation under tension than at no tension, while negative values indicate that lipid-protein contacts decrease by tension.

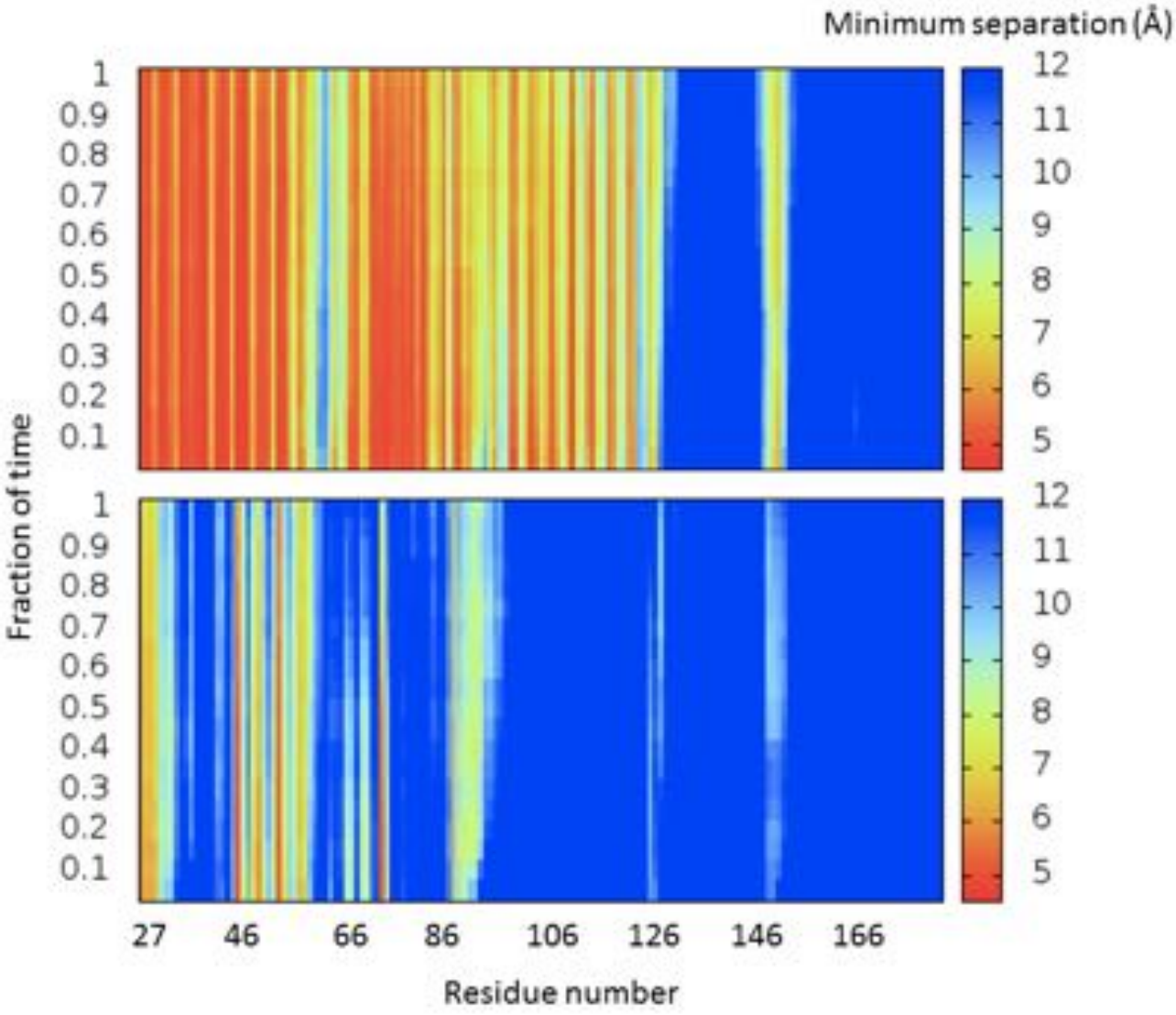

Figure 6.3.22.: Minimum separation to lipid as MscS gates, averaged across chains and trajectories Top: Separation to whole lipid per residue. Bottom: Separation to lipid phosphate per residue.

charged and polar residues on the cytosolic side of the paddle suffer with increased tension. Around residue 90 to 100, the focus of interaction to lipid phosphates 'slides' from residue 90 to residue 93 and broaden by tension**.** This is because TM3a relocates to the pericellular leaflet during gating, as explained in section 6.3.3.3.

Contact analysis (Figure 6.3.23) reveals loss throughout the TM region of both whole lipid and lipid phosphate contacts. However, when I rescale heatmap to capture single contact dynamics at high resolution (Figure 6.3.24), the data show that, while a residue can have numerous contacts and loose some of these contacts, individual lipids often remain in contact throughout gating. With respect to phosphate group interactions (Figure 6.3.24, lower graph), specifically lipid interaction with Tyr27 and Arginines 46, 54 and 88 remain strong, while interactions to Asn50 and Arg74 suffer under tension.

Results for individual chains can be seen in Appendix D.3.

Contact persistence Figure 6.3.25 shows the capacity of each residue to remain in contact with lipid during gating. TM1 shows high propensity to interact with lipid, and the relatively high troughs suggest that lipids solvate most of this helix; not just the residues that face peripherally. This is also the case for the middle of TM2 (residues 72-82). The broad confidence interval around TM3b suggests dynamic lipid interaction; either lipid is not always present in each chain at the start of

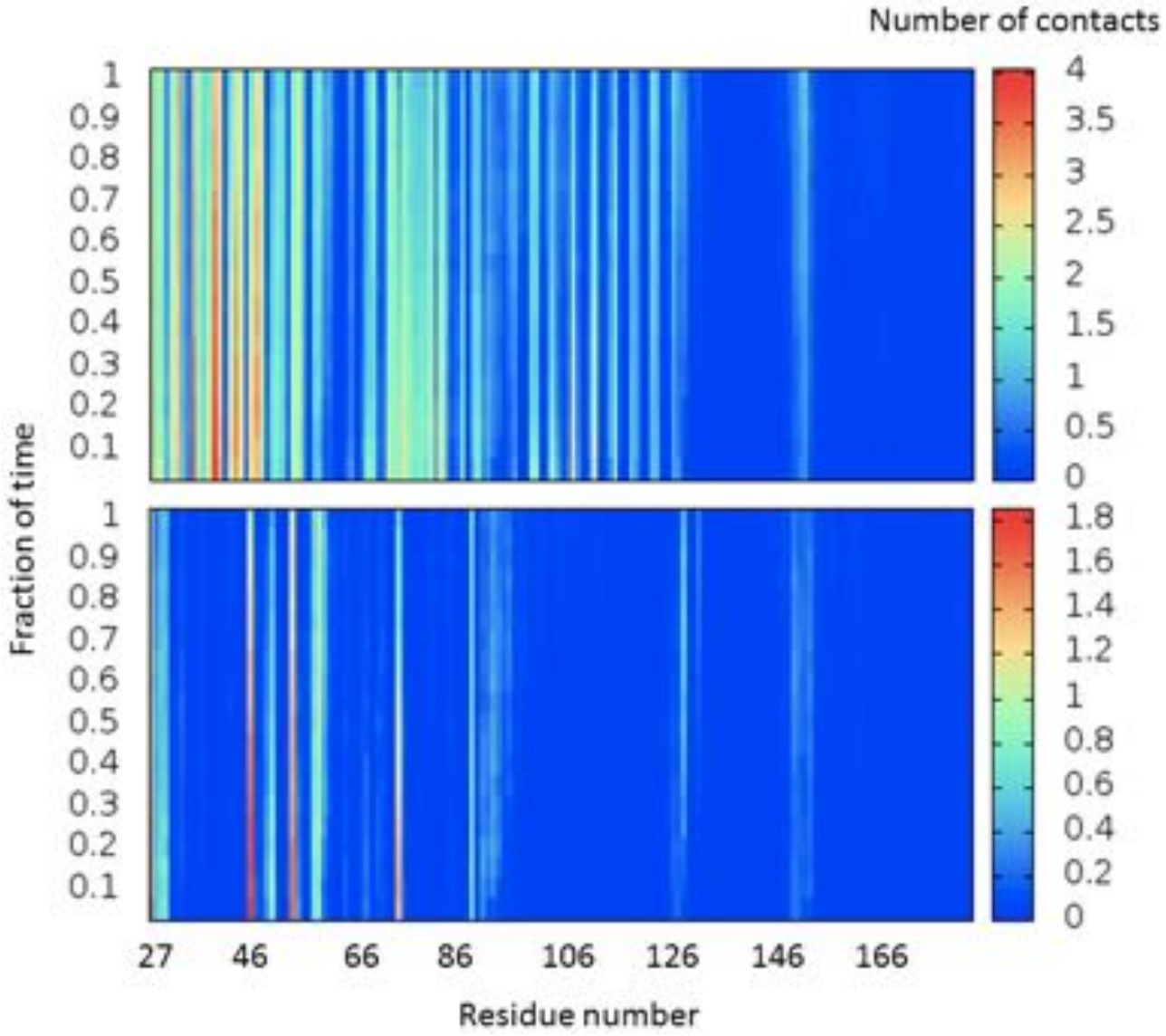

Figure 6.3.23.: Number of contacts to lipid as MscS gates, averaged across chains and trajectories Top: Number of lipids in contact per residue. Bottom: Number of lipid phosphate groups in contact per residue.

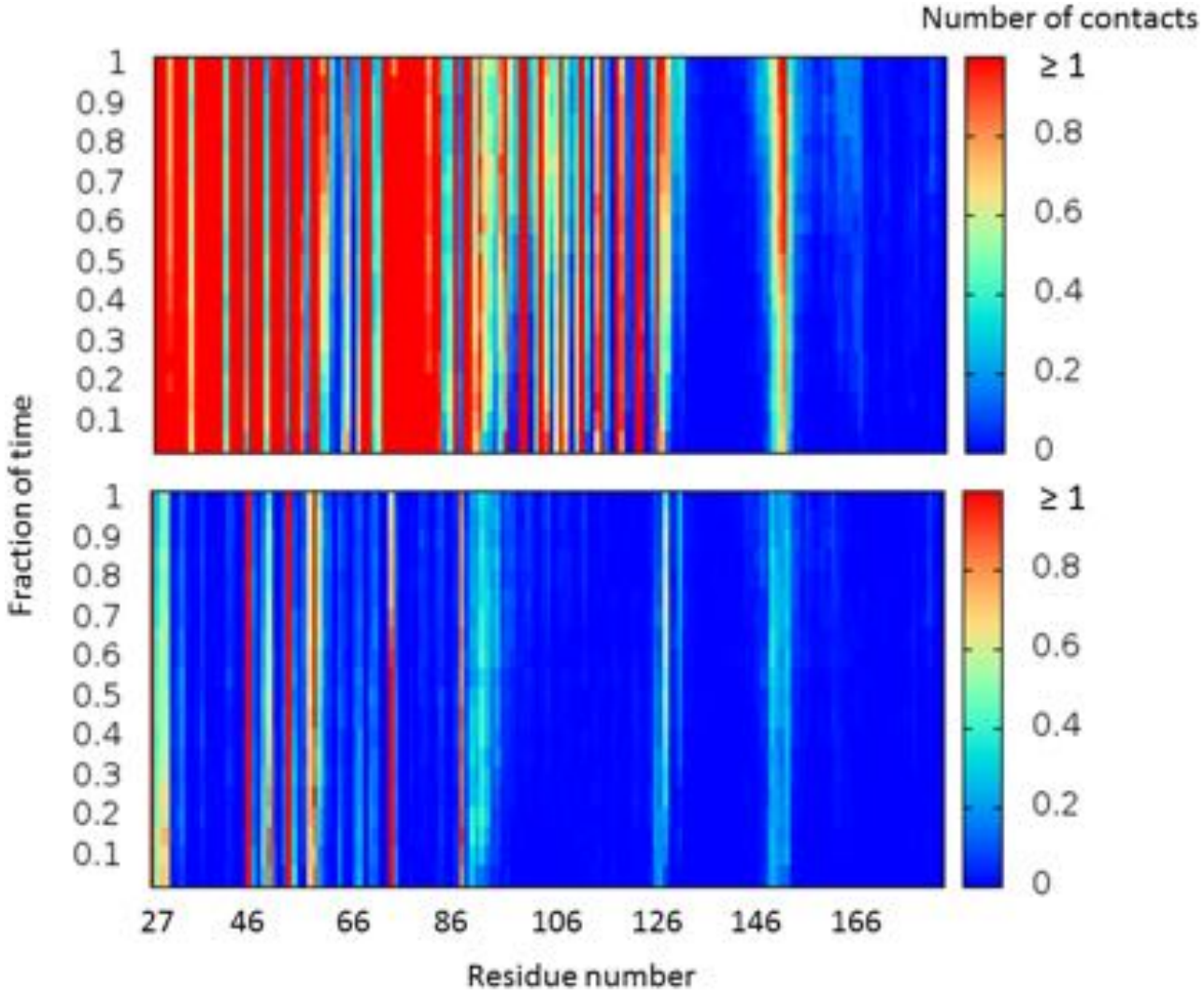

Figure 6.3.24.: Number of contacts to lipid as MscS gates, averaged across chains and trajectories (rescaled)
Top: Number of lipids in contact per residue. Bottom: Number of lipid phosphate groups in contact per residue. The heatmap was rescaled to between 0 and 1 for higher resolution of individual contacts.

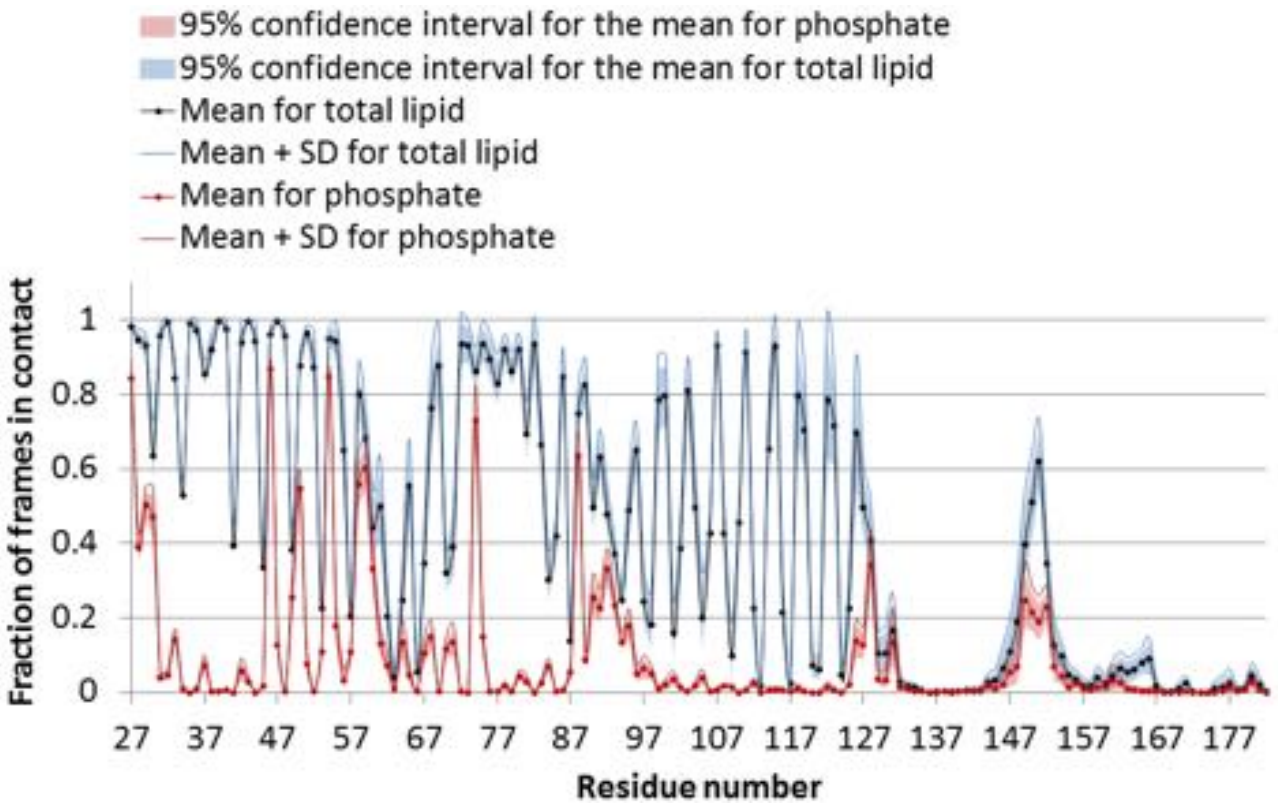

Figure 6.3.25.: Fraction of frames in contact with lipid throughout gating, per residue.

tension, or lipid is lost through tension. Lipid interaction with TM3b is comparable to that of the open crystal structure, but while the alkyl tails of these lipids reach the pore in simulations at zero tension, tails never reach the pore under tension (Figure 5.3.5 **C**; the graph plateau around residue 105 shows lipid accessing the pore).

Peak interactions to phosphate groups have already been discussed earlier for Figure 6.3.24.

Qualitative lipid dynamics around lower TM3 At the start of linear deformation simulations, on average two lipids are bound to the lower TM3 region, in between paddles of adjacent chains (as defined in Chapter 5). As tension increases, membrane curvature decreases and the cytosolic half of the TM domain, including lower TM3, approaches the membrane. This leads to loss of the least cytosolic lower TM3 lipid. As TM3 straightens, the remaining lipid, which was initially positioned parallel to the membrane along membrane-parallel TM3b, follows and adopts an approximately membrane-normal conformation between adjacent chains. This mimics the orientation of lipids in bulk membrane, with which it is in part contact. The lipid remains sheltered within the MscS interior, with its headgroup associated with the C-terminal TM3b and paddle. Only at gating tension, when each TM chain is often closely aligned, which provides little space for lipid to persist, does the remaining lipid have the potential to leave MscS, but does not always do so.

Lipid preference MscS retains more POPE compared to POPG throughout gating, but considering that the membrane lipid composition is 4:1 PE:PG, MscS is biased towards POPG (Figure 6.3.26). More PG is on average retained throughout gating than in equilibrium simulations of either crystal structure (c.f. Figure 5.3.17). Note, however, that the standard deviation relative to mean is high for contacts retained throughout simulations (Figure 6.3.26, lower graph).

Contacts made to MscS as a whole Analysis of the number of lipids in contact with MscS over time (Figure 6.3.27) reveals that more unique lipids come into contact with MscS as tension increases.

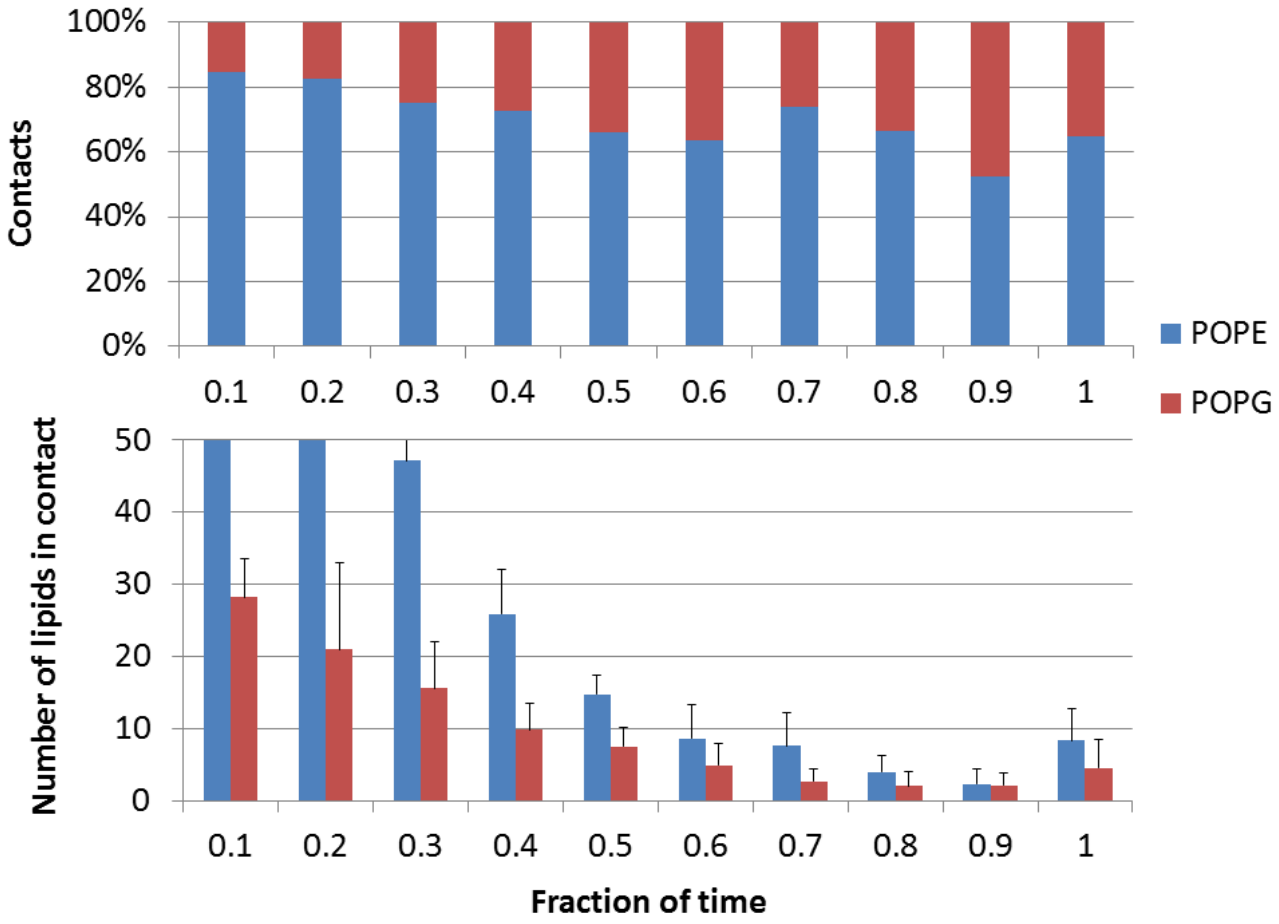

Figure 6.3.26.: The lipid type preference of MscS during gating.
Top: Lipid preference as a percentage of total lipid in contact. Bottom: Exact numbers of POPE and POPG in contact per fraction of time. Error bars indicate +1 SD.

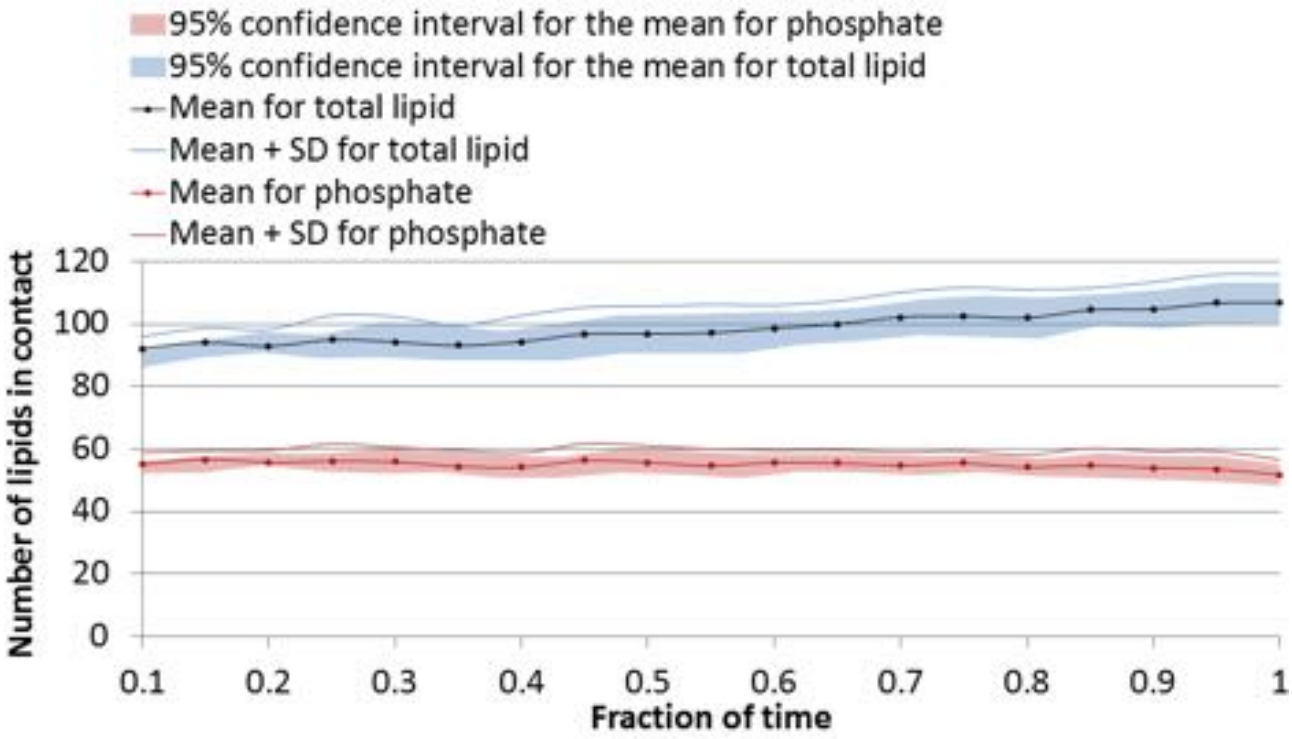

Figure 6.3.27.: Change in number of lipids in contact with MscS during gating
Results for all chains. NB time is not binned, merely made into fractions for compatibility across trajectories of different size.

Circa 20 more lipids (i.e. 3 lipids per chain) make contact with MscS under tension than at under no tension. This increase is seemingly counter-intuitive with respect to the decreasing lipid density in the membrane, but unique contacts does not necessarily mean strong interactions, and this increase could be the result of increased MscS lipid exposure as the TM domain leaves the cytosol and enters the membrane. Contacts to phosphate, on the other hand, remain approximately constant at 55 unique lipid phosphate groups - or 8 per chain - in contact with MscS.

Contacts are made to 0.45 and 0.1 fractions of the 156 residues under investigation for full lipid and lipid phosphate, respectively (Figure 6.3.28). The fraction of MscS in contact with lipid does not change significantly by gating. Therefore, circa 70 residues make contact with a total of 14 unique lipids per chain, while 16 residues contact 8 lipid phosphate group per chain.

6.3.3.4. Differences in lipid interaction between failed and successful gating

Independent t-tests were performed on the sample of trajectories where MscS fails to gate despite high tension, against the open state. Only frames around 600 bar*nm were used from simulations of failed gating.

The type of lipid that contacts MscS does not change significantly as a function of whether MscS fails or succeeds to gate ($p > 0.2$). Separation analysis reveals that MscS that fail to gate have less lipid in the vicinity compared to the open state (see Figure 6.3.29). This lipid loss is primarily observed near TM3 onwards, and the most prominent loss of contacts occurs at TM3b (see Figure 6.3.30). This is notable as this is the area where lipid persists throughout equilibrium simulations of MscS, irrespective of state (see Chapter 5). Results from failed gating suggest that lipid contacts to TM3b could be crucial for successful gating under tension. Interestingly, lipids were found in this area in the new open crystal structure, and their existence is supported by biophysical experiment, as detailed in Chapter 5.

I have also pointed out possibly functional lipid interaction to the cytosolic paddle region, and while

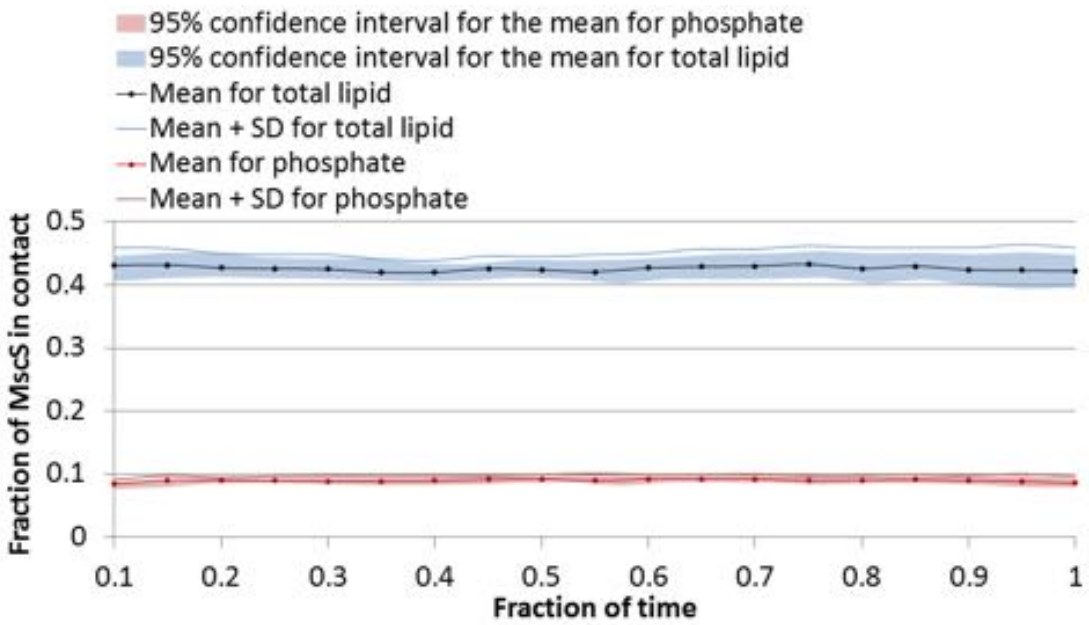

Figure 6.3.28.: Change to the fraction of MscS residues with lipid contacts as MscS gates.

contacts to lipid phosphates are not reduced in simulations of failed gating, there is less contact to whole lipid compared to the open state. Lipid is also further away from the cage domain in failed gating simulations, suggesting that the cage does not approach the membrane to the same extent as in successfully gating MscS. It has been experimentally shown that the cage is involved in gating, but the details remain unknown [Koprowski et al., 2011].

Mid-TM1 is the only region where lipid make more contact to failed than successfully gating MscS (see Figure 6.3.30. This could give rise to higher local pressure, keeping the paddle in a closed conformation. However, on the whole, the failure to gate is largely due to absence of hydrophobic lipid tails. Phosphate-protein contacts are intact.

6.3.3.5. The effect of surface tension application speed on lipid interaction

The open state sample was split into two samples consisting of the five 600 bar*nm instant constant tension simulations and the five linear deformation simulations where MscS gates. Independent t-tests were conducted to investigate differences in lipid interaction (instant constant tension simulations minus linear deformation simulation data), and lipid dynamics were also captured through time from pore diameter 0 to 15 Å.

No statistically significant difference was found with respect to the type of lipid bound ($p > 0.2$). However, separation dynamics between lipids and protein during gating, as depicted in Figure 6.3.31, show that lipid interaction is maintained to a higher degree by rapid than slow tension application. Separation and contact analysis of the open state resulting from fast and slow tension confirm this difference. The area affected is primarily the space between the paddle and the pore-lining helix (mid-TM2 and residues 100 to 113), where lipid is approximately 1-1.5 Å more closely associated by fast tension (Figure 6.3.32). More contacts - on average half a contact more - are also retained in the same area (Figure 6.3.33). Change to contacts between lipid phosphate groups and protein is generally insignificant, apart from slightly more contacts (circa +0.3) to Arginines 46, 74 and 128 by rapid tension application. Importantly, TM2 and 3 residues that show significant contact increases all face the MscS interior; contact differences are not due to bulk lipid effects.

To conclude, it appears that lipid interaction with protein is less dynamic when tension is applied rapidly. Lipid remains more tightly associated with the pore-lining helix and paddle, and contacts to paddle charges are also slightly more firm. Persistent lipids could act as a rigid force transduction

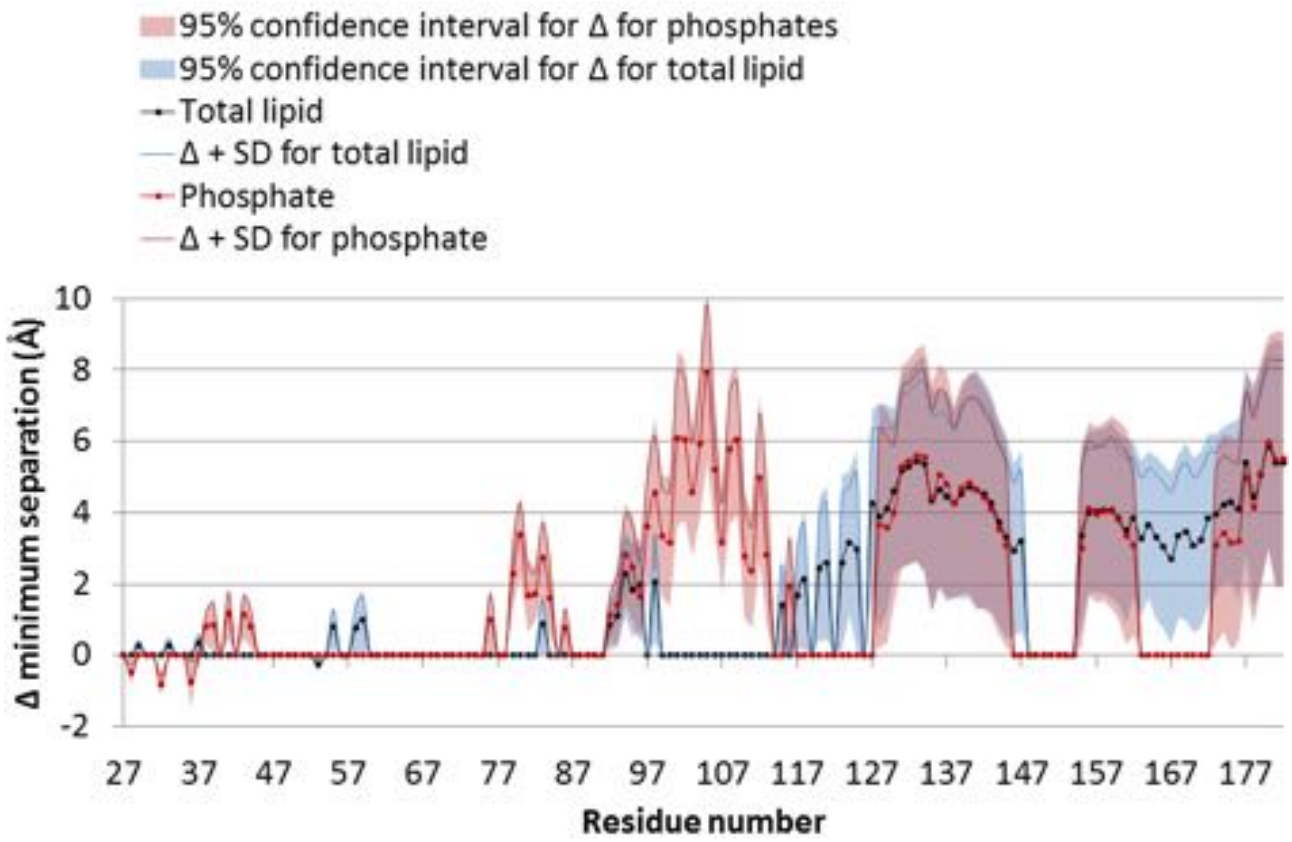

Figure 6.3.29.: Significant change to lipid-protein separation during failed gating.
Independent t-test data for simulations of protein without ENM: lipid interaction data at 600 bar*nm in simulations where MscS fails to gate minus data when the pore of MscS is at 15 Å diameter. Positive values indicate that lipid is further away in failed gating simulations than in successful gating simulations, while negative values indicate that lipid gets closer by failed gating.

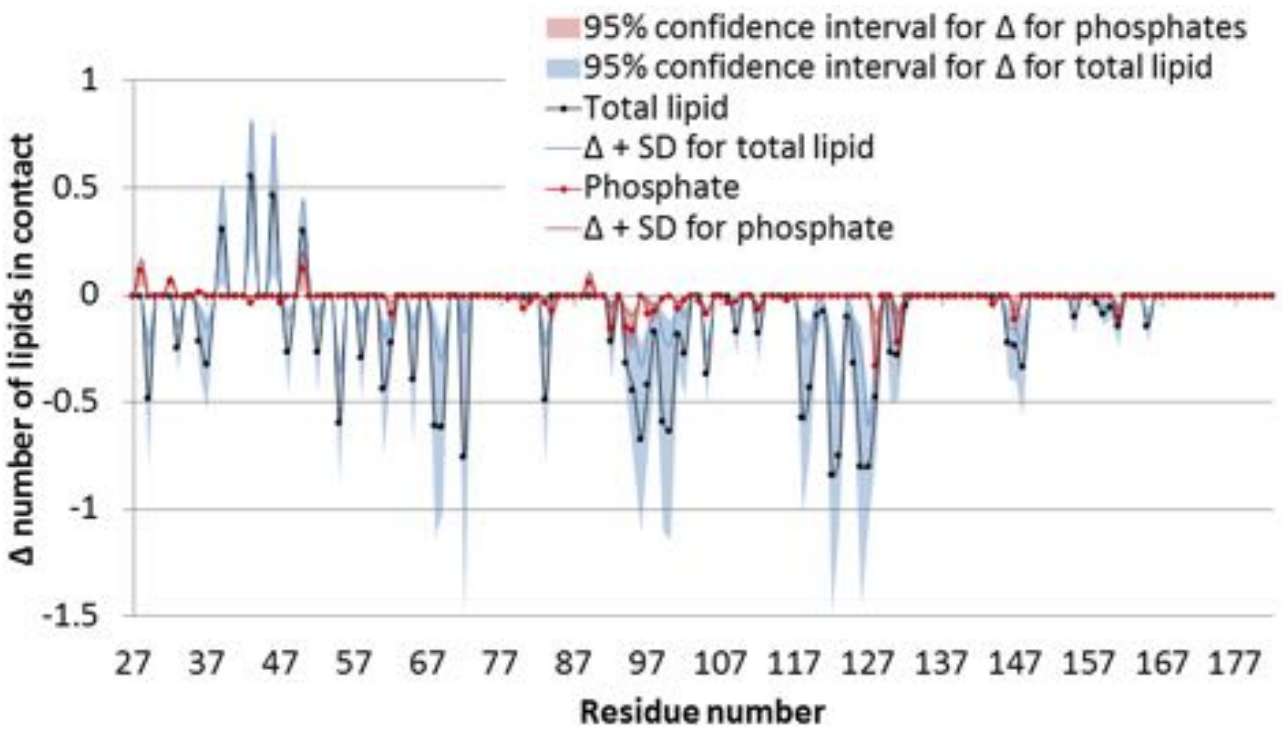

Figure 6.3.30.: Significant change to number of lipids in contact during failed gating.
Independent t-test data for simulations of protein without ENM: lipid interaction data at 600 bar*nm in simulations where MscS fails to gate minus data where the pore of MscS is at 15 Å diameter. Positive values indicate that lipid makes more contact in failed gating simulation than in gating simulations, while negative values indicate that lipid withdraws in failed simulations relative to gating simulations.

network, which would enable MscS to respond rapidly to force. These results suggest that the gating mechanism of MscS is likely to differ between fast and slow tension application.

6.3.3.6. The effect of double the midpoint tension

The effect of twice the midpoint tension was investigated by a dependent t-test of five MscS simulations before and after 110 bar*nm surface tension application. Zero surface tension results was subtracted from 110 bar*nm results. MscS does not gate at this tension (see Figure 6.3.4), so results are compared against results for gating MscS, Figures 6.3.20 and 6.3.21, to investigate how lipid interaction differs.

Contact analysis results reveal that there is insufficient tension to reduce lipid contacts along TM1 (Figure 6.3.34). Also, the increase in contacts observed at the cytosolic part of the paddle is still missing, suggesting that the MscS structure has yet to move its centre of mass closer to the membrane (as is seen under high tension). Lipid is lost from the area between the paddle and the pore-lining helix, as happens in gating MscS. This is important since I and my experimental collaborators suggest that lipid loss in this area gates MscS. These CG simulations at one sixth of the surface tension necessary for gating suggest that lipid loss is a relatively early event in the gating mechanism of MscS, which does not immediately lead to gating. At 110 bar*nm, the pore-lining helix has not yet twisted hydrophobic residues out of the pore (as evidenced by the lack of zig-zag pattern around the hydrophobic gate in Figure 6.3.35). So lipid loss occurs prior to TM3a helix twist. Assuming that PDB id 2VV5 is part of the gating mechanism (even though it was resolved without applied tension), this places 2VV5 further along the gating mechanism than CG simulations at 110 bar*nm.

The loss of lipid contact to TM3b is harder to explain. This loss is not as great in magnitude as in simulations where MscS fails to gate (see Figure 6.3.30) but assuming that lipid presence here is important, the partial loss could still impact the gating capacity of MscS.

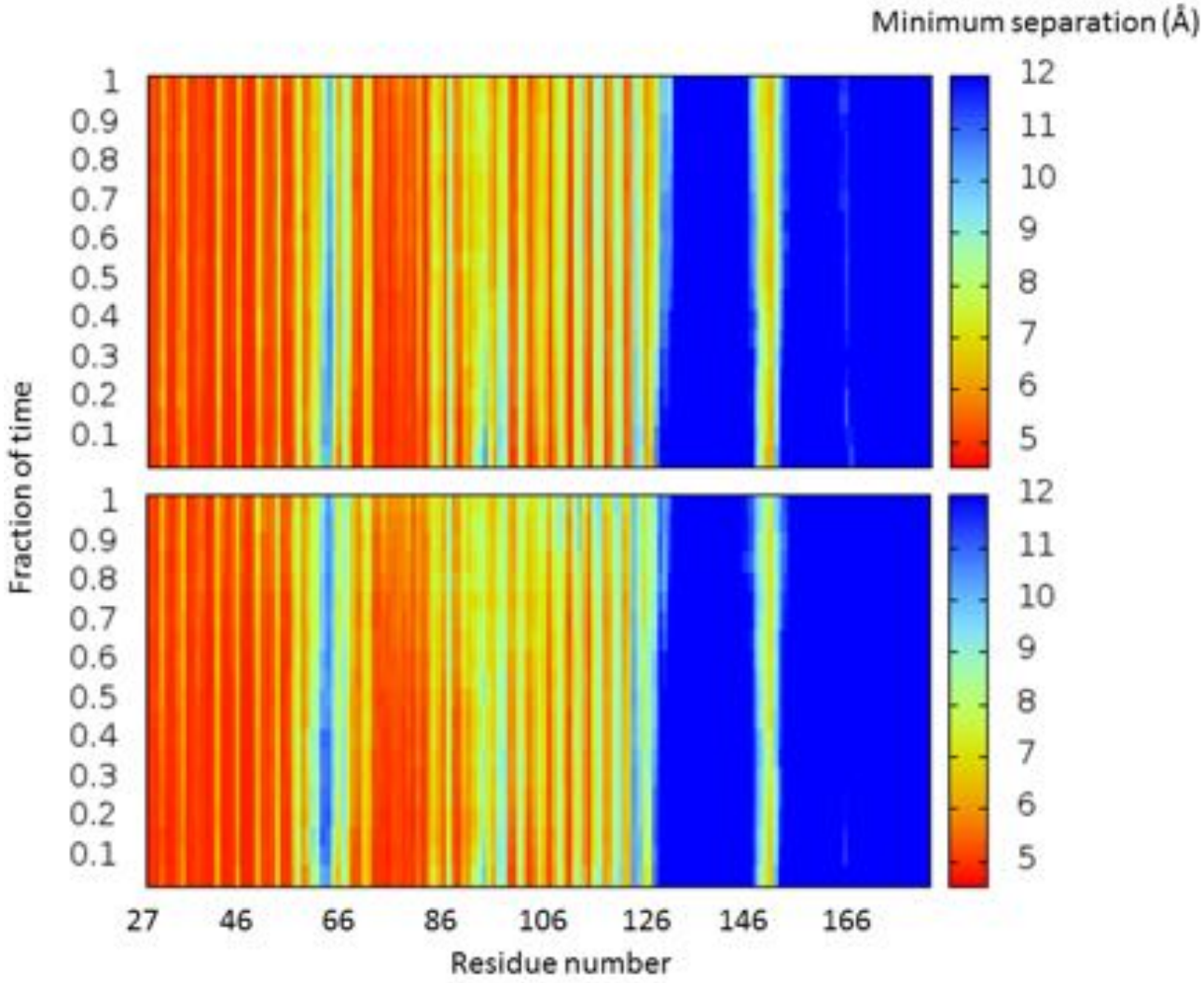

Figure 6.3.31.: Minimum separation to lipid as MscS gates: the effect of tension speed. Dynamic separation to lipid incurred by fast (top) and slow (bottom) tension application.

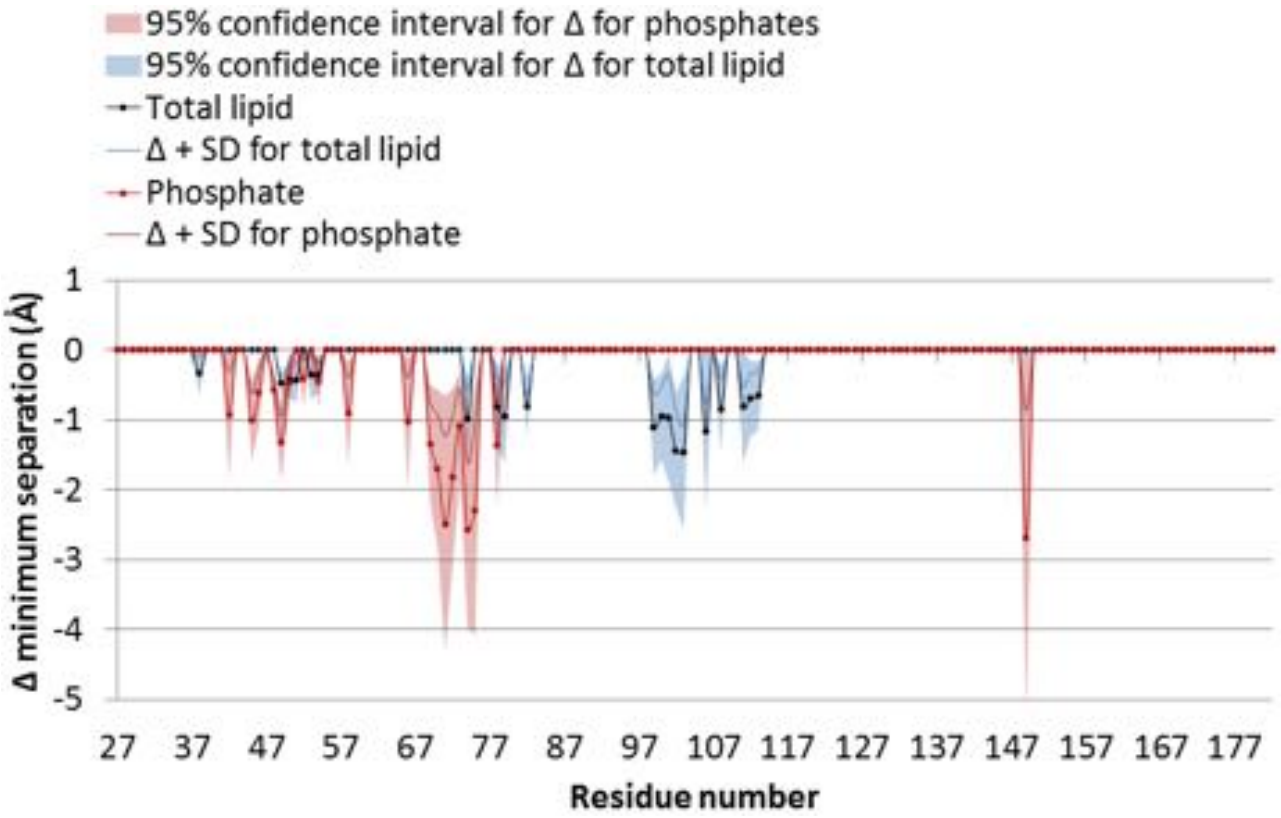

Figure 6.3.32.: Significant change to lipid-protein separation in the open structure by fast versus slow tension application

Positive values indicate that lipid withdraws by fast tension compared to slow tension, while negative values indicate that lipid stays more proximal in fast tension simulations relative to slow tension simulations.

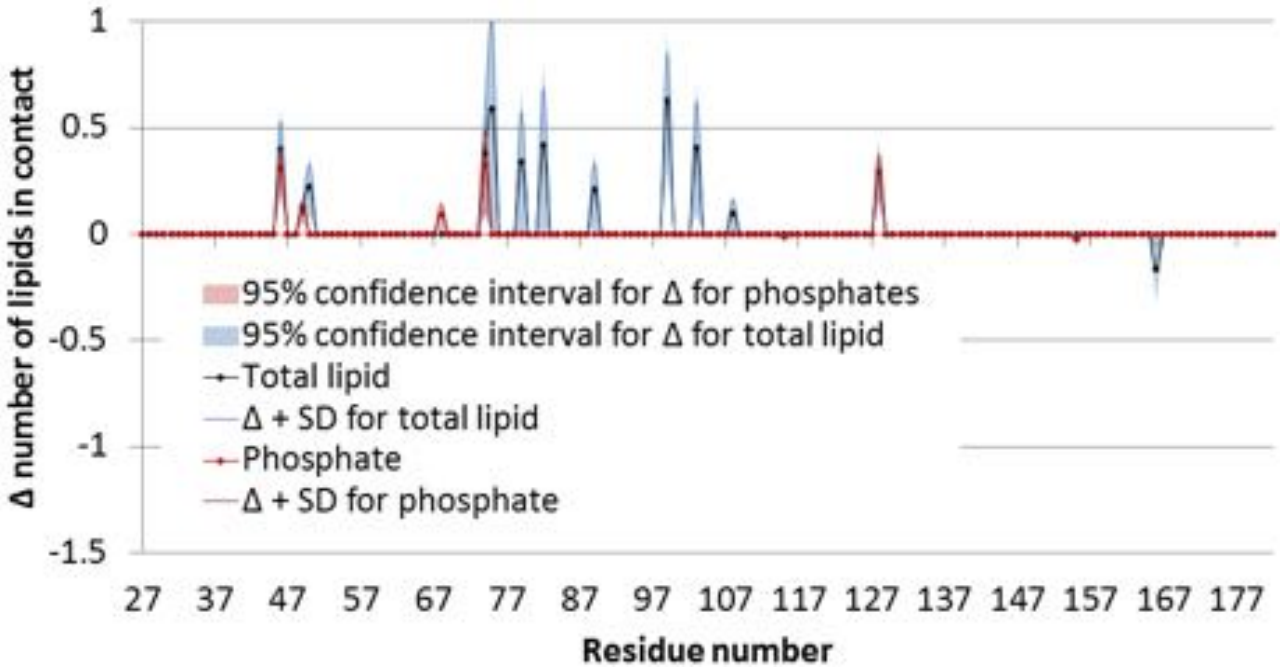

Figure 6.3.33.: Significant change to the number of lipids in contact with the open structure by fast versus slow tension application
Positive values indicate that lipid makes more contact by fast tension application than by slow tension, while negative values indicate that lipid withdraws in fast tension simulations relative to slow tension simulations.

Lastly, separation results suggest that 110 bar*nm tension is insufficient to pull the cage towards to the membrane as in gating MscS (see Figure 6.3.34).

Comparison of lipid contact dynamics to the TM domain across conditions

6.3.3.7. Different tension regimes compared to equilibrium conditions

Figure 6.3.36 shows lipid contact results from all examined tension conditions apart from MscS under complete ENM restraints. Dependent t-tests were made per tension condition with respect to the preceding simulations at zero surface tension. In all cases, zero surface tension data was subtracted from non-zero tension data, and only significantly different data are graphed ($p < 0.05$).

General lipid contact loss is observed along the paddle. Contact gain is seen in linker-regions between helices, more extensively at high than low tension, and notably absent in MscS that does not gate. Lipid loss along TM2 at low tension occurs more cytosolically than at high tension, suggesting that the remaining lipid could be relocating cytosolically during gating. It is important to point out that some TM2 residues face bulk, rather than internalised lipid. Residues that show high lipid loss at low tension all face the paddle-TM3 cavity, while residues that experience contact loss at high tension either face clefts between chains or bulk lipid. Notably, Arg74 faces a cleft between chains, where the headgroups of internalised lipids can make contact, and loss of such lipid by increasing tension is likely the reason for contact loss. However, overall, the exposure of TM2 to bulk lipid makes it a complex indicator of internalised lipid behaviour. TM3, which is at the protein core, is a better indicator of the dynamics of such lipids.

Lipid interaction by fast and slow tension application are closely synchronised, with slow tension resulting in more amplified interaction dynamics (i.e. while a residue might experience contact loss by both slow and fast tension, slow tension increases this contact loss). A notable exception is in the lower TM3 region, where fast tension results in higher contact loss (TM3a) and gain (TM3b). In simulations where gating fails, lipid loss from between the paddle and TM3 is insignificant compared

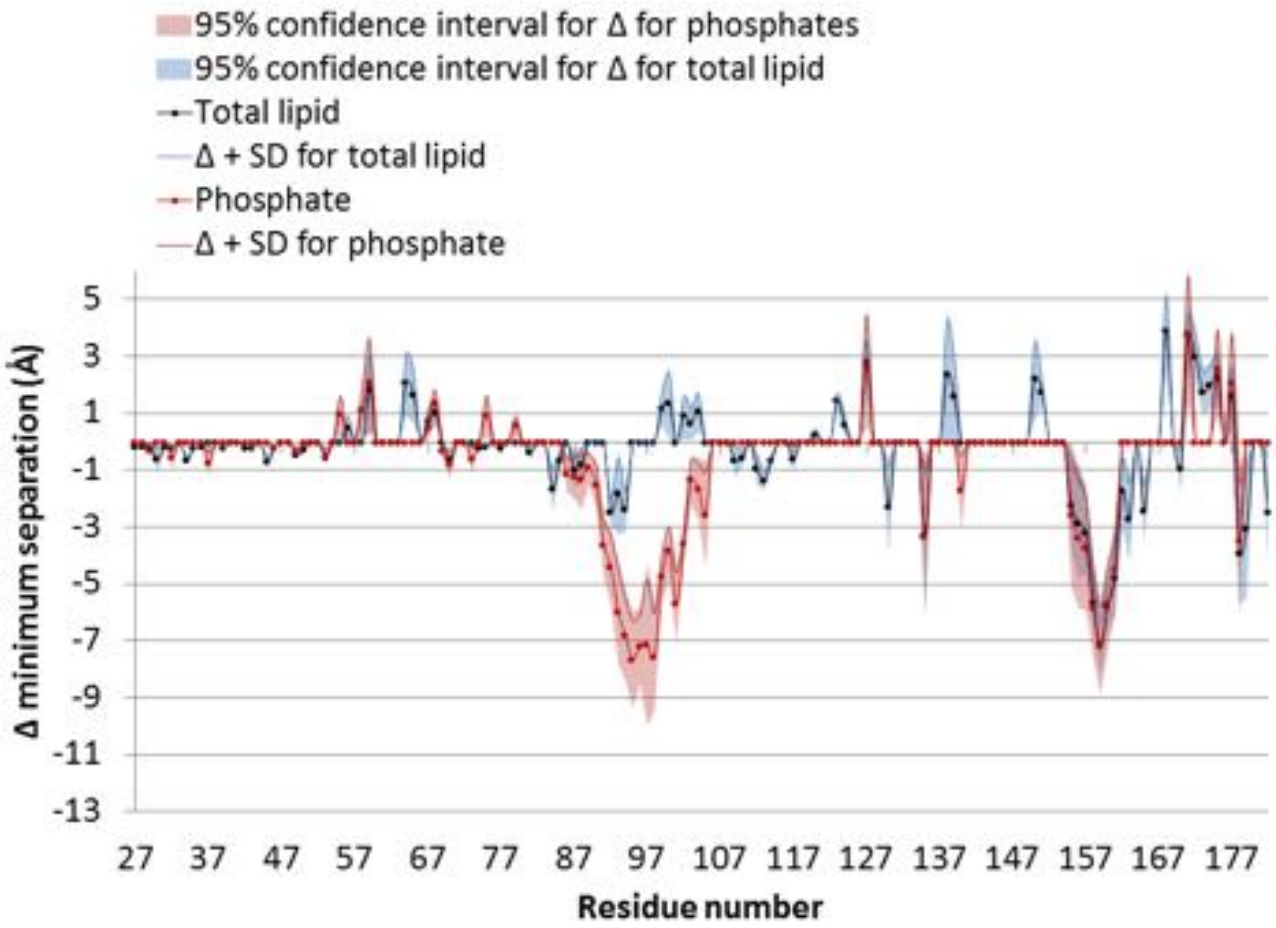

Figure 6.3.34.: Significant change to lipid-protein separation by twice the midpoint tension
Positive values indicate that lipid is further away in the simulation under tension than at no tension, while negative values indicate that lipid gets closer by tension. Axes are scaled according to Figure 6.3.20 for easy comparison.

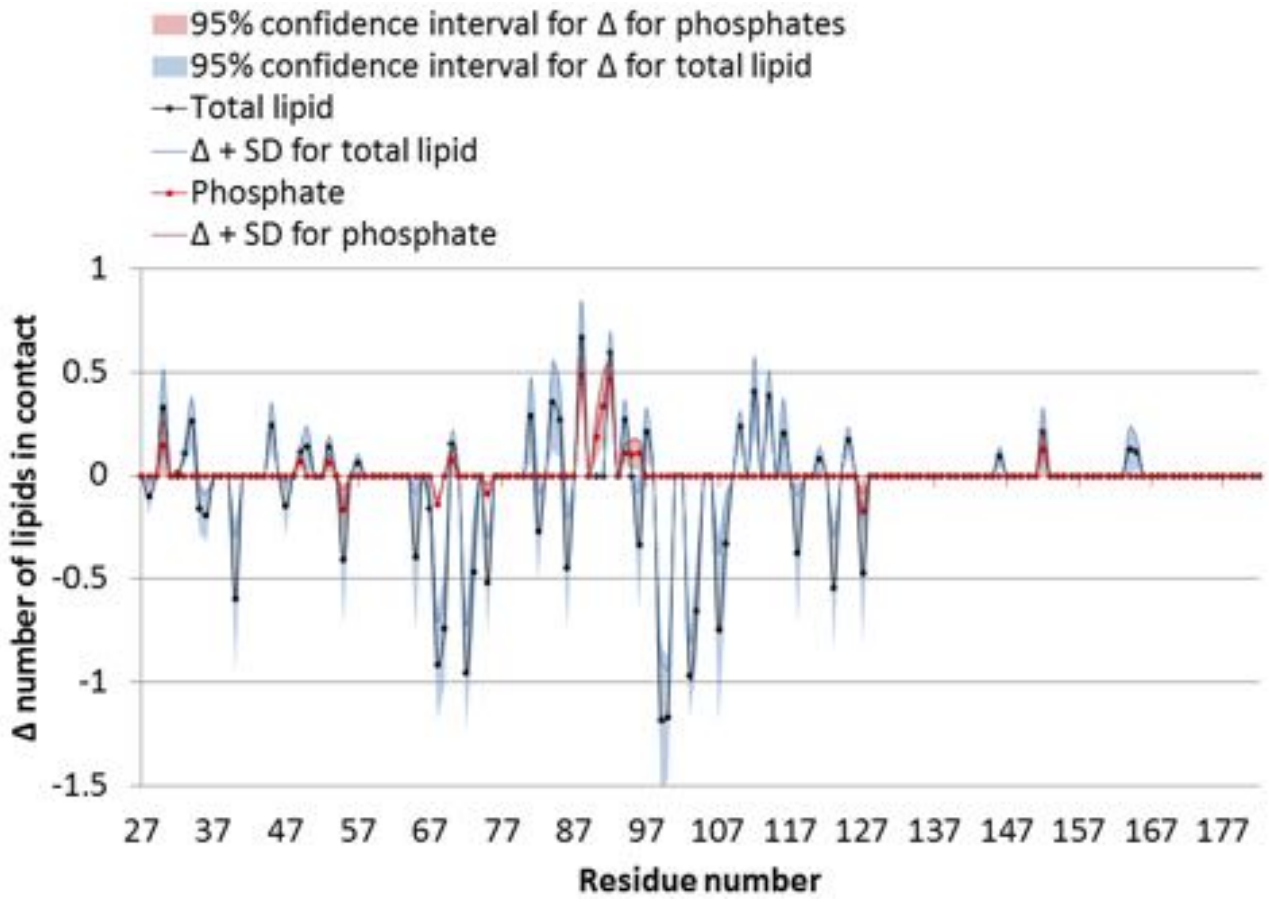

Figure 6.3.35.: Significant change to number of lipids in contact by twice the midpoint tension. Positive values indicate that lipid makes more contact in the simulation under tension than at no tension, while negative values indicate that lipid withdraws by tension. Axes are scaled according to Figure 6.3.21 for easy comparison.

to equilibrium conditions and, notably, the contact gain that is observed in gating simulations to the TM3b area is lacking.

Contacts to lipid phosphate groups are less dynamic than contacts to full lipid. The overall consistency in phosphate contacts suggest strong contacts to MscS. By disregarding areas where failed and successful gating show similar dynamics, I can focus on contact dynamics that are important for gating: loss of contact to Arginines 46 and 74 and gain at Lys60, Gln92 and Arg128. At low tension, only the increased contact to Gln92 is observed; all other dynamics is yet to happen.

6.3.3.8. Exact number of contacts across conditions

Figure 6.3.37 shows the raw contact data used to produce Figure 6.3.36. It is not pre-processed to show only significantly different data, but is useful to see contact dynamics with the additional perspective of contact magnitude; e.g. whether a contact loss is the result of three contacts being reduced to two, or if the only existing contact is lost.

Notably, the contact loss by tension at Tyr27 and Arg46 does not result in contact depletion; both residues retain one lipid phosphate contact, irrespective of tension regime. In gating simulations only, Arg128 makes new contact with lipid at the end of TM3b, having had no contact perviously, and the situation is similar in the TM1-2 linker region. One interesting feature is the loss of lipid between the paddle and TM3 in failed gating simulations, supporting that lipid loss from this area is an early event that is not directly linked to gating. This lipid loss is insignificant with respect to the zero tension state, suggesting high standard deviation (c.f. Figures 6.3.36 and 6.3.37). The small sample size (four observations) could be partly responsible, so I will refrain from concluding that failure to gate is due to failed lipid loss. Instead, failure to make new contacts in linker regions and to Arg128 seems to be the key to successful gating. Statistical insignificance aside, it is interesting to note that the failed gating state is on average the most de-lipidated state throughout the TM domain.

The peaks observed around 72 to 82 of TM2 in the top graph of Figure 6.3.37 are mostly located at residues that face bulk lipid, so do not reflect internalised lipid dynamics. However, the 1-lipid sized plateau observed here reveals that at one lipid remains in the paddle-TM3 area throughout gating.

6.3.4. Conformational change to MscS during gating in the light of lipid interaction

Throughout this discussion, the decimal fractions refer to time leading up until the fully gated state, which is defined as 1.0. 0 is MscS at zero surface tension, at the start of tension simulation.

Key residues that interact with phosphate are visualized

Key contact dynamics between lipid phosphate and MscS residues, that were previously investigated in Section 6.3.3.3, are visualised alongside the conformational change of MscS as it gates in Figures 6.3.38 to 6.3.42. Interaction is classified as constantly strong or reduced, based on results in Figures 6.3.24 and 6.3.25. Contact to at least a single lipid is emphasised; a residue that looses contacts overall but remains in contact with one lipid phosphate throughout gating is therefore classified as a strong contact. Residues are visualised accordingly in Figures 6.3.38 to 6.3.42.

6.3.4.1. Conformational change irrespective of tension application speed

The conical shape of MscS (as defined by PDB id 2OAU) is disrupted immediately at tension onset (by 0.1 of the fully gated state). The in-plane expansion of the cone apex, the pericellular half of the TM domain, which can be tracked through the location of TM1's most N-terminal crystal-resolved residue, Tyr27 (in red in Figures 6.3.38 to 6.3.40) is correlated with the pore's expansion (more details in the sections on slow and fast tension application, below). Another event that occurs immediately at the onset of tension is the disruption of the 'units' made up of one chain's paddle and the adjacent chain's TM3. MscS adopts a state where helices cover the in-plane area in a mesh-like fashion,

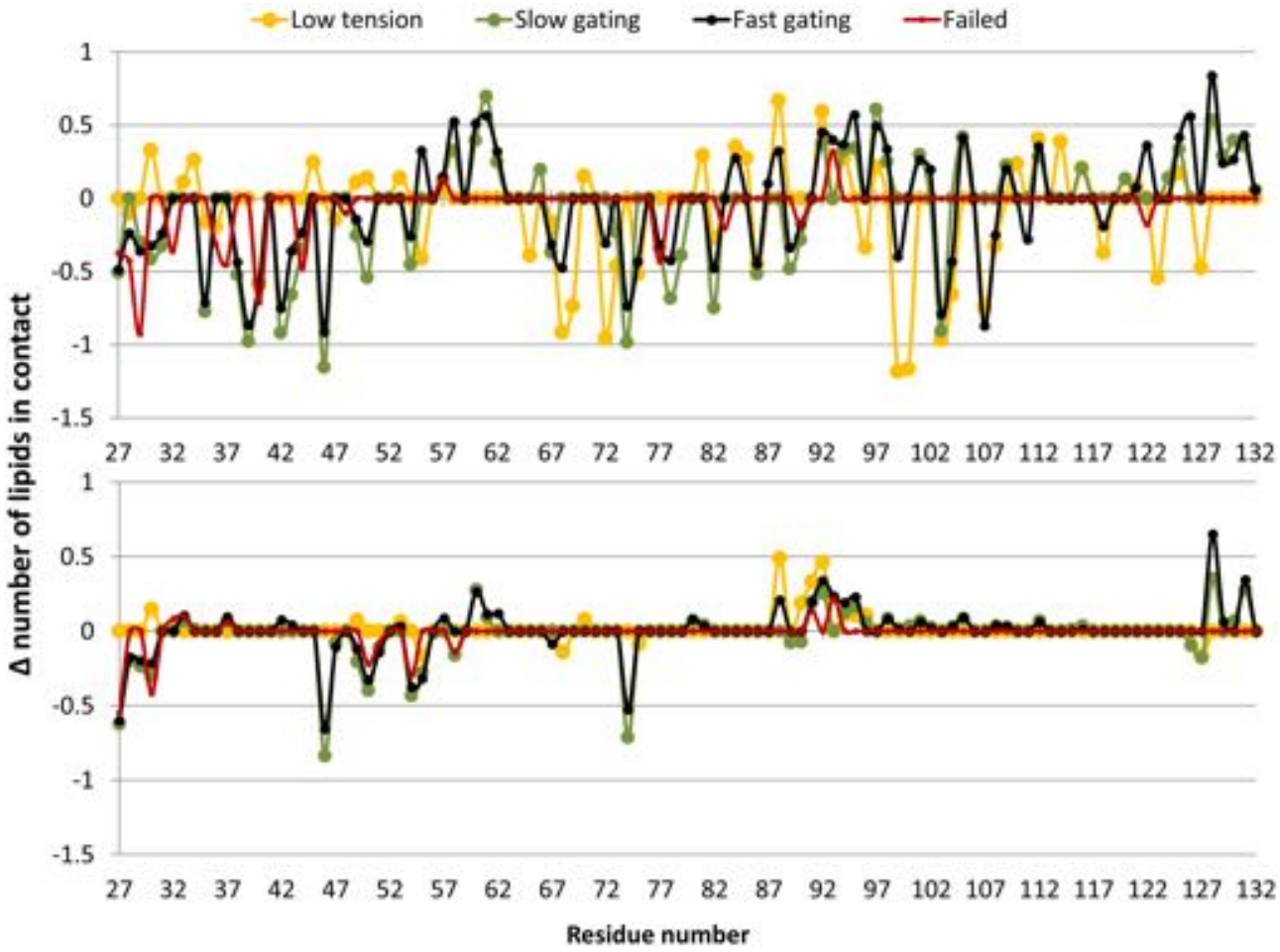

Figure 6.3.36.: Contact comparison across different tension regimes
Top. Contact dynamics between MscS and whole lipid
Bottom. Contact dynamics between MscS and lipid phosphate.
A positive value denotes that tension caused more contacts to form than equilibrium conditions, while negative values mean that tension reduced the number of contacts at this residue. Only significant differences are graphed ($p < 0.05$).
Low tension = the 110 bar*nm state
Slow gating = the 15 Å state of linear deformation simulations
Fast gating = the 15 Å state of 600 bar*nm instant constant tension simulations
Failed = States at 600 bar*nm instant surface tension that do not gate

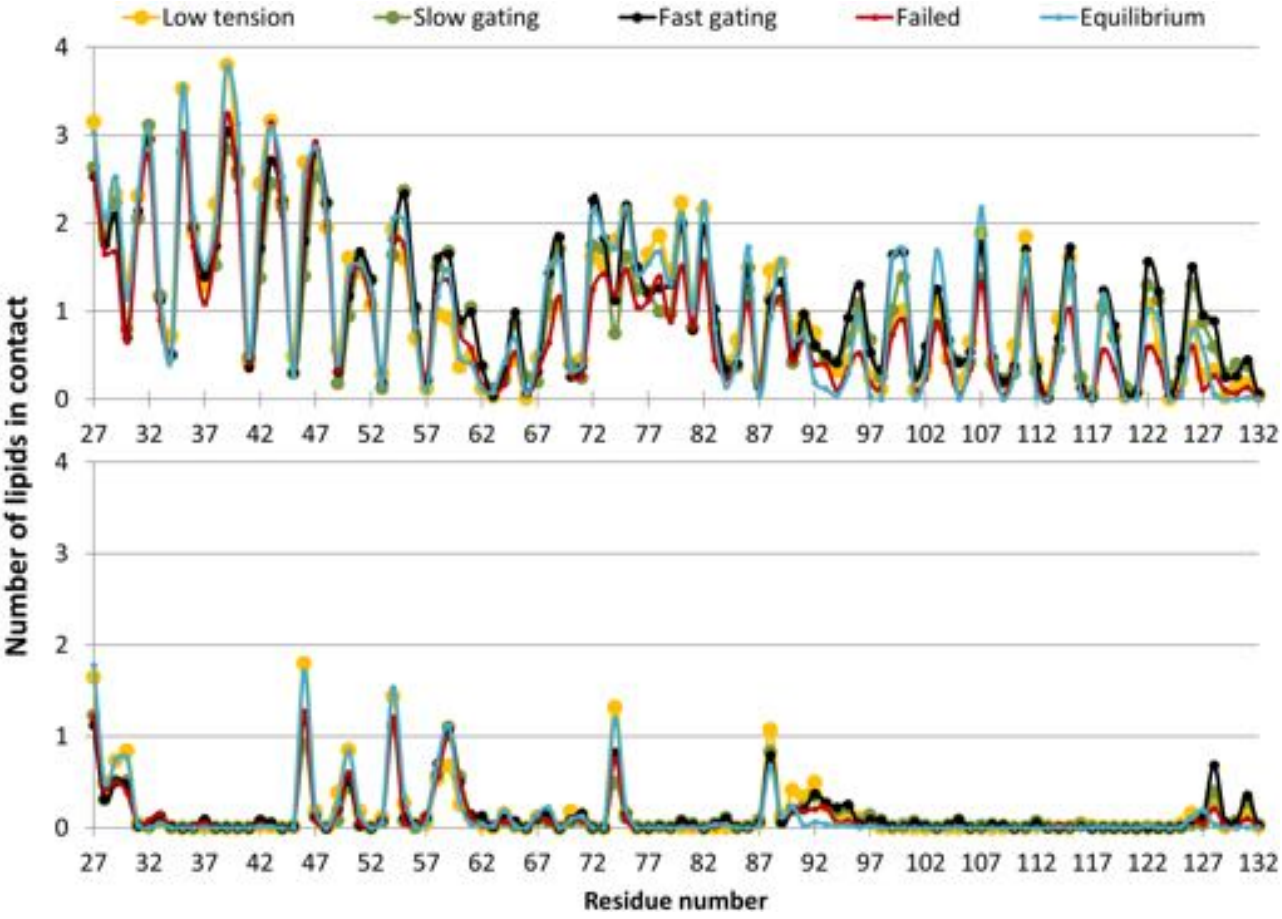

Figure 6.3.37.: Contact comparison across different tension regimes: raw data
Top. Contacts between MscS and whole lipid
Bottom. Contacts between MscS and lipid phosphate.
Low tension = the 110 bar*nm state
Slow gating = the 15 Å state of linear deformation simulations
Fast gating = the 15 Å state of 600 bar*nm instant constant tension simulations
Failed = States at 600 bar*nm instant constant tension that do not gate
Equilibrium = 0 surface tension

rather than retaining the void between units (c.f. Figure 6.3.39, time 0 and 0.1). Charged residues on the cytosolic end of the paddle get into contact with the lower leaflet headgroups as membrane curvature lessens and the TM domain approaches the membrane. Only when Tyr27 is brought in line with the protein periphery and membrane curvature is negligible do MscS helices start to tilt increasingly with respect to the membrane, and continue to do so until gating. The increase in tilt allows (or is mediated by) charges along the cytosolic end of the paddle to make increasing numbers of contacts to lipid headgroups. TM1 even adopts a slightly curved conformation throughout the latter half of gating, which optimises such contacts.

During the last quarter of the time until the fully open state, the paddles, that are lining the circumference of MscS, have all merged into the nw planar bilayer. Paddles group into units of two to four paddles and adopt an asymmetric conformation whereby the groups spread such that they lay approximately diametrically opposite other group(s). Meanwhile, the pore-lining TM3 helices retain a near-symmetric iris shape about the protein centre throughout gating. Its pericellular half twists hydrophobic gating residues out of the gate so that they face the pericellular space, which reduces their vapour lock activity.

Paddle asymmetry characterises the final pore dilation that defines the open state. In order to allow this conformational change, some chains increase the distance between paddle and TM3 (see for example the highlighted helix in Figure 6.3.42).

6.3.4.2. Slow tension application

In linear deformation simulations, initial low tension leads to pore collapse. The conical shape of the closed state TM domain is lost at a slower rate than by fast tension application. It reaches a steady-state where Tyr27 is located half-way between the pore and the MscS periphery, wherefrom it does not move until MscS undergoes further in-plane expansion (0.6 - 0.8), upon which both the previously collapsed hydrophobic gate and Tyr27 move outwards in a coordinated fashion, widening the TM passage to its former, 2OAU crystal structure diameter.

MscS lacking ENM and MscS with ENM limited to the cage behave very similarly in the TM region (c.f. MscS gating with cage-ENM, Figures 6.3.38 to 6.3.40 versus without ENM, Figures D.4.1 and D.4.2 in Appendix), suggesting that the cage has little or no influence on TM conformational change during gating.

6.3.4.3. Fast tension application

In simulations at 600 bar*nm instant constant tension, loss of the initial conical shape of the closed state TM domain happens almost immediately (Figure 6.3.41). No mid-way steady-state for Tyr27 exists; Tyr27 reaches the protein periphery within 0.2 of gating, and the pore widens in a correlated fashion and never undergoes collapse (Figure 6.3.42). All helices undergo rapid tilting with respect to the membrane. This is evident as early as 0.2 and helix tilt continues to increase at a slower rate over time until gating.

6.3.5. Lower TM3 lipid dynamics during conformational change

As MscS gates, lipids associated with the lower TM3 region are persistent. Figure 6.3.43 shows three chains' lower TM3 lipid, coloured according to chain affiliation at zero surface tension. Three lipids are associated with two of the chains, while the middle chain hosts a single lipid. Half-way to gating tension, chains with three lipids loose one lipid each, and only the most strongly associated lipid remains proximal to each chain at three quarters of full gating tension. At gating tension, one chain's strongly associated lipid leaves, while the other two remain. Above these tightly bound lipid, bulk lipid make intermittent contact between helices throughout gating, space permitting. No such individual lipids remain in the cleft for longer than 100ns.

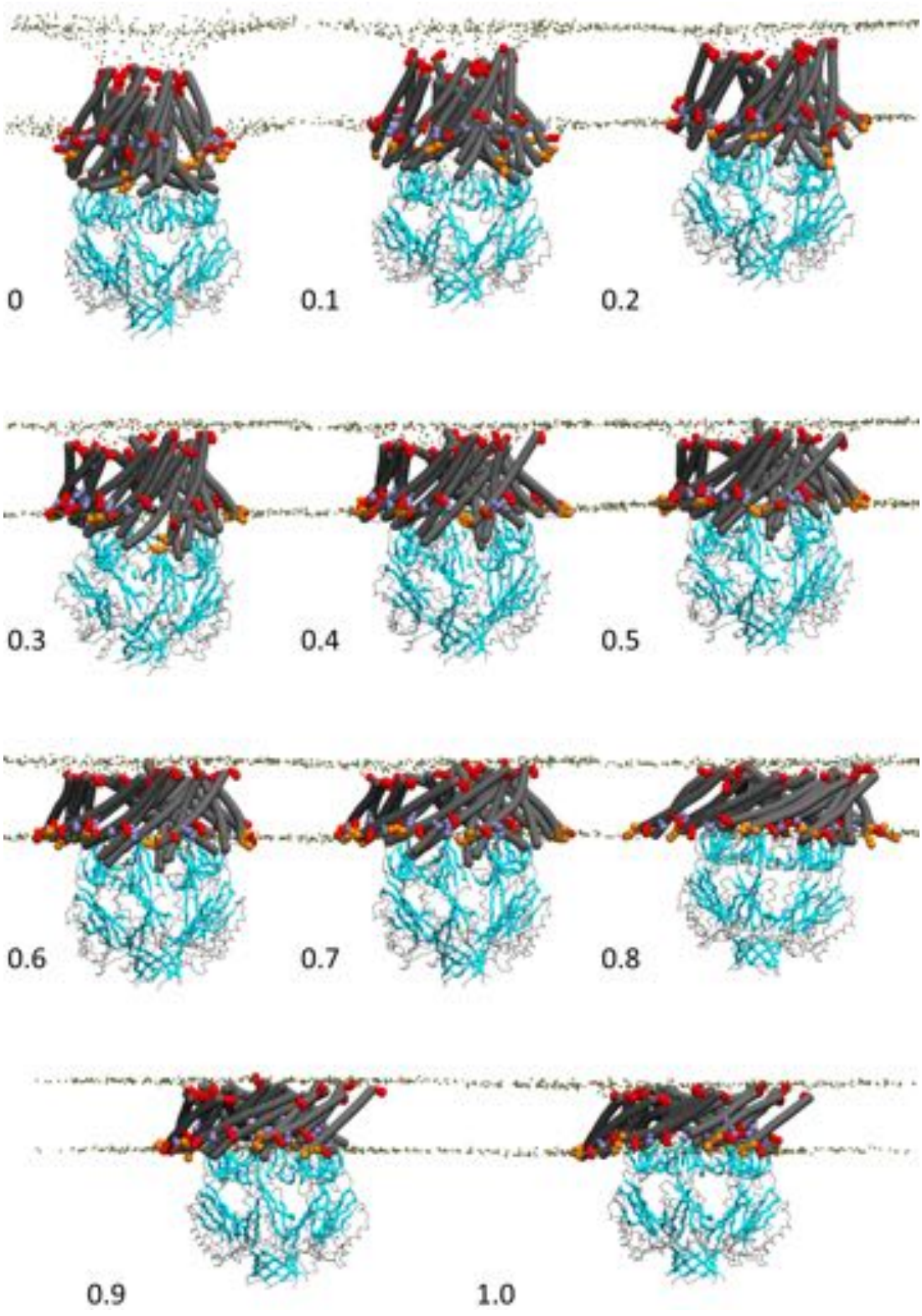

Figure 6.3.38.: Slow gating mechanism: side view
The numbers refers to fractional time of the total simulation run. The MscS TM domain is depicted as gray bendices and cage in the 'Join' backbone representation of Bendix, where β-sheets are represented by turquoise arrows. Tan spheres show lipid phosphate groups of the upper and lower bilayer leaflet. Key residues that make contact to phosphate groups during gating are shown in Van der Waal representations, coloured according to interaction dynamics: red or orange indicates persistent contact throughout gating (Tyr27, Ser58, Arg 46, 54, 59 and 88) and blue indicates decreasing contact by gating (Asn50, Arg74). Categories are based on results in Figures 6.3.24 and 6.3.25. See text for more detail.

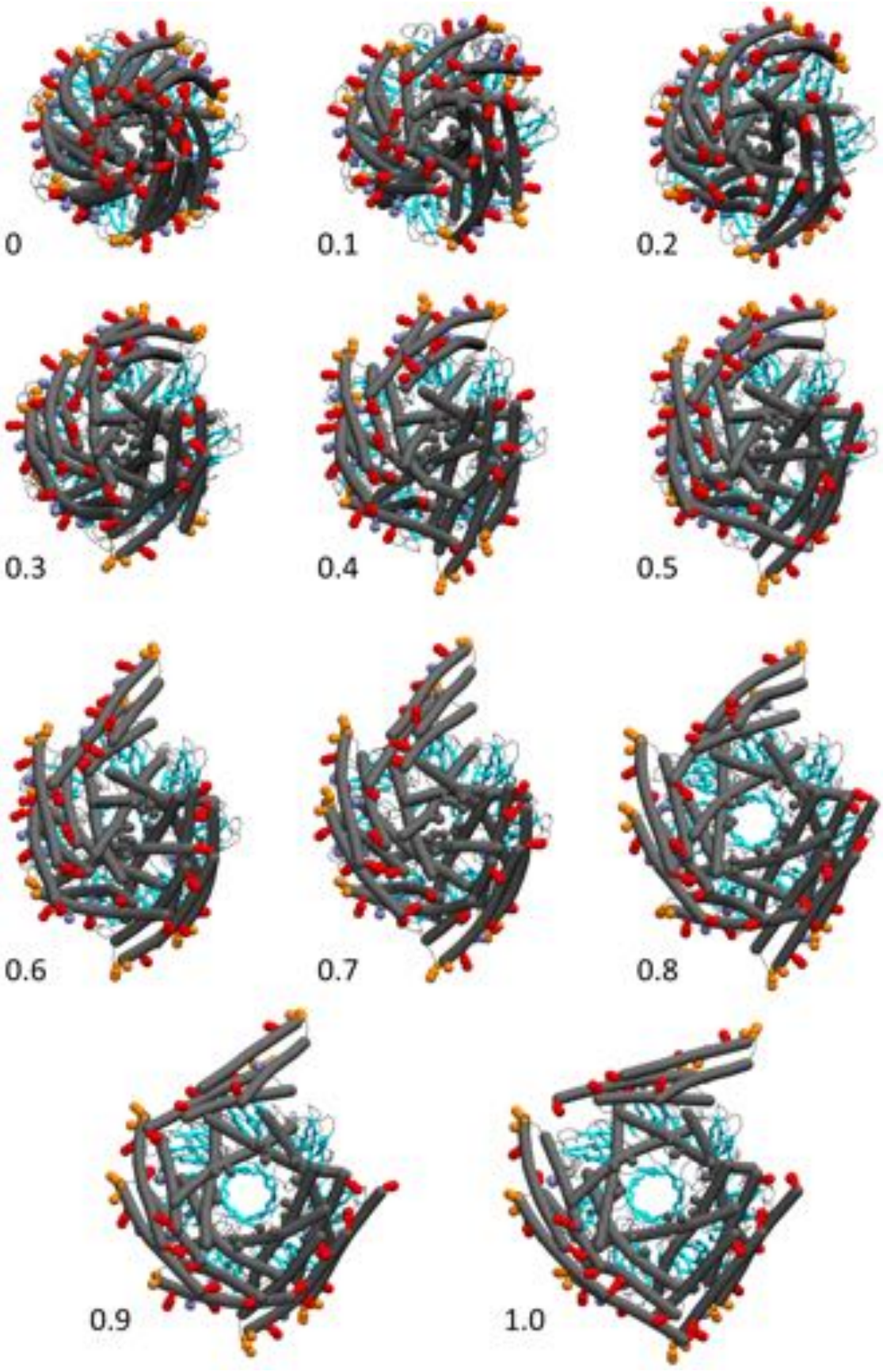

Figure 6.3.39.: Slow gating mechanism: top view
The numbers refers to fractional time of the total simulation run. Same graphical representations as in Figure 6.3.38, only no lipid phosphates are shown, and the leucines that form the hydrophobic gate are shown as gray spheres along the pore-lining helices. Orthographic perspective.

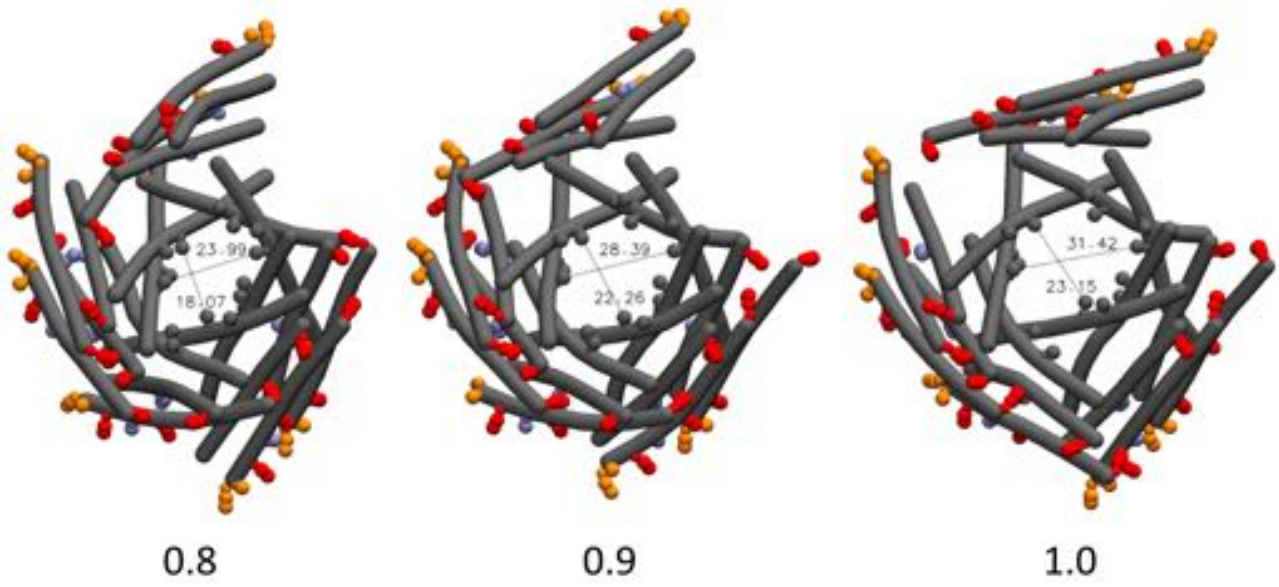

Figure 6.3.40.: Slow gating mechanism: detail
Detail from Figure 6.3.39 including pore dimensions in Å between geometric centers of the hydrophobic gate (reduce by 4.25 Å to retrieve the distance between van der Waal surfaces, as is used in the main text). Orthographic perspective.

At the apex, lipid can remain closely associated throughout gating (not shown in figure). Interactions are made between headgroups and the TM2-3 linker region, as well as the protein N-terminus. However contacts are not maintained as long as lower TM3 region lipid contacts are.

6.3.6. Compilation of results to deduce the MscS gating mechanism

6.3.6.1. Starting structure

At zero surface tension, lipid penetrates deep between and around the paddle, and there is high membrane curvature about MscS. Even though the MscS TM domain is largely located in the cytosol, lipid makes contacts throughout, notably to the lower TM3 region (for more detail about the equilibrium state, refer to Chapter 5).

6.3.6.2. Low tension

As the membrane is put under stress, lipid density decreases, causing lipid withdrawal from MscS. Interpreting results for 110 bar*nm instant constant surface tension as early gating events, lipid is lost from the area between the paddle and TM3a, more precisely around residue 70 and 100 on the paddle and TM3 helix, respectively, which is in the vicinity of the gate. My ENM experiments show that conformational change is absolutely necessary for this lipid loss; the restrained closed crystal structure retains this lipid, even when subjected to lytic tension. The pore collapses, forming an even narrower TM passage than observed in the closed crystal structure. Further, the void observed between adjacent chains' TM helices, which gives the TM domain its distinct 'iris'-shape when viewed from the pericellular space, is lost. Instead, helices rearrange to fill the available space and same-chain helices will eventually align, as seen in the open crystal structure.

6.3.6.3. Transition to open

Membrane curvature and protein shape As tension increases further, membrane curvature decreases, effectively pulling MscS up into the membrane. Notably, the top of the cage's beta-barrel

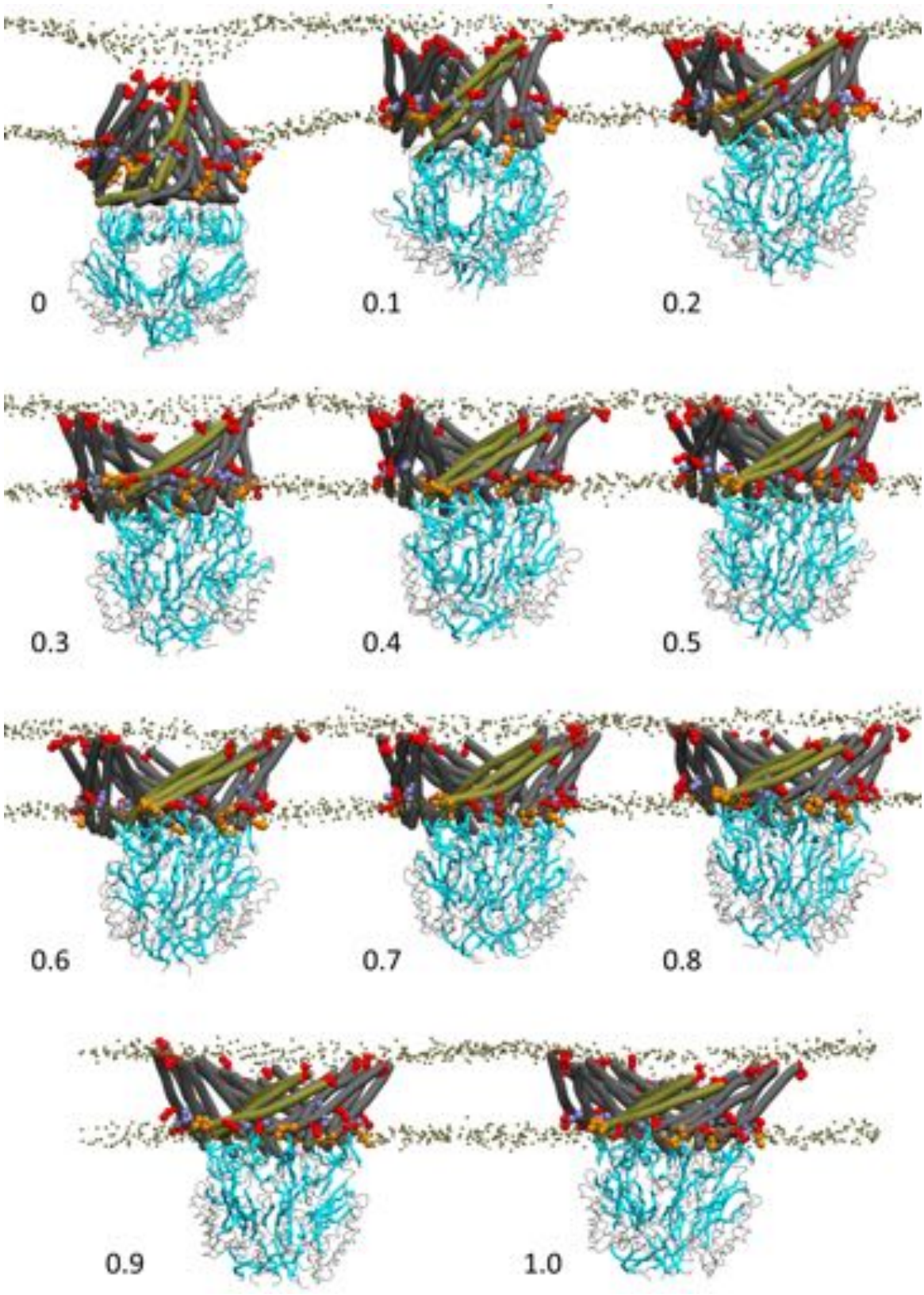

Figure 6.3.41.: Fast gating mechanism: side view
The numbers refers to fractional time of the total simulation run. Same graphical representations as in Figure 6.3.38. One chain's TM domain is marked in tan bendices to visualize its dynamics. Key residues that make contact to phosphate groups during gating are shown in Van der Waal representations, coloured according to interaction dynamics. Constantly strong contacts in red or orange: Tyr27, Ser58, Arg 46, 54, 59 and 88. Reduced number of contacts in ice blue: Asn50, Arg74.

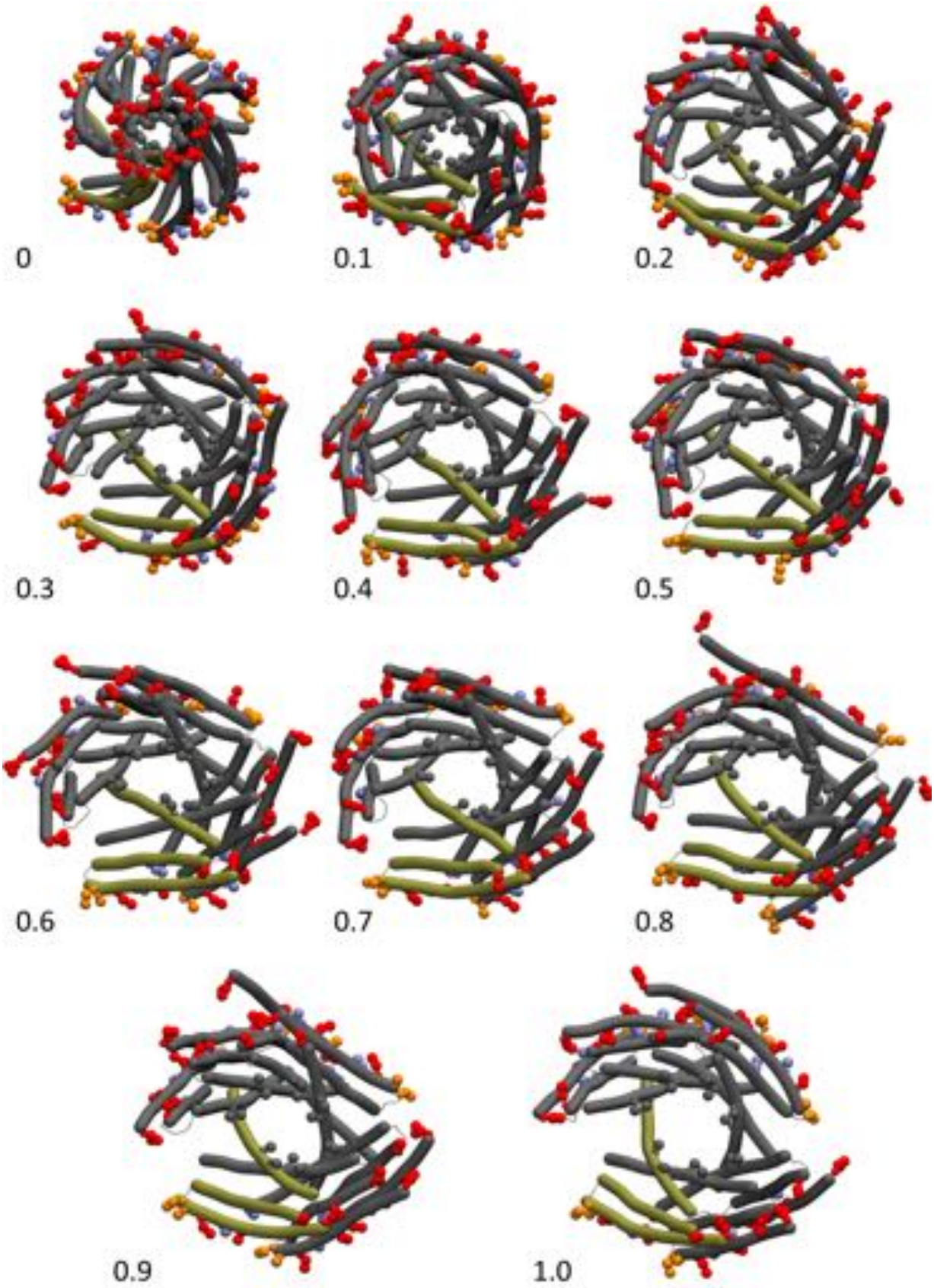

Figure 6.3.42.: Fast gating mechanism: top view
The numbers refers to fractional decimal time of the total simulation run. Same graphical representations as in Figure 6.3.41, only no lipid phosphates are shown, and the leucines that form the hydrophobic gate are shown as gray spheres along the pore-lining helices. Orthographic perspective.

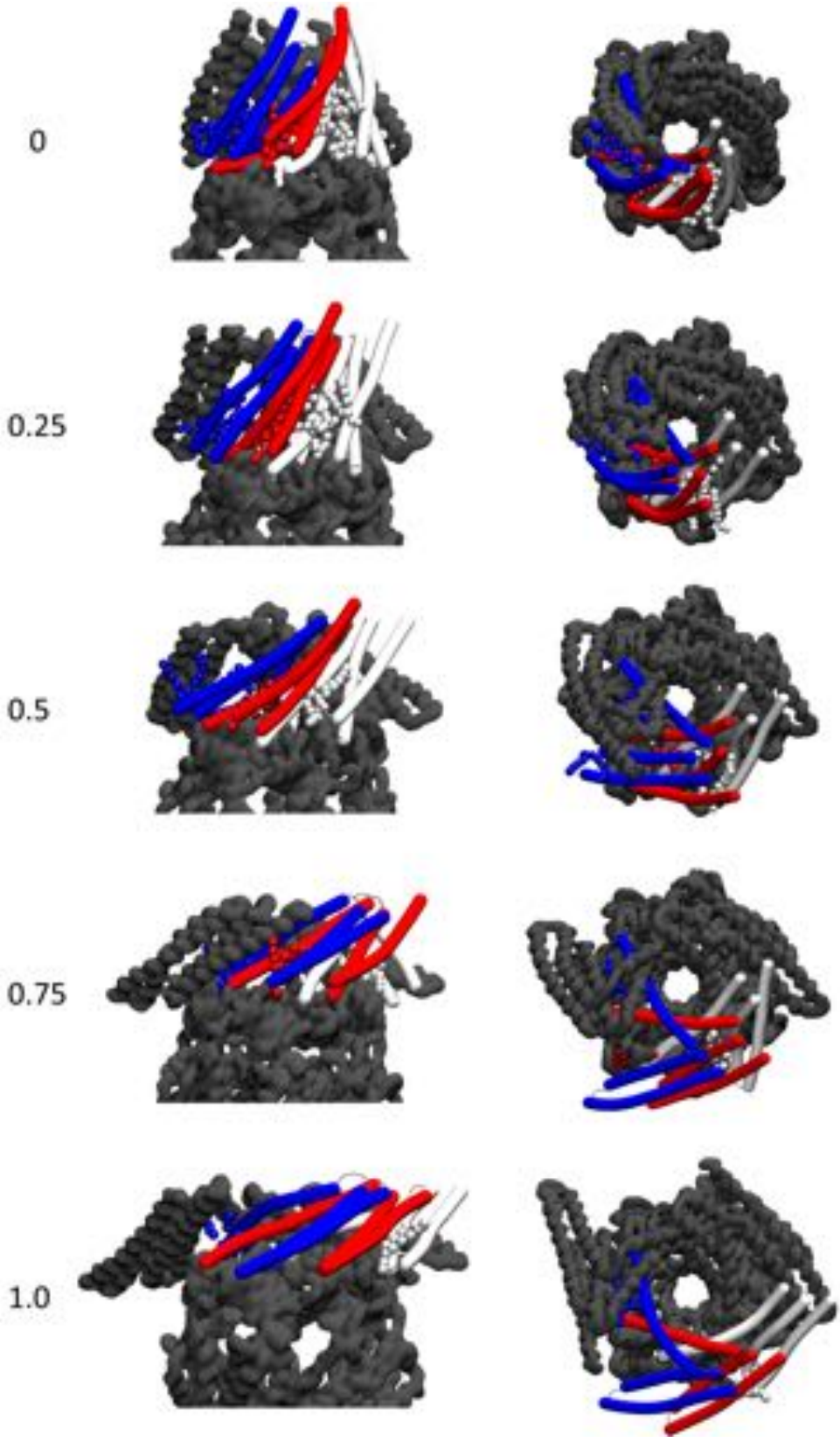

Figure 6.3.43.: Lower TM3-associated lipid dynamics during conformational change
The numbers refers to fractional decimal time of the total linear deformation simulation run. TM domains of chains A, B and G represented as bendices using colours red, white and blue, respectively. Their respective lipids are colour-coded accordingly and shown in space-filling representations. The rest of the protein backbone is shown as gray surface.

makes novel contact with lipid, reflecting its more membrane-proximal location in the open state. The merge with the membrane is done simultaneously as the TM domain looses its conical shape in favour of a cylindrical shape. Throughout gating, the charged and polar residues at the TM apex, Tyr27 and Arg88 remain in contact with pericellular leaflet lipid headgroups, which suggests that radial lipid movement could induce this dramatic change of shape. Tyr27 moves outward to align with the outermost protein circumference, joining and at times even replacing the cytosolic paddle as the most peripherally located part of MscS. At this point, membrane curvature is completely eliminated, and surface tension starts to affect membrane thickness.

Membrane thickness and helix tilt dynamics Membrane thickness remains near-constant until 500 bar*nm, after which is drops rapidly to 3nm over the course of the following 200 bar*nm (after which the membrane lyses). The TM1 helix tilt angle appears mildly correlated with membrane thickness; in linear deformation simulations, it remains approximately constant at 60° until 400-500 bar*nm, at which point it decreases rapidly as surface tension increases further. Interestingly, during this rapid tilt decrease, tilt speed predicts whether MscS gates or not; faster tilt is found in gating MscS. This suggests that rapid tilt response is important for gating. Second, TM1 is invariably more tilted than is necessary for it to avoid hydrophobic mismatch in the membrane. The way in which the high tilt angle is maintained is probably through the positive charges along TM1: Arginines 46, 54 and 59 are all closely associated with lipid phosphate in the open state. I.e. lipid headgroups in the cytosolic membrane leaflet make close contact not just with the tip of the helix, but up until the middle of the helix.

There are insignificant differences to lipid phosphate contacts between MscS that fail and succeed at gating, so these TM1 contacts do not explain the difference in tilt angle. General lipid, on the other hand, does make less contacts to the cytosolic part of the paddle in MscS where gating fails. Non-phosphate lipid parts thus likely play a role in maintaining the high TM1 tilt necessary for gating. Also, lack of contact gain generally reflects failed movement of the full TM domain into the membrane proper. It appears that incorporation of the MscS TM domain into the membrane, where it can optimise its interaction with lipid, is crucial for MscS gating.

The important role of membrane contacts seems to come into play also at TM3b, where only successfully gating MscS gains lipid contacts. Notably Arg128 becomes associated with a lipid phosphate headgroup (see Figure 6.3.36).

General lipid contact dynamics As MscS enters the membrane, more lipids – three per chain – make contact to MscS. The number of contacts to lipid phosphate, however, remain constant. POPG makes more persistent contacts throughout gating than POPE does relative to membrane composition. While the protein as a whole gains lipid contacts, TM helices generally loose contacts as membrane lipid density decreases. As a rule, residues that are in contact with lipid at zero surface tension loose contacts as increasing surface tension decreases lipid density, while cytosolic residues gain contacts.

Pore dynamics At equilibrium, TM3a is in line with the cytosolic leaflet's phosphate headgroups. The maximum angle along TM3, which subdivides it into TM3a and TM3b, is located around Gln112 to Leu115. As TM3 is pulled up into the membrane by increasing surface tension, TM3a relocates to the hydrophobic membrane interior, and the TM2-3 loop gains contacts, reflecting its less shielded position with respect to the membrane as the apex is widened by increasing lipid surface tension.

The maximum TM3 angle value varies inversely and proportionally with membrane surface tension. Its locus moves cytosolically to Leu 115 as tension increases, but a third of the time until gating, TM3 adopts a low, 15° curvature that is distributed evenly about its length. This mild curvature persists until gating.

After TM3 straightening and movement into the membrane, the hydrophobic gate has relocated to a middle or pericellular bilayer location, where lipids' radial force on the TM2-3 loop and adjacent Arg

88 and Tyr27 is more easily translated into pore widening, and therefore mechanotransduction. The important role of the TM apex in gating is supported by the observation that the hydrophobic gate at Leu 105, which is more proximal to the apex, widens before it does so at Leu 109.

The role of lipids in the lower TM3 region As noted in the previous chapter, approximately two lipids locate to the lower TM3 area; one of which lines TM3b and a second lipid that is more loosely associated and lies closer to the lower leaflet, parallel to the TM3b-associated lipid, between adjacent MscS chains. As tension increases and the full TM domain enters the membrane, the loosely associated lipid is lost to bulk membrane. The TM3b-associated lipid persists throughout major conformational change during gating, located between the highly dynamic paddle and pore-lining helix.

Manual inspection of linear deformation simulations suggest that persistent lipids could act as a relay for TM3-paddle interactions, facilitating the parallel helix conformation seen in the open state. At its position near the gate, this lipid could transmit tension to the gate, as I suggested in the previous chapter. Only at gating tension can this lipid leave for bulk lipid, and this does not always happen. Its role therefore seems threefold: facilitate the parallel conformation of same-chain helices, relay paddle outwards movement to the gate and to possibly evacuate the area, causing the pore-lining helices to move out in concert, thus gating the channel.

6.3.6.4. The open state

At high surface tension, approximately 600 bar*nm, the seven paddles form groups that move asymmetrically outwards, while the dilated pore retains its iris shape. The straightened and highly tilted pore-lining helix TM3a twists residue side chains so that the hydrophobic gate partly faces the periplasmic space rather than the pore, reducing its function as a vapour lock.

6.3.6.5. The gating mechanism of MscS is tension rate sensitive

Rapid tension application causes rapid 30° tilt of TM1, and immediate and major pore widening as a near-linear function of TM expansion to form the open structure's cylindrical shape. Tyr27 assumes its peripheral position almost immediately. Contacts to lipid are mostly preserved through gating, and, notably, significantly less lipid is lost from between the paddle-TM3a area, suggesting that speed partly rescues the need for less lipid density in this area. (NB: while this is true for MscS without ENM, MscS with ENM limited to the cage domain shows a more graded gating response, requiring near-lytic tension to gate all instances of MscS. Lipid contacts were not investigated.)

On the contrary, slow tension application causes the pore to collapse and remain unresponsive to the paddle until it moves out by 5 to 10 Å. The unresponsiveness is characterised by a steady-state where Tyr27 is located halfway to the TM periphery, and the TM shape is half-way between a cone and the final cylinder. ENM on the cage domain has in essence no effect on slow gating dynamics.

6.4. Discussion

6.4.1. The 'open' crystal structure in the light of my results

The site of lipid loss; between mid-TM2 and mid-TM3a, is the proposed site of loss in Chapter 5. This chapter shows that lipid loss is a relatively early gating event that does not cause immediate pore widening, contrary to what was proposed in Chapter 5. My results show that the restrained closed crystal structure prevents loss of this lipid. Moreover, MscS goes through a state with rotation of the paddle to align with its own chain TM3, loss of the TM domain iris shape for more mesh-like structure with less space between adjacent chains, a wider apex and increased TM helix tilt. All this is reminiscent of the open crystal structure [Wang et al., 2008], suggesting that it is a state

that is part of the early stages of the gating mechanism. However, in my MD results, the pore has not yet widened and the kink at Gly113 decreases with tension. Likewise, the TM3 twist seen in the crystal is only observed in the final stages of MD gating, and then in the opposite direction compared to the crystal. These contrasts to the open crystal structure likely originate in the crystal's mutation, which forcibly widens the gate. The absence of a bilayer under tension in the crystal setup also calls into question the 'open' crystal structure being the true open state. MD shows that slow tension application first results in asymmetric pore closure, which has been previously demonstrated to occur on a nanosecond timescale in multiple AT MD simulations of the closed crystal structure [Sotomayor and Schulten, 2004, Spronk et al., 2006, Sotomayor et al., 2007, Vásquez et al., 2008a]. I show in the previous chapter that the forcibly widened gate allows lipid to enter the pore under equilibrium conditions, while this chapter's non-zero surface tension prevents entry. As the open structure should be under tension, I conclude that lipid pore access was an artifact of the zero surface tension membrane.

6.4.1.1. Lipid loss in the light of mutation studies

Lipid is lost between the paddle and TM3a, and mutation to this area is generally associated with less MS and conductivity [Nomura et al., 2006, Spronk et al., 2006, Nomura et al., 2008, Belyy et al., 2010a] (c.f. Tables 1.5.1 and 1.5.2). Such mutations could be affecting the propensity of lipid to prevail in this area. While I show that lipid loss is not immediately associated with gating, and occurs in both successful and failed instances of MscS gating, loss of 'surplus lipid' that is not strongly associated with protein could be necessary for better force relay between remainder, stronger lipid contacts and the gate.

6.4.1.2. TM3 helix angle

These MD results show that the crystal structure of MscS that is referred to as 'closed' state can gate when subject to membrane stretch. The bendix settings that I use are valid for analysis of distortion at Gly121, which was proposed to denote the closed state [Akitake et al., 2007], but the TM3 distortion locus notably never relocates there. My results therefore suggests either that PDB id 2OAU depicts a closed state or that gating is possible from an inactive form of MscS, thus passing through an alternative closed state that does not feature a kink at Gly121. However, the angle does decrease and move cytosolically to Leu115, suggesting that this could be an intermediate state between high kinks at Gly121 and Gly113, which allows gating. My results also show that Leu115 remains in close contact to, on average, 1.5 lipids throughout gating, suggesting a concerted, lipid-generated motion. The importance of Leu115 is supported by the fact that it is conserved across bacteria and that mutation to Serine (which would impede contact to lipid tail) causes MscS to become less mechanosensitive [Belyy et al., 2010a].

6.4.2. Decreased annular lipid curvature results is widening of the TM apex, gating MscS

When a membrane's lipid density decreases, as in a membrane under high tension, curvature is not favoured, and lipid that assumes less local curvature may pull on the N-terminal TM1 and the short linker that joins T2 to TM3, where strong links to lipid headgroups exists. This results in a less cone-shaped TM domain with a wider pore mouth, which decreases pore hydrophobicity [Beckstein et al., 2001]. While in-plane expansion of the apex specifically was not plotted against pore dimensions, visual inspection of gating MscS suggests that they are linked. The lag between correlated pore and TM domain expansion seen under slow tension application coincides with the pore collapsing and the apex reaching a steady-state half-way to the periphery before widening simultaneously with the pore at high tension. The apex - pore association is further supported by MD under fast tension, where apex and pore widen simultaneously without an intermediate steady-state.

While correlation does not imply causation, the force involved in merging the TM domain and the bilayer and widening the apex is of lipid origin. Reduction of the membrane curvature about the apex is an early event that is directly associated with apex widening, and the use of membrane geometry as a means to sense change to tension would be sensible for an MSC. Moreover, the apex involvement in gating is supported by mutational data where charge neutralisation or swap to polar Serine at Arg88 results in less MS and conduction [Spronk et al., 2006, Edwards et al., 2008], and Thr93Arg, which could provide additional attachment sites for lipid, results in less tension necessary to gate MscS [Miller et al., 2003a, Edwards et al., 2008]. Mutational studies of the TM1 N-terminal are lacking. While Vásquez et al. [2008b] mutated each TM domain residue in their EPR study of the conformation of MscS, they only scored residues based on percentage survival by downshock; no electrophysiology was performed. This study suggests that electrophysiology of the effects of mutation to the N-terminal would be insightful.

6.4.3. Merging of the TM domain with bilayer facilitates lipid interactions

All TM1s need to merge with the membrane in order for MscS to gate. I hypothesise that membrane merging facilitates contact to lipid, which will affect TM1 tilt - the speed of which is important in gating. The insignificant differences between phosphate contacts to MscS that gate versus fail to gate is perplexing since mutational data show that charge exchange or reduction of polarity at charged residues that are located at the level of lipid phosphate groups does affect gating negatively. Mutation to Asp62, Asp67 or Arg 128 makes MscS less mechanosensitive [Nomura et al., 2008] and mutation to Arg46, 54, 74 or 88 results in less MS and/or conduction [Spronk et al., 2006, Edwards et al., 2008] (however Malcolm et al. [2011] observes wt survival by mutation to Alanine of Arginines 46, 54 or 74; the other residues were not tested). My results do show that significant differences between non/gating MscS were seen for general lipid contacts to the cytosolic paddle region, suggesting that non-phosphate lipid headgroups could play a role.

6.4.4. The role of negatively charged lipid in gating

The +42 charge of the TM domain suggests that it could bind negatively charged lipid moieties. My results show that, relative to system content, POPG is overrepresented in persistent contacts to MscS throughout gating. Because of their relatively poor packing, negatively charged lipids lower the lateral pressure, and neutralising their charge makes MSCs less prone to gate [Ermakov et al., 1998, 2010, Hamill and Martinac, 2001]. I reveal very high annular lipid curvature about MscS, which makes the lateral pressure profile an unsound choice of metric in the immediate protein vicinity. Here, on a molecular level, negatively charged lipid might allow more persistent interaction with MscS, with better force transmission as a result.

6.4.5. Lower TM3 lipid

TM3 is the most conserved region of MscS (see Table 1.5.2). Throughout gating, only lower TM3 lipids remain stably associated, and move in concert with protein. Introduction of polarity to Leu115, a conserved hydrophobic residue in the amphipathic TM3b region which makes contact with lower TM3 lipid, causes MscS to loose MS [Belyy et al., 2010a]. Similarly, mutation to charges on to the cytosolic paddle, that were found to make persistent contacts to lower TM3 lipids in Chapter 5, cause less MS and decreased downshock survival [Nomura et al., 2008]. I also identify contact gain to C-terminal TM3b, where charge neutralisation or reversal to Arg128 and 131 cause less MS and unstable open states [Nomura et al., 2008]. Such mutations should relax interaction to lipid, and my results show that lipid depletion in the lower TM3 region and absence of contact gain to C-terminal TM3b is associated with gating failure. As previously discussed, volume and hydrophobic character conservation in the lower TM3 area is conserved [Belyy et al., 2010a], which should aid lipid adherence. Lastly, lipids'

existence in the new open crystal structure also emphasises its persistence. Taken together, lipid persistence and gain around TM3b appears to be important in MscS gating.

6.4.6. Cage

The cage approaches the TM domain during gating, whether ENM is applied to it or not. While cage compaction in the membrane normal axis is thought to define an inactive state of MscS [Koprowski et al., 2011, Rowe et al., 2014], my MD results suggest that compaction is compatible with an open state. However, the open state necessitates ENM on the cage, as absence distorts the CG cage domain to the point where solute passage is impossible. Similar distortion occurs in the cytosolic domain of a Martini CG model of Kv1.2 after 350 ns, showing backbone RMSD of 8.5$\pm$1.3 [Treptow et al., 2008]. Protein distortion could therefore be a feature of the CG forcefield after a prolongued period of time. CG water molecules' larger size could mean cage dehydration leads to local vacuum and tertiary structure compromise. Alternatively, distortion could be a side-effect of missing residues (as in the case of Kv1.2) or even be a feature in MscS gating. Water would enter from the pericellular space as MscS opens by downshock, and the water potential could be sufficient to expand the collapsed cage domain, widening the portals, which would effectively act as a second gate. The potential of the cage as a gating element has been raised previously [Koprowski et al., 2011, Rowe et al., 2014], and the cage does swell during gating [Machiyama et al., 2009], but the possibility that cage portals could be sealed until opened by water-induced swelling is a new hypothesis.

6.4.7. The sensor paddle realignment optimises capacity for in plane expansion

Some of the effective chain length that is available for extension in the membrane plane is lost to the diagonal arrangement of the sensor paddle. This is therefore relocated to be in line with its own chain's TM3 domain once there is enough tension to break the proposed electrostatic interactions between the Asp62 on the TM domain and the cage [Sotomayor and Schulten, 2004, Nomura et al., 2008]. Lipid headgroups contact Asp62, and may therefore provide it with alternate coordination sites to lower the energy associated with disruption of interaction. Paddle movement is likely also aided by the lipid tails that the paddle confines, that may reduce friction between the sensor paddle and TM3 (by "greasing up" the passageway) and by fluctuating, which yanks at the saltbridge. Alternatively, excess lipid may prohibit paddle movement. In that case, the paddle can only move across to its own chain once bulk lipid is removed (although it is more probable that this happens in a concerted motion).

6.4.8. Surface tension speed

6.4.8.1. Lipid dynamics is a function of tension speed

Membrane lyses at higher tension when tension is applied at a higher speed. When tension is increased in steps (each step being held for "several seconds"), PC lipid membranes lyse at < 100 bar*nm [Rawicz et al., 2000].

A bacterium is likely to be exposed to a wide range of tension speeds in nature, and needs to be capable of surviving all of them. Since bacterial MSCs likely evolved to follow the dynamics of lipid - which changes with loading speed - channels may gate at different tension depending on the loading regime. Therefore, just as bacterial channels' gating tensions are provided as a ratio of MscS and MscL to account for inconsistencies in experimental setups, perhaps gating should also be considered in relationship to lytic tension, to account for tension application speed. This may, in part, explain the discrepancy between the gating tension obtained experimentally and by MD (discussed in more detail in Section 6.4.9).

6.4.8.2. MscS gating as a function of tension speed

In my MD simulations, all instances of MscS gate rapidly by rapid tension, but lags to gate, or do not gate at all, by slow tension. This response has been previously documented in experiments [Akitake et al., 2005], although at velocities available to wetlab, which are typically a thousand times slower than those available to MD. A tension-velocity sensitive response could be advantageous to a bacterium; rapid tension caused by water influx suggests the requirement for rapid gating to avoid bursting. On the contrary, a slower increase in tension is suggestive of a water potential that could be offset by more stress-sensitive channels with smaller diameter gates. If MscS gating can be avoided, it is beneficial for the bacterium as it looses fewer metabolites in a situation that is not life-threatening [Akitake et al., 2005].

My results suggest that slow tension gives intercalated lipids the possibility to escape their lodged positions within MscS, while rapid tension gates MscS without complete lipid loss. The protein conformational change also shows tension speed sensitivity; the pore collapses under slow tension, while no collapse is observed at rapid tension; the pore only expands.

6.4.9. In defence of the high tension necessary to gate MscS

The experimentally determined gating tension of MscS is approximately 55 bar*nm [Sukharev, 2002]; half the surface tension necessary to lyse pure *E.coli* membrane (without protein or peptidoglycan cell wall) [Anishkin and Sukharev, 2009]. The surface tension which gates MscS in this study, 600bar*nm, is much higher than both the experimental gating and membrane lysis tensions. Put in perspective of previous MD of MSCs , however, the tension is acceptable. 600 bar*nm is lower than than required to gate MscL in a similar MD setup, and higher than tensions that have failed to gate MscS previously (c.f Tables A.1.1 and A.2.1). At the extreme, in one study, CG MscL only gated once 5000 bar*nm surface tension was applied; more than two orders-of-magnitude larger than the experimentally required tension magnitude [Colombo et al., 2003]. Since membrane behaviour is a function of surface tension application speed (as discussed above), the discrepancy between experimental and MD tension values may in part reflect the much higher speed at which tension is applied in MD compared to experimental settings.

6.4.9.1. MD simulations require excessive stimuli to achieve an effect in limited time

In an experimental setup, MscS gates within milliseconds [Levina et al., 1999]. Researchers note that, in order to observe effects on MD timescales, the magnitudes of stimuli such as tension and voltage must exceed the physiological range [Gullingsrud et al., 2001, Tieleman et al., 2001, Sotomayor and Schulten, 2004, Spronk et al., 2006]. To apply 10 times higher stimulation magnitude than experimentally determined is not uncommon, which precisely fits the discrepancy between the gating tension observed by experimental and MD methodology.

6.5. Conclusions

Coarse-grained MscS in a physiological environment This chapter has investigated the gating mechanism of MscS using CG MD methodology. This is the first model of MscS using the Martini CG model, which accesses longer time-scales and has successfully been used to gate MscL. Both the membrane and solvent were chosen to approximate physiological *E.coli* conditions; MscS has previously only been simulated in membrane made of POPC; a lipid which does not exist in *E.coli*.

Gating by surface tension The pore diameter to characterise the open state, 15 Å, was chosen to correspond with both previous MD studies, experiment and theory. Surface tension alone has never before gated MscS in MD simulations; additional steered MD, which necessarily biases a protein's

conformational change, was always required. I successfully gate MscS using 600 bar*nm surface tension. This gating tension fits between published tensions used to unsuccessfully gate MscS, and tensions successfully used to gate the less tension-sensitive channel MscL, which suggests that it is a fair value given the MD methodology.

Novel, more physiological surface tension methodology In addition to the standard, instant constant surface tension method, I use linear deformation, which has never previously been adopted by the MSC MD field, and develop a highly accurate tension quantification methodology. Linear deformation better approximates patch clamp experiments than the instant constant surface tension method, where the system is rapidly brought to the choice surface tension, so this technique should be valuable for future MD research. In this thesis, both tension regimes were used in order to compare the response of MscS to tension velocity.

MscS gating mechanism My results show how MscS undergoes large conformational change in order to achieve the necessary pore size necessary for solute passage. Tension causes the MscS TM domain to merge with the membrane, where TM helices undergo high tilt, voids between chains are reduced, TM3 straightens and the TM in-plane area expands, attaining a cylindrical shape. The conical apex is completely lost. At high tension, the iris formed by TM3 helices always persists, and paddles form an approximately symmetric arrangement about the pore, but in groups. TM3 partly twists leucines out of the gate (ENM permitting), and judging from available crystal structures, the direction of twist appears to be related to TM3's tilt angle. Tilt angle is also found to be important in TM1, where I find a relationship between fast tilt response to surface tension and successful gating.

Speed of tension has significant effect on the gating mechanism, and determines whether MscS undergoes pore collapse, responsiveness to TM in-plane expansion and the dynamics of lipid interaction. Indeed, MscS behaves as a Maxwell solid with dynamic viscosity depending on stimulus speed. MscS with and without ENM on the cage displays similar gating dynamics by slow tension application, but slows pore and TM expansion during fast tension application.

Conformational change is correlated with associated lipid dynamics Strikingly, I find that the conformational changes that MscS undergo are closely correlated with annular and intercalated lipid dynamics. Early gating events are associated with lipid loss from between the paddle and pore-lining helix (as predicted in Chapter 5), apex opening involves increased interaction between lipid and helix linker-regions and persistent lower TM3 associated lipids closely adapt to protein conformational change. As MscS enters the membrane under tension, one or two intercalated lipids leave the MscS interior adjacent to the hydrophobic gate for bulk lipid, probably predisposing MscS to an open state. The one remaining lipid fits the electron density in the new open crystal structure, and remains in close vicinity of sensor paddle and pore.

Unchanging, permanent contacts to lipid also have the potential to be functional. Chapter 5 identified contacts to lipid phosphate groups as particularly strong, and several lipid-facing residues stay in contact throughout the large conformational change of MscS as it gates.

Lipid links the pore to the bulk lipid under tension: a structural element of MscS' gating mechanism My results present a molecular dimension to the *dashpot* model suggested by Akitake et al. [2005], where lipid take on a prominent, structural role. Akitake et al. [2005] and Belyy et al. [2010a] assume that paddle and pore contact is necessary for force transmission, and suggests that their uncoupling is the reason for inactivation. I show that all conserved hydrophobic residues identified in Belyy et al. [2010a] make contact to intercalated lipid in MscS on the way to gating (Chapter 5 results for the new open crystal structure), that lipid deficiency from between paddle and pore is associated with failed gating and that increased lipid retention is a general feature in fast tension application. Lipid is, literally, the missing link between the sensor paddle and the pore.

This model is still compatible with that proposed by Belyy et al. [2010a]; lipid which leaves the lower TM3 region at high tension should allow increased contact between increasingly parallel TM helices. However, during the majority of the gating mechanism, a single lipid persists between paddle and pore, where it is in a prime position to mediate force transmission.

The role of lipid as a structural element in MscS is a new insight into mechanosensitive gating mechanisms. My results underscores the potential for lipid to augment the protein structure in the protein structure-function paradigm.

7. Conclusions and future directions

This thesis work concerned reduction of complexity in MD systems, as well as molecular level mechanosensitivity. I developed accessible tools to make abstraction and statistical analysis manageable beyond this work, and applied them in my studies of mechanosensitivity in the bacterial mechanosensitive channel MscS. Below follows discussions of results presented in this thesis, and proposed extensions.

7.1. Software development in order to reduce MD complexity

MscS in its physiological environment is a large and complicated system. In order to reduce complexity both in terms of visualisation and quantitative analysis, I wrote two programs; *Bendix* and *InterQuant*. Both programs are highly customizable and take advantage of the popular and widely distributed molecular graphics package *VMD* in order to visualize their results.

7.1.1. Bendix

Current abstractions of protein helices often use a ribbon that incorporates each backbone particle. While this representation is precise, it has much redundant detail that can be overwhelming when dealing with a protein with as many helices as MscS. The alternative is to use rigid cylinders in the place of helices, but these do not disclose helix distortions, which can be functional, as in the case of TM3a, the pore-lining helix of MscS. Bendix detects and abstracts protein helices into cylinders that follow the helix axis. It also quantifies helix distortion over time, which is of immediate use in my studies of the gating mechanism of MscS. The software features are equally applicable to CG systems, where secondary structure visualisation is very limited. Bendix is hailed as a very useful tool on the Martini CG website, and will ship as part of the core analysis programs in the next VMD release.

I demonstrate the capabilities of Bendix on the transporter Mhp1, where helix distortion is intimately linked to its transport function. Bendix allows me to identify a state that was previously proposed. Further, I hypothesise that substrate presence confers rigidity to Mhp1, allowing its gates to be structurally linked during transport in a way that is not possible when substrate has been released inwards.

Future work

Console version A console version of Bendix, that generates analytical and pictorial output according to user settings, would be very useful. This benefits computational researchers that develop script pipelines for their work.

Bendix for Pymol A Bendix plugin for PyMol was the most requested feature at the Biophysics conference in 2012, where Bendix was first presented publicly.

Surface graphics that scale better and support variable radius Bendix representation surface generation using vertices rather than spheres and cylinders, that are computationally demanding to use at high resolution. A second reason for using triangulation is that triangulated bendices handle

transparency and vector graphics output better. At the moment, spheres and cylinders obstruct transparent visualization, particularly at high resolution, and vector graphics output is not smooth. Triangulation also lends itself to variable helix radius, which could be used to indicate e.g. the thinner 3_{10} helix conformation.

Choice of spline Add a choice of splines, in particular the choice to use a beta spline in place of the Hermite spline. The Hermite spline moves through input coordinates (the computed helix axis, in the case of Bendix), whereas the beta spline has a parameter to control accordance to input coordinates, offering the choice of smoother line output. The Hermite spline is therefore more compliant with inputs, which should be beneficial. However axis point generation is imperfect, and the Hermite spline is inherently prone to introduce sinusoidal noise in the output. Another reason to allow the user to choose the beta spline alternative is that this spline is implemented for NewCartoon representations in VMD. When atomistic backbone and beta sheets are hooked up to bendices, there should be better agreement between representations at joints if they use the same spline and spline parameters.

Application to molecular knots Molecular knots require a high level of abstraction in order to be understandable. During Bendix evaluation, I found that representing a full protein as a a single bendix substantially reduced visual complexity and produced visual output that was particularly pertinent to the area of molecular knots. I would like to explore this area of research further, armed with the visual analytics of the Bendix software, and present it to relevant researchers in the field.

Individual helices as selections Currently, Bendix only supports representation of a single selection. Extension to allow multiple selections would enable each selection to be individually treated for bendix settings, as well as material and colouration.

Display the maximum detected angle in the GUI When helix geometry is evaluated and depicted onto the relevant helix surface using a heatmap, the angle values corresponding to different colours are not shown. Quantitative data are produced only when saving to file. However, the data is available at the Bendix backend, and it would be advantageous to display the maximum value variable as text in the GUI.

Such a feature is already implemented in the InterQuant GUI.

7.1.2. InterQuant

InterQuant was developed in order to quantify reproducibility and system differences within MD data using statistical inference methods. The program offers insights with quantified certainty into the parameters of the underlying population that contains all possible simulation replicates, and therefore supplies results that are ultimately comparable to macroscopic, experimental data. InterQuant is very versatile, and when used to characterize solute-solvent interactions, it facilitates analysis of solvent accessibility and binding preference, where solvent could be e.g. water, ion, lipid or a drug molecule.

Future work

Screening and support for non-normally distributed data Whereas the current InterQuant version takes the required measures to provide the correct output for un/equal variance and input sample sizes, it does assume normality in the underlying population(s). Normality is hard to test for in small samples, but by using the most robust tests, I would like to be able to offer users the option to test for normality prior to statistical analysis. Examples of tests for normality include the Shapiro-Wilk and

Kolmogorov-Smirnov tests. A non-normally distributed sample requires a nonparametric test to compare population parameters, such as the Wilcoxon tests, which would be beneficial to implement. However, as previously mentioned, the x^2 shaped potentials about equilibrium that exists for most interactions, advocates the assumption of normally distributed observations.

Support for analysis of a chosen subset of non-bonded interactions, e.g. hydrogen bonds At the moment, all Origin and Targets are considered for interactions. It would be beneficial to only interrogate subsets of interest, e.g. hydrogen bonds, salt bridges or hydrophobic interactions. While the user can currently customize selections so that Origin and Target only contain, for example, hydrogen bond donors or acceptors (or vice versa), analysis of only explicit types of interaction would be good.

More automated analyses, unrelated to selection separation InterQuant supplies a relatively large amount of types of analyses for the proximity and contact family of analyses. However, more types of analyses could be implemented, for example radius of gyration or bond angles. Ideally a way for a user to supply his or her own analysis script to compute and collect results centrally and produce text files that are easily loaded into IIQ.

Notify the user of heterogenous variance in dependent samples As mentioned previously, there is no Welch's test equivalent test for dependent samples with heterogenous variances. However dependent samples' homogeneity can be tested using the Pitman-Morgan test, and the user could be cautioned when interpreting the data.

Stand-alone suite of loadable scripts to retrieve statistical significance When a creative user has results for analyses that are currently not yet implemented in InterQuant, he or she could paste those results into existing IQ files and run IIQ analysis on those edited files to compare samples statistically. However, I would like to implement a stand-alone user interface for statistical analysis of custom metrics. This would allow the user to input samples and retrieve relevant statistics for any analyses; not just proximity and contact-based data. This would be implemented in a user-friendly manner for VMD in order to gain the visual advantage that is not possible using existing statistics software.

Report the number of samples required for a desired confidence interval While a user will obtain a certain CI given his or her sample size and its variance, one can also estimate how many observations would be needed for desired CI, given the current sample variance. The method assumes that sample variation does not change as more observations are collected, so needs to be a good estimate of σ^2. I believe that this would be useful information to an MD researcher who plans future MD simulations. For a one-sample t-test,

$$N_{needed} = \frac{s^2 t^2_{\alpha,df}}{d^2}$$

where d is the half-width of the desired CI [Zar, 2010]. For example, if the user requires an average separation difference that is no larger than 0.1 Å compared to the reference, $d = 0.05$ Å . Generally, the higher sample variance and smaller the range of the CI, the more observations are required. The solution is not trivial since df depends on N_{needed}, so the equation is solved iteratively, which could be challenging for a beginner student, but is a good feature for an automated, computational statistics tool.

Likewise, the sample size needed to ensure a certain probability of rejecting $H0$ when it is in fact true, a probability of accepting $H0$ when it is false (type I and II errors) as well as capacity to resolve differences between the reference and sample's population mean, is also iteratively computable.

Similar relationships are available for two-sample t-test results, and, if implemented in InterQuant, should help plan MD studies

Implementation in C++ IQ is optimised with conditionals so that it only runs necessary functions and loops, but it is still a relatively large software. When I worked as part of a larger team to implement another VMD plugin, I learnt how to call local C++ code from Tcl. If I had the time, I would like to re-write the key, computationally expensive parts of InterQuant in C++ to speed up execution, while keeping the interface and software backbone in Tcl.

Implement the option to correct for multiple testing The t-test p-value is only statistically valid as a statistical confidence measure when a single test is performed, e.g. when one residue is tested for change to separation between samples. When the same sample is used for several tests, the chance of false positives increase [McDonald, 2014].

There are several techniques that reduce the number of false positives, but it is often the case that these techniques produce false negatives instead. Commonly used multiple testing correction algorithms, their applications and pros and cons are listed in Appendix C. I would like to research the available techniques further and implement at least one of the procedures for IIQ execution.

7.2. Study of mechanosensitivity at the molecular level using MscS

In my work on MscS, I have generated a CG model of MscS and used it to simulate MscS in a physiological environment for longer than ever done before. I compared and contrasted membrane behaviour about resolved crystal structures in MD systems without applied tension, and examined my results in the light of new structural and biophysical data. I then pursued MscS gating mechanism using instant constant surface tension as well as linear deformation - a technique that has never been used on MSCs before and which better approximates the experimental setup. Lipid is shown to interact distinctively with different states of MscS, and since membrane alone drives MscS state change, the differential lipid interaction that I show explains molecular level MS in a CG model. In simulations of MscS under tension, I used a range of tension application speeds, and my results suggest a molecular basis for the previously published dashpot mechanism of MscS gating.

Lipid as a structural element in membrane proteins Although my results for MscS provides a particularly dramatic illustration of pressure sensing, it has been shown that many other ion channels are in fact modulated by membrane tension. Lipids that repartition between bulk membrane and protein constitute a natural source of tension relay for a protein with closely associated lipid. This is particularly the case if such lipid has a role in the protein's structural integrity. My collaborators and I suggest that this is a general model for such proteins, a class which includes eukaryotic potassium channels [Schmidt et al., 2012]. Lipid may thus plays a previously unidentified mechanical role in mechanotransduction by participating in the structural integrity of the channel itself.

Future work

Addition of TM0 The one domain that was not included in this study is the first transmembrane helix, TM0, is not resolved in available crystal structures. However, it has since been studied using EPR of the closed state, which led to the generation of approximate TM0 coordinates [Vásquez et al., 2008b]. The influence of TM0 on the behaviour of MscS in membrane remains to be examined in silico.

Further biophysical data MD results suggests many sites of lipid proximity that were not tested in quenching experiments. It would be beneficial to have this additional data to see whether experiment supports MD results.

Possible improvement of crystal data My results suggest that electron density is not all alkyl tail. Rather, electron density along the TM3b C-terminus is probably lipid headgroup. By refitting the electron density using this MD result, the overall crystal structure resolution could be improved. I am currently in discussion with my collaborators about this.

Lipid accessibility as a means to assess promising sites for Drug Discovery Lipid works as a probe to explore differences to accessibility between states. If the goal is to stabilise the open state by introduction of a drug, significant differences in lipid binding between the open and closed states - in particular sites that are significantly stronger bound to lipid in the open state - suggest sites to explore further.

Additional MD to verify the role of lipid in the structural integrity of MscS My collaborators and I hypothesise that lipid presence stabilise the paddle's orientation relative to the TM3 domain. However, stability was not tested as heavy restraints were imposed on the protein structure to retain the crystal structure conformation. Preliminary MD data show that the paddle domain does not close in on the TM3 domain during a 100 ns AT simulation of the closed state without restraints (data not shown), but these are only data from a single simulation.

Another simulation that would be interesting to do would be to remove a lipid from the TM2 region in the closed state, to see whether this does destabilise the closed state and leads to pore widening.

Voltage alters the cavity geometry: unknown effects on lipid content MscS is sensitive to voltage. An earlier study found that depolarisation, as would result from more mechanosensitive Msc family members opening at lower tension, prior to MscS, resulted in increased conductance and a lower tension threshold [Martinac et al., 1987]. More recent work suggests that voltage only affects deactivation [Akitake et al., 2005]. Spronk *et al.* investigate the molecular detail of voltage sensitivity of MscS, and show that the paddle moves 0.3nm in the direction of the pericellular space upon depolarisation from -1.1V. Seeing as the paddle is in close interaction with lipid in the paddle-TM3 cavity, the effect on this lipid would be interesting to explore further.

Mutant behaviour in MD under tension This study of MscS behaviour under tension identifies many residues that interact with lipid during channel opening. Mutation to such residues and subsequent simulation to identify aberrant opening behaviour would be a good start to further investigation into what lipid interaction is important to the function of MscS. Example mutants would be double mutant Arginine 59 and 60 to negative or neutral, or TM2-3 linker neutralisation to see whether in-plane expansion is negatively affected. A mutational study could likewise be broadened to include any of the published mutations that I compiled in Table 1.5.1 and 1.5.2 to investigate the molecular detail of the phenotype.

Introduce membrane curvature The curvature-inducing lipid LPC, which features a large head and small tail, opens MscL when introduced into the pericellular membrane leaflet [Moe and Blount, 2005]. Curvature is therefore probably a reason for why MscL opens wider at 670 bar*nm, when in a high curvature environment [Louhivuori et al., 2010] than at the higher tension of 770 bar*nm in a planar bilayer [Yefimov et al., 2008]. MscS also opens in the presence of lysolipids, but the molecular details of how the conformation of MscS is affected are unknown [Martinac et al., 1990]. Introduction of lysolipid into the MD system could therefore be revealing.

Compute forces felt by individual particles While InterQuant details lipid interaction with MscS in terms of separation, my initial vision was to incorporate the underlying forcefield to compute the forces felt by protein. I have written rudimentary Matlab scripts to visualize the forces computed in GROMACS, but a utility to measure how lipid affects MscS as tension is increased could expose the importance of individual residues and lipids.

More physiological membrane An accurate representation of the membrane environment is important in membrane protein studies, and particularly so when studying MSCs where membrane under tension is sufficient for gating. The membrane that I use here is more physiological than the POPC bilayers used in previous MscS MD work (particularly since POPC does not exist in *E.coli*), and was used because both AT and CG parameters existed for them. However, authors have pointed out that the *E.coli* membrane is more diverse, and should incorporate 5% cardiolipin [Raetz, 1986] and lipids with acyl chain cyclic moieties [Pandit and Klauda, 2012]. The latter makes the membrane more rigid and thinner, which would probably affect the function of MscS.

A. Appendix: Compiled data on previous MD results

Tables A.1.1 and A.2.1 summarises results from previous MD studies of MscS and MscL, respectively. Studies are ordered according to achieved gate diameter magnitude.

A.1. Previous MD results for MscS

A.2. Previous MD results for MscL

ϕ_{Max} (Å)	MscS settings	System	MD settings	Reference
16.5	2OAU, sensor paddle aligned with own chain, compacted paddle-TM3 cavities, simulating only residues 1-130, 137-140, 147-154, 162-175	POPC bilayer	4 ns AT MD, 500 bar*nm surface tension plus steered MD pull of the sensor paddle peripherally towards previously published, computationally designed structure (see [Anishkin et al., 2008b])	Belyy et al. [2010a]
16	Residues 1–130, 137–140, 147–154, and 162–175, with soft α carbon restraints of the C-terminal residues	POPC bilayer, 200mM salt	15ns AT MD, 100 bar*nm surface tension plus steered MD of the full protein towards sevenfold symmetric positions	Anishkin et al. [2008a]
15.7	Full original (unrefined) closed crystal structure (PDB id 1MXM)	POPC bilayer, 200mM salt	5ns AT MD at 400 bar*nm surface tension followed by steered MD (20pN) applied radially outwards to Cα of TM3a	Sotomayor et al. [2006]
13	No restraints	POPC bilayer	10 ns AT MD at +1.2V	Sotomayor et al. [2007]
11.6	Full original (unrefined) closed crystal structure (PDB id 1MXM)	POPC bilayer, 200mM salt	16ns AT MD , 500 bar*nm surface tension followed by steered MD (40pN) applied radially outwards to the Cα of Tyr27, Val29, Arg46, Arg54, Arg59, Arg74, Arg128	Sotomayor et al. [2006]
7.5	No restraints	POPC bilayer	3.7 ns AT MD, 200 bar*nm surface tension	Sotomayor and Schulten [2004]
"No pore opening"	2OAU, sensor paddle aligned with own chain, compacted paddle-TM3 cavities, simulating only residues 1-130, 137-140, 147-154, 162-175	POPC bilayer	16ns AT MD, 100 bar*nm	Belyy et al. [2010a]
"No pore opening"	2OAU, sensor paddle aligned with own chain, compacted paddle-TM3 cavities, simulating only residues 1-130, 137-140, 147-154, 162-175	POPC bilayer	12ns AT MD, 500 bar*nm	Belyy et al. [2010a]

Table A.1.1.: Previously achieved pore diameters using different MD setups
MD studies, where an objective was to increase the pore diameter using different stimuli, are shown in decreasing order of achieved diameter. ϕ_{Max} refers to the maximum achieved pore diameter, measured at the pore bottleneck. NB. Presence of hydrogen atoms and the use of van der Waals radii to compute inter-atomic distances will affect the exact diameter.

ϕ_{Max} (Å)	Tension (bar*nm)	MD settings / Notes	Reference
30-32	120 or 300	Used tension in addition to EPR-based steered MD	[Deplazes et al., 2012]
24	5,000	22ns	[Colombo et al., 2003]
12.4	720	338 K to accelerate MscL dynamics	[Yefimov et al., 2008]
11.6$\pm$0.8	670	16nm diameter DOPC liposome system.	[Louhivuori et al., 2010]
8	770	Further pore expansion caused disintegration of the TM domain	[Yefimov et al., 2008]
6	600	100 ps AT MD in vacuum	[Gullingsrud et al., 2001]
4	730	Custom CG particles for the protein	[Yefimov et al., 2008]
"No pore opening"	$\leq$ 600	Custom CG particles for the protein	[Yefimov et al., 2008]
"No pore opening"	300	POPC bilayer	[Deplazes et al., 2012]
"No pore opening"	120	POPC bilayer	[Deplazes et al., 2012]

Table A.2.1.: Previously achieved pore diameters in MscL using different MD setups
MD studies, where an objective was to increase the pore diameter using different stimuli, are shown in decreasing order of achieved pore diameter. ϕ_{Max} refers to the maximum achieved pore diameter, measured at the pore bottleneck. NB. Presence of hydrogen atoms and the use of van der Waals radii to compute inter-atomic distances will affect the exact diameter.

B. Appendix: Statistics concepts

Background to and description of the statistical inference techniques that are implemented in InterQuant (see Chapter 4).

Statistical resources used

Rather than citing multiple books throughout this appendix, I wish to acknowledge the following sources for the fundamental statistical methodologies and equations used:

"Biostatistical Analysis, 5th edition" by Jerrold H. Zar (Pearson 2010)
"Statistics in a nutshell: A desktop quick reference" by Sarah Boslaugh (O'Reilly 2013)
"Elementary statistics, 11th edition" by Mario F. Triola (Addison-Wesley 2012)
"Schaum's Outline of Statistics, 5th Edition" by Murray Spiegel (McGraw-Hill Professional 2014)

B.1. Sample probability distributions and their relationship to the population

In statistics, the *population* is the complete set of all observations. In reality, it is not always possible to measure the full population, as explained in the chapter introduction. Instead, we use a *sample*; a small, random subset of the population, for which we compute a *statistic*. A *probability distribution* of a sample's values for a statistic is used to infer something about the population *parameter*; the population equivalent of the statistic. The graphed probability density shows probability as a function of values for a given statistic, and the total area under the curve is 1. Most datasets have distribution curves that are bell-shaped and centered about the the statistic's average value. Because it is so common, this distribution shape is called 'normal'. The standardised normal distribution is centered about 0.

If all observations were accessible for a normally distributed population, the curve peak would indicate the parameter's true mean. *Statistical inference* is the study of the probabilistic relationship between samples and their unknown population. Inference methods often rely on knowledge about the underlying population's distribution shape, and this can be derived form large sample sizes. For small sample sizes, one must assume that data come from a population of a certain distribution shape. One infers knowledge about the population parameter by investigating the spread of its samples' distributions.

B.2. Normally distributed data spread: Variance and standard deviation

Variance and standard deviation (SD) are measures of the spread of a distribution of observations. Each observation's deviation from the parameter mean is squared, and the average of all such squared deviations is the population variance.

$$\sigma^2 = \frac{\sum_{i=1}^{N} \left[(x_i - \mu)^2\right]}{N}$$

The *sample variance* is higher than the *population variance*; rather than dividing by the sample size N when computing the average, one divides by $N - 1$.

$$s^2 = \frac{\sum_{i=1}^{N} \left[(x_i - \bar{x})^2\right]}{N - 1}$$

SD is the square root of variance. $\pm 1SD$ contains approximately $\frac{2}{3}$ of all observations (68.26%). The *sample SD* uses the sample variance rather than the population variance, so shows a higher spread than the population SD.

$$\sigma = \sqrt{\sigma^2}$$

$$s = \sqrt{s^2}$$

B.3. Sampling theory

B.3.1. Sample size

A small sample is one where there are not enough observations to determine its underlying distribution [Student, 1908]. It is commonly defined as consisting of less than 30 observations. This cutoff is set by the *central limit theorem*, which states that if the means for any statistic are pooled from 30 or more samples to give the *sampling distribution* for that statistic, the sampling distribution will have a Normal shape. The shape will be Normal irrespective of the shape of the underlying individual samples' distributions. The ability to be able to rely on the distribution curve to be Normal allows population parameters to be estimated reliably. However, if the underlying samples' distributions are known to be normal, an even sample size smaller than 30 will still return a normal sampling distribution, and the sample can be used for statistical inference about its population parameters. The t-test, discussed below, makes use of this. Unless one is certain that the underlying distribution is normal for the metric of interest, one should aim to collect 30 or more observations to invoke the central limit theorem.

B.3.2. How to pick observations: averaging over time is good statistical practice

Parameter inference methods operate on the assumption that the observations of a sample are randomly picked and independent. It is therefore incorrect to use single frames of a trajectory as individual observations. IQ statistics that are not explicit time-series are therefore averaged over the full trajectory. An average value per simulation is better practice, and this average could be over a full simulation or over select frames from the most relevant part of a simulation, e.g. once interactions of interest are equilibrated. The exception to this is when statistic dynamics over time is sought, in which case time-series are explicitly required.

B.4. The t-test

The t-test family of analyses facilitates statistical comparison between samples. It is the most commonly used method for hypothesis testing about means [Zar, 2010]. The t-test was invented in 1908 by William Sealy Gosset, publishing under the pseudonym 'Student', thus the alternate name 'Student's t-test'. Gosset developed the method so that it would handle in particular small sample sizes, but it is equally applicable to larger data sets. It assumes no knowledge of the population other than that it is normally distributed [Lehmann, 1999], which is a safe assumption in most cases [Boslaugh,

Symbol	Meaning
x	Sample observation
$\bar{x}$	Sample mean
μ	Population mean
s^2	Sample variance
σ^2	Population variance
s	Sample SD
σ	Population SD
N_{sample}	Sample size
$z_{\bar{x}}$	Samples' distribution statistic, where z = any statistic
α	Significance
df	Degrees of freedom
$H0$	Null hypothesis
d	Difference between samples
$\bar{d}$	Mean difference between samples

Table B.3.1.: Statistical notation used in this thesis.
When a symbol has a subscript number, this refers to the sample that it belongs to.

2013]. However, it is relatively robust to non-normality; only large deviations from normality invalidates its use [Student, 1908, Lehmann, 1999]. This makes the t-test a good choice for statistical treatment of MD data, and it has previously been successfully used for this purpose in individual studies [Likic et al., 2005].

Using sample statistics in the place of population parameters, sample data are fitted onto a second dataset - whether a potential parameter or another sample - using the t distribution, which is a unimodal curve with total area 1, centered about 0, just like the standardised normal probability distribution. However, the t distribution's shape depends on the degrees of freedom (df) of the data, which is correlated with the sample size. At low df, the t distribution's tails are longer and the peak is lower to accommodate the increased tail area to give a total area of 1. At high df, the t distribution shape approaches the normal distribution. The difference between the two means of the samples, divided by the standard error of the mean (a measure of variability that incorporates the sample size and spread of observations) generates a *t-value.* The probability of there being no statistically significant difference between the samples is computed using this t-value and df. df adjusts the probability distribution so that the tails have the correct area, and the tail area beyond t is the *p-value*; the probability that this sample discrepancy could be obtained by chance.

B.4.1. One- and two-tailed t-tests

The t-distribution is symmetric about $t = 0$, and in a *two-tailed t-test*, both tails are used and t can be either positive or negative. The sign of t denotes which sample has the larger value for the queried statistic. In a *one-tailed t-test*, only the positive half of the t distribution is used to determine whether samples are different with any statistical significance; it does not reveal which sample has the higher value. InterQuant implements two-tailed t-tests, so one-tailed t-test settings will not be discussed further.

B.4.2. Critical values, statistical significance and the Null hypothesis

For a given number of df , the tail area shifts and so does the t-value that contains a particular area. The t-values corresponding to areas 0.05, 0.01 and 0.005 are examples of *critical values* of t; if the computed t value is beyond that critical value, sample discrepancy is considered statistically significant. The level of the critical value is up to the user, but 0.05 is commonly used.

The p-value, tail area and α, the significance level, are all used relatively synonymously to refer to the probability of rejecting the *Null hypothesis* $(H0)$ when it is in fact true. For example, if a computed t corresponds to $\alpha = 0.05$, that difference between sample means is likely to occur by chance 5% of the time. In a t-test, the null hypothesis states that the two inputs originate from the same population; i.e. that there is no difference between their underlying parameters. If $|t \geq t_{critical}|$, $H0$ is rejected in preference for the alternative hypothesis that the difference between the samples *is* statistically significant. So a higher absolute t value indicates higher significant difference between means.

Ultimately, statistical significance does not necessarily equal practical significance. Only the user can decide whether a statistically significant difference is of practical value for the case at hand.

B.4.3. Confidence Interval

The confidence interval (CI) is the precision of any computed statistic relative to the true population parameter. A wide confidence interval signifies that there is a high probability of retrieving a very different sample mean if a new sample is generated, compared to the sample that the CI was computed for. A narrow CI, on the other hand, signifies that all samples that are drawn ought to have relatively similar sample means.

The 95% CI is the most commonly used interval [Boslaugh, 2013], and means that if one collected infinite numbers of samples and computed their CIs, 95% of these CIs would contain the true population mean. The 95% CI will have a narrower spread about the statistic's mean than the 99% CI for the same sample, since the 99% CI indicates more certainty that the actual population parameter lies within its bounds.

Equation for the prediction of a population parameter

Given a chosen level of confidence, the CI for the population mean is given by :

$$CI = \overline{x} \pm \frac{t_{0.5\alpha,df} * s}{\sqrt{N}}$$

where $t_{0.5\alpha,\,df}$ is the upper critical value for t at half the chosen confidence level, using the relevant df (retrieved from tables), and

$$df = N_{samples} - 1\,.$$

B.4.4. One-sample t-test

The one-sample t-test is used when one wants to compare a sample to a pre-specified reference population. This t-test computes whether the sample belongs to the reference population.

One-sample t-tests are useful for comparison between crystallographic and MD data. When protein crystal structures co-purify with unidentified non-protein electron density, MD of the experimental setup can attempt to resolve this electron density's identity. As exemplified in Chapter 5, IIQ provides a means of retrieving statistics for protein interaction with non-protein system parts across multiple MD simulations, and to compare those results to the crystal structure data. The crystal structure is therefore treated as the reference parameter that the MD sample statistics are compared against. Insignificant differences between the sample and reference support that MD has elucidated the source of the electron density.

How to calculate t in a one-sample t-test

$$t = \frac{\overline{x} - \mu_0}{\frac{s}{\sqrt{N_{sample}}}}$$

where μ_0 is the known reference population parameter, and sample SD and degrees of freedom are computed as

$$s = \sqrt{\frac{\sum(x - \overline{x})^2}{N_{sample} - 1}}$$

$$df = N_{sample} - 1\,.$$

The sign of t reflects whether the reference or the sample has the highest mean value (the same is true for the two-sample t-tests, but samples are compared, no reference). However, the statistical significance that the value of t corresponds to is absolute; e.g. a t-value of -5 has the same significance as 5.

Confidence interval for the one-sample t-test The lower and upper boundary of the CI for the population mean is computed by

$$CI = \bar{x} \pm \frac{s * t_{0.5\alpha,\,df}}{\sqrt{N}}\,.$$

B.4.5. Two-sample t-test

The two-sample t-test determines whether two samples belong to the same population. Samples can be *independent* or *dependent* (also referred to as *paired* or *matched* samples), and requires the *independent t-test* and the *dependent t-test*, respectively. Independent samples are unrelated; an observation from the first sample does not influence observations in the second sample. If one sample does influence the outcome of the other sample, e.g. if the second sample was obtained by subjecting the first sample to different conditions, the samples are dependent. An MD example of dependent samples is analysis of protein-lipid interactions in the same set of simulations with and without protein Cα restraints. Dependent samples have necessarily equal sample sizes, while the sample size for independent samples can vary between samples. Two-sample t-test results are more reliable the more equal sample sizes are, and the larger the sample size.

The two-sample t-test for independent samples

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2 \left(\frac{1}{N_{sample,1}} + \frac{1}{N_{sample,2}}\right)}}$$

where s_p^2 is the pooled variance; an estimate of the two samples' common variance, which equals

$$s_p^2 = \frac{(N_{sample,1} - 1)\, s_1^2 + (N_{sample,2} - 1)\, s_2^2}{df}$$

and

$$df = N_{sample,1} + N_{sample,2} - 2\,.$$

Confidence interval for the difference between the true population means of two independent samples

$$CI = (\bar{x}_1 - \bar{x}_2) \pm \sqrt{s_p^2 \left(\frac{1}{N_{sample,1}} + \frac{1}{N_{sample,2}}\right)} * t_{0.5\alpha,\,df}$$

where s_p^2 is the pooled variance, calculated as shown previously. If no significant difference is found between two samples, the CI for the difference between population means will include 0.

The two-sample t-test for dependent samples

Dependent (or paired) samples have necessarily the same sample size, and

$$t = \frac{\bar{d}}{\frac{s_d}{\sqrt{N_{pairs}}}}$$

where $\bar{d}$ is the mean of the difference scores, and N_{pairs} is the number of pairs of observations:

$$\bar{d} = \frac{1}{N_{pairs}} \sum_{i=1}^{N_{pairs}} (x_{i,1} - x_{i,2})$$

$$N_{pairs} = \frac{1}{2}(N_{sample,1} + N_{sample,2})$$

s_d is the standard deviation of the difference scores:

$$s_d = \sqrt{\frac{\sum(d - \bar{d})^2}{N_{pairs} - 1}}$$

and dependent samples' degrees of freedom are

$$df = N_{pairs} - 1\,.$$

Confidence interval for the difference between the means of populations of dependent samples The lower and upper boundary of the CI is computed by

$$CI = \bar{d} \pm \frac{s_d * t_{0.5\alpha,\,d.f.}}{\sqrt{N_{pairs}}}$$

The equation is very similar to the CI equation for the one-sample t-test.

B.4.6. Welch's t-test for independent samples with unequal sample variances

A sample's variance is the spread of its observation values. When two samples are compared, their inferred population distributions are compared, and their fit determines whether the samples are statistically different. This fit does not accommodate differences in variance, as both samples' variances are pooled in the standard t-test. The t-test therefore operates on the assumption of *homogeneity of variance*, i.e. that variance in one sample approximately equals the variance of the other sample. If samples have unequal variances; *heterogenous variances*, the risk of erroneous conclusions about samples increases and gives rise to the Behrens-Fisher problem [Boslaugh, 2013].

Homogeneity of variance can be tested by Levene's Test, the Brown-Forsythe test, Bartlett's test and Hartley's F-test. In the event that the variances of independent samples are found to be unequal, their means should be compared using Welch's t-test, which accommodates heterogenous variances. Use of the Welch t-test reduces cases where $H0$ is wrongfully rejected [Zar, 2010].

Hartley's F_{max} test

To compute the variance heterogeneity, the ratio of the variances of the two samples, F, is first computed. The larger of the two variances are used in F's nominator, so

$$F = \begin{cases} s_1^2/s_2^2 & if\; s_1^2 > s_2^2 \\ s_2^2/s_1^2 & otherwise\,. \end{cases}$$

The F statistic is compared to the two-tailed critical value of the relevant F-distribution using $\alpha = 0.05$ and df , which is computed separately for the numerator and denominator using

$$df = N_{sample} - 1$$

[Hartley, 1950]. A larger ratio means a higher variance heterogeneity, which outputs a smaller α_F . If $\alpha_F > 0.05$, homogenous variance is accepted.

Welch's t-test

The major difference between equations for the t-test and the Welch t-test is that the latter does not pool variance. Rather, the Welch algorithm computes each variance separately and uses this to deduce both t and CIs [Zar, 2010].

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{s_1^2}{N_{sample,1}} + \frac{s_2^2}{N_{sample,2}}\right)}} .$$

The df is calculated according to the Welch-Satterthwaite equation:

$$df = \frac{\left(\frac{s_1^2}{N_{sample,1}} + \frac{s_2^2}{N_{sample,2}}\right)^2}{\frac{s_1^4}{N_{sample,1}^2\left(N_{sample,1}-1\right)} + \frac{s_2^4}{N_{sample,2}^2\left(N_{sample,2}-1\right)}} .$$

If df is a non-integer value, it is rounded down to the nearest integer value.

Confidence interval for the difference in means of the two underlying populations with heterogenous variance

$$CI = (\bar{x}_1 - \bar{x}_2) \pm \sqrt{\left(\frac{s_1^2}{N_{sample,1}} + \frac{s_2^2}{N_{sample,2}}\right)} * t_{0.5\alpha,\, df} .$$

B.4.7. Note on the reliability of the t-test for dependent samples

While homogeneity of variance can be tested for dependent samples, there is no Welch's t-test equivalent for dependent samples with heterogenous variances. Indeed, similar variances are often assumed for dependent samples [Morgan, 1939]. However, the other concern when using the t-test, sample-size, is not an issue for dependent samples, since dependent samples naturally have the same sample size. Equal sample sizes have been shown to increase t-test robustness to variance heterogeneity ([Howell, 2007] via [Zar, 2010]). Therefore standard t-test results from paired samples are relatively dependable.

C. Appendix: The problem of multiple tests

This section concerns the problem of performing multiple tests on the same two samples.

C.1. Overview of the problem

The p-value is only statistically valid as a statistical confidence measure when a single test is performed, e.g. when one residue is tested for change to separation between samples. When the same sample is used for several tests, the chance of false positives increase, so p needs to be adjusted [McDonald, 2014].

In practice, this means that when a population is t-tested against itself, $p = 0.05$ means that there is a 5% chance that the t-test will show that the means are significantly different, despite there being no difference. If instead the sample contains many data points that are used for testing, as in a protein where each residue is tested, more than 5 tests will show false positive results at the $p = 0.05$ level, even though the sample is tested against itself.

C.1.1. Application for reduction of false positives

P-value correction therefore becomes increasingly important for techniques that employ large-scale datasets, as is the case in microarray technology, MRI brain scans of blood flow, and in genomic or proteomic studies [McDonald, 2014]. To quote William Noble: "In a nutshell, the property that makes these experiments so attractive—their massive scale—also creates many opportunities for spurious discoveries, which must be guarded against" [Noble, 2009].

C.2. How to correct for multiple tests

The goal of multiple testing correction is to reduce the number of false positives.

C.2.1. The Bonferroni correction

The simplest adjustment is the Bonferroni correction (BC). Given a significance threshold α (e.g. 0.05) and a number of 'similar tests' performed on the same sample (explained below), N_{tests}, the new p-value significance threshold for significance at that α is given by

$$p_{BC} = \frac{\alpha}{N_{tests}}$$

[Dunn, 1959]. At $\alpha = 0.05$, BC allows one to be 95% sure that scores would not be observed by chance if $p \leq p_{BC}$.

C.2.1.1. When to use Bonferroni correction

BC is appropriate when

- few tests are done
- only one or two significant differences are sought
- a single false positive in a set of tests would be a problem

C.2.1.2. Issues with the Bonferroni correction

In tests with many plausible significant differences, BC may lead many false negatives. For example, in a study of lipid interaction of 500 residues between a system at 0 bar*nm and the same system at 600 bar*nm, dozens or hundreds of residues may show differential interaction levels. BC demands that $p \leq \frac{0.05}{500}$, i.e. $p \leq 0.0001$ in order to be significant at $\alpha = 0.05$, which would require very large differences in interaction. In such a case, one should decide on whether false positives are more important than false negatives, and consider alternative correction procedures.

A second issue with BC is the definition of 'similar tests'. For example, the two tests for significant differential separation to lipid for residues A and B would belong to the same "family" of tests. By comparison, a test for protein diffusion should probably not be grouped with separation to lipid of residue A. But what tests are similar is debatable, and no firm rules exist. In an extreme example of this, published significant differences need to be withdrawn if the same sample is used to test another variable for another publication, where it can be argued that the variable belongs to the same family as previously published tests [McDonald, 2014].

C.2.2. False discovery rate estimation: an alternative to the Bonferroni correction

The Benjamini-Hochberg procedure identifies the proportion of "discoveries" (significant results) that are false positives. The user decides on a *false discovery rate*, $rate_{FDiscovery}$, the % of discoveries that are deemed ok to be false positives, e.g. 10%. McDonald [2014] notes that it is a common error is to confuse $rate_{FDiscovery}$ with the significance threshold α, and argues that it is wrong to set $rate_{FDiscovery}$ to 0.05 for many experiments as it is too low.

Original p-values are ordered from smallest to largest and ranked starting with rank 1 for the smallest p-value, and finishing with rank N, the number of tests, for the largest p-value. Then the Benjamini-Hochberg critical value, BH_c is computed per p-value i as follows

$$BH_{c,i} = \left(\frac{rank_i}{N}\right) rate_{FDiscovery}$$

The highest rank p-value that is smaller than its corresponding $BH_{c,i}$ is noted, and this test and all tests with lower rank are deemed significantly different, irrespective of how they compare to their respective $BH_{c,i}$.

The Benjamini-Hochberg procedure is less strict than BC, and preferable to BC when dealing with large data sets. The Benjamini-Hochberg algorithm is most applicable when identified significant discoveries can be supported with additional follow-up tests conducted on a new sample [Noble, 2009, McDonald, 2014].

C.2.3. Alternative correction techniques

Other techniques exist for reducing false positives in multiple t-tests. Storey [2002] takes into account the distribution of p-values by binning them and noting when the result graph tails off at high p-values. This number of p-values is used as a cutoff to determine the ratio of low p-values to discard. No false discovery rate needs to be set, which is a benefit over the Benjamini-Hochberg procedure.

One could also take into consideration the fact that, depending on the metric sought, results for residues A and B may be dependent if these residues are located adjacent to one another. A similar situation arises in microarray data, where gene networks produce expression dependencies. False discovery rate is then handled by a method developed by Reiner et al. [2003].

C.3. When to correct for false positives

The techniques discussed above may increase the number of false negatives, to different degrees. The cost of a false negative should be compared to that of a false positive before proceeding with

either technique. Generally, one should be cautious and do additional tests on another sample to substantiate any claim. On the contrary, it has also been argued that if when there is strong support for expecting that a t-test result is true, correcting for multiple tests is unnecessary [Rothman, 1990].

D. Additional MscS MD data

D.1. The kink in TM3 does not change magnitude or position as a function of state

Since the angle of the pore-lining helix TM3 has been associated with the MscS gating mechanism, helix geometry was analysed in all three structures. A qualitative overview of the full protein is seen in Fig. D.1.1. Note the high degree of curvature that appears half-way along TM3. Adjacent to this is where the hydrophobic gate is located, composed of Leucines 105 and 109. These, and other residues involved in the helix kink are shown in the TM3 close-up in Fig. D.1.2. The angles along TM3 were measured qualitatively using bendix, and values for all chains, per structure, are shown in Fig. D.1.3. All crystals display similar angle values and location of the kink. This confirms earlier qualitative geometric analyses of the closed and open states [Wang et al., 2008]. Note that angle deviation across helices is higher in the closed structure compared to the open structures. This is probably related to the unsymmetric pore of the closed crystal (c.f. Fig. 5.1.1).

D.2. Lipid preference per residue: detailed discussion

D.2.1. Phosphates of either POPE or POPG

Common strong contacts in both the closed and open state Tyr27, Arg46 and Arg54 had strong contacts to lipid phosphates in both closed and open states. These residues are all located on TM1, the most peripheral of the TM helices.

Significant differences in contacts between the closed and open state Five residues were found to have significantly different interactions to phosphates between the closed and open state: Ser58, Arg59, Lys60, Thr64 and Gln149. Out of these, Thr64 and Gln149 only contacted phosphate groups for maximally 0.2 of the full trajectory, so are discarded as weak contacts. Ser58 (sample mean difference = -0.164 ± 0.039), Arg59 (sample mean difference = -0.278 ± 0.077) and Lys60 (sample mean difference = -0.227 ± 0.058) all contacted lipid phosphates significantly longer in the open state than they do in the closed state. These residues are located on the C-terminus of TM1 and on the linker region between TM1 and TM2.

Arg74 has insignificantly more average contacts to phosphates in the closed than open state, but only Arg74 of the closed state has strong contacts to phosphate, on average.

It is interesting to note that lipid headgroups can make contacts as far away from the membrane COM as the first β-domain of the cage region, even though such contacts are weak. Val148, Gln149 and Ile150 all form part of the first β-domain, and interact more strongly with lipid phosphate groups in the closed than open state (but only significantly so for Gln149).

D.2.2. Amine of POPE or hydroxyl groups of POPG

No common strong contacts Neither amine groups of POPEs nor hydroxyl groups of POPGs bind strongly to either state of MscS. The least weak common interactions to lipid non-phosphate headgroups are made by Tyr27 and Arginines 46, 54 and 59, whereof the strongest is Arg59 to the open state. These are all charged or polar residues along TM1, and they bind to non-phosphate lipid headgroups for approximately two fifths of the trajectory.

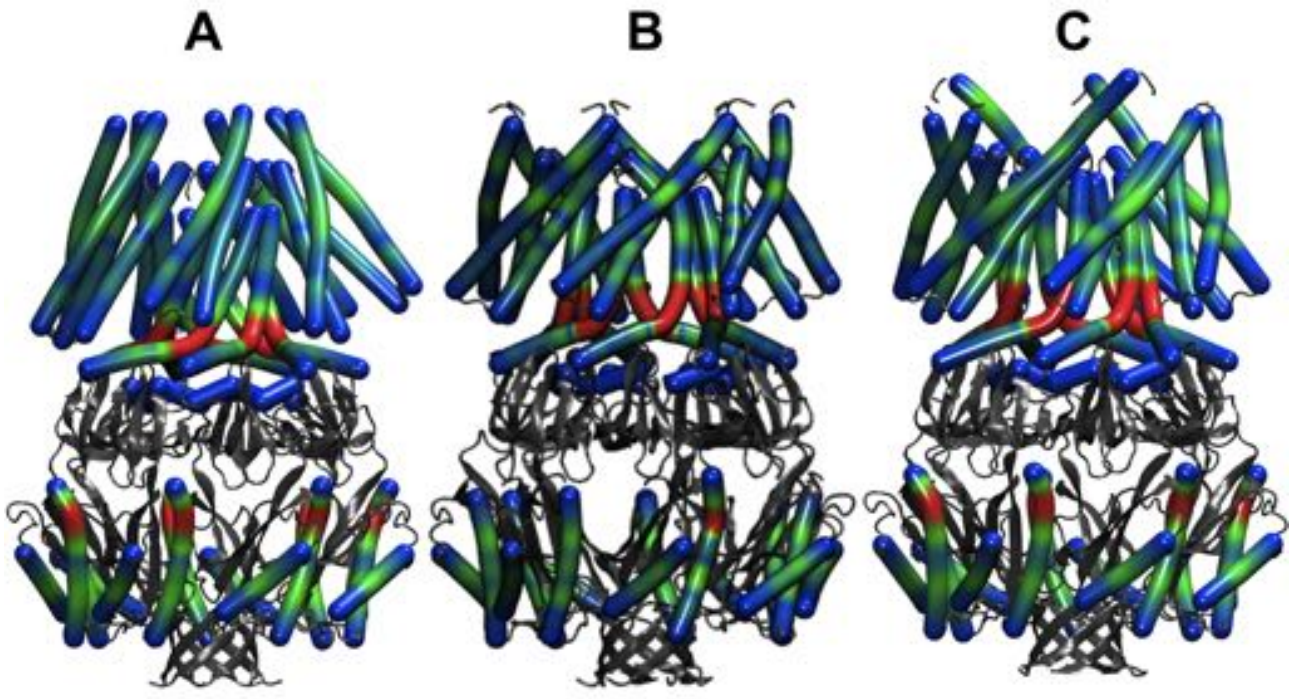

Figure D.1.1.: Available crystal structures showing helix angles.
Comparative helix angles across available crystal structures. The structures were rendered using Bendix with angle-indicative colour for helices. The heatmap used is BGR; blue indicates low curvature, red indicates high curvature. For quantitative angle measurements, please refer to the text. **A.** The closed structure with PDB id 2OAU. **B.** The open structure with PDB id 2VV5. STRIDE's helix assignment of TM3 into two separate helices was manually overridden in Bendix to evaluate the helix kink. **C.** The new open structure, yet unpublished.

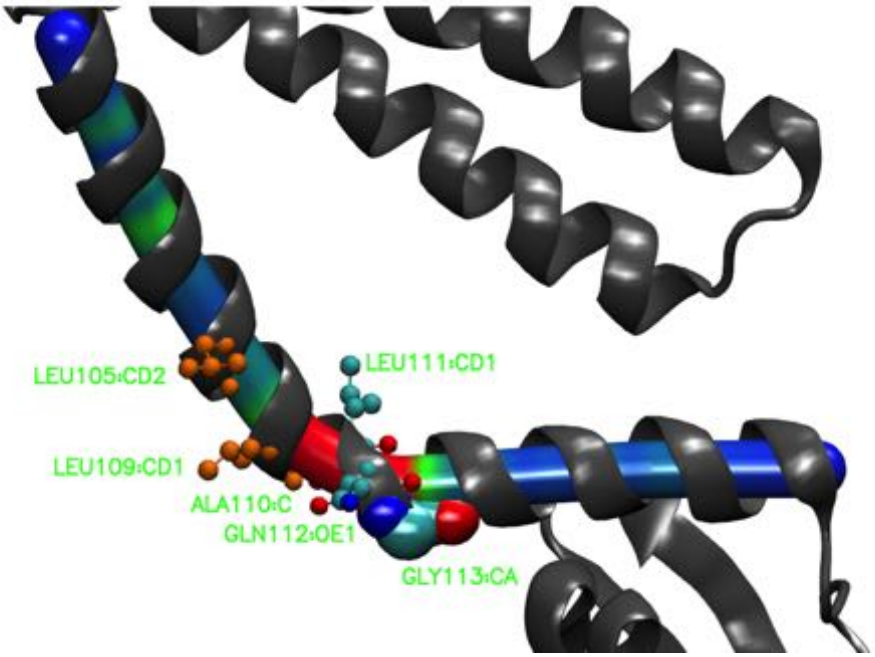

Figure D.1.2.: Close-up of the kink in the pore-lining helix.
View of the pore-lining helix of new open structure in both cartoon and angle-indicative bendix rendering. The hydrophobic gate, formed by Leucines 105 and 109, are shown in orange. Residues about the helix kink, residues 110 to 113, are shown in atom-type indicative color. Glycine 113, where the kink reportedly is, is enlarged. Residue and kink positioning in pore-lining helices of the other structures are highly similar (data not shown).

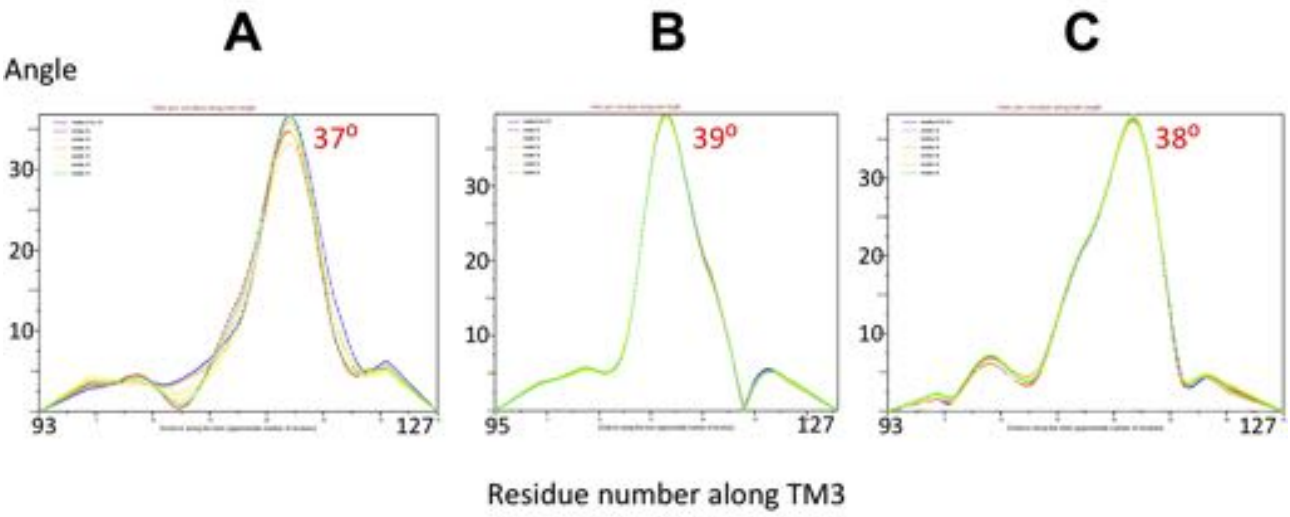

Figure D.1.3.: All MscS structures have the same degree of kink in the pore-lining helix.
Bendix analysis of the geometry of the pore-lining helices (TM3s) shows that all three crystal structures share approximately the same degree of kink at the same site. The peak angle is written out in red next to the peak. **A.** The closed structure. **B.** The published open structure. STRIDE's helix assignment of TM3 into two separate helices was manually overridden in Bendix to evaluate the helix kink. **C.** The new open structure

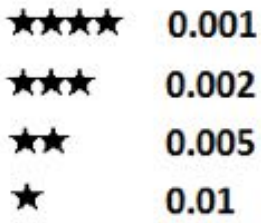

Figure D.2.1.: Star symbolism used to denote significance, p.

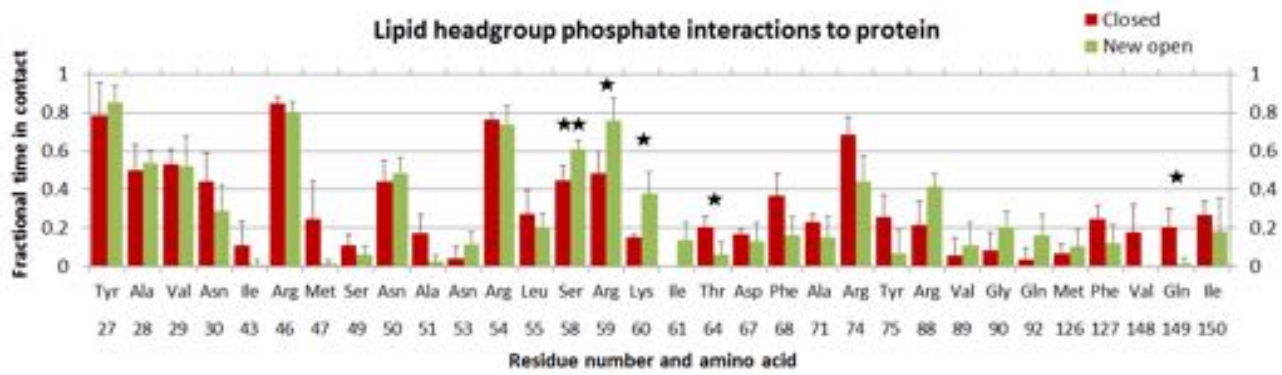

Figure D.2.2.: Differential interaction to lipid phosphates in closed and open state MscS.
Only residues that feature interaction with phosphate for ≥ 0.1 of the trajectories of at least one of the states are shown. Error bars denote +1 SD.

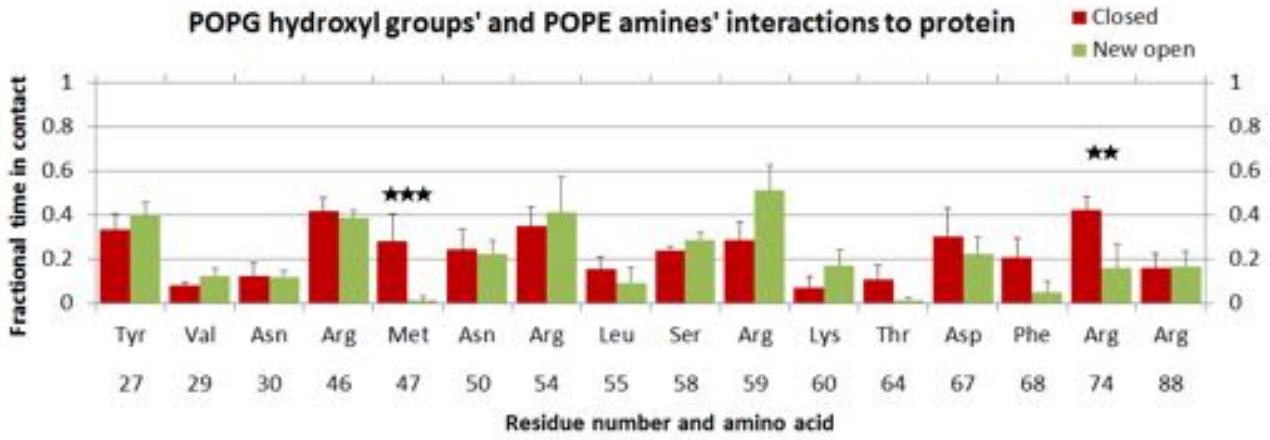

Figure D.2.3.: Differential interaction to lipid headgroup non-phosphates in closed and open state MscS.
Only residues that feature interaction with hydroxyl or amine for ≥ 0.1 of the trajectories in at least one of the states are shown. Error bars denote +1 SD.

Significant differences in interactions across states exist, but the interactions are weak MscS states show significant differences in binding to Met47 (sample mean difference = 0.272 ± 0.042) and Arg74 (sample mean difference = 0.265 ± 0.06). However, neither state binds non-phosphate lipid headgroups strongly. Out of the two, Arg74 of the closed state binds the least weak, on average throughout 0.4 of the trajectory.

D.2.3. The amine group of POPE

No common strong contacts Neither state has strong contacts to the amine group of POPE. The strongest amine binding is by Arg59 and Tyr27, that have an average contact time of less than a quarter of the simulation.

Significant differences in interactions across states exist, but the interactions are weak Only Leu55 (sample mean difference = 0.107 ± 0.03, p = 0.01) binds the POPE amine group significantly longer in the closed than open state. This binding is weak (on average circa 0.1 of the trajectory in the closed state) so is discarded.

D.2.4. The hydroxyl groups of POPG

No common strong contacts Neither state has strong contacts to POPG's hydroxyl groups. The strongest hydroxyl binding is by Arginines 46, 54 and 59, that have an average contact time of approximately 0.3 of the trajectory time.

Significant differences in interactions across states exist, but the interactions are weak Asp67 (sample mean difference = 0.1 ± 0.028, p = 0.01) and Arg74 (sample mean difference = 0.217 ± 0.05, p = 0.005) bind significantly tighter to POPG's hydroxyl groups in the closed than in the open state. Out of these two residues, Asp67 only binds hydroxyl for 0.1 of the closed state trajectory, so is disregarded as a very weak contact. Arg74 binds POPG hydroxyl groups throughout, on average, a third of the closed state trajectory, compared to a tenth of the open state trajectory.

D.2.5. Lipid preference per state

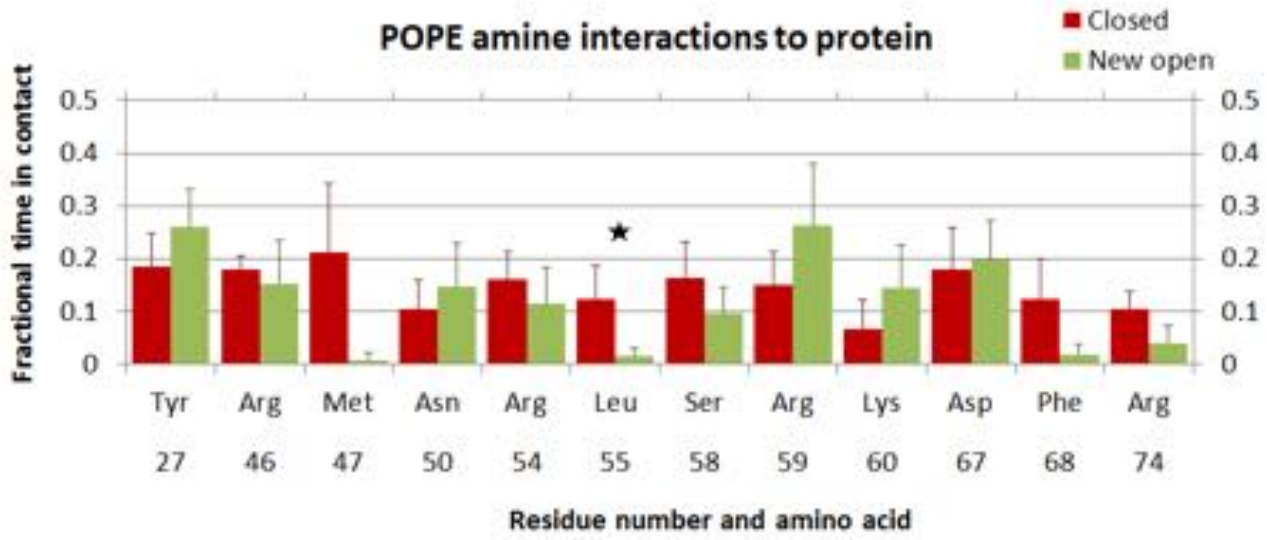

Figure D.2.4.: Differential interaction to POPE amine in closed and open state MscS.
Only residues that feature interaction with amine for ≥ 0.1 of the trajectories in at least one of the states are shown. Error bars denote +1 SD.

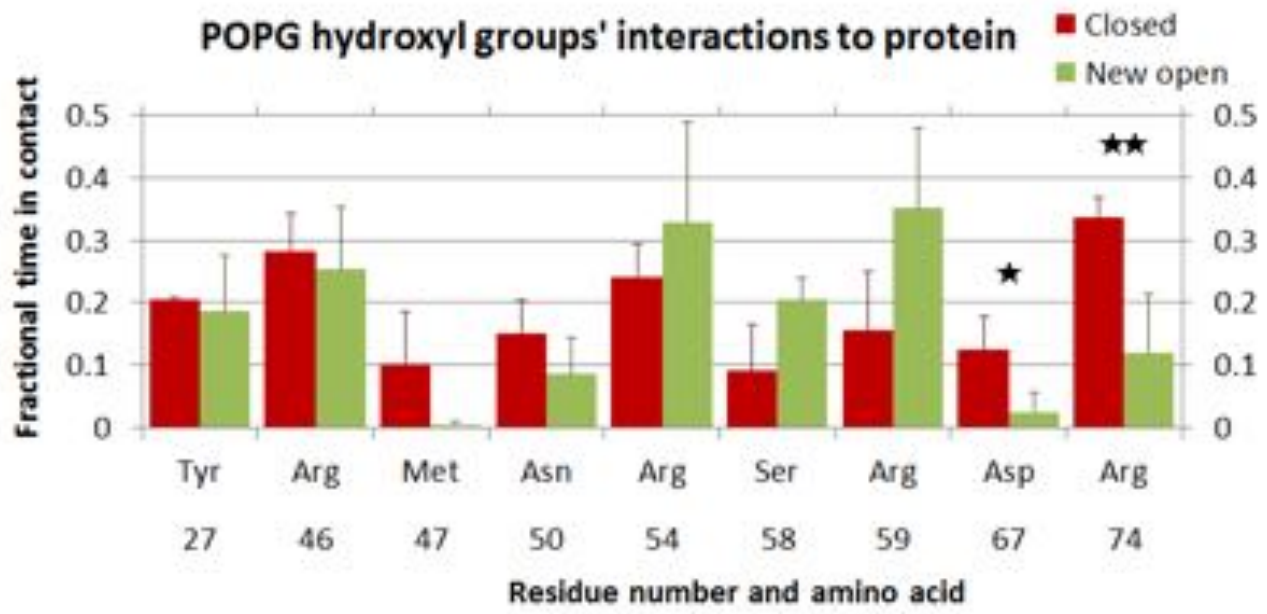

Figure D.2.5.: Differential interaction to POPG hydroxyls in closed and open state MscS.
Only residues that feature interaction with hydroxyl groups for ≥ 0.1 of the trajectories in at least one of the states are shown. Error bars denote +1 SD.

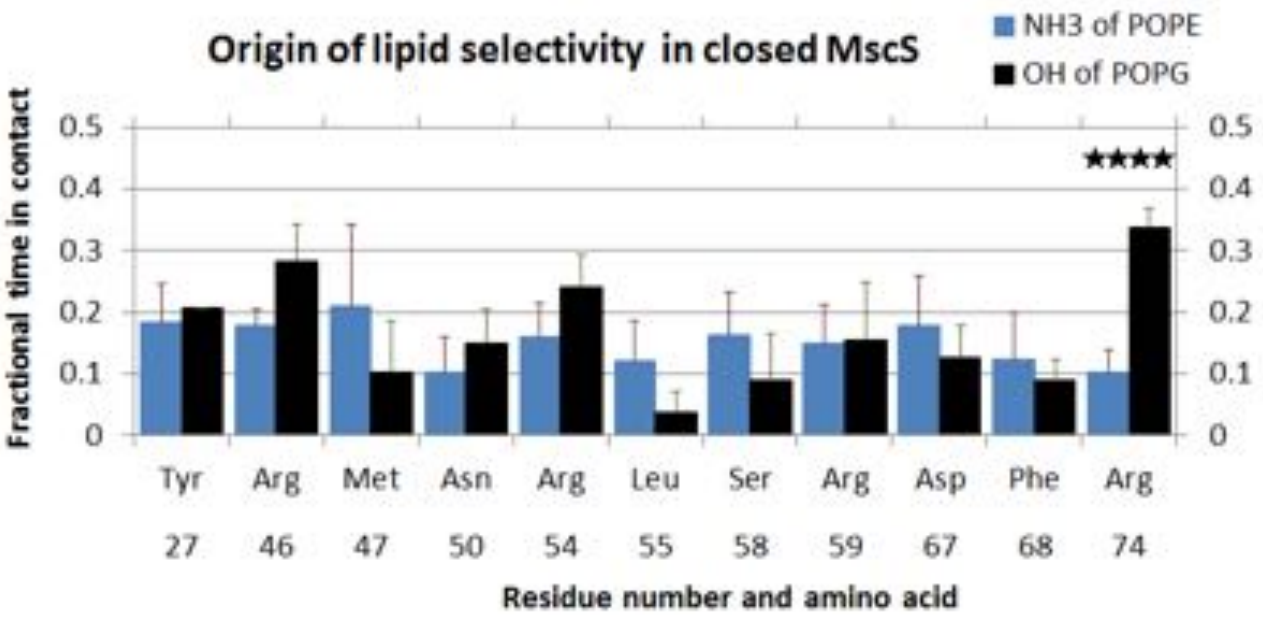

Figure D.2.6.: Differential interaction to non-phosphate lipid headgroups in closed state MscS. Only residues that feature interaction with hydroxyl groups or amine for ≥ 0.1 of the trajectories in at least one of the states are shown. Error bars denote +1 SD.

Closed state: Significant but weak preference for POPG Only one residue displays significant lipid selectivity between open and closed states: Arg74 (sample mean difference = -0.233 ± 0.024, p = 0.001). Although the selectivity is highly significant, binding is not strong (POPG is on average bound 0.35 of the trajectory). The selectivity of positive Arginine 74 for negative POPG makes sense from an electrostatic point of view, but this can not be the only contributing factor to the interaction since other arginines do not display significant selectivity (see e.g. Args 46, 54 and 59).

Open state: Significant but weak preference Two residues show significant selectivity between POPE and POPG: Ser58 (sample mean difference = -0.107 ± 0.027, p = 0.005) and Asp67 (sample mean difference = 0.173 ± 0.036, p = 0.002). Ser58 favours POPG and Asp67 favours POPE. However their binding is weak; they bind their respective lipid for a fifth of the trajectory maximally.

Arginines 46, 54, 59 and 74 all display stronger contacts to POPG's hydroxyl groups than they do to POPE's amine group, but there is high variation on contact time between trajectories, so neither of these residues show significant selectivity for either lipid.

D.3. Investigation into asymmetric gating: interaction between lipids and individual chains

Interaction between lipids and individual chains can be seen in Figures D.3.1 to D.3.3.

D.4. Key phosphate interaction dynamics in MscS gating without ENM

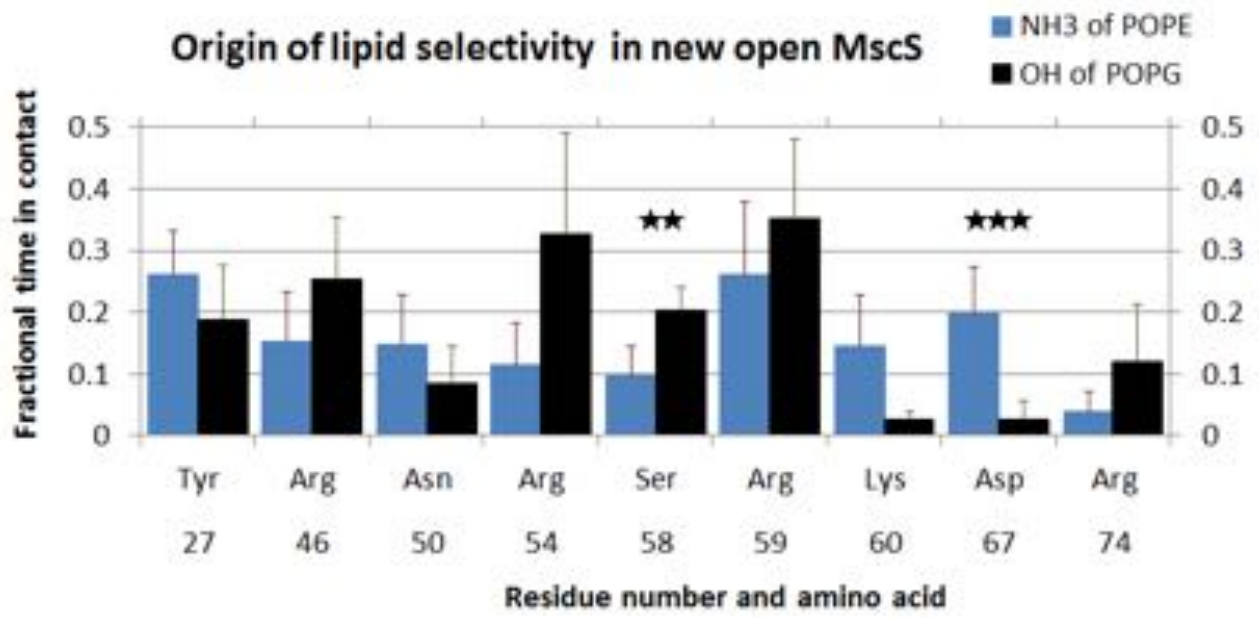

Figure D.2.7.: Differential interaction to non-phosphate lipid headgroups in open state MscS. Only residues that feature interaction with hydroxyl groups or amine for ≥ 0.1 of the trajectories in at least one of the states are shown. Error bars denote +1 SD.

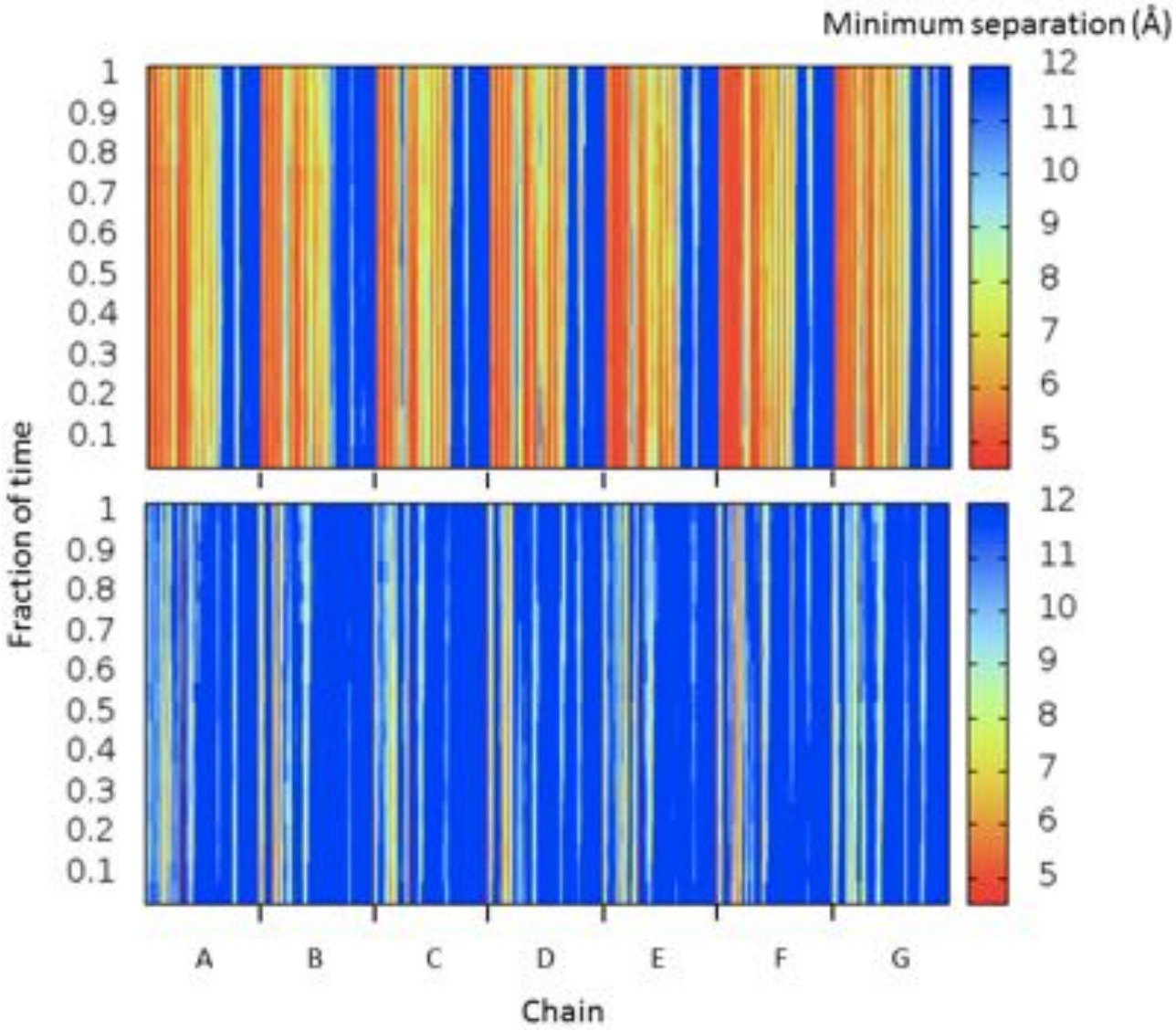

Figure D.3.1.: Separation between lipid and individual chains during gating

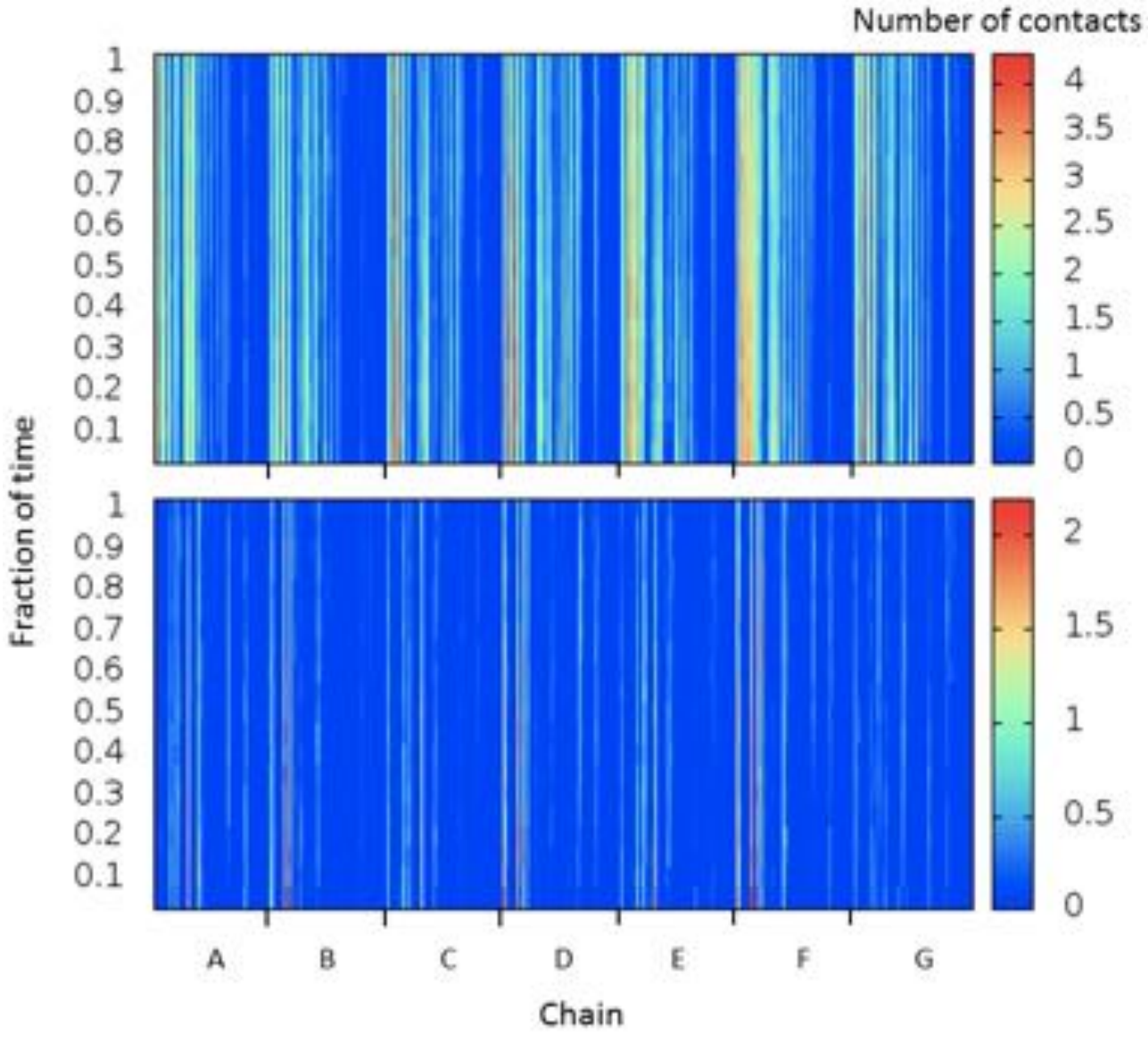

Figure D.3.2.: Contacts between lipids and individual chains during gating

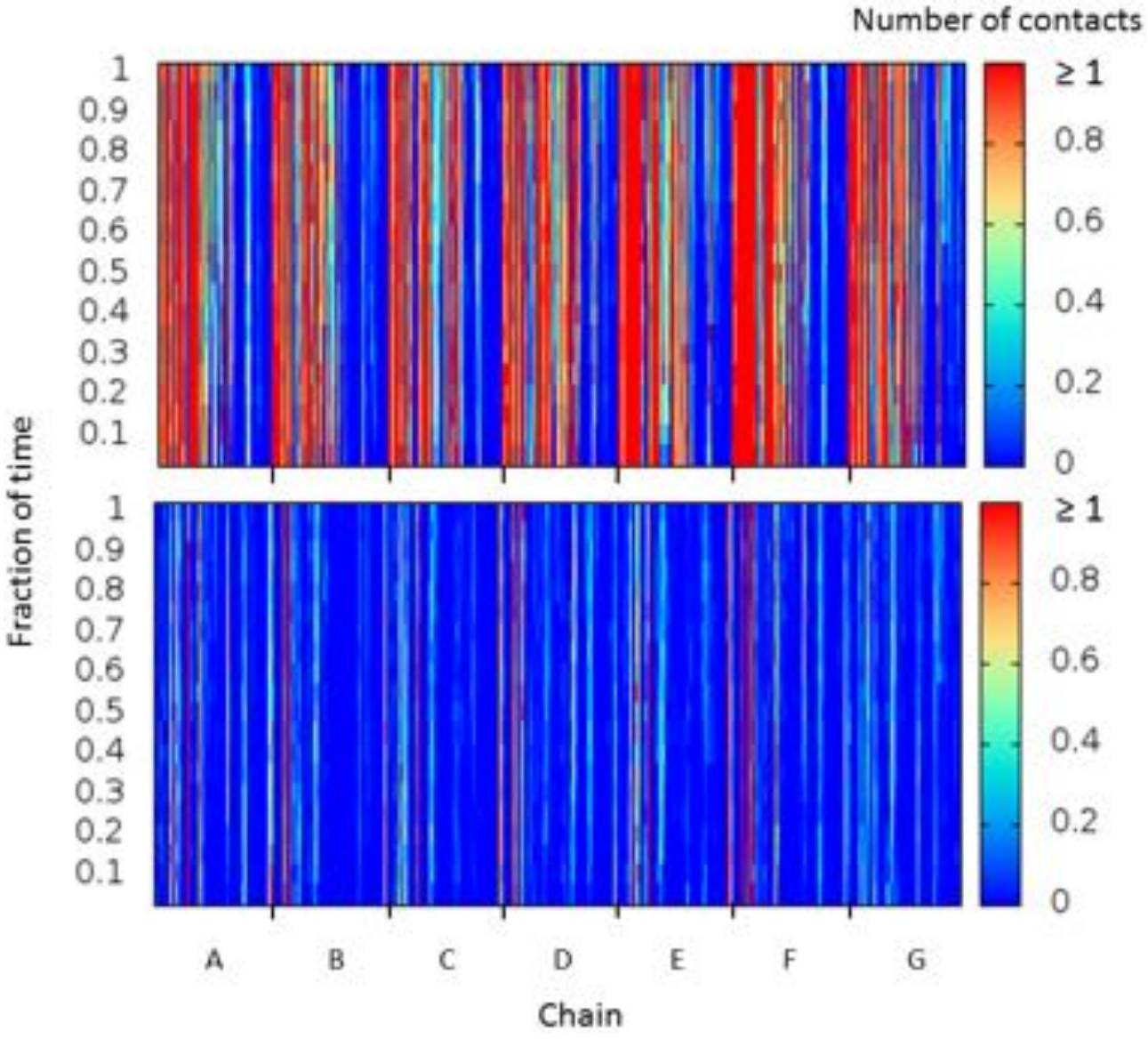

Figure D.3.3.: Contacts between lipids and individual chains during gating, rescaled

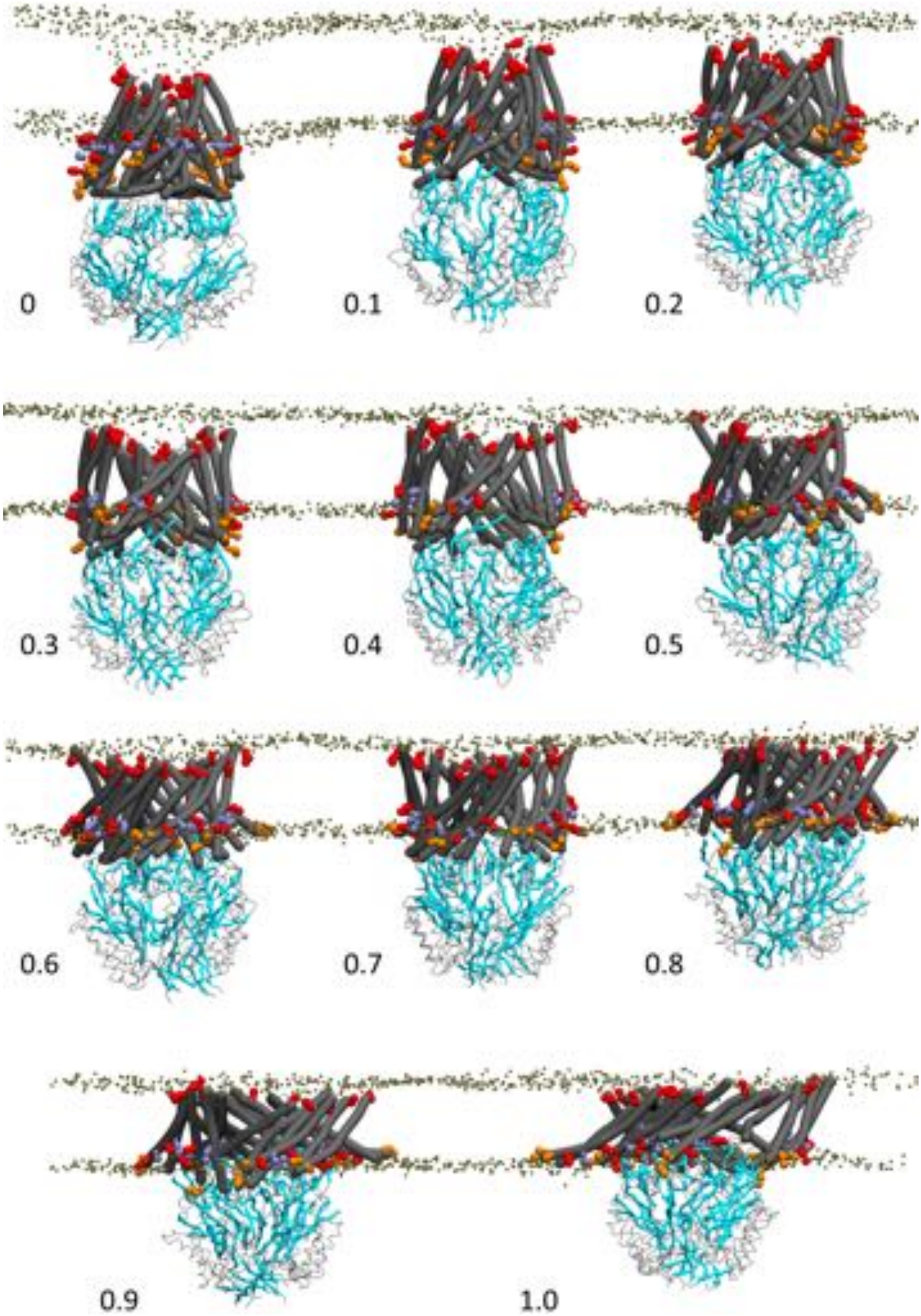

Figure D.4.1.: The slow gating mechanism of MscS: side view
The numbers refers to fractional time of the total simulation run. MscS TM domain as gray bendices and cage in the 'Join' backbone representation of Bendix, where β-sheets are represented by turquoise arrows. Tan spheres show lipid phosphate groups of the upper and lower bilayer leaflet. Key residues that make contact to phosphate groups during gating are shown in Van der Waal representations, coloured according to interaction dynamics. See text for details. Orthographic perspective.

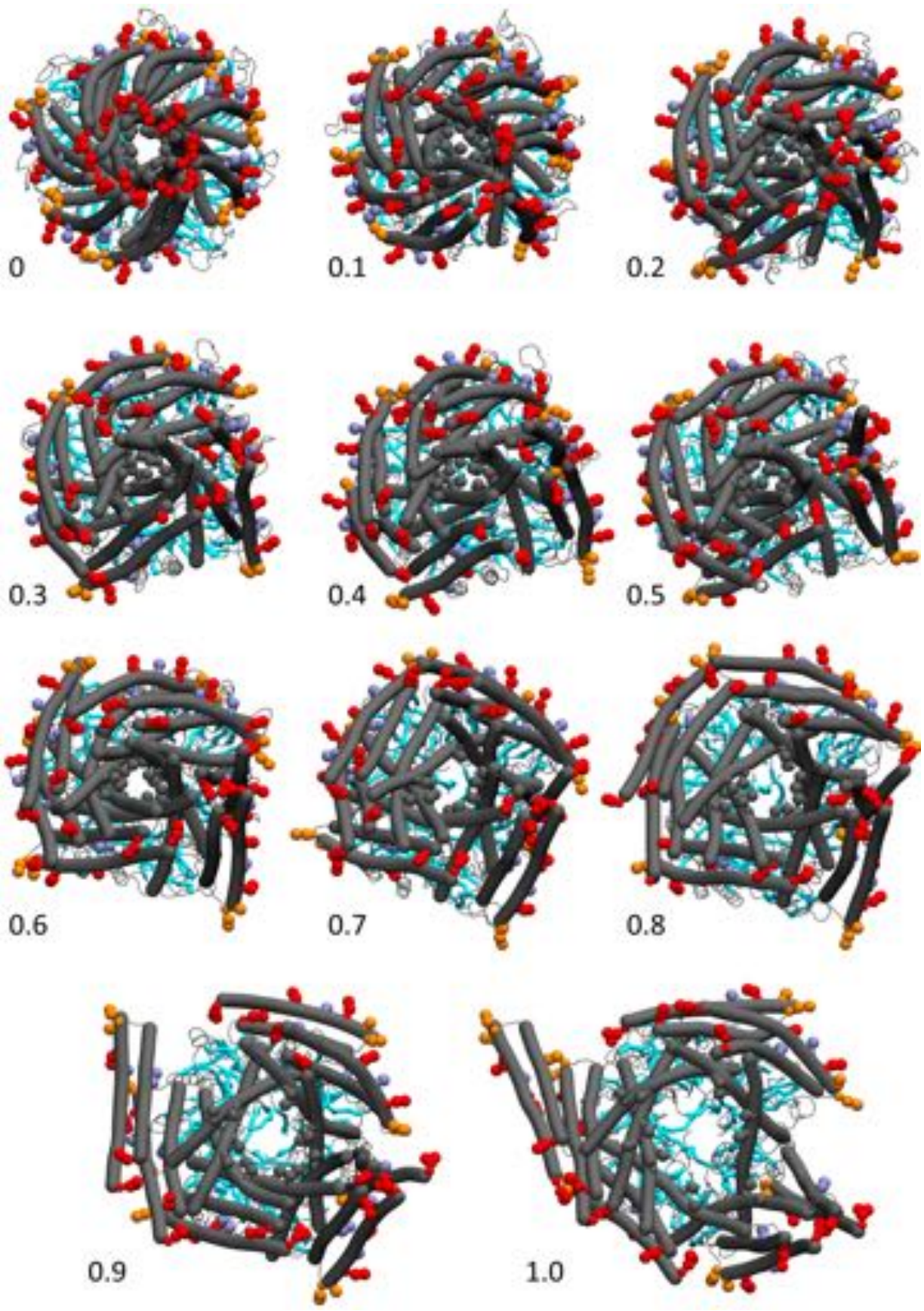

Figure D.4.2.: The slow gating mechanism of MscS: top view
The numbers refers to fractional time of the total simulation run. Same graphical representations as in Figure D.4.1, only no lipid phosphates are shown, and the leucines that form the hydrophobic gate are shown as gray spheres along the pore-lining helices. Orthographic perspective.

Bibliography

J. Abramson and E. M. Wright. Structure and function of na(+)-symporters with inverted repeats. *Curr. Opin. Struct. Biol.*, 19:425, 2009.

B. Akitake, A. Anishkin, and S. Sukharev. The dashpot mechanism of stretch-dependent gating in mscs. *J. Gen. Physiol.*, 125:143–154, 2005.

B. Akitake, A. Anishkin, N. Liu, and S. Sukharev. Straightening and sequential buckling of the pore-lining helices define the gating cycle of mscs. *Nat. Struct. Mol. Biol.*, 14:1141–1149, 2007.

A. Alam and Y. Jiang. High-resolution structure of the open nak channels. *Nat. Struct. Biol.*, 16: 30–34, 2009.

A. Anishkin and C. Kung. Stiffened lipid platforms at molecular force foci. *PNAS*, 110:4886–4892, 2013.

A. Anishkin and S. Sukharev. Water dynamics and dewetting transitions in the small mechanosensitive channel mscs. *Biophys. J.*, 86:2883–2895, 2004.

A. Anishkin and S. Sukharev. State-stabilizing interactions in bacterial mechanosensitive channel gating and adaptation. *J. Biol. Chem.*, 284:19153–19157, 2009.

A. Anishkin, B. Akitake, and S. Sukharev. Characterization of the resting mscs: modeling and analysis of the closed bacterial mechanosensitive channel of small conductance. *Biophys. J.*, 94:1252–1266, 2008a.

A. Anishkin, K. Karnaraju, and S. Sukharev. Mechanosensitive channel mscs in the open state: modeling of the transition, explicit simulations, and experimental measurements of conductance. *J. Gen. Physiol.*, 132:67–83, 2008b.

G. S. Ayton and G. A. Voth. Multiscale computer simulation of the immature hiv-1 virion. *Biophys. J.*, 99:2757–2765, 2010.

D. Balleza and F. Gomez-Lagunas. Conserved motifs in mechanosensitive channels mscl and mscs. *Eur. Biophys. J.*, 38:1013–1027, 2009.

M. Bansal, S. Kumar, and R. Velavan. Helanal - a program to characterise helix geometry in proteins. *J Biomol Struct Dyn.*, 17(5):811–819, 2000.

R. Baradaran, J. M. Berrisford, G. S. Minhas, and L. A. Sazanov. Crystal structure of the entire respiratory complex i. *Nature*, 494:443–448, 2013.

D. J. Barlow and J. M. Thornton. Ion-pairs in proteins. *J Mol Biol*, 168(4):867–885, 1983.

D. J. Barlow and J. M. Thornton. Helix geometry in proteins. *J Mol Biol*, 201:601–619, 1988.

R. B. Bass, P. Strop, M. Barclay, and D. Rees. Crystal structure of escherichia coli mscs, a voltage-modulated and mechanosensitive channel. *Science*, 298:1582–1587, 2002.

A. L. Beberg, D. Ensign, G. Jayachandran, S. Khaliq, and V. S. Pande. Folding@home: lessons from eight years of volunteer distributed computing. In *Proc. 2009 IEEE Intl Symp Parallel Distributed Processing (IPDPS 09) Washington, DC: IEEE Comp Soc*, 2009.

O. Beckstein and M. S. P. Sansom. The influence of geometry, surface character, and flexibility on the permeation of ions and water through biological pores. *Phys. Biol.*, 1:42–52, 2004.

O. Beckstein, P. C. Biggin, and M. S. P. Sansom. A hydrophobic gating mechanism for nanopores. *J. Phys. Chem. B*, 105:12902–12905, 2001.

V. Belyy, A. Anishkin, K. Kamaraju, N. Liu, and S. Sukharev. The tension-transmitting 'clutch' in the mechanosensitive channel mscs. *Nat. Struct. Mol. Biol.*, 17:451–458, 2010a.

V. Belyy, K. Kamaraju, B. Akitake, A. Anishkin, and S. Sukharev. Adaptive behavior of bacterial mechanosensitive channels is coupled to membrane mechanics. *J Gen Physiol*, 135:641–652, 2010b.

H. J. Berendsen, D. van der Spoel, and R. van Drunen. Gromacs: A message-passing parallel molecular dynamics implementation. *Comp. Phys. Comm.*, 91:43–56, 1995.

H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81:3684–3690, 1984.

C. Berrier, M. Besnard, B. Ajouz, A. Coulombe, and A. Ghazi. Multiple ms ion channels from e.coli, activated at different thresholds of applied pressure. *J Membr Biol*, 151:175–187, 1996.

I. Bertini, A Giachetti, C. Luchinat, G. Parigi, M. V. Petoukhov, R. Pierattelli, E. Ravera, and D. I. Svergun. Conformational space of flexible biological macromolecules from average data. *JACS*, 132:13553–13558, 2010.

F. Bezanilla and E. Perozo. Force and voltage sensors in one structure. *Science*, 298:1562–1563, 2002.

J. B. Birks. *Photophysics of aromatic molecules*. John Wiley & Sons, Ltd, London, 1970.

P. Blount and P. C. Moe. Bacterial mechanosensitive channels: integrating physiology, structure and function. *Trends Microbiol.*, 7:420–424, 1999.

P. J. Bond and M. S. P. Sansom. Insertion and assembly of membrane proteins via simulation. *J. Am. Chem. Soc.*, 128:2697–2704, 2006.

P. J. Bond, C. L. Wee, and M. S. P. Sansom. Coarse-grained molecular dynamics simulations of the energetics of helix insertion into a lipid bilayer. *Biochemistry*, 47:11321–11331, 2008.

S. Boslaugh. *Statistics in a nutshell: A desktop quick reference*. O'Reilly, 2013.

R. J. Britten and F. T. McClure. The amino acid pool of escherichia coli. *Microbiol. Mol. Biol. Rev.*, 26:292–335, 1962.

S. G. Brohawn, J. del Mármol, and R. MacKinnon. Crystal structure of the human k2p traak, a lipid- and mechano-sensitive k+ ion channel. *SCIENCE*, 335:436–441, 2012.

B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. Charmm: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*, 4:187–217, 1983.

G. Bussi, D. Donadio, and M. Parrinello. Canonical sampling through velocity rescaling. *J. Chem. Phys.*, 126:014101, 2007.

Z. Cao and J. U. Bowie. Shifting hydrogen bonds may produce flexible transmembrane helices. *PNAS*, 109:8121–8126, 2012.

M. Carson and C. Bugg. Algorithm for ribbon models of proteins. *J. Mol. Graph.*, 4:121–122, 1986.

D. A. Case, T. E. C. Darden, H. Gohlke, R. Luo, K. M. M. Onufriev, C. Simmerling, B. Wang, and R. J. Woods. The amber biomolecular simulation programs. *J. Comput. Chem.*, 26:1668–1688, 2005.

L. S. D. Caves, J. D. Evanseck, and M. Karplus. Locally accessible conformations of proteins: Multiple molecular dynamics simulations of crambin. *Protein Sci.*, 7:649–666, 1998.

M. A. Ceruso and H. Weinstein. Structural mimicry of proline kinks: tertiary packing interactions support local structural distortions. *J. Mol. Biol.*, 318:1237–1249, 2002.

G. Chang, R. H. Spencer, A. T. Lee, M. T. Barclay, and D. C. Rees. Structure of the mscl homolog from mycobacterium tuberculosis: A gated mechanosensitive ion channel. *Science*, 282:2220–2226, 1998.

A. Chattopadhyay and E. London. Parallax method for direct measurement of membrane penetration depth utilizing fluorescence quenching by spin-labeled phospholipids. *Biochemistry*, 26:39–45, 1987.

M. Chavent, T. Reddy, J. Goose, A. C. E Dahl, J. E. Stone, B. Jobard, and M. S. P. Sansom. Methodologies for the analysis of instantaneous lipid diffusion in md simulations of large membrane systems. *Faraday Discussions*, 169:1–20, 2014.

A. Cheng, A. N. van Hoek, M. Yeager, A. S. Verkman, and A. K. Mitra. Three-dimensional organization of a human water channel. *Nature*, 387:627–630, 1997.

S. Choe and M. Grabe. Conformational dynamics of the inner pore helix of voltage-gated potassium channels. *J. Chem. Phys.*, 130:215103, 2009.

B. Colombo, S.-J. Marrink, and M. S. P. Mark. Simulation of mscl gating in a bilayer under stress. *Biophys J*, 83:2331–2337, 2003.

B. Corry and B. Martinac. *Mechanosensitive ion channels*. Springer, 2008.

C. Cui and J. Adler. Effect of mutation of potassium-efflux system, kefa, on mechanosensitive channels in the cytoplasmic membrane of escherichia coli. *J. Membr. Biol.*, 150:143–152, 1996.

C. J. B. daCosta et al. A distinct mechanism for activating uncoupled nicotinic acetylcholine receptors. *Nat. Chem. Biol.*, 9:701–707, 2013.

A. C. E. Dahl and M. Thompson. *Comprehensive Biotechnology, 2nd edition.*, chapter Mechanobiology of Bone, pages 217–236. A Pergamon Title, Elsevier, 2011.

J. A. Dalton, I. Michalopoulos, and D. R. Westhead. Calculation of helix packing angles in protein structures. *Bioinformatics*, 19(10):1298–1299, 2003.

T. A. Darden, D. M. York, and L. G. Pedersen. Particle mesh ewald: An n log(n) method for ewald sums in large systems. *J. Chem. Phys.*, 98:10089–10092, 1993.

G. B. Dawe, M. Musgaard, E. D. Andrews, B. A. Daniels, M. R. P. Aurousseau, P. C. Biggin, and D. Bowie. Defining the structural relationship between kainate-receptor deactivation and desensitization. *Nat. Struct. Mol. Biol.*, 20:1054–1061, 2013.

E. Deplazes, M. Louhivuori, D. Jayatilaka, S. J. Marrink, and B. Corry. Structural investigation of mscl gating using experimental data and coarse grained md simulations. *PLoS Comp Biol*, 8:e1002683, 2012.

M. Dimitrov, J.-M. Alattia, T. Lemmin, R. Lehal, A. Fligier, J. Houacine, I. Hussain, F. Radtke, M. Dal Peraro, D. Beher, and P. C. Fraering. Alzheimer's disease mutations in app but not gamma-secretase modulators affect epsilon-cleavage-dependent aicd production. *Nature comm.*, 4, 2013.

C. Domene, D. A. Doyle, and C. Venien-Bryan. Modeling of an ion channel in its open conformation. *Biophys. J.*, 89:L01–L03, 2005.

R. O. Dror, R. M. Dirks, J. P. Grossman, H. Xu, and D. E. Shaw. Biomolecular simulation: A computational microscope for molecular biology. *Annu Rev Biophys*, 41:429–452, 2012.

O. J. Dunn. Estimation of the medians for dependent variables. *Ann. Math. Statist.*, 30:192–197, 1959.

M. D. Edwards, Y. Li, S. Kim, S. Miller, W. Bartlett, and S. et al. Black. Pivotal role of the glycine-rich tm3 helix in gating the mscs mechanosensitive channel. *Nat. Struct. Mol. Biol.*, 12:113–119, 2005.

M. D. Edwards, W. Bartlett, and I. R. Booth. Pore mutations of the escherichia coli mscs channel affect desensitization but not ionic preference. *Biophys J*, 94:3003–3013, 2008.

A. Elofsson and L. Nilsson. How consistent are molecular dynamics simulations? comparing structure and dynamics in reduced and oxidized escherichia coli thioredoxin. *J. Mol. Biol.*, 233:766–780, 1993.

Y. A. Ermakov, A. Z. Averbakh, A. B. Arbuzova, and S. I. Sukharev. Lipid and cell membranes in the presence of gadolinium and other ions with high affinity to lipids. 2. a dipole component of the boundary potential on membranes with different surface charge. *Membr. Cell. Biol.*, 12:411–426, 1998.

Y. A. Ermakov, K. Kamaraju, K. Sengupta, and S. Sukharev. Gadolinium ions block mechanosensitive channels by altering the packing and lateral pressure of anionic lipids. *Biophys J.*, 98:1018–1027, 2010.

U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. A smooth particle mesh ewald method. *J. Chem. Phys.*, 103:8577–8593, 1995.

E. Evans and V. Heinrich. Dynamic strength of fluid membranes. *C. R. Physique*, 4:265–274, 2003.

M. W. Evans. *Advances in Chemical Physics, Volume 81*. Wiley, 1992.

N. et al Fabelo. Severe alterations in lipid composition of frontal cortex lipid rafts from parkinson's disease and incidental parkinson's disease. *Mol. Med.*, 17:1107–1118, 2011.

K. M. Fagerbakke, S. Norland, and M. Heldal. The inorganic ion content of native aquatic bacteria. *Can J Microbiol*, 45:304–311, 1999.

P. L. Freddolino, A. S. Arkhipov, S. B. Larson, A. McPherson, and K. Schulten. Molecular dynamics simulations of the complete satellite tobacco mosaic virus. *Structure*, 14:437–449, 2006.

R. Gamini, M. Sotomayor, C. Chipot, and K. Schulten. Cytoplasmic domain filter function in the mechanosensitive channel of small conductance. *Biophys J.*, 101:80–89, 2011.

E. Giudice, A. E. Molza, Y. Laurin, A. Nicolas, E. Le Rumeur, and O. Delalande. Molecular clues about the dystrophin-neuronal nitric oxide synthase interaction: A theoretical approach. *Biochemistry*, 52(44):7777–7784, 2013.

E. Glaasker and B. Poolman. Regulation of compatible solute accumulation in bacteria. *Molecular Microbiology*, 29:397, 1998.

T. D. Goddard and T. E. Ferrin. Visualization software for molecular assemblies. *Curr. Opin. Struct. Biol.*, 17:587–595, 2007.

F. Guharay and F. Sachs. Stretch-activated single ion channel currents in tissue-cultured embryonic chick skeletal muscle. *J. Physiol.*, 352:685–701, 1984.

J. R. Gullingsrud and Schulten. Lipid bilayer pressure profiles and mechanosensitive channel gating. *Biophys. J.*, 86:3496–3509, 2004.

J. R. Gullingsrud, D. Kosztin, and K. Schulten. Structural determinants of mscl gating studied by molecular dynamics simulations. *Biophys. J.*, 80:2074–2081, 2001.

K. Gunasekaran, H. A. Nagarajaram, C. Ramakrishnan, and P. Balaram. Stereochemical punctuation marks in protein structures: glycine and proline containing helix stop signals. *J Mol Biol*, 275:917–932, 1998.

M. C. Gustin, X. L. Zhou, B. Martinac, and C. Kung. A mechanosensitive ion channel in the yeast plasma membrane. *Science*, 242:762–765, 1988.

O. Guvench, S. S. Mallajosyula, E. P. Raman, E. Hatcher, K. Vanommeslaeghe, T. J. Foster, F. W. Jamison, and A. D. MacKerell Jr. Charmm additive all-atom force field for carbohydrate derivatives and its utility in polysaccharide and carbohydrate- protein modeling. *J Chem Theory Comput.*, 7: 3162–3180, 2011.

L. Hakes, S. C. Lovell, S.G. Oliver, and D. L. Robertson. Specificity in protein interactions and its relationship with sequence diversity and coevolution. *PNAS*, 104:7999–8004, 2007.

B. A. Hall and M. S. P. Sansom. Coarse-grained md simulations and protein-protein interactions: The cohesion-dockerin system. *J. Chem. Theory Comput.*, 5:2465–2471, 2009.

B. A. Hall, K. Bariyyah Abd Halim, A. Buyan, B. Emmanouil, and M. S. P. Sansom. Sidekick for membrane simulations: Automated ensemble molecular dynamics simulations of transmembrane helices. *JCTC*, 2014.

O. P. Hamill and B. Martinac. Molecular basis of mechanotransduction in living cells. *Physiol Rev*, 81:685–740, 2001.

S. B. Hansen, X. Tao, and R. MacKinnon. Structural basis of pip2 activation of the classical inward rectifier k+ channel kir2.2. *Nature*, 477:495–498, 2011.

R. C. Hardie and K. Franze. Photomechanical responses in drosophila photoreceptors. *Science*, 338:260–263, 2012.

H. O. Hartley. The maximum f-ratio as a short-cut test for heterogeneity of variances. *Biometrika*, 37: 308–312, 1950.

C. C. Hase, A. C. Le Dain, and B. Martinac. Purification and functional reconstitution of the recombinant large mechanosensitive ion channel (mscl) of escherichia coli. *J. Biol. Chem.*, 270: 18329–18334, 1995.

E. S. Haswell and E. M. Meyerowitz. Mscs-like proteins control plastid size and shape in arabidopsis thaliana. *Curr. Biol.*, 16:1–11, 2006.

E. S. Haswell, R. Phillips, and D. C. Rees. Mechanosensitive channels: What can they do and how do they do it? *Structure*, 19:1356–1369, 2011.

M. Heinig and D. Frishman. Stride: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucl. Acids Res.*, 32:W500–502, 2004.

J. Hermans, H. J. C. Berendsen, W. F. van Gunsteren, and J. P. M. Postma. A consistent empirical potential for water-protein interactions. *Biopolymers*, 23:1513–1518, 1984.

B. Hess. Convergence and sampling in protein simulations. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, 65:031910, 2002.

B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije. Lincs: A linear constraint solver for molecular simulations. *J. Comput. Chem.*, 18:1463–1472, 1997.

B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl. Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Th. Comput.*, 4:435–447, 2008.

E. Honoré. The neuronal background k2p channels: focus on trek1. *Nature Reviews Neuroscience*, 8:251–261, 2007.

D. C. Howell. *Statistical methods for psychology*. Thomson Wadsworth, 2007.

A. J. Hudspeth. The ionic channels of a vertebrate hair cells. *Hearing Res.*, 22:21–27, 1986.

W. Humphrey, A. Dalke, and K. Schulten. Vmd: Visual molecular dynamics. *Dynamics J. Mol. Graphics*, 14:33–38, 1996.

Y. Jiang, A. Lee, J. Chen, M. Cadene, B. T. Chait, and R. MacKinnon. The open pore conformation of potassium channels. *Nature*, 417:523–526, 2002.

W. L. Jorgensen and J. Tirado-Rives. The opls (optimised potentials for liquid simulations) potential functions for proteins, energy minimisations for crystals of cyclic peptides and crambin. *J. A*, 110: 1657–1666, 1988.

W. Kabsch and C. Sander. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22:2577–2637, 1983.

P. Kaprowski and A. Kubalski. C-termini of the e.coli ms channel move apart upon the channel opening. *J. Biol. Chem*, 278:11237–11245, 2003.

M. Karplus and J. Kuriyan. Molecular dynamics and protein function. *PNAS*, 102:6679–6685, 2005.

M. Karplus and J. A. McCammon. Molecular dynamics simulations of biomolecules. *Nature Struct. Biol.*, 9:646–651, 2002.

J. A. Killian and G. von Heijne. How proteins adapt to a membrane-water interface. *Trends*, 25: 429–434, 2000.

J. A. Killian, I. Salemink, M. R. de Planque, G. Lindblom, R. E. Koeppe II, and D. V. Greathouse. Induction of nonbilayer structures in diacylphosphatidylcholine model membranes by transmembrane alpha-helical peptides: importance of hydrophobic mismatch and proposed role of tryptophans. *Biochemistry*, 35:1037–1045, 1996.

J. B. Klauda, R. M. Venable, J. A. Freites, J. W. O'Connor, D. J. Tobias, and C. et al Mondragon-Ramirez. Update of the charmm all-atom additive force field for lipids: validation on six lipid types. *J. Phys. Chem. B*, 114:7830–7843, 2010.

J. Klepeis, K. L. Larsen, R. Dror, and D. Shaw. Long-timescale molecular dynamics simulations of protein structure and function. *Current Opinion in Structural Biology*, 19:120–127, 2009.

A. L. Koch. Microbial physiology and ecology of slow growth. *Microbiol. Mol. Biol. Rev*, 61:305–318, 1997.

P. Koprowski, W. Grajkowski, E. Y. Isacoff, and A. Kubalski. Genetic screen for potassium leaky small mechanosensitive channels (mscs) in escherichia coli: recognition of cytoplasmic beta domain as a new gating element. *J Biol Chem.*, 286:877–888, 2011.

M. Kospach and M. Trautner Kromann. Distributed perl archive network statistial distributions library. http://www.cpan.org/, 2003. URL `cpansearch.perl.org/src/MIKEK/Statistics-Distributions-1.02/Distributions.pm`.

H. Krishnamurthy, C. L. Piscitelli, and E. Gouaux. Unlocking the molecular secrets of sodium-coupled transporters. *Nature*, 459:347, 2009.

W. Kuhlbrandt, D. N. Wang, and Y. Fujiyoshi. Atomic model of plant light-harvesting complex by electron crystallography. *Nature*, 367:614–621, 1994.

S. Kumar and M. Bansai. Geometrical and sequence characteristics of alpha-helices in globular proteins. *Biophys J*, 75:1935–1944, 1998.

C. Kung. A possible unifying principle for mechanosensation. *Nature*, 436:647–654, 2005.

J. Y. Lai, Y. Shuen Poon, J. T. Kaiser, and D. C. Rees. Open and shut: Crystal structures of the dodecylmaltoside solubilized mechanosensitive channel of small conductance from escherichia coli and helicobacter pylori at 4.4 a and 4.1 a resolutions. *Prot Sci*, 22:502–509, 2013.

I. Lauritzen, J. Chemin, E. Honoré, M. Jodar, N. Guy, M. Lazdunski, and A. J. Patel. Cross-talk between the mechano-gated k2p channel trek-1 and the actin cytoskeleton. *EMBO Rep*, 6:642–648, 2005.

A. R. Leach. *Molecular Modelling - principles and applications*. Pearson Prentice Hill, 2001.

A. G. Lee. Lipid-protein interactions in biological membranes: a structural perspective. *Biochim. Biophys. Acta*, 1612:1–40, 2003.

C. Lee, H. J. Kang, C. von Ballmoos, S. Newstead, P. Uzdavinys, D. L. Dotson, S. Iwata, O. Beckstein, Cameron A. D., and D. Drew. A two-domain elevator mechanism for sodium/proton antiport. *Nature*, 501(7468):573–577, 2013.

E. L. Lehmann. "student" and small-sample theory. *Statist. Sci.*, 14:418–426, 1999.

N. Levina, S. Totemeyer, N. R. Stokes, P. Louis, M. A. Jones, and I.R Booth. Protection of escherichia coli cells against extreme turgor by activation of mscs and mscl mechanosensitive channels: identification of genes required for mscs activity. *EMBO J*, 18:1730–1737, 1999.

M. Levitt. The birth of computational structural biology. *Nat. Str*, 8:392–393, 2001.

L. Li, X. Shi, X. Guo, H. Li, and C. Xu. Ionic protein-lipid interaction at the plasma membrane: what can the charge do? *Trends Biochem Sci.*, 39:130–140, 2014.

S. Li, X. Zhang, and W. Wang. Coarse-grained model for mechanosensitive ion channels. *J. Phys. Chem. B*, 113:14431–14438, 2009.

V. A. Likic, P. R. Gooley, T. P. Speed, and E. E. Strehler. A statistical approach to the interpretation of molecular dynamics simulations of calmodulin equilibrium dynamics. *Protein Science*, 14:2955–2963, 2005.

E. Lindahl and M. S. P. Sansom. Membrane proteins: molecular dynamics simulations. *Curr. Opin. Struct. Biol.*, 18:425–431, 2008.

M. Lingenheil, R. Denschlag, R. Reichold, and P. Tavan. The "hot-solvent/cold-solute" problem revisited. *J. Chem. Theory Comput.*, 4:1293–1306, 2008.

Z. Liu, C. S. Gandhi, and D. C. Rees. Structure of a tetrameric mscl in an expanded intermediate state. *Nature*, 461:120–124, 2009.

E. London and G. W. Feigenson. Fluorescence quenching in model membranes. 1. characterization of quenching caused by a spin-labeled phospolipid. *Biochemistry*, 20:1932–1938, 1981.

C. A. Lopez, Z. Sovova, F. J. van Eerden, A. H. de Vries, and S. J. Marrink. Martini force field parameters for glycolipids. *J. Chem. Theory Comput.*, 9:1694–1708, 2013.

M. H. Louhivuori, J. Risselada, E. van der Giessen, and S. J. Marrink. Release of content through mechano-sensitive gates in pressurized liposomes. *PNAS*, 107:19856–60, 2010.

S. Loukin, X. L. Zhou, Z. Su, Y Saimi, and C. Kung. Wild-type and brachyolmia-causing mutant trpv4 channels respond directly to stretch force. *J. Biol. Chem*, 285:27176–27181, 2010.

M. W. MacArthur and J. M. Thornton. Influence of proline residues on protein conformation. *J Mol Biol.*, 218:397–412, 1991.

H. Machiyama, H. Tatsumi, and M. Sokabe. Structural changes in the cytoplasmic domain of the mechanosensitive channel mscs during opening. *Biophys J.*, 97:1048–1057, 2009.

D. et al. MacKerell. All-atom empirical poteintial for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, 102:3586–3616, 1998.

F. Maingret, M. Fosset, F. Lesage, M. Lazdunski, and E. Honoré. Traak is a mammalian neuronal mechano-gated k+ channel. *J. Biol. Chem.*, 274:1381–1387, 1999.

H. R. Malcolm, Y. Y. Heo, D. E. Elmore, and J. A. Maurer. Defining the role of the tension sensor in the mechanosensitive channel of small conductance. *Biophys J.*, 101:345–352, 2011.

S. E. Mansoor, M. A. DeWitt, and D. L. Farrens. Distance mapping in proteins using fluorescence spectroscopy: The tryptophan-induced quenching (triq) method. *Biochemistry*, 49:9722–9731, 2010.

S. J. Marrink, H. J. Risselada, S. Yefimov, and A. H. de Vries. The martini force field: Coarse grained model for biomolecular simulations. *J Phys Chem B*, 111:7812–7824, 2007.

B. Martinac and O. P. Hamill. Gramicidin a channels switch between stretch activation and stretch inactivation depending on bilayer thickness. *PNAS*, 99:4308–4312, 2002.

B. Martinac, M. Buechner, A. H. Delcour, J. Adler, and C. Kung. Pressure-sensitive ion channel in escherichia coli. *Proc Natl Acad Sci U S A.*, 84:2297–2301, 1987.

B. Martinac, J. Adler, and C. Kung. Mechanosensitive ion channels of e. coli activated by amphipaths. *Nature*, 348:261–263, 1990.

J.H. McDonald. *Handbook of Biological Statistics*. Sparky House Publishing, Baltimore, Maryland., 2014.

M. Mezei and M. Filizola. Trajelix: a computational tool for the geometric characterization of protein helices during molecular dynamics simulations. *J. Comput. Aided Mol. Des.*, 20:97–107, 2006.

S. Miller, W. Bartlett, S. Chandrasekaran, S. Simpson, M. Edwards, and I. R. Booth. Domain organization of the mscs mechanosensitive channel of escherichia coli. *EMBO J.*, 22:36–46, 2003a.

S. Miller, M.D. Edwards, C. Ozdemir, and I. R. Booth. The closed structure of the mscs mechanosensitive channel - cross linking of single cysteine mutants. *J. Biol. Chem.*, 278:32246–32250, 2003b.

P. Moe and P. Blount. Assessment of potential stimuli for mechano-dependent gating of mscl: Effects of pressure, tension, and lipid headgroups. *Biochemistry*, 44(36):12239–12244, 2005. PMID: 16142922.

L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman, and S.-J. Marrink. The martini coarse-grained force field: Extension to proteins. *Journal of Chemical Theory and Computation*, 4 (5):819–834, 2008.

W. A. Morgan. A test for the significance of the difference between two variances in a sample from a normal bivariate distribution. *Biometrika*, 31:13–19, 1939.

D. M. Mount. *Bioinformatics: Sequence and Genome Analysis (2nd ed.)*. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2004.

J. H. Naismith and I. R. Booth. Bacterial mechanosensitive channels - mscs: Evolution's solution to creating sensitivity in function. *Annu. Rev. Biophys.*, 41:7.1–7.21, 2012.

V. A. Ngo, R. K. Kalia, A. Nakano, and P. Vashishta. Supercrystals of dna-functionalized gold nanoparticles: A million-atom molecular dynamics simulation study. *J. Phys. Chem. C*, 116:19579–19585, 2012.

W. S. Noble. How does multiple testing correction work? *Nature Biotechnology*, 27:1135–1137, 2009.

T. Nomura, M. Sokabe, and K. Yoshimura. Lipid-protein interaction of the mscs mechanosensitive channel examined by scanning mutagenesis. *Biophys J*, 91:2874–2881, 2006.

T. Nomura, M. Sokabe, and K. Yoshimura. Interaction between the cytoplasmic and transmembrane domains of the mechanosensitive channel mscs. *Biophysical Journal*, 94:1638, 2008.

J. Norberg and L. Nilsson. On the truncation of long-range electrostatic interactions in dna. *Biophys. J.*, 79:1537–1553, 2000.

H. Ohmori. Studies of ionic currents in the isolated vestibular hair cell of the chick. *J. Physiol., Lond.*, 350:561–581, 1984.

K. Okada, P. C. Moe, and Blount P. Functional design of bacterial mechanosensitive channels. comparisons and contrasts imlluminated by random mutagenesis. *J. Biol. Chem.*, 277:27682–27688, 2002.

O. H. S. Ollila and I. Vattulainen. *Lateral Pressure Profiles in Lipid Membranes: Dependence on Molecular Composition*, chapter 2, pages 26–55. RSC Publishing, 2010.

S. Ollila, M. Louhivuori, S. J. Marrink, and I. Vattulainen. Protein shape change has a major effect on the gating energy of a mechanosensitive channel. *Biophys J*, 100:1651–1659, 2011.

C. Oostenbrink, A. Villa, A. E. Mark, and W. F. van Gunsteren. A biomolecular force field based on the free enthalphy of hydration and solvation: The gromos force-field parameter sets 53a5 and 53a6. *J. Comput. Chem.*, 25:1656–1676, 2004.

J. P. Overington, B Al-Lazikani, and A. L. Hopkins. How many drug targets are there? *Nat Rev Drug Discovery*, 5:993–996, 2006.

K. R. Pandit and J. B. Klauda. Membrane models of e. coli containing cyclic moieties in the aliphatic lipid chain. *BBA Biomembranes*, 1818:1205–1210, 2012.

M. Parrinello and A. Rahman. Polymorphic transitions in single-crystals: A new molecular-dynamics method. *J. Appl. Phys.*, 52:7182–7190, 1981.

A. J. Patel, E. Honoré, F. Maingret, F. Lesage, M. Fink, F. Duprat, and M. Lazdunski. A mammalian two pore domain mechanogated s-like k+ channel. *EMBO J*, 17:4283–4290, 1998.

M. Patra, M. Karttunen, M. T. Hyvönen, E. Falck, P. Lindqvist, and I. Vattulainen. Molecular dynamics simulations of lipid bilayers: major artifacts due to truncating electrostatic interactions. *Biophys. J.*, 84:3636–3645, 2003.

L. Pauling, R. B. Corey, and H. R. Branson. The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. *PNAS*, 4:205–211, 1951.

PDB. Yearly growth of total structures, August 2014a. URL `www.rcsb.org/pdb/statistics/contentGrowthChart.do?content=total&seqid=100`.

PDB. Mechanosensitive channel of small conductance (mscs). Website, July 2014b. URL `www.rcsb.org/pdb/explore/remediatedSequence.do?structureId=2OAU&bionumber=1`.

E. Pebay-Peyroula, G. Rummel, J. P. Rosenbusch, and E. M. Landau. X-ray structure of bacteriorhodopsin at 2.5 angstroms from microcrystals grown in lipidic cubic phases. *Science*, 277: 1676–1681, 1997.

E. Perozo and D. C. Rees. Structure and mechanism in prokaryotic mechanosensitive channels. *Current Opinion in Structural Biology*, 13:432–442, 2003.

R. Phillips, T. Ursell, P. Wiggins, and P. Sens. Emerging roles for lipids in shaping membrane-protein function. *Nature*, 459:379–385, 2009.

D. Picot, P. J. Loll, and R. M. Garavito. The x-ray crystal structure of the membrane protein prostaglandin h2 synthase-1. *Nature*, 367:243–249, 1994.

C. D. Pivetti, M. R. Yen, S. Miller, W. Busch, Y. Tseng, and I. R. et al. Booth. Two families of mechanosensitive channel proteins. *Microbiol. Mol. Biol. Rev*, 67:66–85, 2003.

C. Pliotas, R. Ward, E. Branigan, A. Rasmussen, G. Hagelueken, H. Huang, S. S. Black, I. R. Booth, O. Schiemann, and Naismith J. H. Conformational state of the mscs mechanosensitive channel in solution revealed by pulsed electron-electron double resonance (peldor) spectroscopy. *Proc Natl Acad Sci U S A*, 109:2675–2682, 2012.

A. M. Powl, J. N. Wright, and A. G. East, J. M. amd Lee. Identification of the hydrophobic thickness of a membrane protein using fluorescence spectroscopy: studies with the mechanosensitive channel mscl. *Biochemistry*, 44:5713–5721, 2005.

A. J. Rader, C. Chennubhotla, L.-W. Yang, and I. Bahar. *Normal mode analysis: theory and applications to biological and chemical systems*, chapter The Gaussian network model: theory and applications, pages 41–64. Chapman & Hall / CRC, 2006.

C. R. Raetz. Molecular genetics of membrane phospholipid synthesis. *Annu. Rev. Genet.*, 20:253–295, 1986.

A. Rasmussen, T. Rasmussen, M. D. Edwards, D. Schauer, and U. et al. Schumann. The role of tryptophan residues in the function and stability of the mechanosensitive channel mscs from escherichia coli. *Biochemist*, 46:10899–10908, 2007.

W. Rawicz, K. C. Olbrich, T. McIntosh, D. Needham, and E. Evans. Effect of chain length and unsaturation on elasticity of lipid bilayers. *Biophys. J.*, 79:328–339, 2000.

K. Rayavara and S. A. Desai. Genetic disruption of a mechanosensitive ion channel in plasmodium falciparum. *Am. J. Trop. Med. Hyg.*, 79:65–66, 2008.

A. Reiner, D Yekutieli, and Y. Benjamini. Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics*, 19:368–375, 2003.

Y. M. Rhee, E. J. Sorin, G. Jayachandran, E. Lindahl, and V. S. Pande. Simulations of the role of water in the protein-folding mechanism. *PNAS*, 101:6456–6461, 2004.

J. S. Richardson. Anatomy and taxonomy of protein structures. *Advances in Protein Chemistry*, 34: 167–339, 1981.

R. P. Riek, I. Rigoutsos, J. Novotny, and R. M. Graham. Non-alpha-helical elements modulate polytopic membrane protein architecture. *J. Mol. Biol.*, 306:349–362, 2001.

K. J. Rothman. No adjustments are needed for multiple comparisons. *Epidemiology*, 1:43–46, 1990.

I. Rowe, A. Andriy Anishkin, K. Kamaraju, K. Yoshimura, and S. Sukharev. The cytoplasmic cage domain of the mechanosensitive channel mscs is a sensor of macromolecular crowding. *J. Gen. Physiol.*, 143:543–557, 2014.

S. Samsonov, J. Teyra, and M. T. Pisabarro. A molecular dynamics approach to study the importance of solvent in protein interactions. *Proteins*, 73:515–525, 2008.

M. S. P. Sansom and H. Weinstein. Hinges, swivels and switches: the role of prolines in signalling via transmembrane alpha-helices. *Trends Pharmacol. Sci.*, 21:445–451, 2000.

D. Schmidt and R. MacKinnon. Voltage-dependent k+ channel gating and voltage sensor toxin sensitivity depend on the mechanical state of the lipid membrane. *Proceedings of the National Academy of Sciences*, 105(49):19276–19281, 2008.

D. Schmidt, J. del Marmol, and R. MacKinnon. Mechanistic basis for low threshold mechanosensitivity in voltage-dependent k+ channels. *PNAS*, 109:10352–10357, 2012.

L. D. Schuler, X. Daura, and W. F. van Gunsteren. An improved gromos96 force field for aliphatic hydrocarbons in the condensed phase. *J. Comput. Chem.*, 22:1205–1218, 2001.

U. Schumann, M. D. Edwards, C. Chan Li, and I. R. Booth. The conserved carboxy-terminus of the mscs ms channel is not essential but increase stability and activity. *FEBS Lett.*, 572:233–237, 2004.

K. A. Scott, P. J. Bond, A. Ivetac, A. P. Chetwynd, S. Khalid, and M. S. P. Sansom. Coarse-grained md simulations of membrane protein-bilayer self-assembly. *Structure*, 16 (4):621–630, 2008.

W. R. P. Scott, P. H. Hunenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, A. E. Torda, T. Huber, P. Kruger, and W. F. van Gunsteren. The gromos biomolecular simulation program package. *J. Phys. Chem. A*, 103:3596–3607, 1999.

M. K. Sener, J. D. Olsen, C. N. Hunter, and K. Schulten. Atomic-level structural and functional model of a bacterial photosynthetic membrane vesicle. *PNAS*, 104:15723–15728, 2007.

R. B. Sessions, N. Gibbs, and C. E. Dempsey. Hydrogen bonding in helical polypeptides from molecular dynamics simulations and amide hydrogen exchange analysis: alamethicin and melittin in methanol. *Biophys. J.*, 74:138–152, 1998.

D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan, and W. Wriggers. Atomic-level characterization of the structural dynamics of proteins. *Science*, 330:341–346, 2010.

N. Shi, S. Ye, A. Alam, L. Chen, and Y. Jiang. Atomic structure of a na+ - and k+-condicting channels. *Nature*, 440:570–574, 2006.

T. Shimamura, S. Weyand, O. Beckstein, N. G. Rutherford, J. M. Hadden, D. Sharples, M. S. Sansom, S. Iwata, P. J. Henderson, and A. D. Cameron. Molecular basis of the alternating access model of membrane transport by the sodium-hydantoin transporter, mhp1. *Science*, 328:470–473, 2010.

M. Sotomayor and K. Schulten. Molecular dynamics study of gating in the mechanosensitive channel of small conductance mscs. *Biophys. J.*, 87:3050–3065, 2004.

M. Sotomayor, T. A. van der Straaten, U. Ravaioli, and K. Schulten. Electrostatic properties of the mechanosensitive channel of small conductance mscs. *Biophys. J.*, 90:3496–3510, 2006.

M. Sotomayor, V. Vásquez, E. Perozo, and K. Schulten. Ion conduction though mscs as determined by electrophysiology and simulation. *Biophys. J.*, 92:886–902, 2007.

S. A. Spronk, D. E. Elmore, and D. A. Dougherty. Voltage-dependent hydration and conduction properties of the hydrophobic pore of the mechanosensitive channel of small conductance,. *Biophys. J.*, 90:3555–3569, 2006.

P. J. Stansfeld and M. S. P. Sansom. From coarse-grained to atomistic: A serial multiscale approach to membrane protein simulations. *J. Chem. Theory Comput.*, 7 (4):1157–1166, 2011.

S. Steinbacher, R. Bass, P. Strop, and D. C. Rees. Structures of the prokaryotic mechanosensitive channels mscl and mscs. *Curr. Top. Membr.*, 58:1–24, 2007.

L. S. Stelzl, P. W. Fowler, M. S. Sansom, and O. Beckstein. Flexible gates generate occluded intermediates in the transport cycle of lacy. *Journal of Molecular Biology*, 426(3):735–751, 2013.

N.R. Stokes, H.D. Murray, C. Subramaniam, R.L. Gourse, P. Louis, and Bartlett W. et al. A role for mechanosensitive channels in survival of stationary phase: Regulation of channel expression by rpos. *PNAS*, 100:15959–15964, 2003.

J. D. Storey. A direct approach to false discovery rates. *J. R. Statist. Soc. B*, 64:479–498, 2002.

Student. The probable error of a mean. *Biometrika*, 6:1–25, 1908.

H. Sugeta and T. Miyazawa. General method for calculating helical parameters of polymer chains from bond lengths, bond angles, and internal-rotation angles. *Biopolymers*, 5(7):673–679, 1967.

B. C. Suh and B. Hille. Pip2 is a necessary cofactor for ion channel function: how and why? *Annu. Rev. Biophys.*, 37:175–195, 2008.

S. Sukharev. Purification of the small mechanosensitive channel of escherichia coli (mscs): the subunit structure, conduction, and gating characteristics in liposomes. *Biophys. J.*, 83:290–98, 2002.

S. I. Sukharev, B. Martinac, V. Y. Arshavsky, and C. Kung. Two types of mechanosensitive channels in the escherichia coli cell envelope: solubilization and functional reconstitution. *Biophys. J.*, 65: 177–83, 1993.

S. I. Sukharev, W. J. Sigurdson, C. Kung, and F. Sachs. Energetic and spatial parameters for gating of the bacterial large conductance mechanosensitive channel. *J. Gen. Physiol.*, 113:525–540, 1999.

S. Suzuki and P. J. Henderson. The hydantoin transport protein from microbacterium liquefaciens. *J Bacteriol.*, 188:3329–3336, 2006.

SWISS-MODEL. Principles of protein structure, comparative protein modelling, and visualisation, June 2011. URL `http://swissmodel.expasy.org/course/text/chapter1.htm`.

D. P. Tieleman, H. J. Berendsen, and M. S. Sansom. Voltage-dependent insertion of alamethicin at phospholipid/water and octane/water interfaces. *Biophys. J.*, 80:331–346, 2001.

D. P. Tieleman, H. Leontiadou, A. E. Mark, and S.-J. Marrink. Simulation of pore formation in lipid bilayers by mechanical stress and electric fields. *J. Am. Chem. Soc.*, 125:6382–6383, 2003.

I. G. Tikhonova, B. Selvam, A. Ivetac, J. Wereszczynski, and J. A. McCammon. Simulations of biased agonists in the beta2 adrenergic receptor with accelerated molecular dynamics. *Biochemistry*, 52(33):5593–5603, 2013.

W. Treptow, S.-J. Marrink, and M. Tarek. Gating motions in voltage-gated potassium channels revealed by coarse-grained molecular dynamics simulations. *J. Phys. Chem. B*, 112:3277–3282, 2008.

D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. Gromacs: Fast, flexible, and free. *Journal of Computational Chemistry*, 26 (16):1701–1718, 2005.

D. van der Spoel, E. Lindahl, and B. et al Hess. *The GROMACS user manual, v 4.5.4*. University of Groeningen, 4.5.4 edition, 2010.

T. A. Van Der Straaten, G. Kathawala, A. Trellakis, R. S. Eisenberg, and U. Ravaioli. Biomoca - a boltzmann transport monte carlo model for ion channel simulations. *Molecular Simulation*, 31: 151–171, 2005.

W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hunenberger, P. Kruger, A. E. Mark, W. R. P. Scott, and I. G. Trioni. *Biomolecular simulation: The GROMOS96 Manual and User Guide*. vdf Hochschulverlag AG an der ETH Zurich, 1996.

T. Vora, B. Corry, and S. H. Chung. Brownian dynamics investigations into the conductance state of the mscs channel crystal structure. *Biochim. Biophys. Acta*, 1758:730–737, 2006.

V. Vásquez, M. Sotomayor, J. Cordero-Morales, K. Schulten, and E. Perozo. A structural mechanism for mscs gating in lipid bilayers. *Science*, 321:1210–1214, 2008a.

V. Vásquez, M. Sotomayor, D. M. Cortes, B. Roux, K. Schulten, and E. Perozo. Three-dimensional architecture of membrane-embedded mscs in the closed conformation. *J Mol Biol.*, 378:55–70, 2008b.

T. Walz, T. Hirai, K. Murata, J. B. Heymann, K. Mitsuoka, Y. Fujiyoshi, B. L. Smith, P. Agre, and A. Engel. The three-dimensional structure of aquaporin-1. *Nature*, 387:624–627, 1997.

W. Wang, S. S. Black, M. D. Edwards, S. Miller, E. L. Morrison, W. Bartlett, C. Dong, J. H. Naismith, and I. R. Booth. The structure of an open form of an e. coli mechanosensitive channel at 3.45 a resolution. *Science*, 321:1179–1183, 2008.

X. Wang, F. Xu, J. Liu, B. Gao, Y. Liu, Y. Zhai, J. Ma, K. Zhang, T. S. Baker, K. Schulten, D. Zheng, H. Pang, and F. Sun. Atomic model of rabbit hemorrhagic disease virus by cryo-electron microscopy and crystallography. *PLoS Pathogens*, 9:e1003132, 2013.

B. B. Welch and K. Jones. *Practical Programming in Tcl and Tk, 4th Ed.* Prentice Hall PTR, 2003.

S. Weyand, T. Shimamura, S. Yajima, S. Suzuki, O. Mirza, K. Krusong, E. P. Carpenter, N. G. Rutherford, J. M. Hadden, J. O'Reilly, P. Ma, M. Saidijam, S. G. Patching, R. J. Hope, H. T. Norbertczak, P. C. Roach, S. Iwata, P. J. Henderson, and A. D. Cameron. Structure and molecular mechanism of a nucleobase-cation-symport-1 family transporter. *Science*, 322:709, 2008.

P. L. Yeagle, M. Bennett, V. Lemaître, and A. Watts. Transmembrane helices of membrane proteins may flex to satisfy hydrophobic mismatch. *Biochim Biophys Acta.*, 1768:530–537, 2007.

S. Yefimov, E. van der Giessen, P. Onck, and S. Marrink. Mechanosensitive membrane channels in action. *Biophys J*, 94:2994–3002, 2008.

K. Yoshimura and M. Sokabe. Mechanosensitivity of ion channels based on protein-lipid interactions. *J. R. Soc. Interface*, 7:S307–S320, 2010.

J. H. Zar. *Biostatistical Analysis*. Pearson, 2010.

G. Zhao, J. R. Perilla, E. L. Yufenyuy, X. Meng, B. Chen, J. Ning, J. Ahn, A. M. Gronenborn, K. Schulten, C. Aiken, and P. Zhang. Mature hiv-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature*, 497:643–646, 2013.

M. Zink and H. Grubmuller. Mechanical properties of the icosahedral shell of southern bean mosaic virus: a molecular dynamics study. *Biophysical Journal*, 96:1350–1363, 2009.