

Computational studies of lipopolysaccharide interactions with membrane transporters



Joshua Sauer

*Department of Biochemistry and
St Cross College, Oxford*

A thesis submitted in partial fulfilment of the requirements for
the degree of Doctor of Philosophy at the University of Oxford
Michaelmas 2021

Acknowledgements

I would like to extend my sincerest gratitude to my supervisors **Prof Phill Stansfeld** and **Prof Dame Carol Robinson**. I would like to thank Phill for supporting me throughout the entire PhD process; I am exceedingly grateful for his patience with me (especially though all my computational incompetence), his extremely helpful advice as well as answering all my relentless questions/emails. Phill has been a fantastic, relaxed and easy-going supervisor and made the research process both very fun and highly enjoyable! Carol has also, been a great support to me, I particularly have enjoyed our conversations about the start-up/spin out/biotech life and have really valued her encouragement and advice throughout my PhD. Thank you Phill and Carol!

I owe a huge debt of gratitude to **Dr Robin Corey**, **Dr Owen Vickery** and **Dr Keith Cassidy**. Robin has been exceptionally helpful in sound boarding ideas as well as trouble shooting especially irksome simulations. Owen has helped me extensively with all my energy calculation studies and has been extremely generous with his time working through a great deal of technical issues. I'd also like to thank Keith for many useful discussions about my research.

Dr Jani Reddy Bolla and **Dr Francesco Fiorentino** have also been extremely gracious with me with their time, and I have loved working with them on exciting collaborative projects! I'm glad to be part of the 'JaniMass' team for life. Both Jani and Francesco are complete legends and it's been a pleasure working with them!

I'd like to thank **Michael Horrell** who has been an absolute babe in the lab sorting our technical issues with workstations as well as troubleshooting the cluster. Thanks for taking care of 'BamB' Michael!

I've also loved working in the SBCB and the CVR office, and I want to thank everyone in both communities for making it such a welcoming and fun workplace.

And, of course, I must acknowledge **the Lord** and His provisional hand with all my PhD results; *"in all your ways acknowledge Him, and He will make your paths straight."* (Proverbs 3.6). And He certainly has!

Abstract

Lipopolysaccharide (LPS) lipids compose the outer leaflet of the outer membrane in pathogenic Gram-negative bacteria and are chiefly responsible for host-immune responses. Therefore, the membrane proteins which interact with LPS as well as the transport system for LPS, the Lpt system, represents a class of extremely important drug targets. The focus of this thesis is to employ molecular dynamics studies to examine the effects of LPS binding upon the Lpt complex at every stage of the transportation process, validating and conferring recent mass spectrometry and cryo-electron microscopy (Cryo-EM) studies.

Molecular dynamics (MD) studies in this thesis reveal that LPS binds to other proteins found in the inner membrane in a similar manner to the transporter subcomplex LptB₂FGC. Within the transporter itself, simulations of recently resolved Cryo-EM structure reveal that LPS shifts position to facilitate transportation across the periplasmic protein bridging complex. Binding energy calculations between the acyl chains of LPS and the bridging complex reveal a generally flat energetic landscape before reaching the terminal outer membrane subcomplex of the Lpt system, LptDE. Here, MD studies validate and explain hydrogen-deuterium exchange mass spectrometry results which capture the conformational changes required for LPS insertion into the outer membrane. With an increasing prevalence of antibiotic resistant pathogens, this thesis concerns characterising LPS interactions with its transport machinery suggesting means for developing antimicrobial therapeutics which target this essential system.

Table of Contents

CHAPTER 1 Introduction	26
1.1 Membrane and lipid biology	26
1.1.1 The significance and diversity of membrane proteins	28
1.1.2 Lipopolysaccharide and Bacterial membranes	33
1.1.3 The Lpt System	41
1.2 Molecular dynamics.....	47
1.2.1 Molecular dynamics and membrane proteins	47
1.2.2 Molecular dynamics and free energy calculations	50
1.3 Mass spectrometry and structural biology	52
1.3.1 Native MS	52
1.3.2 Native MS theory and background.....	54
1.3.3 Hydrogen deuterium exchange mass spectrometry	58
1.4 Aims and objectives of this thesis	64
CHAPTER 2 Theory and Methods	66
2.1 Molecular dynamics simulations	66
2.1.1 An overview of the principles of molecular dynamics.....	66
2.1.2 Molecular mechanics force field	67
2.1.3 The particle mesh Ewald method.....	72
2.2 Force fields within MD	73
2.2.1 Atomistic (AT) Forcefields	73
2.2.2 Coarse Grain (CG) force fields	75
2.3 Energy minimisation	84
2.4 Integrator algorithms in Molecular Dynamics	85
2.5 Settings and parameters of MD simulations.....	87
2.5.1 Temperature and pressure.....	87
2.6 Periodic boundary conditions	89
2.7 Water models	89
2.8 Potential of Mean Force (PMF) calculations	92
CHAPTER 3 Characterising LPS binding with inner membrane proteins	95
3.1 The role of LPS within the Inner Membrane	95
3.1.1 Case study 1: PL-binding protein AmtB	97
3.1.2 Case study 2: LPS-binding protein BtuCD	101
3.1.3 LptB ₂ FGC soluble binding contacts with LPS	109
3.2 Characterising LPS interactions with LptB₂FGC	114
3.2.1 Structural studies of LptB ₂ FGC with its substrate.....	115
3.2.2 Lipid interactions with LptB ₂ FGC	117
3.2.3 The conformational dynamics of the LptB ₂ FGC β -taco domains.....	134
3.3 PaLptC interaction with PaLptA	144
3.4 Discussion.....	147

3.5 Methods	150
3.5.1 Modification of Lpt structures used	150
3.5.2 The IM sub complex LptB ₂ FGC	150
3.5.3 Constructing the PaLptAC complex	153
3.6 Atomistic (AT) simulations	153
3.6.1 PaLptB ₂ FGC AT simulations	153
3.6.2 PaLptAC AT simulations.....	154
3.7 CG-MD simulations	154
3.7.1 IM proteins CG simulations	154
3.8 Lipid contact analysis	155
3.9 Author contributions	155
CHAPTER 4 The thermodynamics of lipid A transport through the Lpt system	157
4.1 Background to the periplasmic LPS chute.....	157
4.2 Opening up the β -taco subunits	159
4.3 The binding energy of lipid A within each of the isolated subunits of the periplasmic chute is variable	161
4.4 Binding energy of lipid A across multiple LptA subunits	182
4.5 Binding energy of lipid A across the periplasmic chute LptA ₃ CDF.....	188
4.6 The binding energy of lipid A through LptA ₄	194
4.7 Building the model for a continuous chute of LptA units	198
4.7.1 Thermodynamic landscape of lipid A through the continuous LptA chute	199
4.7.2 Thermodynamic landscape of lipid A through the continuous LptA chute with lipid neighbours ..	204
4.8 Discussion.....	210
4.9 Methods	214
4.9.1 Modification of Lpt structures used	214
4.9.2 The periplasmic bridging proteins LptA ₃ CDF	215
4.9.3 Constructing the LptA ₄ continuous chute	216
4.9.4 Lpt periplasmic subunits CG-MD simulations.....	217
4.10 Steered MD pulling experiments.....	217
4.11 Author contributions	218
CHAPTER 5 Defining the conformational changes of LptDE with LPS and an inhibitor	220
5.1 Background to the LptDE translocon complex	220
5.2 LptDE conformational change with Re-LPS	224
5.2.1 HDX observations of LptDE substrate binding effects.....	224
5.2.2 MD approach for interpreting HDX results.....	226
5.3 Differential contact plot of LptDE bound with Re-LPS.....	232
5.4 LptE solvent differential HDX plots.....	247
5.5 LptDE and thanatin differential HDX plot	248
5.6 Conformation dynamics of the complex with both the substrate LPS and the inhibitor thanatin	254
5.7 Hydrogen bond analysis of the β -taco	257

5.8 Discussion.....	264
5.9 Methods.....	268
5.9.1 Modification of Lpt structures used	268
5.9.2 The OM translocon heterodimer LptDE	268
5.9.3 KpLptDE AT simulations.....	271
5.10 Differential solvent uptake analysis	273
5.11 Author contributions	278
CHAPTER 6 Summary and future directions	279
CHAPTER 7 References.....	285

List of figures

FIGURE 1.1 SCHEMATIC OF THE PROKARYOTIC GRAM-NEGATIVE CELL MEMBRANE. THE MEMBRANE IS A DYNAMIC ENTITY, COMPRISING PROTEINS INTERACTING WITH DIFFERENT TYPES OF LIPIDS.....	27
FIGURE 1.2 EXAMPLES OF MEMBRANE PROTEINS. OmpT (PDB ID: 1I78), OmpF (PDB ID: 3K1B) AND THE BAM (PDB ID: 5LJO) COMPLEX ARE EXAMPLES OF B-BARREL OUTER MEMBRANE PROTEINS. AcrABZ-TolC (PDB ID: 5O66) AS WELL AS AND P-GLYCOPROTEIN (PDB ID: 4KSC) IS AN MULTI DRUG EFFLUX PUMP. LeuT (PDB ID: 3F3A) AND SemiSWEET (PDB ID: 4X5M) ARE RESPONSIBLE FOR TRANSPORTING LEUCINE AND SUGARS. Mscl (PDB ID: 4Y7K) IS A MECHANOSENSITIVE CHANNEL AND AqpZ (PDB ID: 2ABM) IS A WATER CHANNEL. B2-AR IS AN A GPCR PROTEIN IN COMPLEX WITH ITS G PROTEIN (PDB ID: 3SN6).	30
FIGURE 1.3 SCHEMATIC OF GRAM-NEGATIVE BACTERIAL MEMBRANE, FIGURE MODIFIED FROM FLORENTINO ET AL ⁵¹ . (A) CARTOON SCHEMATIC OF GRAM-NEGATIVE BACTERIA MEMBRANE DEPICTING THE INNER MEMBRANE SHIELDED BY THE PEPTIDOGLYCAN LAYER. THE OUTER MEMBRANE'S OUTER LEAFLET FEATURE LPS EXCLUSIVELY. (B) A SKELETAL STRUCTURE OF LPS, ORGANISED IN THREE PARTS: LIPID A, LIPID A GLYCOSYLATED WITH TWO KETO-DEOXYOCTULOSONATE MOIETIES (LIPID A-KDO ₂ OR Re-LPS) INCLUDING THE REST OF THE CORE-OLIGOSACCHARIDE AND THE O-ANTIGEN REPEAT UNIT.	34
FIGURE 1.4 THE Re-LPS BIOSYNTHESIS PATHWAY. THE MODIFICATIONS TO THE PROCEEDING STRUCTURE ARE MADE IN RED. MODIFICATIONS THAT ARE MADE BY LpxL AND LpxM ARE LABELLED IN RED AND BLUE RESPECTIVELY. THIS FIGURE IS MODIFIED FROM BERTANI ET AL. ⁶⁹	39
FIGURE 1.5 LPT SYSTEM SCHEMATIC. (A) CARTOON DEPICTING THE LPT SYSTEM SPANNING TWO BILAYERS. LPS IS INITIALLY SYNTHESISED ON THE INNER LEAFLET OF THE IM AND IS TRANSLOCATED BY THE ABC TRANSPORTER MsbA TO THE PERIPLASMIC SIDE. THE TRANSPORT OF LPS FROM THE IM TO THE OM IS HANDLED BY THE LPT SYSTEM, CONSISTING OF LptA-G. LPS IS EXTRACTED FROM THE IM BY LptB ₂ FG, HANDED TO LptC WHICH FORMS A PERIPLASMIC BRIDGE WITH THE SOLUBLE PROTEIN LptA AND THE N-TERMINAL REGION OF LptD. LPS MOVES THROUGH THIS BRIDGE WHERE IT THEN REACHES THE OM AND INSERTED BY THE TRANSLOCON COMPLEX LptDE. (B) CRYSTAL STRUCTURE OF THE LPT SYSTEM COMPRISING: EcLptB2FGC HOMOLOGY MODEL FROM <i>E. CLOACAE</i> (PDB ID: 6MIT) AS WELL AS THE EcLptDE HOMOLOGY MODEL OF KpLptDE (PDB ID: 5IV9). LptA CONNECTS THE TWO SUBCOMPLEX (PDB ID: 2R19) (WHERE FOUR UNITS ARE ARBITRARILY CHOSEN HERE). THE CHUTE IS REPRESENTED AS A CARTOON AND WAS GENERATED VIA ALIGNMENT OF THE BRIDGE SUBUNITS IN PYMOL.	41
FIGURE 1.6 LptB2FGC SYSTEM SCHEMATIC. (A) CARTOON DEPICTING LptB ₂ FGC WHICH EXTRACTS AND PUSHES LPS UP THROUGH THE SYSTEM (B) CRYSTAL STRUCTURE OF EcLptB2FGC HOMOLOGY MODEL FROM <i>E. CLOACAE</i> (PDB ID: 6MIT).....	43
FIGURE 1.7 LptCFaD SYSTEM SCHEMATIC. (A) LEFT: CARTOON STRUCTURE OF TamB (PDB ID: 5VTG) THE FIRST RECOGNISED EXAMPLE OF A 'B-TACO'. RIGHT: NEXT TO A CARTOON REPRESENTATION (B) CARTOON DEPICTING THE LPT BRIDGING COMPLEX LptCFaD, WHICH SHUTTLES LPS ACROSS THE PERIPLASM (C) STRUCTURES OF THE B-TACO PROTEIN BRIDGING SUBUNITS: KpLptF (PDB ID: 5L75), EcLptC (PDB ID: 3MY2), EcLptA (PDB ID: 2R19) AND KpLptD (PDB ID: 5IV9).	45
FIGURE 1.8 LptDE SCHEMATIC. (A) CARTOON DEPICTING THE TRANSLOCON COMPLEX LptDE WHICH INSERTS LPS INTO THE OM WITH PORE DIAMETER COMPARED TO THE SIZE OF LPS (B) STRUCTURE OF THE EcLptDE HOMOLOGY MODEL OF KpLptDE (PDB ID: 5IV9).	46
FIGURE 1.9 NATIVE MS AND STRUCTURAL BIOLOGY. NATIVE MS IS A POWERFUL ANALYTICAL TECHNIQUE USED TO CHARACTERISE COMPLEX STOICHIOMETRY AND LIGAND OR SMALL MOLECULE BINDING. IT IS OFTEN USED TO SCREEN AND OPTIMISE CONDITIONS FOR X-RAY CRYSTALLOGRAPHY OR CYRO-EM PROTOCOLS. FIGURE OBTAINED FROM LIKO ET AL ¹⁵⁹	54
FIGURE 1.10 FIGURE 9: ELECTROSPRAY AND QUADRUPOLE MASS ANALYSERS. (A) AN OVERVIEW OF THE ELECTROSPRAY PROCESS WHERE THE APPLICATION OF A VOLTAGE BETWEEN THE CAPILLARY AND THE ENTRY OF THE INSTRUMENT DISTORTS THE ANALYTE TO RELEASE DROPLETS, WHICH BREAK UP AND RESULT IN THE GENERATION OF GASEOUS IONS FOR MASS ANALYSIS. (B) THESE IONS (GREEN) ENTER A FIELD PRODUCED BY THE QUADRUPOLE. HERE, THE YELLOW IONS ARE UNSTABLE AND CRASH WITH THE ELECTRODES HOWEVER THE BLUE IONS CAN STABLY TRAVERSE THE FIELD AND MAKE IT THE NEXT STAGE OF THE INSTRUMENT. 57	
FIGURE 1.11 DIFFERENT HDX KINETIC REGIMES. THE PANEL ON THE LEFT DEPICTS THE DIFFERENCE BETWEEN THE EX1 AND EX2 REGIMES. THE MOST COMMON REGIME, EX2, SHOWS AN ISOTOPIC POPULATION GRADUALLY INCREASING IN <i>m/z</i> OVER TIME. FOR EX1, A BIMODAL DISTRIBUTION IS PRESENT, WITH THE INTENSITY OF THE HIGHER <i>m/z</i> POPULATION RISING OVER TIME. THE RIGHT SIDE OF THE PANEL SHOWS THE EXX REGIME WHICH IS A MIXTURE OF BOTH EX1 AND EX2 ISOTOPIC MASS DISTRIBUTIONS. FIGURE ADAPTED FROM WIES ET AL. ¹⁸²	61
FIGURE 1.12 EXPERIMENTAL HDX WORKFLOW. THE PROTEIN IS INCUBATED IN A DEUTERATED BUFFER FOR AN EXCHANGE REACTION, WHICH IS SUBSEQUENTLY QUENCHED BY ACID AFTER A SPECIFIC TIME INTERVAL. THIS PROCESS IS REPEATED WITH THE PROTEIN IN COMPLEX WITH A LIGAND. THE COMPLEX IS THEN DIGESTED, AND THE PEPTIDES ARE SEPARATED AND ANALYSED BY LC-MS. THE PROTECTION LEVEL OF EACH PEPTIDE IS MONITORED AT SPECIFIC TIME INTERVALS FOR BOTH THE APO AND LIGAND BOUND STATE.	63

FIGURE 2.1 THE POTENTIAL OF BONDED INTERACTIONS. THE FOUR MAIN TYPES OF POTENTIAL THROUGH BONDED INTERACTIONS ARE REPRESENTED AS A SCHEMATIC.....	69
FIGURE 2.2 THE POTENTIAL OF NON-BONDED INTERACTIONS. THE TWO MAIN TYPES OF POTENTIAL THROUGH NON-BONDED INTERACTIONS BETWEEN NUCLEI /AND /ARE REPRESENTED ABOVE, LEFT DESCRIBES THE LENNARD JONES POTENTIAL, RIGHT: THE COULOMBIC POTENTIAL.....	72
FIGURE 2.3 AT AND CG REPRESENTATIONS. TOP: COMPARISON OF THE LPTC-SKEWED <i>pALptB₂FGC</i> PROTEIN IN BOTH ITS ATOMISTIC REPRESENTATION (CHARMM FORCEFIELD) AND ITS COARSE-GRAINED REPRESENTATION (MARTINI FORCEFIELD). BOTTOM: POPG LIPID IN THE ATOMISTIC AND CG VERSIONS OF THE STRUCTURE WITH CHARMM AND MARTINI FORCEFIELDS RESPECTIVELY.	80
FIGURE 2.4 AT AND CG WATER MODELS. TOP: SHOWS THE SCHEMATIC OF A THREE-POINT AND FOUR-POINT AT WATER MODEL, WHERE THE OXYGEN IS COLOURED RED, AND THE HYDROGENS COLOURED WHITE. FOR A THREE-POINT MODEL CHARGES ARE LOCATED ON THE ATOMS RESPECTIVELY WHEREAS THE FOUR-POINT MODEL INCLUDED A VIRTUAL SITE BETWEEN THE HYDROGENS. BOTTOM: CG MODEL OF MARTINI WATER IN BOTH ITS STANDARD MODEL AND POLARIZABLE MODEL.....	91
FIGURE 3.1 OBSERVING AcrB AND AmtB LIPID INTERACTIONS WITH NATIVE-MS. FIGURE MODIFIED FROM URNER <i>ET AL</i> ²⁸³ , ALL WORK HERE WAS CONDUCTED BY COLLABORATORS (LHU, FF, DS, MA, TRE, JRB, CVR) FULLY DELIPIDATED AcrB AND AmtB WERE MIXED WITH EQUIMOLAR AMOUNTS OF RE-LPS, LIPID A, CARDIOLIPIN, POPE AND POPG UNDER THE SAME CONDITIONS AND THEIR PROTEIN-LIPID INTERACTIONS WERE MONITORED USING NATIVE-MS. BOTH AmtB AND AcrB DEMONSTRATED HIGHER RELATIVE PROTEIN-LIPID INTENSITIES FOR RE-LPS, LIPID A AND CARDIOLIPIN COMPARED TO PROTEIN-PL INTENSITIES. THIS SUGGESTS THAT CO-PURIFIED LIPIDS FOR EITHER AmtB OR AcrB ARE INDICATIVE OF THEIR LOCAL LIPID ENVIRONMENTS WITHIN THE NATIVE MEMBRANE.	99
FIGURE 3.2 % LIPID OCCUPANCY PLOTS FOR AmtB. AVERAGE % LIPID CONTACTS FOR RE-LPS, LIPID A, CARDIOLIPIN AND POPG WERE MONITORED FOR AmtB IN A BILAYER OF COMPOSITION 83 % POPE, 5 % CARDIOLIPIN, 10 % POPG AND 2 % OF EITHER LIPID A OR RE-LPS. AS WITH THE EXPERIMENTAL RESULTS, CONTACTS WAS OBSERVED FOR LIPID A OR RE-LPS DESPITE AmtB BEING DESCRIBED AS HAVING PL-BINDING BINDING PROPERTIES. IF GIVEN THE OPPORTUNITY LPS DERIVATIVES WILL BIND TO AmtB IF THESE LIPIDS ARE PROXIMAL WITHIN THE ENVIRONMENT. LPS IS FOUND WITHIN THE IM BEFORE TRANSPORT, BUT IT IS UNKNOWN WHETHER THEY FREELY INTERACT WITH IMPs IN THE CELL ITSELF.	100
FIGURE 3.3 FUNCTIONAL AND STRUCTURAL BEHAVIOUR OF BTUCD WITH LPS INTERACTION. FIGURE MODIFIED FROM URNER <i>ET AL</i> ²⁸³ , ALL WORK HERE FOR (A),(C-D) WAS CONDUCTED BY COLLABORATORS (LHU, FF, DS, MA, TRE, JRB, CVR) WHEREAS (B) WAS WORK CARRIED OUT BY MYSELF. (A) NORMAL DELIPIDATION METHODS FAILED TO COMPLETELY REMOVE LPS FROM BTUCD DERIVED FROM 'CORE' LPS-PRODUCING <i>E. COLI</i> STRAINS. (B) % RE-LPS OCCUPANCY CONTACT SITES FOR LPS BINDING WITH BTUCD FOR BOTH THE TRANSMEMBRANE DOMAIN BTUC (ORANGE) AND CYTOPLASMIC DOMAIN BTUD (BLUE) RESPECTIVELY, WHERE THE CA OF THE BACK BACKBONE ARE REPRESENTED AS SPHERES AND ARE SCALED BY % LIPID CONTACT. (C) BTUCD WAS EXPRESSED AND PURIFIED FROM CLEARCOLI®, A CONSTRUCT WHICH FEATURES LIPID IVA INSTEAD OF LPS IN ITS CELL WALL, IN ORDER TO FULLY DELIPIDATE THE PROTEIN. (D) BTUCD ATPASE ACTIVITY WAS QUANTIFIED WITH LPS DERIVATIVES WITH INCREASING LENGTH OF THE GLYCAN HEADGROUP: LIPID A (N = 0) < RE-LPS (N = 2) < 'CORE' LPS (2 < N ≤ 10). THE ATPASE ACTIVITY OF BTUCD INCREASES AS THE NUMBER OF GLYCAN GROUPS INCREASE.....	103
FIGURE 3.4 LIPID BINDING ACROSS BTUCD. (A) % LIPID OCCUPANCY HIGHLIGHTED BY COLOUR ORANGE AND BLUE FOR TRANSMEMBRANE AND CYTOPLASMIC DOMAINS RESPECTIVELY. CA SPHERES ARE SCALED BY LIPID OCCUPANCY FROM 0 TO 100%. (B) RESIDUES HIGHLIGHTED SHOW ONLY RESIDUES WHICH ARE BOTH ABOVE 25% OCCUPANCY (ARBITRARY CUT-OFF) AND IN COMMON BETWEEN EACH LIPID SPECIES (LIPID A, RE-LPS AND CORE-LPS). AMINO ACID SEQUENCES INVOLVED IN BINDING TO 'LPS CAN BE DESCRIBED AS 'FINGERPRINTS' AS THEY FOLLOW A REPETITIVE TREND.	104
FIGURE 3.5 QUANTIFYING LPS CONTACTS ACROSS THE TRANSMEMBRANE DOMAIN OF BTUCD (BTUC). FIGURE ADAPTED FROM URNER <i>ET AL</i> ²⁸³ , ALL WORK HERE COMPLETED BY MYSELF. % OCCUPANCY PLOTS OF LIPID A (N = 0), RE-LPS (N = 2), 'CORE' LPS (2 < N ≤ 10) CONTACTS WITH BTUC. ERROR BARS SHOWN IN GREY REPRESENT THE STANDARD DEVIATION BETWEEN FIVE REPLICATES. CARDIOLIPIN AND POPG % OCCUPANCY PLOTS SHOWN HERE AS A CONTROL.	105
FIGURE 3.6 QUANTIFYING LPS CONTACTS ACROSS THE SOLUBLE DOMAIN OF BTUCD (BTUD). FIGURE ADAPTED FROM URNER <i>ET AL</i> ²⁸³ , ALL WORK HERE COMPLETED BY MYSELF. % OCCUPANCY OF LIPID CONTACTS ACROSS BTUD AND ERROR BARS SHOWN IN GREY REPRESENT THE STANDARD DEVIATION BETWEEN FIVE REPLICATES. CONTACTS INCREASE WITH THE NUMBER OF GLYCAN UNITS (N) ATTACHED TO LIPID A IN THE FOLLOWING ORDER: LIPID A (N = 0), RE-LPS (N = 2), 'CORE' LPS (2 < N ≤ 10).....	107
FIGURE 3.7 BTUCD CYTOPLASMIC BINDING CONTACTS BETWEEN RE-LPS AND CORE-LPS WITH THE ATP-BINDING POCKET. FIGURE ADAPTED FROM URNER <i>ET AL</i> ²⁸³ , ALL WORK HERE COMPLETED BY MYSELF. THE ATP-BINDING POCKET IN BTUD IS HIGHLIGHTED IN SPHERES AND DEFINED AS THOSE RESIDUES 4 Å AWAY FROM THE BOUND NUCLEOTIDE WHICH IS DRAWN AS STICKS AND COLOURED CYAN. RESIDUES WHICH CONTACT THE SACCHARIDE UNITS OF LPS ARE COLOURED BLUE, WHERE INTERACTIONS WITH THE ATP POCKET INCREASE AS THE NUMBER OF GLYCAN UNITS (N) INCREASE IN THE ORDER: LIPID A (N = 0) < RE-LPS (N = 2) < 'CORE' LPS (2 < N ≤ 10). RESIDUES ARG121 AND GLN126 (OF THE ATP BINDING POCKET) ARE SHOWN TO CONTACT THE SACCHARIDE UNITS OF LPS MAY BE PARTICULARLY RELEVANT FOR THE ACTIVITY OF THE TRANSPORTER.	108

FIGURE 3.8 INTERACTION OF LIPOSACCHARIDE WITH LPTB ₂ FGC. FIGURE ADAPTED FROM LI <i>ET AL.</i> ⁸⁶ . SURFACE AND CROSS-SECTIONAL VIEW OF THE LPTB ₂ FG CRYO-EM DENSITY WITH THE LPS MOLECULE INSIDE THE HYDROPHOBIC CAVITY OF LPTG (BLUE) AND LPTF (ORANGE). INSERT SHOWS THE ATOMIC MODEL OF LPS SUPERIMPOSED WITH THE CRYO-EM DENSITY (GREY).	109
FIGURE 3.9 CARTOON SCHEMATIC DESCRIBING THE LPS TRANSPORT PATHWAY. FIGURE ADAPTED FROM FIORENTINO <i>ET AL.</i> ⁵¹ , ALL WORK HERE COMPLETED BY MYSELF AND FF. (A) A SKELETAL REPRESENTATION OF LPS WHERE ITS TRIPARTITE STRUCTURE IS DESCRIBED: LIPID A, CORE OLIGOSACCHARIDE AND THE O-ANTIGEN REPEATING UNIT. THE ORANGE BOX HIGHLIGHTS A DERIVATIVE OF LPS, CALLED 'RE-LPS' WHICH CONSISTS OF THE LIPID A UNIT CONNECTED TO TWO KETODEOXYOCTONIC ACID (Kdo) MOIETIES. THE RED BOX SHOWS THE LIPID A UNIT AND THE PURPLE BOX OUTLINES THE MONO-PHOSPHATE LIPID A (MPLA) (B) CARTOON SCHEMATIC OF THE LPT SYSTEM: LPS IS EXTRACTED FROM THE INNER MEMBRANE (IM) BY THE ABC TRANSPORTER COMPLEX LPTB ₂ FGC, WHERE IT IS SUBSEQUENTLY SHUTTLED THROUGH THE PERIPLASM VIA A SOLUBLE BRIDGE OF LPTA UNITS TO BE FINALLY INSTALLED IN THE OUTER MEMBRANE (OM) BY LPTDE. (C) RECENTLY OBTAINED CRYO-EM STRUCTURE BY COLLABORATORS (FF, JRB, CVR) OF LPTB ₂ FGC FROM <i>PSEUDOMONAS AERUGINOSA</i> . THE SOLUBLE B-TACO DOMAIN OF LPTC IS KNOWN TO INTERACT WITH LPTA TO FORM THE PERIPLASMIC BRIDGE. THE TRANSMEMBRANE DOMAINS LPTF, LPTG AND TM _c (WHICH I MODELLED IN BY USING THE RESOLVED STRUCTURE OF <i>E. CLOACAE</i> AS A TEMPLATE) FORM A HYDROPHOBIC CAVITY WHICH LATERALLY EXTRACTS LPS FROM THE IM TO BE EXTRUDED THROUGH THE PERIPLASMIC BRIDGE. THE TRANSPORT OF LPS IS POWERED BY ATP WHICH IS HYDROLYSED BY THE ABC TRANSPORTER CASSETTE PROTEIN LPTB.	110
FIGURE 3.10 INVESTIGATING LIPID CONTACTS ACROSS THE CYTOPLASMIC DOMAIN OF LPTB ₂ FGC. % OCCUPANCY OF LIPID CONTACTS ACROSS LPTB AND ERROR BARS SHOWN IN GREY REPRESENT THE STANDARD DEVIATION BETWEEN FIVE REPLICATES. CONTACTS INCREASE WITH THE NUMBER OF GLYCAN UNITS (N) ATTACHED TO LIPID A IN THE FOLLOWING ORDER: LIPID A (N = 0) AND RE-LPS (N = 2). REMOVING ONE OF THE PHOSPHATE GROUPS ALSO REDUCES THE NUMBER OF CONTACTS OF THE LIPID WITH THE SOLUBLE DOMAIN.	112
FIGURE 3.11 INTERACTIONS BETWEEN LPS AND THE ATP-BINDING POCKET OF LPTB ₂ FGC. THE ATP-BINDING POCKET IN LPTB IS HIGHLIGHTED IN SPHERES AND DEFINED AS THOSE RESIDUES 4 Å AWAY FROM THE BOUND ATP-PNP WHICH IS DRAWN AS STICKS AND COLOURED CYAN. RESIDUES WHICH CONTACT THE SACCHARIDE UNITS OF LPS ARE COLOURED BLUE, WHERE INTERACTIONS WITH THE ATP POCKET INCREASE AS THE NUMBER OF GLYCAN UNITS (N) INCREASE IN THE ORDER: MPLA (N = 0) < LIPID A (N = 0) < RE-LPS (N = 2). RESIDUE TYR13 (OF THE ATP BINDING POCKET) ARE SHOWN TO CONTACT THE SACCHARIDE UNITS OF LPS MAY BE PARTICULARLY RELEVANT FOR THE ACTIVITY OF THE TRANSPORTER.....	113
FIGURE 3.12 STRUCTURES OF THE 'LPTC-SKEWED' AND 'LPTC-UPRIGHT' CONFORMATION OF LPTB ₂ FGC TRANSPORTER. (A) CRYO-EM STRUCTURE OF THE 'LPTC-SKEWED' CONFORMATION OF LPTB ₂ FGC WHERE THE TM _c HELIX HAS BEEN MODELLED IN USING THE RESOLVED STRUCTURE OF <i>E. CLOACAE</i> AS A TEMPLATE. (B) PALPTB ₂ FGC HOMOLOGY MODEL OF 'LPTC-UPRIGHT' CONFORMATION (C) 'LPTC-SKEWED' AND 'LPTC-UPRIGHT' CONFORMATIONS ALIGNED ON TOP OF EACH OTHER.....	116
FIGURE 3.13 LIPID CONTACTS WITH THE HYDROPHOBIC CAVITY OF THE SKEWED-LPTC CONFORMATION. AVERAGE % OCCUPANCY OF RESIDUE CONTACTS BETWEEN THE LPTC-SKEWED CONFORMATION OF LPTB ₂ FGC WITH RE-LPS, LIPID A, MPLA AND CARDIOLIPIN WITHIN THE HYDROPHOBIC CAVITY. AVERAGE % OCCUPANCY CONTACTS WITH LIPIDS: RE-LPS, LIPID A, MPLA AND CARDIOLIPIN MADE BETWEEN (A) LPTF (B) LPTG AND (C) LPTC. ERROR BARS REPRESENT STANDARD DEVIATION OVER THE THREE REPLICATES. (D) TM HELICES IN CONTACT WITH LIPIDS INDICATED ON CONTACT MAPS AND LABELLED ONTO STRUCTURE.	118
FIGURE 3.14 LIPID CONTACTS WITH THE HYDROPHOBIC CAVITY OF THE UPRIGHT-LPTC CONFORMATION. AVERAGE % OCCUPANCY OF RESIDUE CONTACTS BETWEEN THE LPTC-UPRIGHT CONFORMATION OF LPTB ₂ FGC WITH RE-LPS, LIPID A, MPLA AND CARDIOLIPIN WITHIN THE HYDROPHOBIC CAVITY. AVERAGE % OCCUPANCY CONTACTS WITH LIPIDS: RE-LPS, LIPID A, MPLA AND CARDIOLIPIN MADE BETWEEN (A) LPTF (B) LPTG AND (C) LPTC. ERROR BARS REPRESENT STANDARD DEVIATION OVER THE THREE REPLICATES. (D) TM HELICES IN CONTACT WITH LIPIDS INDICATED ON CONTACT MAPS AND LABELLED ONTO STRUCTURE.	119
FIGURE 3.15 CONTACT MAP OF LIPID INTERACTION WITH THE HYDROPHOBIC CAVITY OF THE SKEWED-LPTC CONFORMATION. LIPID CONTACTS MAPPED ONTO THE LPTC-SKEWED CONFORMATION OF LPTB ₂ FGC FOR TOP LEFT: RE-LPS, TOP RIGHT: LIPID A, BOTTOM LEFT: MPLA AND BOTTOM RIGHT: CARDIOLIPIN WHERE GREATER LIPID CONTACTS ARE HIGHLIGHTED IN DARKER SHADES OF RED.	120
FIGURE 3.16 CONTACT MAP OF LIPID INTERACTION WITH THE HYDROPHOBIC CAVITY OF THE UPRIGHT-LPTC CONFORMATION. LIPID CONTACTS MAPPED ONTO THE LPTC-UPRIGHT CONFORMATION OF LPTB ₂ FGC FOR TOP LEFT: RE-LPS, TOP RIGHT: LIPID A, BOTTOM LEFT: MPLA AND BOTTOM RIGHT: CARDIOLIPIN WHERE GREATER LIPID CONTACTS ARE HIGHLIGHTED IN DARKER SHADES OF RED.	121
FIGURE 3.17 TM _c INTERACTIONS WITH LIPIDS IN THE CAVITY OF THE LPTB ₂ FGC TRANSPORTER. % OCCUPANCY CONTACTS OF THE LPTC TRANSMEMBRANE HELIX INTERACTIONS WITH RE-LPS, LIPID A, MPLA AND CARDIOLIPIN WITHIN THE CAVITY OF THE (A) LPTC-SKEWED AND (B) LPTC-UPRIGHT CONFORMATION OF LPTB ₂ FGC. ERROR BARS REPRESENT STANDARD DEVIATION OVER THE THREE REPLICATES. (C) DIFFERENCE PLOT OF % LIPID CONTACTS MOVING FROM LPTC-SKEWED AND LPTC-UPRIGHT CONFORMATIONS. (D) TM _c OF THE LPTC-SKEWED AND LPTC-UPRIGHT CONFORMATIONS WHERE RESIDUES 16, 19-20, 23-24, 26 COMMON TO BOTH CONFORMATIONS AND ARE HIGHLIGHTED.	122

FIGURE 3.18 HEADGROUP INTERACTIONS OF CARDIOLIPIN VERSUS RE-LPS WITH TM _c . (A) LPTC TM HELIX % CONTACTS OF BOTH THE LPTC-SKEWED AND LPTC-UPRIGHT CONFORMATION WITH PHOSPHATE HEAD GROUP OF CARDIOLIPIN COMPARED TO (B) THE KDO SUGARS OF RE-LPS, WHERE ERROR BARS REPRESENT STANDARD DEVIATION OVER THE THREE REPLICATES. (C) RENDERED IMAGES OF RESIDUES WHICH CONTACT THE TM _c HELIX FOR BOTH CARDIOLIPIN PHOSPHATE HEADGROUP AND THE KDO SUGARS OF RE-LPS.	124
FIGURE 3.19 LPTG INTERACTIONS WITH LIPIDS IN THE CAVITY OF THE LPTB ₂ FGC TRANSPORTER. % OCCUPANCY CONTACTS OF LPTG INTERACTIONS WITH RE-LPS, LIPID A, MPLA AND CARDIOLIPIN WITHIN THE CAVITY OF THE (A) LPTC-SKEWED AND (B) LPTC-UPRIGHT CONFORMATION OF LPTB ₂ FGC. ERROR BARS REPRESENT STANDARD DEVIATION OVER THE THREE REPLICATES. (C) DIFFERENCE PLOT OF % LIPID CONTACTS MOVING FROM LPTC-SKEWED AND LPTC-UPRIGHT CONFORMATIONS, WHERE +VE % CONTACTS REPRESENT MORE INTERACTION WITH THE LPTC-SKEWED CONFORMATION AND -VE % CONTACTS REPRESENT MORE INTERACTION WITH THE LPTC-UPRIGHT CONFORMATION. (D) TM HELICES IN CONTACT WITH LIPIDS INDICATED ON CONTACT MAPS AND LABELLED ONTO STRUCTURE.....	125
FIGURE 3.20 LPTF INTERACTIONS WITH LIPIDS IN THE CAVITY OF THE LPTB ₂ FGC TRANSPORTER. % OCCUPANCY CONTACTS OF LPTF INTERACTIONS WITH RE-LPS, LIPID A, MPLA AND CARDIOLIPIN WITHIN THE CAVITY OF THE (A) LPTC-SKEWED AND (B) LPTC-UPRIGHT CONFORMATION OF LPTB ₂ FGC. ERROR BARS REPRESENT STANDARD DEVIATION OVER THE THREE REPLICATES. (C) DIFFERENCE PLOT OF % LIPID CONTACTS MOVING FROM LPTC-SKEWED AND LPTC-UPRIGHT CONFORMATIONS, WHERE +VE % CONTACTS REPRESENT MORE INTERACTION WITH THE LPTC-SKEWED CONFORMATION AND -VE % CONTACTS REPRESENT MORE INTERACTION WITH THE LPTC-UPRIGHT CONFORMATION. (D) TM HELICES IN CONTACT WITH LIPIDS.	127
FIGURE 3.21 PHOSPHATE HEADGROUP CONTACTS WITH THE LPTB ₂ FGC HYDROPHOBIC CAVITY. (A) % AVERAGE OCCUPANCY PHOSPHATE HEADGROUP CONTACTS ACROSS LPTF WITH RE-LPS, LIPID A, MPLA AND CARDIOLIPIN WITHIN CAVITY FOR THE LPTC-SKEWED (LEFT) AND LPTC-UPRIGHT (RIGHT) CONFORMATIONS OF LPTB ₂ FGC. ERROR BARS REPRESENT STANDARD DEVIATION OVER THE THREE REPLICATES. (B) % AVERAGE OCCUPANCY OF PHOSPHATE HEADGROUP CONTACTS ACROSS LPTG WITH RE-LPS, LIPID A, MPLA AND CARDIOLIPIN WITHIN CAVITY FOR THE LPTC-SKEWED (LEFT) AND LPTC-UPRIGHT CONFORMATIONS OF LPTB ₂ FGC (RIGHT).	128
FIGURE 3.22 KEY BINDING CONTACTS OF THE LPTB ₂ FGC HYDROPHOBIC CAVITY WITH THE PHOSPHATE HEADGROUP OF RE-LPS. (A) RESIDUES WHICH CONTACT PHOSPHATE GROUPS OF RE-LPS WITHIN CAVITY ARE DEPICTED AS STICKS IN BOTH THE LPTC-SKEWED AND (B) LPTC-UPRIGHT CONFORMATIONS OF LPTB ₂ FGC.....	129
FIGURE 3.23 KEY BINDING CONTACTS OF THE LPTB ₂ FGC HYDROPHOBIC CAVITY WITH THE KDO ₂ GLYCAN UNITS. (A) % AVERAGE OCCUPANCY KDO ₂ SUGAR CONTACTS OF RE-LPS WITHIN CAVITY FOR THE LPTC-SKEWED. ERROR BARS REPRESENT STANDARD DEVIATION OVER THE THREE REPLICATES. (B) LPTC-UPRIGHT CONFORMATIONS OF LPTB ₂ FGC. KDO ₂ SUGAR CONTACTS ARE DEPICTED AS STICKS FOR BOTH THE LPTC-SKEWED AND LPTC-UPRIGHT CONFORMATIONS OF LPTB ₂ FGC.....	131
FIGURE 3.24 THE COM OF LIPIDS IN THE CAVITY FOR THE LPTC-SKEWED STRUCTURE. 2D KERNEL DENSITY ESTIMATES PLOTTING THE CENTRE OF MASS (COM) POSITION OF LIPID SPECIES WITHIN THE HYDROPHOBIC CAVITY OF THE LPTC-SKEWED STRUCTURE RELATIVE TO THE COM POSITIONS OF LPTF AND LPTG. TOP LEFT: RE-LPS, TOP RIGHT: LIPID A, BOTTOM LEFT: MPLA, BOTTOM RIGHT: CARDIOLIPIN.	132
FIGURE 3.25 THE COM OF LIPIDS IN THE CAVITY FOR THE LPTC-UPRIGHT STRUCTURE. 2D KERNEL DENSITY ESTIMATES PLOTTING THE CENTRE OF MASS (COM) POSITION OF LIPID SPECIES WITHIN THE HYDROPHOBIC CAVITY OF THE LPTC-UPRIGHT STRUCTURE RELATIVE TO THE COM POSITIONS OF LPTF AND LPTG. TOP LEFT: RE-LPS, TOP RIGHT: LIPID A, BOTTOM LEFT: MPLA, BOTTOM RIGHT: CARDIOLIPIN.	133
FIGURE 3.26 THE LPTF AND LPTG B-TACO DOMAIN VECTOR. SCHEMATIC ILLUSTRATING THE VECTOR ACROSS LPTF AND LPTG B-TACO DOMAIN AND THE PRINCIPAL X, Y, Z AXIS OF THE LPTC-SKEWED AND LPTC-UPRIGHT CONFORMATION.	135
FIGURE 3.27 LPTF AND LPTG B-TACO DOMAINS ANGLE PLOT IN THE LPTC-SKEWED CONFORMATION. (A) AVERAGE X, Y, Z ANGLE PLOT OVER TIME FOR THE B-TACO DOMAINS OF LPTF (LEFT) AND LPTG (RIGHT) WITH NO LIPID PRESENT OR RE-LPS, LIPID A, MPLA AND CARDIOLIPIN PRESENT IN THE CAVITY. SHADED GREY AREA REPRESENTS THE STANDARD DEVIATION OVER THREE REPLICATES (B) THE SAME X, Y, Z ANGLE DATA FOR THE B-TACO DOMAINS OF LPTF (LEFT) AND LPTG (RIGHT) PLOTTED AS A KERNEL DENSITY ESTIMATE.....	136
FIGURE 3.28 DISPLACEMENT OF LPTF B-TACO DOMAIN IN THE LPTC-SKEWED CONFORMATION. (A) 1D X, Y DISPLACEMENT OF THE LPTF B-TACO DOMAIN IN THE LPTC-SKEWED CONFORMATION PLOTTED AS A KERNEL DENSITY ESTIMATE (KDE) WITH NO LIPID PRESENT OR RE-LPS, LIPID A, MPLA AND CARDIOLIPIN PRESENT IN THE CAVITY. WHERE THE POSITIONS X = 0 NM AND Y = 0 NM ARE SET AS THE INITIAL POSITION OF THE B-TACO DOMAIN IN EACH OF THE STATES RESPECTIVELY. (B) 2D KDE PLOT OF X, Y DISPLACEMENT ACROSS LPTF B-TACO WITH: TOP LEFT: NO LIPID PRESENT, TOP RIGHT: RE-LPS, MIDDLE LEFT: LIPID A, MIDDLE RIGHT: MPLA AND BOTTOM LEFT: CARDIOLIPIN PRESENT IN THE CAVITY.	137
FIGURE 3.29 DISPLACEMENT OF LPTG B-TACO DOMAIN IN THE LPTC-SKEWED CONFORMATION. (A) 1D X, Y DISPLACEMENT OF THE LPTG B-TACO DOMAIN IN THE LPTC-SKEWED CONFORMATION PLOTTED AS A KERNEL DENSITY ESTIMATE (KDE) WITH NO LIPID PRESENT OR RE-LPS, LIPID A, MPLA AND CARDIOLIPIN PRESENT IN THE CAVITY. WHERE THE POSITION X = 0 NM AND Y = 0 NM IS SET AS THE INITIAL POSITION OF THE B-TACO DOMAIN IN EACH OF THE STATES RESPECTIVELY. (B) 2D KDE PLOT OF X, Y	

DISPLACEMENT ACROSS LPTG B-TACO WITH: TOP LEFT: NO LIPID PRESENT, TOP RIGHT: RE-LPS, MIDDLE LEFT: LIPID A, MIDDLE RIGHT: MPLA AND BOTTOM LEFT: CARDIOLIPIN PRESENT IN THE CAVITY.	138
FIGURE 3.30 LPTF AND LPTG B-TACO DOMAINS ANGLE PLOT IN THE LPTC-UPRIGHT CONFORMATION. (A) AVERAGE X, Y, Z ANGLE PLOT OVER TIME FOR THE B-TACO DOMAINS OF LPTF (LEFT) AND LPTG (RIGHT) WITH NO LIPID PRESENT OR RE-LPS, LIPID A, MPLA AND CARDIOLIPIN PRESENT IN THE CAVITY. SHADED AREAS REPRESENT THE STANDARD DEVIATION OVER THREE REPLICATES (B) THE SAME X, Y, Z ANGLE DATA FOR THE B-TACO DOMAINS OF LPTF AND LPTG PLOTTED AS A KERNEL DENSITY ESTIMATE.	140
FIGURE 3.31 DISPLACEMENT OF LPTF B-TACO DOMAIN IN THE LPTC-UPRIGHT CONFORMATION. (A) 1D X, Y DISPLACEMENT OF THE LPTF B-TACO DOMAIN IN THE LPTC-UPRIGHT CONFORMATION PLOTTED AS A KERNEL DENSITY ESTIMATE (KDE) WITH NO LIPID PRESENT OR RE-LPS, LIPID A, MPLA AND CARDIOLIPIN PRESENT IN THE CAVITY. WHERE THE POSITIONS $x = 0$ NM AND $y = 0$ NM ARE SET AS THE INITIAL POSITION OF THE B-TACO DOMAIN IN EACH OF THE STATES RESPECTIVELY. (B) 2D KDE PLOT OF X, Y DISPLACEMENT ACROSS LPTF B-TACO WITH: TOP LEFT: NO LIPID PRESENT, TOP RIGHT: RE-LPS, MIDDLE LEFT: LIPID A, MIDDLE RIGHT: MPLA AND BOTTOM LEFT: CARDIOLIPIN PRESENT IN THE CAVITY.	142
FIGURE 3.32 DISPLACEMENT OF LPTG B-TACO DOMAIN IN THE LPTC-UPRIGHT CONFORMATION. (A) 1D X, Y DISPLACEMENT OF THE LPTG B-TACO DOMAIN IN THE LPTC-SKEWED CONFORMATION PLOTTED AS A KERNEL DENSITY ESTIMATE (KDE) WITH NO LIPID PRESENT OR RE-LPS, LIPID A, MPLA AND CARDIOLIPIN PRESENT IN THE CAVITY. WHERE THE POSITION $x = 0$ NM AND $y = 0$ NM ARE SET AS THE INITIAL POSITION OF THE B-TACO DOMAIN IN EACH OF THE STATES RESPECTIVELY. (B) 2D KDE PLOT OF X, Y DISPLACEMENT ACROSS LPTG B-TACO WITH: TOP LEFT: NO LIPID PRESENT, TOP RIGHT: RE-LPS, MIDDLE LEFT: LIPID A, MIDDLE RIGHT: MPLA AND BOTTOM LEFT: CARDIOLIPIN PRESENT IN THE CAVITY.	143
FIGURE 3.33 CHEMICAL STRUCTURES OF CROSSLINKERS. ILLUSTRATION OF DISUCCINIMIDYL SULFOXIDE (DSSO) AND DISUCCINIMIDYL DIBUTYRIC UREA (DSBU).	144
FIGURE 3.34 LYSINE DISTANCE PLOT. (A) AVERAGE CA-CA DISTANCES MEASURED BETWEEN LYSINE 131 OF LPTC WITH LYSINE 22 OF LPTA AS WELL AS (B) THE AVERAGE DISTANCE BETWEEN LYSINE 176 OF LPTC WITH LYSINE 22 OF LPTA. THE SHADED GREY AREA REPRESENTS THE STANDARD DEVIATION OVER THREE REPLICATES. (C) STRUCTURE OF PALPTAC COMPLEX WITH THE RELEVANT LYSINES LABELLED.	145
FIGURE 3.35 PALPTAC HYDROGEN BOND COUNT. (A) AVERAGE TOTAL HYDROGEN BONDS ACROSS THE PALPTAC COMPLEX OVER THE SIMULATION AND (B) AVERAGE TOTAL HYDROGEN BONDS ACROSS THE LPTAC INTERFACE OVER THE SIMULATION TIME. THE SHADED GREY AREA REPRESENTS THE STANDARD DEVIATION OVER THREE REPLICATES. (C) THE KEY RESIDUES WHICH FEATURE IN THE HYDROGEN BOND NETWORK BETWEEN LPTA AND LPTC ARE REPRESENTED AS STICKS. HYDROGEN BONDS ARE DRAWN ON WITH DASHED LINES.	146
FIGURE 4.1 CARTOON SCHEMATIC DESCRIBING THE LPS TRANSPORT PATHWAY. FIGURE ADAPTED FROM FIORENTINO <i>ET AL</i> , ⁵¹ ALL WORK HERE COMPLETED BY MYSELF AND FF. (A) CARTOON SCHEMATIC OF THE LPT SYSTEM: LPS IS EXTRACTED FROM THE INNER MEMBRANE (IM) BY THE ABC TRANSPORTER COMPLEX LPTB ₂ FGC, WHERE IT IS SUBSEQUENTLY SHUTTLED THROUGH THE PERIPLASM VIA A SOLUBLE BRIDGE OF LPTA UNITS TO BE FINALLY INSTALLED IN THE OUTER MEMBRANE (OM) BY LPTDE. (B) STRUCTURES OF THE 'B-TACO' PROTEIN BRIDGING SUBUNITS: LPTF, LPTC, LPTA AND LPTD (C) A SKELETAL REPRESENTATION OF LPS WHERE ITS TRIPARTITE STRUCTURE IS DESCRIBED: LIPID A, CORE OLIGOSACCHARIDE AND THE O-ANTIGEN REPEATING UNIT WHERE THE RED BOX HIGHLIGHTS THE LIPID A UNIT.	158
FIGURE 4.2 DISTANCE PLOT OF THE SIDES OF THE KPLPTD B-TACO. (A) AVERAGE DISTANCE BETWEEN BACKBONE ATOMS OF THE TWO RESIDUES (ASN91, ASN135) ON THE EITHER SIDE OF THE LPTD B-TACO OVER FIVE REPLICATES. B-TACO CLOSURE OCCURS WITHIN 1 NS OF SIMULATION TIME. (B) RESIDUES DEPICTED ON SCHEMATIC OF KPLPTD.	160
FIGURE 4.3 SCHEMATIC OF STEERED MD. LIPID A BEING PULLED OUT FROM THE LPTC B-TACO IN BOTH THE FORWARD (TOWARDS THE OM) AND REVERSE DIRECTION (TOWARDS THE IM).	162
FIGURE 4.4 BINDING ENERGY OF LIPID A WITH THE LPTF B-TACO. (A) PMF PROFILE OF LIPID A MOVING THROUGH LPTF, AREAS SHADED IN GREY INDICATES THE ERROR AND IS OBTAINED BY BOOTSTRAPPING THE PROFILE 200 TIMES. CARTOON REPRESENTATION OF THE LPTF SUBUNIT ALIGNED WITH THE ENERGY PROFILE IN THE BACKGROUND FOR REFERENCE. (B) TOTAL COUNTS SUMMED ACROSS THE COLLECTIVE VARIABLE OF LIPID A BEING PULLED FROM THE B-TACO INTO SOLVENT FROM THE RAW HISTOGRAMS, WHERE THE DATASET IS NORMALISED TO 1. THE RED LINE IS A 'RULE OF THUMB' MARKER FOR ACCEPTABLE SAMPLING POSITIONED AT 0.2 OF TOTAL COUNTS. (C) THE NUMBER OF OVERLAPPING HISTOGRAMS FROM THE UMBRELLA SAMPLING ACROSS THE COLLECTIVE VARIABLE. A COUNT IS CONSIDERED WHEN THE OVERLAP IS >10% OF THE TOTAL HISTOGRAM HEIGHT. THE RED LINE CORRESPONDS TO A 'RULE OF THUMB' FOR SUFFICIENT OVERLAP SET TO 3. (D) THE HISTOGRAMS FROM THE WHAM ANALYSIS, WHERE EVERY HISTOGRAM CORRESPONDS TO A SEPARATE SAMPLING POSITION OF LIPID A.	164
FIGURE 4.5 LPTF PMF CONVERGENCE ANALYSIS. (A) PMF PROFILES OF INCREASING BLOCKS SIMULATION TIME BY 100 NS FOR UP TO ONE MICROSECOND. COLOUR SCALE FROM BLUE TO RED ILLUSTRATES HOW THE PROFILE EVOLVES THROUGH TIME. THE FIRST 200 NS WAS REMOVED FROM EACH WINDOW TO ENSURE THAT THE SYSTEM HAD FULLY EQUILIBRATED FIRST. (B) THE BINDING VALUES (THE LOWEST ENERGY VALUE) OF EACH PMF PROFILE OVER THE COURSE OF THE SIMULATION TIME FOR INCREASING BLOCKS OF 100 NS. ERRORS CALCULATED BY BOOTSTRAPPING EACH PMF PROFILE 10 TIMES. FOR INSTANCE, 200 – 300 NS,	

200 – 400 ns, ..., 200 – 1000 ns. (c) THE BINDING VALUES (THE LOWEST ENERGY VALUE) OF EACH PMF PROFILE OVER THE COURSE OF THE SIMULATION TIME FOR CONSECUTIVE BLOCKS OF 100 ns. FOR INSTANCE, 200 – 300 ns, 300 – 400 ns, ..., 900 – 1000 ns.	165
FIGURE 4.6 BINDING ENERGY OF LIPID A WITH THE LPTC B-TACO. (A) PMF PROFILE OF LIPID A MOVING THROUGH LPTC, AREAS SHADED IN GREY INDICATES THE ERROR AND IS OBTAINED BY BOOTSTRAPPING THE PROFILE 200 TIMES. CARTOON REPRESENTATION OF THE LPTC SUBUNIT ALIGNED WITH THE ENERGY PROFILE IN THE BACKGROUND FOR REFERENCE. (B) TOTAL COUNTS SUMMED ACROSS THE COLLECTIVE VARIABLE OF LIPID A BEING PULLED FROM THE B-TACO INTO SOLVENT FROM THE RAW HISTOGRAMS, WHERE THE DATASET IS NORMALISED TO 1. THE RED LINE IS A ‘RULE OF THUMB’ MARKER FOR ACCEPTABLE SAMPLING POSITIONED AT 0.2 OF TOTAL COUNTS. (C) THE NUMBER OF OVERLAPPING HISTOGRAMS FROM THE UMBRELLA SAMPLING ACROSS THE COLLECTIVE VARIABLE. A COUNT IS CONSIDERED WHEN THE OVERLAP IS >10% OF THE TOTAL HISTOGRAM HEIGHT. THE RED LINE CORRESPONDS TO A ‘RULE OF THUMB’ FOR SUFFICIENT OVERLAP SET TO 3. (D) THE HISTOGRAMS FROM THE WHAM ANALYSIS, WHERE EVERY HISTOGRAM CORRESPONDS TO A SEPARATE SAMPLING POSITION OF LIPID A.	167
FIGURE 4.7 LPTC PMF CONVERGENCE ANALYSIS. (A) PMF PROFILES OF INCREASING BLOCKS SIMULATION TIME BY 100 ns FOR UP TO ONE MICROSECOND. COLOUR SCALE FROM BLUE TO RED ILLUSTRATES HOW THE PROFILE EVOLVES THROUGH TIME. THE FIRST 200 ns WAS REMOVED FROM EACH WINDOW TO ENSURE THAT THE SYSTEM HAD FULLY EQUILIBRATED FIRST. (B) THE BINDING VALUES (THE LOWEST ENERGY VALUE) OF EACH PMF PROFILE OVER THE COURSE OF THE SIMULATION TIME FOR INCREASING BLOCKS OF 100 ns. ERRORS CALCULATED BY BOOTSTRAPPING EACH PMF PROFILE 10 TIMES. FOR INSTANCE, 200 – 300 ns, 200 – 400 ns, ..., 200 – 1000 ns. (c) THE BINDING VALUES (THE LOWEST ENERGY VALUE) OF EACH PMF PROFILE OVER THE COURSE OF THE SIMULATION TIME FOR CONSECUTIVE BLOCKS OF 100 ns. FOR INSTANCE, 200 – 300 ns, 300 – 400 ns, ..., 900 – 1000 ns.	169
FIGURE 4.8: SNAPSHOTS OF LIPID A ALONG THE REACTION COORDINATE. TOP: LOOKING DOWN ON THE LPTC B-TACO. MIDDLE: SIDE VIEW OF THE LPTC B-TACO AS LIPID A IS PULLED THROUGH THE SUBUNIT. IN BOTH PERSPECTIVES THE B-TACO IS MADE TRANSPARENT FOR CLARITY. BOTTOM: PMF PROFILE OVERLAYED FOR REFERENCE	170
FIGURE 4.9 BINDING ENERGY OF LIPID A WITH THE LPTA B-TACO. (A) PMF PROFILE OF LIPID A MOVING THROUGH LPTD, AREAS SHADED IN GREY INDICATES THE ERROR AND IS OBTAINED BY BOOTSTRAPPING THE PROFILE 200 TIMES. CARTOON REPRESENTATION OF THE LPTA SUBUNIT ALIGNED WITH THE ENERGY PROFILE IN THE BACKGROUND FOR REFERENCE. (B) TOTAL COUNTS SUMMED ACROSS THE COLLECTIVE VARIABLE OF LIPID A BEING PULLED FROM THE B-TACO INTO SOLVENT FROM THE RAW HISTOGRAMS, WHERE THE DATASET IS NORMALISED TO 1. THE RED LINE IS A ‘RULE OF THUMB’ MARKER FOR ACCEPTABLE SAMPLING POSITIONED AT 0.2 OF TOTAL COUNTS. (C) THE NUMBER OF OVERLAPPING HISTOGRAMS FROM THE UMBRELLA SAMPLING ACROSS THE COLLECTIVE VARIABLE. A COUNT IS CONSIDERED WHEN THE OVERLAP IS >10% OF THE TOTAL HISTOGRAM HEIGHT. THE RED LINE CORRESPONDS TO A ‘RULE OF THUMB’ FOR SUFFICIENT OVERLAP SET TO 3. (D) THE HISTOGRAMS FROM THE WHAM ANALYSIS, WHERE EVERY HISTOGRAM CORRESPONDS TO A SEPARATE SAMPLING POSITION OF LIPID A.	171
FIGURE 4.10 LPTA PMF CONVERGENCE ANALYSIS. (A) PMF PROFILES OF INCREASING BLOCKS SIMULATION TIME BY 100 ns FOR UP TO ONE MICROSECOND. COLOUR SCALE FROM BLUE TO RED ILLUSTRATES HOW THE PROFILE EVOLVES THROUGH TIME. THE FIRST 200 ns WAS REMOVED FROM EACH WINDOW TO ENSURE THAT THE SYSTEM HAD FULLY EQUILIBRATED FIRST. (B) THE BINDING VALUES (THE LOWEST ENERGY VALUE) OF EACH PMF PROFILE OVER THE COURSE OF THE SIMULATION TIME FOR INCREASING BLOCKS OF 100 ns. ERRORS CALCULATED BY BOOTSTRAPPING EACH PMF PROFILE 10 TIMES. FOR INSTANCE, 200 – 300 ns, 200 – 400 ns, ..., 200 – 1000 ns. (c) THE BINDING VALUES (THE LOWEST ENERGY VALUE) OF EACH PMF PROFILE OVER THE COURSE OF THE SIMULATION TIME FOR CONSECUTIVE BLOCKS OF 100 ns. FOR INSTANCE, 200 – 300 ns, 300 – 400 ns, ..., 900 – 1000 ns.....	173
FIGURE 4.11 LIPID A BINDING CONFORMATION WITHIN THE LPTA B-TACO. (A) PMF PROFILE OF LIPID A MOVING THROUGH LPTA, AREAS SHADED IN GREY INDICATES THE ERROR AND IS OBTAINED BY BOOTSTRAPPING THE PROFILE 200 TIMES. CARTOON REPRESENTATION OF THE LPTA SUBUNIT ALIGNED WITH THE ENERGY PROFILE IN THE BACKGROUND FOR REFERENCE. (B) LIPID A BINDING POSE AT -1 nm AND AT 0 nm LABELLED ONTO THE LPTA PMF BINDING PROFILE.....	174
FIGURE 4.12 BINDING ENERGY OF LIPID A WITH THE LPTD B-TACO. (A) PMF PROFILE OF LIPID A MOVING THROUGH LPTD, AREAS SHADED IN GREY INDICATES THE ERROR AND IS OBTAINED BY BOOTSTRAPPING THE PROFILE 200 TIMES. CARTOON REPRESENTATION OF THE LPTD SUBUNIT ALIGNED WITH THE ENERGY PROFILE IN THE BACKGROUND FOR REFERENCE. (B) TOTAL COUNTS SUMMED ACROSS THE COLLECTIVE VARIABLE OF LIPID A BEING PULLED FROM THE B-TACO INTO SOLVENT FROM THE RAW HISTOGRAMS, WHERE THE DATASET IS NORMALISED TO 1. THE RED LINE IS A ‘RULE OF THUMB’ MARKER FOR ACCEPTABLE SAMPLING POSITIONED AT 0.2 OF TOTAL COUNTS. (C) THE NUMBER OF OVERLAPPING HISTOGRAMS FROM THE UMBRELLA SAMPLING ACROSS THE COLLECTIVE VARIABLE. A COUNT IS CONSIDERED WHEN THE OVERLAP IS >10% OF THE TOTAL HISTOGRAM HEIGHT. THE RED LINE CORRESPONDS TO A ‘RULE OF THUMB’ FOR SUFFICIENT OVERLAP SET TO 3. (D) THE HISTOGRAMS FROM THE WHAM ANALYSIS, WHERE EVERY HISTOGRAM CORRESPONDS TO A SEPARATE SAMPLING POSITION OF LIPID A.	176

FIGURE 4.13 LptD PMF CONVERGENCE ANALYSIS. (A) PMF PROFILES OF INCREASING BLOCKS SIMULATION TIME BY 100 NS FOR UP TO ONE MICROSECOND. COLOUR SCALE FROM BLUE TO RED ILLUSTRATES HOW THE PROFILE EVOLVES THROUGH TIME. THE FIRST 200 NS WAS REMOVED FROM EACH WINDOW TO ENSURE THAT THE SYSTEM HAD FULLY EQUILIBRATED FIRST. (B) THE BINDING VALUES (THE LOWEST ENERGY VALUE) OF EACH PMF PROFILE OVER THE COURSE OF THE SIMULATION TIME FOR INCREASING BLOCKS OF 100 NS. ERRORS CALCULATED BY BOOTSTRAPPING EACH PMF PROFILE 10 TIMES. FOR INSTANCE, 200 – 300 NS, 200 – 400 NS, ..., 200 – 1000 NS. (C) THE BINDING VALUES (THE LOWEST ENERGY VALUE) OF EACH PMF PROFILE OVER THE COURSE OF THE SIMULATION TIME FOR CONSECUTIVE BLOCKS OF 100 NS. FOR INSTANCE, 200 – 300 NS, 300 – 400 NS, ..., 900 – 1000 NS.....	178
FIGURE 4.14 LIPID A BINDING CONFORMATION WITHIN THE LptD B-TACO. (A) PMF PROFILE OF LIPID A MOVING THROUGH LptD, AREAS SHADED IN GREY INDICATES THE ERROR AND IS OBTAINED BY BOOTSTRAPPING THE PROFILE 200 TIMES. CARTOON REPRESENTATION OF THE LptA SUBUNIT ALIGNED WITH THE ENERGY PROFILE IN THE BACKGROUND FOR REFERENCE. (B) LIPID A BINDING POSE AT -1.5 NM, +1 NM AND AT +2.5 NM LABELLED ONTO THE LptA PMF BINDING PROFILE.	179
FIGURE 4.15 WIDTH OF LptD SUBUNIT DURING THE UMBRELLA SAMPLING. TOP: DISTANCE OVER THE B-TACO MEASURED BETWEEN ARG96 AND ARG142 OVER THE 1 μ S SAMPLING TIME FOR THE WINDOW SAMPLING AT +1 NM. BOTTOM: REPRESENTATIVE SNAPSHOTS OF LIPID A SAMPLING OF THE 0-100 NS, 200-700 NS AND 900-1000 NS SECTIONS OF THE SIMULATION.	180
FIGURE 4.16 PMF PROFILES FOR THE BRIDGING SUBUNITS. PMF CALCULATIONS FROM THE LATERAL ‘CARD SWIPING’ THROUGH EACH OF THE B-TACO UNITS (A) LptF, (B) LptC, (C) LptA AND (D) LptD. AREAS SHADED IN GREY INDICATES THE ERROR AND IS OBTAINED BY BOOTSTRAPPING EACH PROFILE RESPECTIVELY.	181
FIGURE 4.17 BINDING ENERGY OF LIPID A WITH THE LptA ₃ PORTION OF THE CHUTE. (A) PMF PROFILE OF LIPID A MOVING THROUGH LptA ₃ , AREAS SHADED IN GREY INDICATES THE ERROR AND IS OBTAINED BY BOOTSTRAPPING THE PROFILE 200 TIMES. CARTOON REPRESENTATION OF THE LptA ₃ TRIMER ALIGNED WITH THE ENERGY PROFILE IN THE BACKGROUND FOR REFERENCE. (B) TOTAL COUNTS SUMMED ACROSS THE COLLECTIVE VARIABLE OF LIPID A BEING PULLED FROM THE B-TACO INTO SOLVENT FROM THE RAW HISTOGRAMS, WHERE THE DATASET IS NORMALISED TO 1. THE RED LINE IS A ‘RULE OF THUMB’ MARKER FOR ACCEPTABLE SAMPLING POSITIONED AT 0.2 OF TOTAL COUNTS. (C) THE NUMBER OF OVERLAPPING HISTOGRAMS FROM THE UMBRELLA SAMPLING ACROSS THE COLLECTIVE VARIABLE. A COUNT IS CONSIDERED WHEN THE OVERLAP IS >10% OF THE TOTAL HISTOGRAM HEIGHT. THE RED LINE CORRESPONDS TO A ‘RULE OF THUMB’ FOR SUFFICIENT OVERLAP SET TO 3. (D) THE HISTOGRAMS FROM THE WHAM ANALYSIS, WHERE EVERY HISTOGRAM CORRESPONDS TO A SEPARATE SAMPLING POSITION OF LIPID A FROM THE LATERAL ‘CARD SWIPING’ THROUGH THE LptA ₃ TRIMER.	184
FIGURE 4.18 LptA ₃ PMF CONVERGENCE ANALYSIS. (A) PMF PROFILES OF INCREASING BLOCKS SIMULATION TIME BY 100 NS FOR UP TO ONE MICROSECOND. COLOUR SCALE FROM BLUE TO RED ILLUSTRATES HOW THE PROFILE EVOLVES THROUGH TIME. THE FIRST 200 NS WAS REMOVED FROM EACH WINDOW TO ENSURE THAT THE SYSTEM HAD FULLY EQUILIBRATED FIRST. (B) THE BINDING VALUES (THE LOWEST ENERGY VALUE) OF EACH PMF PROFILE OVER THE COURSE OF THE SIMULATION TIME FOR INCREASING BLOCKS OF 100 NS. ERRORS CALCULATED BY BOOTSTRAPPING EACH PMF PROFILE 10 TIMES. FOR INSTANCE, 200 – 300 NS, 200 – 400 NS, ..., 200 – 1000 NS. (C) THE BINDING VALUES (THE LOWEST ENERGY VALUE) OF EACH PMF PROFILE OVER THE COURSE OF THE SIMULATION TIME FOR CONSECUTIVE BLOCKS OF 100 NS. FOR INSTANCE, 200 – 300 NS, 300 – 400 NS, ..., 900 – 1000 NS.....	186
FIGURE 4.19 LIPID A BINDING CONFORMATION WITHIN THE LptA ₃ TRIMER. (A) PMF PROFILE OF LIPID A MOVING THROUGH LptA ₃ , AREAS SHADED IN GREY INDICATES THE ERROR AND IS OBTAINED BY BOOTSTRAPPING THE PROFILE 200 TIMES. CARTOON REPRESENTATION OF THE LptA SUBUNIT ALIGNED WITH THE ENERGY PROFILE IN THE BACKGROUND FOR REFERENCE. (B) LIPID A BINDING POSE AT -1.8 NM, +2.7 NM AND AT +4.7 NM ARE LABELLED ONTO THE LptA ₃ PMF BINDING PROFILE, WHERE IN EACH INSTANCE, THE BINDING POSE HAS BEEN HIGHLIGHTED. THE RESIDUES WHICH INTERACT WITH THE PHOSPHATE GROUPS ARE HIGHLIGHTED ON THE STRUCTURE.	188
FIGURE 4.20 BINDING ENERGY OF LIPID A WITH THE LptA ₃ CDF PERIPLASMIC CHUTE. (A) PMF PROFILE OF LIPID A MOVING THROUGH LptA ₃ CDF, AREAS SHADED IN GREY INDICATES THE ERROR AND IS OBTAINED BY BOOTSTRAPPING THE PROFILE 200 TIMES. CARTOON REPRESENTATION OF THE LptA ₃ CDF COMPLEX IT ALIGNED WITH THE ENERGY PROFILE IN THE BACKGROUND FOR REFERENCE. (B) TOTAL COUNTS SUMMED ACROSS THE COLLECTIVE VARIABLE OF LIPID A BEING PULLED FROM THE B-TACO INTO SOLVENT FROM THE RAW HISTOGRAMS, WHERE THE DATASET IS NORMALISED TO 1. THE RED LINE IS A ‘RULE OF THUMB’ MARKER FOR ACCEPTABLE SAMPLING POSITIONED AT 0.2 OF TOTAL COUNTS. (C) THE NUMBER OF OVERLAPPING HISTOGRAMS FROM THE UMBRELLA SAMPLING ACROSS THE COLLECTIVE VARIABLE. A COUNT IS CONSIDERED WHEN THE OVERLAP IS >10% OF THE TOTAL HISTOGRAM HEIGHT. THE RED LINE CORRESPONDS TO A ‘RULE OF THUMB’ FOR SUFFICIENT OVERLAP SET TO 3. (D) THE HISTOGRAMS FROM THE WHAM ANALYSIS, WHERE EVERY HISTOGRAM CORRESPONDS TO A SEPARATE SAMPLING POSITION OF LIPID A.....	190
FIGURE 4.21 LptA ₃ CDF PMF CONVERGENCE ANALYSIS. (A) PMF PROFILES OF INCREASING BLOCKS SIMULATION TIME BY 100 NS FOR UP TO ONE MICROSECOND. COLOUR SCALE FROM BLUE TO RED ILLUSTRATES HOW THE PROFILE EVOLVES THROUGH TIME. THE FIRST 200 NS WAS REMOVED FROM EACH WINDOW TO ENSURE THAT THE SYSTEM HAD FULLY EQUILIBRATED FIRST. (B) THE BINDING VALUES (THE LOWEST ENERGY VALUE) OF EACH PMF PROFILE OVER THE COURSE OF THE SIMULATION TIME FOR INCREASING BLOCKS OF 100 NS. ERRORS CALCULATED BY BOOTSTRAPPING EACH PMF PROFILE 10 TIMES. FOR INSTANCE, 200	

– 300 ns, 200 – 400 ns, ..., 200 – 1000 ns. (c) THE BINDING VALUES (THE LOWEST ENERGY VALUE) OF EACH PMF PROFILE OVER THE COURSE OF THE SIMULATION TIME FOR CONSECUTIVE BLOCKS OF 100 ns. FOR INSTANCE, 200 – 300 ns, 300 – 400 ns, ..., 900 – 1000 ns.....	190
FIGURE 4.22 LIPID A BINDING CONFORMATION WITHIN THE LPTA ₃ CDF COMPLEX. (A) PMF PROFILE OF LIPID A MOVING THROUGH LPTA ₃ CDF, AREAS SHADED IN GREY INDICATES THE ERROR AND IS OBTAINED BY BOOTSTRAPPING THE PROFILE 10 TIMES. CARTOON REPRESENTATION OF THE LPTA SUBUNIT ALIGNED WITH THE ENERGY PROFILE IN THE BACKGROUND FOR REFERENCE. (B) LIPID A BINDING POSE AT -6.0 nm, -2.5 nm, +0.5 nm, +2.5 nm, +6.0 nm, +8.0 nm, +10.0 nm AND AT +11.2 nm LABELLED ONTO THE LPTA ₃ CDF PMF BINDING PROFILE, WHERE FOR POSITIONS -6.0 nm, 0.5 nm AND +8 nm THE BINDING POSE HAS BEEN HIGHLIGHTED. THE RESIDUES WHICH INTERACT WITH THE PHOSPHATE GROUPS ARE LABELLED ONTO THE STRUCTURE.	193
FIGURE 4.23 BINDING ENERGY OF LIPID A WITH THE LPTA ₄ PORTION OF THE CHUTE. (A) PMF PROFILE OF LIPID A MOVING THROUGH LPTA ₄ , AREAS SHADED IN GREY INDICATES THE ERROR AND IS OBTAINED BY BOOTSTRAPPING THE PROFILE 200 TIMES. CARTOON REPRESENTATION OF THE LPTA ₄ TETRAMER COMPLEX IT ALIGNED WITH THE ENERGY PROFILE IN THE BACKGROUND FOR REFERENCE. (B) TOTAL COUNTS SUMMED ACROSS THE COLLECTIVE VARIABLE OF LIPID A BEING PULLED FROM THE B-TACO INTO SOLVENT FROM THE RAW HISTOGRAMS, WHERE THE DATASET IS NORMALISED TO 1. THE RED LINE IS A ‘RULE OF THUMB’ MARKER FOR ACCEPTABLE SAMPLING POSITIONED AT 0.2 OF TOTAL COUNTS. (C) THE NUMBER OF OVERLAPPING HISTOGRAMS FROM THE UMBRELLA SAMPLING ACROSS THE COLLECTIVE VARIABLE. A COUNT IS CONSIDERED WHEN THE OVERLAP IS >10% OF THE TOTAL HISTOGRAM HEIGHT. THE RED LINE CORRESPONDS TO A ‘RULE OF THUMB’ FOR SUFFICIENT OVERLAP SET TO 3. (D) THE HISTOGRAMS FROM THE WHAM ANALYSIS, WHERE EVERY HISTOGRAM CORRESPONDS TO A SEPARATE SAMPLING POSITION OF LIPID A.....	195
FIGURE 4.24 LPTA ₄ PMF CONVERGENCE ANALYSIS. (A) PMF PROFILES OF INCREASING BLOCKS SIMULATION TIME BY 100 ns FOR UP TO ONE MICROSECOND. COLOUR SCALE FROM BLUE TO RED ILLUSTRATES HOW THE PROFILE EVOLVES THROUGH TIME. THE FIRST 200 ns WAS REMOVED FROM EACH WINDOW TO ENSURE THAT THE SYSTEM HAD FULLY EQUILIBRATED FIRST. (B) THE BINDING VALUES (THE LOWEST ENERGY VALUE) OF EACH PMF PROFILE OVER THE COURSE OF THE SIMULATION TIME FOR INCREASING BLOCKS OF 100 ns. ERRORS CALCULATED BY BOOTSTRAPPING EACH PMF PROFILE 10 TIMES. FOR INSTANCE, 200 – 300 ns, 200 – 400 ns, ..., 200 – 1000 ns. (C) THE BINDING VALUES (THE LOWEST ENERGY VALUE) OF EACH PMF PROFILE OVER THE COURSE OF THE SIMULATION TIME FOR CONSECUTIVE BLOCKS OF 100 ns. FOR INSTANCE, 200 – 300 ns, 300 – 400 ns, ..., 900 – 1000 ns.....	197
FIGURE 4.25 BINDING ENERGY OF LIPID A WITH CONTINUOUS LPTA CHUTE. (A) PMF PROFILE OF LIPID A MOVING THROUGH LPTA ₄ , AREAS SHADED IN GREY INDICATES THE ERROR AND IS OBTAINED BY BOOTSTRAPPING THE PROFILE 200 TIMES. (B) TOTAL COUNTS SUMMED ACROSS THE COLLECTIVE VARIABLE OF LIPID A BEING PULLED FROM THE B-TACO INTO SOLVENT FROM THE RAW HISTOGRAMS, WHERE THE DATASET IS NORMALISED TO 1. THE RED LINE IS A ‘RULE OF THUMB’ MARKER FOR ACCEPTABLE SAMPLING POSITIONED AT 0.2 OF TOTAL COUNTS. (C) THE NUMBER OF OVERLAPPING HISTOGRAMS FROM THE UMBRELLA SAMPLING ACROSS THE COLLECTIVE VARIABLE. A COUNT IS CONSIDERED WHEN THE OVERLAP IS >10% OF THE TOTAL HISTOGRAM HEIGHT. THE RED LINE CORRESPONDS TO A ‘RULE OF THUMB’ FOR SUFFICIENT OVERLAP SET TO 3. (D) THE HISTOGRAMS FROM THE WHAM ANALYSIS, WHERE EVERY HISTOGRAM CORRESPONDS TO A SEPARATE SAMPLING POSITION OF LIPID A.	200
FIGURE 4.26 CONTINUOUS LPTA CHUTE PMF CONVERGENCE ANALYSIS. (A) PMF PROFILES OF INCREASING BLOCKS SIMULATION TIME BY 100 ns FOR UP TO ONE MICROSECOND. COLOUR SCALE FROM BLUE TO RED ILLUSTRATES HOW THE PROFILE EVOLVES THROUGH TIME. THE FIRST 200 ns WAS REMOVED FROM EACH WINDOW TO ENSURE THAT THE SYSTEM HAD FULLY EQUILIBRATED FIRST. (B) THE BINDING VALUES (THE LOWEST ENERGY VALUE) OF EACH PMF PROFILE OVER THE COURSE OF THE SIMULATION TIME FOR INCREASING BLOCKS OF 100 ns. ERRORS CALCULATED BY BOOTSTRAPPING EACH PMF PROFILE 10 TIMES. FOR INSTANCE, 200 – 300 ns, 200 – 400 ns, ..., 200 – 1000 ns. (C) THE BINDING VALUES (THE LOWEST ENERGY VALUE) OF EACH PMF PROFILE OVER THE COURSE OF THE SIMULATION TIME FOR CONSECUTIVE BLOCKS OF 100 ns. FOR INSTANCE, 200 – 300 ns, 300 – 400 ns, ..., 900 – 1000 ns.	202
FIGURE 4.27 LIPID A BINDING CONFORMATION WITHIN THE LPTA CONTINUOUS CHUTE. (A) PMF PROFILE OF LIPID A MOVING THROUGH LPTA CONTINUOUS CHUTE AREAS SHADED IN GREY INDICATES THE ERROR AND IS OBTAINED BY BOOTSTRAPPING THE PROFILE 200 TIMES. CARTOON REPRESENTATION OF THE LPTA ₄ TETRAMER ALIGNED WITH THE ENERGY PROFILE IN THE BACKGROUND FOR REFERENCE. (B) LIPID A BINDING POSE AT -6.0 nm, -2.5 nm, +11.2 nm LABELLED ONTO THE LPTA CONTINUOUS CHUTE PMF BINDING PROFILE.	203
FIGURE 4.28 BINDING ENERGY OF LIPID A WITH CONTINUOUS, FULLY LIPIDATED LPTA CHUTE. (A) PMF PROFILE OF LIPID A MOVING THROUGH A FULLY LIPIDATED LPTA ₄ , AREAS SHADED IN GREY INDICATES THE ERROR AND IS OBTAINED BY BOOTSTRAPPING THE PROFILE 10 TIMES. (B) TOTAL COUNTS SUMMED ACROSS THE COLLECTIVE VARIABLE OF LIPID A BEING PULLED FROM THE B-TACO INTO SOLVENT FROM THE RAW HISTOGRAMS, WHERE THE DATASET IS NORMALISED TO 1. THE RED LINE IS A ‘RULE OF THUMB’ MARKER FOR ACCEPTABLE SAMPLING POSITIONED AT 0.2 OF TOTAL COUNTS. (C) THE NUMBER OF OVERLAPPING HISTOGRAMS FROM THE UMBRELLA SAMPLING ACROSS THE COLLECTIVE VARIABLE. A COUNT IS CONSIDERED WHEN THE OVERLAP IS >10% OF THE TOTAL HISTOGRAM HEIGHT. THE RED LINE CORRESPONDS TO A ‘RULE OF THUMB’ FOR SUFFICIENT	

OVERLAP SET TO 3. (D) THE HISTOGRAMS FROM THE WHAM ANALYSIS, WHERE EVERY HISTOGRAM CORRESPONDS TO A SEPARATE SAMPLING POSITION OF LIPID A.	206
FIGURE 4.29 CONTINUOUS, FULLY LIPIDATED LPTA CHUTE PMF CONVERGENCE ANALYSIS. (A) PMF PROFILES OF INCREASING BLOCKS SIMULATION TIME BY 100 NS FOR UP TO ONE MICROSECOND. COLOUR SCALE FROM BLUE TO RED ILLUSTRATES HOW THE PROFILE EVOLVES THROUGH TIME. THE FIRST 200 NS WAS REMOVED FROM EACH WINDOW TO ENSURE THAT THE SYSTEM HAD FULLY EQUILIBRATED FIRST. (B) THE BINDING VALUES (THE LOWEST ENERGY VALUE) OF EACH PMF PROFILE OVER THE COURSE OF THE SIMULATION TIME FOR INCREASING BLOCKS OF 100 NS. ERRORS CALCULATED BY BOOTSTRAPPING EACH PMF PROFILE 10 TIMES. FOR INSTANCE, 200 – 300 NS, 200 – 400 NS, ..., 200 – 1000 NS. (C) THE BINDING VALUES (THE LOWEST ENERGY VALUE) OF EACH PMF PROFILE OVER THE COURSE OF THE SIMULATION TIME FOR CONSECUTIVE BLOCKS OF 100 NS. FOR INSTANCE, 200 – 300 NS, 300 – 400 NS, ..., 900 – 1000 NS.	207
FIGURE 4.30 LIPID A BINDING CONFORMATION WITHIN THE FULLY LIPIDATED LPTA CONTINUOUS CHUTE. (A) PMF PROFILE OF LIPID A MOVING THROUGH LPTA CONTINUOUS CHUTE AREAS SHADED IN GREY INDICATES THE ERROR AND IS OBTAINED BY BOOTSTRAPPING THE PROFILE 200 TIMES. CARTOON REPRESENTATION OF THE LPTA ₄ TETRAMER ALIGNED WITH THE ENERGY PROFILE IN THE BACKGROUND FOR REFERENCE. (B) LIPID A BINDING POSE AT -4.0 NM AND +0.8 NM LABELLED ONTO THE FULLY LIPIDATED LPTA CONTINUOUS CHUTE PMF BINDING PROFILE.	208
FIGURE 4.31 LPTA CONTINUOUS CHUTE. RENDERED IMAGES OF THE CHUTE TAKEN EVERY 90° ROTATION THROUGH THE CHUTE'S PRINCIPAL AXIS. RESTRICTIONS ALONG THE CHUTE ARE NOTED BY ARROWS AND BINDING POCKETS ARE HIGHLIGHTED IN CIRCLES.	209
FIGURE 4.32 OPENING LPTA WITH ALCHEMBED: A BEFORE (LEFT) AND AFTER (RIGHT) B-TACO OPENING OF THE LPTA MONOMER	216
FIGURE 5.1 CARTOON SCHEMATIC DESCRIBING THE LPS TRANSPORT PATHWAY. FIGURE ADAPTED FROM FIORENTINO <i>ET AL</i> ⁵¹ , ALL WORK HERE COMPLETED BY MYSELF AND FF. (A) A SKELETAL REPRESENTATION OF LPS WHERE ITS TRIPARTITE STRUCTURE IS DESCRIBED: LIPID A, CORE OLIGOSACCHARIDE AND THE O-ANTIGEN REPEATING UNIT. THE ORANGE BOX HIGHLIGHTS A DERIVATIVE OF LPS, CALLED 'RE-LPS' WHICH CONSISTS OF THE LIPID A UNIT CONNECTED TO TWO KETODEOXYOCTONIC ACID (KDO) MOIETIES. (B) CARTOON SCHEMATIC OF THE LPT SYSTEM: LPS IS EXTRACTED FROM THE INNER MEMBRANE (IM) BY THE ABC TRANSPORTER COMPLEX LPTB ₂ FGC, WHERE IT IS SUBSEQUENTLY SHUTTLED THROUGH THE PERIPLASM VIA A SOLUBLE BRIDGE OF LPTA UNITS TO BE FINALLY INSTALLED IN THE OM BY LPTDE. (C) THE LPTDE X-RAY CRYSTAL STRUCTURE (PDB ID: 5IV9) ³¹⁹ FROM <i>KLEBSIELLA PNEUMONIAE</i> . THE SOLUBLE B-TACO DOMAIN OF LPTD IS KNOWN TO INTERACT WITH LPTA TO FORM THE PERIPLASMIC BRIDGE AND THE B-BARREL ACCOMMODATES LPTD INSIDE TO FORM THE PLUG-AND-BARREL COMPLEX. INSET: CARTOON STRUCTURE OF THE ANTIMICROBIAL MOLECULE THANATIN (PDB ID: 5XO4).	222
FIGURE 5.2 SCHEMATIC DESCRIBING PROPOSED MECHANISM OF LPS INSTALLATION INTO THE OM. FIGURE ADAPTED FROM FIORENTINO <i>ET AL</i> ⁵¹ , ALL WORK HERE COMPLETED BY MYSELF AND FF. THE OPEN B-TACO FORMS AN INTRAMEMBRANE HOLE IN THE INNER LEAFLET OF THE OUTER MEMBRANE TO ALLOW THE LPS TAILS TO BE THREADED THROUGH AS THE HEADGROUP MOVES THROUGH THE LUMEN OF THE B-BARREL. THE LATERAL GATE IS PROPOSED TO BE OPEN TO ACCOMMODATE THE LIPID'S LATERAL EXTRACTION INTO THE OM.	223
FIGURE 5.3 CONFORMATIONAL DYNAMICS OF LPTDE WITH RE-LPS. FIGURE MODIFIED FROM FIORENTINO <i>ET AL</i> ⁵¹ , ALL WORK HERE COMPLETED BY COLLABORATORS (FF, XQ, SM, JRB, CVR). (A) LEFT: PEPTIDES ILLUSTRATING INCREASED DEUTERIUM UPTAKE (105–115, B-TACO) AND DECREASED DEUTERIUM UPTAKE (232–247, B-BARREL) AND THEIR CORRESPONDING LABELLING TIME (0.167–420 MIN) PLOTTED FOR APO-LPTDE (PURPLE CIRCLES) AND LPS-BOUND LPTDE (ORANGE SQUARES). HERE THE ERROR BARS SPECIFY STANDARD DEVIATION. ($N_{\text{BIOLOGICAL}} = 2$; $N_{\text{TECHNICAL}} = 3$). THESE PEPTIDES ARE MAPPED UPON THE LPTDE STRUCTURE ACCORDING TO THE DIFFERENCE IN DEUTERIUM UPTAKE AT THE 16.67-MIN LABELLING TIME, WHERE ONLY PEPTIDES WHICH DEMONSTRATE A SIGNIFICANT DIFFERENCE IN DEUTERATION ARE SHOWN. RESPECTIVELY, RED PEPTIDES EMPHASISE AN INCREASE IN HDX AND BLUE PEPTIDES REPRESENT A DECREASE. (B) MASS SPECTRA FOR PEPTIDES 105–115 ARE ILLUSTRATED HERE FOR THE 'APO' LPTDE AND 'LPS-BOUND' STATES. THE BEST FIT FOR THE SPECTRA IS TWO BINOMIAL ISOTOPIC ENVELOPES, PRODUCING LOW-MASS (GREEN) AND HIGH-MASS (LIGHT BLUE) POPULATIONS.	225
FIGURE 5.4 OVERVIEW OF PROTOCOL TO PRODUCE DIFFERENTIAL SOLVENT MAPS. FIGURE ADAPTED FROM FIORENTINO <i>ET AL</i> ⁵¹ , ALL WORK HERE COMPLETED BY MYSELF. (A) A RESIDUE-BY-RESIDUE DIFFERENTIAL SOLVENT UPTAKE MAP IS GENERATED BY SUBTRACTING OCCUPANCY VALUES FOR THE BOUND STATE AWAY FROM THE APO STATE, INCLUDING ALL <95% SIGNIFICANT VALUES. (B) EXPERIMENTALLY DETERMINED PEPTIDE COVERAGE OF B-TACO. EACH PEPTIDE IS COLOURED AND NUMBERED (1-9) FOR CLARITY. AVERAGING THE VALUES OF THE DIFFERENTIAL SOLVENT CONTACTS ACROSS THESE PEPTIDES (1-9) PRODUCES (C) WHICH IS TERMED THE 'PEPTIDE DIFFERENTIAL SOLVENT UPTAKE MAP' AND INCLUDES ALL <95% SIGNIFICANT VALUES. THE PEPTIDES WHICH HAVE FOUND TO BE AT LEAST 95% STATISTICALLY SIGNIFICANT ARE COLOURED IN RED, THOSE THAT DO NOT ARE SHOWN IN PINK OR LIGHT BLUE. (D) THE RESULTANT DIFFERENTIAL SOLVENT UPTAKE MAP AFTER FILTERING OUT PEPTIDES WHICH DO NOT PASS THE 95% CONFIDENCE INTERVAL TEST.	227
FIGURE 5.5 PREDICTED CONFORMATIONAL CHANGE ON BINDING TO RE-LPS. FIGURE ADAPTED FROM FIORENTINO <i>ET AL</i> ⁵¹ , ALL WORK HERE COMPLETED BY MYSELF. PREDICTED CONFORMATIONAL CHANGE ON BINDING TO RE-LPS. LEFT: APO 'CLOSED' STATE AND RIGHT: LIPID BOUND 'OPEN' STATE. CYSTEINE BONDS BETWEEN CYS7–CYS696 AND CYS149–CYS697 ARE REPRESENTED AS YELLOW SPHERES. RE-LPS COLOURED ORANGE AND SHOWN AS STICKS.	229

FIGURE 5.6 BLOCK CONVERGENCE ANALYSIS. FIGURE ADAPTED FROM FIORENTINO <i>ET AL</i> ⁵¹ , ALL WORK HERE COMPLETED BY MYSELF. (A) CONVERGENCE BLOCK ANALYSIS OF THE OPEN Re-LPS BOUND STATE WITHOUT PLUMED RESTRAINTS WHERE CONVERGENCE HAS CONSIDERED TO BE REACHED AT 150 NS. LPTDE AVERAGE SOLVENT CONTACT IS CALCULATED BY AVERAGING THE TOTAL SOLVENT CONTACTS OVER ALL THE RESIDUES OF THE PROTEIN. HERE, THE ERROR BARS REPRESENT THE STANDARD DEVIATION OVER THE FIVE REPEATS (B) CONVERGENCE BLOCK ANALYSIS OF THE OPEN STATE INCORPORATING PLUMED UPPER AND LOWER WALL RESTRAINTS UPON THE LATERAL GATE, WHERE A CONVERGENCE IS CONSIDERED AFTER 50 NS.	230
FIGURE 5.7 B-BARREL LATERAL GATE DISTANCE ANALYSIS. FIGURE ADAPTED FROM FIORENTINO <i>ET AL</i> ⁵¹ , ALL WORK HERE COMPLETED BY MYSELF. (A) THE DISTANCE OF THE LATERAL LPTDE B-BARREL EXIT GATE LPTDE WAS MEASURED OVER FIVE REPLICATES WITHOUT Re-LPS BOUND TO THE B-TACO. OVER THE SIMULATION PERIOD REPLICATE 1 LATERAL GATE DISTANCE WAS CONSISTENT OVER AFTER SAMPLING A DISTANCE ABOVE 2.25 NM. (B) THE LATERAL GATE DISTANCE OF THE B-BARREL WITHOUT (LEFT) AND WITH (RIGHT) PLUMED RESTRAINTS ³²³	232
FIGURE 5.8 DIFFERENTIAL SOLVENT PLOT OF LPTDE ON BINDING TO Re-LPS. FIGURE ADAPTED FROM FIORENTINO <i>ET AL</i> ⁵¹ , ALL WORK HERE COMPLETED BY MYSELF. (A) CARTOON SCHEMATIC REPRESENTING THE OPEN AND CLOSED STATES OF THE STRUCTURE DEPICTED WITH THEIR RESPECTIVE PROTEIN STRUCTURES WITH RESULTANT DIFFERENTIAL CONTACT MAP. (B) HERE THE DIFFERENTIAL CONTACT PLOT IS MAPPED UPON THE LPTDE STRUCTURE GENERATED FROM THE DIFFERENCE BETWEEN THE 'OPEN' AND 'CLOSED' STATE OF THE COMPLEX. SIMILAR TO THE EXPERIMENTAL HDX PLOTS, REGIONS WHICH ARE MORE EXPOSED TO SOLVENT ARE COLOURED RED WHERE REGIONS WHICH ARE SHIELDED AND LESS EXPOSED ARE COLOURED BLUE. THE LATERAL GATE OF THE PLOT IS HIGHLIGHTED HERE TO DRAW ATTENTION TO THE PARTICULARLY SIGNIFICANT RESIDUES WHICH CONTRIBUTE MOST THE PEPTIDE BEING DEPROTECTED. SIMULATION SNAPSHOT OF THE LATERAL GATE INTERACTING WITH Re-LPS(1-3) (REPRESENTED IN STICKS) WHICH CONTRIBUTES TO SOLVENT SHIELDING IS SHOWN. (C) LIKEWISE, FOR THE B-TACO, RESIDUES HIGHLIGHTED AS STICKS ARE PARTICULARLY INVOLVED IN INCREASING SOLVENT CONTACT FOR THE DOMAIN. A REPRESENTATIVE RENDERING OF Re-LPS BOUND WITHIN THE B-TACO IS ALSO DEPICTED HERE TO ILLUSTRATE HOW THE LIPID OPENS UP THE DOMAIN.	233
FIGURE 5.9 RESIDUE-BY-RESIDUE DIFFERENTIAL CONTACT MAP FOR LPTD PRODUCED BY THE DIFFERENCE BETWEEN 'OPEN' AND 'CLOSED' STATE. FIGURE ADAPTED FROM FIORENTINO <i>ET AL</i> ⁵¹ , ALL WORK HERE COMPLETED BY MYSELF. DIFFERENT ANGLES OF THE RESIDUE-BY-RESIDUE SOLVENT CONTACT MAP. RESIDUES HIGHLIGHTED IN STICKS HAVE PASSED A GREATER THAN 95% CONFIDENCE INTERVAL.	234
FIGURE 5.10 Re-LPS CONTACTS AND POSITION WITHIN THE B-TACO. FIGURE ADAPTED FROM FIORENTINO <i>ET AL</i> ⁵¹ , ALL WORK HERE COMPLETED BY MYSELF. (A) THE AVERAGE % OCCUPANCY OF Re-LPS (HEAVY ATOMS) CONTACTS MADE BETWEEN THE LIPID AND B-TACO WITHIN 4 Å. CONTACTS MADE BY THE DETERGENT IN THE 4Q35 STRUCTURE ARE MARKED ON WITH AN ASTERISK. 90% OF THESE CONTACTS CORRELATE WITH THOSE MADE BY Re-LPS. THIS SHOWS GOOD AGREEMENT BETWEEN Re-LPS DOCKED INTO THE B-TACO IN MY MODEL AND THE POSITION OF THE DETERGENT FROM THE STRUCTURAL DATA. (B) TOP: THE CENTRE OF MASS (COM) POSITION PLOTTED AS A KERNAL DENSITY ESTIMATE OVER THE FIVE REPEATS REVEAL MINIMAL MOVEMENT OF Re-LPS DURING THE SIMULATIONS. MIDDLE: SIMILARLY, PARTIAL DENSITY OF Re-LPS ACROSS THE LENGTH OF THE B-TACO REVEALS THAT THE LIPID POSITION IS CONSISTENT AND NON-DIFFUSE ACROSS THE FIVE REPLICATES. THIS ACCOUNTS FOR THE POOR SAMPLING OF Re-LPS WITHIN THE B-TACO WHERE THE LIPID DOES NOT SAMPLE DIFFERENT POSES WITHIN THE DOMAIN AND CONSEQUENTLY DOES NOT REPRODUCE FULLY THE DEPROTECTION OBSERVED IN EXPERIMENT (FIGURE 5.11). BOTTOM: CARTOON REPRESENTATION OF THE B-TACO FOR REFERENCE.	236
FIGURE 5.11 COMPARISON WITH THE Re-LPS CONTACTS ON THE B-TACO WITH THE DIFFERENTIAL SOLVENT PLOT. FIGURE ADAPTED FROM FIORENTINO <i>ET AL</i> ⁵¹ , ALL WORK HERE COMPLETED BY MYSELF. (A) STRUCTURE OF THE B-TACO INDICATING THE CONTACTS MADE WITH Re-LPS. THOSE RESIDUES HIGHLIGHTED IN STICKS CORRESPOND TO RESIDUES WHICH HAVE ALSO INCREASED IN THEIR EXPOSURE TO SOLVENT AND THEREFORE CONTRIBUTE TO THE DEPROTECTION OF THESE PEPTIDES. (B) AS BEFORE, TO COMPARE TO THE CONTACT MAP IN (A) THE RESIDUE DIFFERENTIAL CONTACT IS PLOTTED HERE WHERE ALL RESIDUES WITH LESS THAN 95% SIGNIFICANCE ARE VISUALISED. (C) THE EXPERIMENTAL PEPTIDE COVERAGE OF THE B-TACO. EACH PEPTIDE IS COLOURED AND NUMBER 1-9 TO DISTINGUISH BETWEEN AREAS WHICH HAVE BEEN AVERAGED OVER TO PRODUCE (D). THE PEPTIDES WHICH PASS A 95% CONFIDENCE INTERVAL TEST ARE FINALLY PLOTTED ON THE RESULTANT DIFFERENTIAL CONTACT MAP (E). HERE, I CAN OBSERVE THAT PEPTIDE 4 REVEALS NO OVERALL INCREASE IN AVERAGE SOLVENT CONTACT WHICH CAN BE RATIONALISED BY COMPARING WITH THE COMPARATIVELY LOW Re-LPS CONTACTS IN THIS AREA, WHICH HAS BEEN SHOWN TO INCREASE SOLVENT ACCESSIBILITY. SLIGHT INCREASE IN SOLVENT CONTACTS IN THE REGION MOSTLY LIKELY IS 'AVERAGED OUT' OVER 29 RESIDUES AND THEREFORE SHOWS NOT OVERALL INCREASE ACROSS THE PEPTIDE. LOW CONTACTS IN THIS REGION ARE ASSOCIATED WITH POOR SAMPLING OF Re-LPS WITHIN THE B-TACO WHICH HAS SHOWN TO BE RIGID OVER THE SHORT TIMESCALES OF THE SIMULATION (SEE FIGURE 5.10). ALSO, I OBSERVE THAT PEPTIDE 9 REVEALS A SMALL OVERALL PROTECTION WHICH IS ATTRIBUTED TO TRANSIENT INTERACTIONS WITH THE INNER LEAFLET OF THE OM (FIGURE 5.13).	238

FIGURE 5.12 RE-LPS CONTACTS MADE WITH LPTD. FIGURE ADAPTED FROM FIORENTINO <i>ET AL</i> ⁵¹ , ALL WORK HERE COMPLETED BY MYSELF. AVERAGE % OCCUPANCY OF CONTACTS MADE WITH RE-LPS AND LPTD OVER FIVE REPEATS. HERE, THOSE CONTACTS MADE WITH RE-LPS(1-3) REFERENCED IN (FIGURE 8B, 13D) ARE MARKED WITH AN ASTERISK.....	241
FIGURE 5.13 CONTACT MAP OF RE-LPS PLOTTED ON LPTDE STRUCTURE. FIGURE ADAPTED FROM FIORENTINO <i>ET AL</i> ⁵¹ , ALL WORK HERE COMPLETED BY MYSELF. (A) RE-LPS % OCCUPANCY CONTACTS AVERAGED OVER FIVE REPEATS AND MAPPED ONTO LPTDE (B) RESIDUES WHICH CONTACT WITH THREE RE-LPS LIPIDS REFERENCED IN FIGURE 5.8 (RE-LPS(1-3)) ARE REPRESENTED AS STICKS AND MAPPED ONTO BOTH THE CLOSED AND OPEN STATES OF THE B-BARREL FOR COMPARISON. (C) SIDE PROFILE VIEW OF LPTDE WHERE THE C-TERMINAL STRAND OF THE B-TACO INTERACTS WITH THE INNER LEAFLET OF THE OUTER MEMBRANE (FIGURE 5.11). (D) RE-LPS LIPIDS CONTACT RESIDUES 186 AND 187 ASSOCIATED WITH THE LATERAL GATE OF THE B-BARREL. THE LATERAL GATE APPEARS TO PULL RE-LPS LIPIDS (COLOURED YELLOW AND PINK) INTO THE B-BARREL WHICH CAUSES THEM TO SINK LOWER INTO THE INNER LEAFLET AND FACILITATES A TRANSIENT INTERACTION WITH THE TERMINAL STRAND OF THE B-TACO.	242
FIGURE 5.14 DISTANCE OF RE-LPS AWAY FROM LATERAL GATE. FIGURE ADAPTED FROM FIORENTINO <i>ET AL</i> ⁵¹ , ALL WORK HERE COMPLETED BY MYSELF. TOP: RENDERED IMAGE OF SIMULATION CAPTURING THE INTERACTION OF RE-LPS(1-3) WITH THE LATERAL GATE PEPTIDE (232-251). BOTTOM: DISTANCE FROM THE GEOMETRIC CENTRE OF THE PEPTIDE (232-251) TO THE GEOMETRIC CENTRE OF THE LIPIDS PLOTTED OVER THE FIVE REPEATS FOR RE-LPS (1, 2 AND 3). THE STANDARD DEVIATION IS SHADED GREY BETWEEN THE REPLICATES.	244
FIGURE 5.15 VOL-MAP DENSITY OF RE-LPS. DENSITY OF RE-LPS PLOTTED AS A VOL-MAP OVER THE FIRST HALF OF THE SIMULATION (0 - 150 NS) COLOURED IN PURPLE COMPARED TO THE DENSITY IN THE SECOND HALF OF THE SIMULATION (150 - 300 NS) COLOURED IN YELLOW. THE CIRCLE HIGHLIGHTS THE DENSITY OF THE RE-LPS LIPIDS ENGAGING WITH THE LATERAL GATE....	245
FIGURE 5.16 RESIDUE-RESIDUE DISTANCE PLOTS. FIGURE ADAPTED FROM FIORENTINO <i>ET AL</i> , ⁵¹ ALL WORK HERE COMPLETED BY MYSELF. (A) LEFT: SCHEMATIC REPRESENTATION OF THE APO CLOSED STATE (GREEN) AND THE RE-LPS BOUND STATE (PURPLE). RIGHT: PLOTS MEASURING THE DISTANCE BETWEEN RESIDUES MET240 AND MET247 FROM THE B1/B2 LOOP OF THE B-BARREL. (B) LEFT: RESIDUE DISTANCE PLOTS BETWEEN MET240 AND ASN215 (FROM THE B1/B2 LOOP), MET247 AND ASN215 AS WELL AS BETWEEN MET240 AND MET247 WHERE THEY ARE BEING DRAWN CLOSER TOGETHER. RIGHT: STRUCTURES OF THE OPEN AND CLOSED STATE CONFORMATION OF LPTD. HERE MET240 AND MET247 ARE COLOURED IN ORANGE AND ASN215 AND THE B1/B2 LOOP ARE COLOURED YELLOW.	246
FIGURE 5.17 HDX PLOT OF LPTDE BINDING TO THE INHIBITOR THANATIN. FIGURE MODIFIED FROM FIORENTINO <i>ET AL</i> ⁵¹ , ALL WORK HERE COMPLETED BY COLLABORATORS (FF, XQ, SM, JRB, CVR. AS BEFORE, REPRESENTATIVE PEPTIDES (27–45, B-TACO N TERMINUS; 105–115, B-TACO CENTRAL REGION) DEUTERIATION UPTAKE PLOTTED AS A FUNCTION OF LABELLING TIME FOR APO AND THANATIN-BOUND LPTDE (PURPLE CIRCLES AND GREEN SQUARES RESPECTIVELY). THE ERROR BARS REPRESENT STANDARD DEVIATION (NBIOLOGICAL = 2; NTECHNICAL = 3). THESE PEPTIDES ARE MAPPED UPON THE LPTDE STRUCTURE ACCORDING TO THE DIFFERENCE IN DEUTERIUM UPTAKE AT THE 16.67 MIN LABELLING TIME, WHERE ONLY PEPTIDES WHICH DEMONSTRATE A SIGNIFICANT DIFFERENCE IN DEUTERIATION ARE SHOWN. RESPECTIVELY, RED PEPTIDES EMPHASISE AN INCREASE IN HDX AND BLUE PEPTIDES REPRESENT A DECREASE.	249
FIGURE 5.18 STABILITY OF THANATIN BOUND TO THE CLOSED STATE OF LPTDE. FIGURE ADAPTED FROM FIORENTINO <i>ET AL</i> ⁵¹ , ALL WORK HERE COMPLETED BY MYSELF. (A) RMSD PLOT OF THANATIN POSITION AT N-TERMINAL (i) OF THE B-JELLYROLL WITHOUT RE-LPS BOUND (CLOSED CONFORMATION) COMPARED TO BEING POSITIONED AT THE C-TERMINAL (ii). HERE, STANDARD DEVIATION IS COLOURED GREY CALCULATED OVER THE FIVE REPLICATES. (B) LEFT: B-TACO VECTOR DRAWN THROUGH THE C _α OF RESIDUES ASP33 AND PRO190, WHERE X, Y DISPLACEMENT IS DEPICTED AS RED ARROWS. RIGHT: 2D KERNEL DENSITY PLOT OF THE X AND Y DISPLACEMENT OF THE B-TACO VECTOR MAPPED RELATIVE TO THE ORIGIN 0.0 NM, 0.0 NM (WHICH IS DEFINED AS THE 'RESTING' POSITION OF B-TACO IN THE CRYSTAL STRUCTURE). THE PLOT MAPS THE CONFORMATIONAL SPACE SAMPLED BY B-TACO IN DIFFERENT STATES, WHERE THE APO CLOSED STATE IS COLOURED PURPLE AND THE (i) BOTTOM AND (ii) TOP THANATIN DOCKED STATES ARE COLOURED GREEN AND YELLOW RESPECTIVELY.....	250
FIGURE 5.19 STABILITY OF THANATIN BOUND TO THE OPEN STATE OF LPTDE. FIGURE ADAPTED FROM FIORENTINO <i>ET AL</i> ⁵¹ , ALL WORK HERE COMPLETED BY MYSELF. (A) RMSD PLOT OF THANATIN POSITION AT N-TERMINAL (i) OF THE B-JELLYROLL WITH RE-LPS BOUND (OPEN CONFORMATION) COMPARED TO BEING POSITIONED AT THE C-TERMINAL (ii). HERE, STANDARD DEVIATION IS COLOURED GREY CALCULATED OVER THE FIVE REPLICATES. (B) LEFT: B-TACO VECTOR DRAWN THROUGH THE C _α OF RESIDUES ASP33 AND PRO190, WHERE X, Y DISPLACEMENT IS DEPICTED AS RED ARROWS. RIGHT: 2D KERNEL DENSITY PLOT OF THE X AND Y DISPLACEMENT OF THE B-TACO VECTOR MAPPED RELATIVE TO THE ORIGIN 0.0 NM, 0.0 NM (WHICH IS DEFINED AS THE 'STARTING' POSITION OF B-TACO IN THE OPEN STATE STRUCTURE). THE PLOT MAPS THE CONFORMATIONAL SPACE SAMPLED BY B-TACO IN DIFFERENT STATES, WHERE THE APO CLOSED STATE IS COLOURED PURPLE AND THE (i) BOTTOM AND (ii) TOP THANATIN DOCKED STATES ARE COLOURED GREEN AND YELLOW RESPECTIVELY.	252
FIGURE 5.20 DIFFERENTIAL CONTACT PLOT OF LPTDE BINDING TO THE INHIBITOR THANATIN. FIGURE ADAPTED FROM FIORENTINO <i>ET AL</i> ⁵¹ , ALL WORK HERE COMPLETED BY MYSELF. (A) CARTOON SCHEMATIC REPRESENTING THE APO CLOSED STATE AND THANATIN BOUND STATE OF THE COMPLEX DEPICTED WITH THEIR RESPECTIVE PROTEIN STRUCTURES AND RESULTANT DIFFERENTIAL CONTACT MAP. (B) TOP: THE FINAL FRAME OF THE SIMULATION WITH THANATIN BOUND TO THE N-TERMINAL OF THE B-TACO,	

WHERE THE HYDROGEN BONDS BETWEEN THE INHIBITOR (COLOURED GREEN) AND THE HYDROGEN BONDS ARE HIGHLIGHTED BY YELLOW DASHED LINES WITH THE RESIDUES DRAWN AS STICKS. BOTTOM: HERE THE DIFFERENTIAL CONTACT PLOT IS MAPPED UPON THE LptDE STRUCTURE GENERATED FROM THE DIFFERENCE BETWEEN THE APO CLOSED STATE AND THANATIN BOUND STATE OF THE COMPLEX. LIKE THE EXPERIMENTAL HDX PLOTS, REGIONS WHICH ARE MORE EXPOSED TO SOLVENT ARE COLOURED RED WHERE REGIONS WHICH ARE SHIELDED AND LESS EXPOSED ARE COLOURED BLUE. THE N-TERMINAL REGION OF THE B-TACO OF THE PLOT IS HIGHLIGHTED HERE TO DRAW ATTENTION TO THE PARTICULARLY SIGNIFICANT RESIDUES WHICH CONTRIBUTE MOST THE PEPTIDE BEING DEPROTECTED WHICH ARE THE SAME AS THOSE INVOLVED WITH THE HYDROGEN BONDING WITH THANATIN.	253
FIGURE 5.21 HDX PLOT OF LptDE BINDING TO THE INHIBITOR THANATIN AND Re-LPS. FIGURE MODIFIED FROM FIORENTINO <i>ET AL</i> ⁵¹ , ALL WORK HERE COMPLETED BY COLLABORATORS (FF, XQ, SM, JRB, CVR). AS BEFORE, REPRESENTATIVE PEPTIDES (105–115, B-TACO AND 232–247, B-BARREL) DEUTERIATION UPTAKE PLOTTED AS A FUNCTION OF LABELLING TIME FOR APO AND THANATIN-BOUND LptDE (PURPLE CIRCLES AND GREEN SQUARES RESPECTIVELY). THE ERROR BARS REPRESENT STANDARD DEVIATION (NBIOLOGICAL = 2; NTECHNICAL = 3). THESE PEPTIDES ARE MAPPED UPON THE LptDE STRUCTURE ACCORDING TO THE DIFFERENCE IN DEUTERIUM UPTAKE AT THE 16.67-MIN LABELLING TIME, WHERE ONLY PEPTIDES WHICH DEMONSTRATE A SIGNIFICANT DIFFERENCE IN DEUTERIATION ARE SHOWN. RESPECTIVELY, RED PEPTIDES EMPHASISE AN INCREASE IN HDX AND BLUE PEPTIDES REPRESENT A DECREASE.	254
FIGURE 5.22 DIFFERENTIAL CONTACT PLOT OF LptDE BINDING TO THE INHIBITOR THANATIN IN THE OPEN STATE. FIGURE ADAPTED FROM FIORENTINO <i>ET AL</i> ⁵¹ , ALL WORK HERE COMPLETED BY MYSELF. (A) THE DIFFERENTIAL SOLVENT CONTACT MAP GENERATED FROM THE DIFFERENCE BETWEEN THE SOLVENT CONTACTS OF THE OPEN Re-LPS BOUND STATE WITH AND WITHOUT THANATIN DOCKED TO THE BOTTOM OF THE B-TACO. (B) SIDE-BY-SIDE COMPARISON OF THE RESIDUE BY RESIDUE (LEFT) AND THE PEPTIDE (RIGHT) DIFFERENTIAL SOLVENT MAP OF THE B-TACO. RESIDUES WHICH CONTRIBUTE TO THE MOST SOLVENT DEPROTECTION HIGHLIGHTED AS STICKS.	256
FIGURE 5.23 DIFFERENTIAL CONTACT PLOT OF LptDE BETWEEN THE CLOSED STATE AND LptDE BINDING TO THE INHIBITOR THANATIN IN THE OPEN STATE. FIGURE ADAPTED FROM FIORENTINO <i>ET AL</i> ⁵¹ , ALL WORK HERE COMPLETED BY MYSELF. LEFT: CARTOON SCHEMATIC ILLUSTRATING THE CONFORMATIONAL CHANGE IN ADOPTING Re-LPS FROM THE CLOSED STATE, AND THEN FINALLY THE PROPOSED FULLY BOUND SUBSTRATE AND INHIBITOR OPEN STATE. RIGHT: HDX DIFFERENTIAL PLOT BETWEEN THE FULLY BOUND SUBSTRATE AND INHIBITOR OPEN STATE (STATE 2) AND THE CLOSED APO STATE (STATE 1). PEPTIDES HIGHLIGHTED ON B-TACO AND B-BARREL WHICH SHOW PROTECTION WHERE SIGNIFICANT RESIDUES THAT CONTRIBUTE TO SOLVENT SHIELDING ARE DRAWN AS STICKS.	257
FIGURE 5.24 KDE OF HYDROGEN BOND NUMBER. FIGURE ADAPTED FROM FIORENTINO <i>ET AL</i> ⁵¹ , ALL WORK HERE COMPLETED BY MYSELF. HYDROGEN-BOND NUMBER PLOTTED AS KERNEL DENSITY ESTIMATE (KDE) PER FRAME OF SIMULATION FOR EACH STATE OF THE B-TACO (CLOSED, OPEN, CLOSED WITH THANATIN AND OPEN WITH THANATIN). LEFT: THE HYDROGEN-BOND NETWORK PERTAINING TO THE B-TACO ITSELF. RIGHT: THE LOWER PLOT CONSIDERS THE HYDROGEN BONDS WHICH THE B-TACO MAKES WITH THE SOLVENT.	258
FIGURE 5.25 HYDROGEN BOND OCCUPANCY. FIGURE ADAPTED FROM FIORENTINO <i>ET AL</i> ⁵¹ , ALL WORK HERE COMPLETED BY MYSELF. (A) HYDROGEN BOND % OCCUPANCY OF THE B-TACO BETWEEN RESIDUES 28 TO 182 (I.E. NOT CONSIDERING THE TERMINAL STRAND WHICH FORMS HYDROGEN BONDS WITH THANATIN). THE % HYDROGEN BOND OCCUPANCY WAS AVERAGED OVER FIVE REPEATS FOR THE FOUR DIFFERENT STATES OF THE B-TACO (CLOSED, OPEN, CLOSED WITH THANATIN AND OPEN WITH THANATIN). (B) % DIFFERENCE OCCUPANCY PLOTS FOR HYDROGEN BONDS BETWEEN DIFFERENT STATES OF THE B-TACO. FROM TOP TO BOTTOM: THE PLOTS CORRESPOND TO THE DIFFERENCE BETWEEN THE CLOSED STATE OF THE B-TACO WITH AND WITHOUT THANATIN BOUND, THE OPEN STATE OF THE B-TACO WITH AND WITHOUT THANATIN BOUND, THE OPEN STATE WITH THANATIN BOUND VERSUS THE CLOSED STATE AND FINALLY THE DIFFERENCE BETWEEN THE OPEN AND CLOSED STATE. FOR EACH PLOT THE DASHED LINES REPRESENT THE ARBITRARY CUT-OFF AT 25% AND -25% OCCUPANCY DIFFERENCE.	261
FIGURE 5.26 HYDROGEN BOND POSITION. FIGURE ADAPTED FROM FIORENTINO <i>ET AL</i> ⁵¹ , ALL WORK HERE COMPLETED BY MYSELF. POSITION OF THE HYDROGEN BONDS WHICH OCCUPY MORE THAN 25% IN ONE STATE OF THE B-TACO COMPARED TO THE OTHER STATE AS REFERENCED IN THE DIFFERENCE PLOTS FIGURE 25B. TOP LEFT: DIFFERENCE BETWEEN THE CLOSED STATE OF THE B-TACO WITH AND WITHOUT THANATIN BOUND. TOP RIGHT: DIFFERENCE BETWEEN THE OPEN STATE OF THE B-TACO WITH AND WITHOUT THANATIN BOUND. BOTTOM LEFT: DIFFERENCE BETWEEN THE OPEN STATE WITH THANATIN BOUND AND THE CLOSED STATE. BOTTOM RIGHT: DIFFERENCE BETWEEN THE OPEN AND CLOSED STATE.	262
FIGURE 5.27 CARTOON SCHEMATIC OF LptDE CONFORMATIONS INDUCED BY ITS SUBSTRATE AND INHIBITOR. FIGURE ADAPTED FROM FIORENTINO <i>ET AL</i> , ⁵¹ ALL WORK HERE COMPLETED BY MYSELF AND FF. (i) BINDING OF THANATIN TO THE APO STATE OF THE COMPLEX CAUSES B-TACO STABILISATION. (ii) THE INTERACTION BETWEEN THE SUBSTRATE LPS AND LptDE OPENS THE B-TACO AND THEREFORE PRIES OPEN THE B STRANDS BETWEEN b26 AND b1 OF THE LATERAL GATE, COORDINATE BY CYSTEINE BONDS. THIS STRAND SEPARATION ACCOMMODATES AN LPS LIPID WITHIN THE B-BARREL. THIS CONFORMATIONAL CHANGE CREATES ENOUGH SPACE TO ACCOMMODATE AN LPS MOLECULE WITHIN THE B-BARREL FOR THE LIPID TO BE LATERALLY INSERTED INTO THE OM. THANATIN CAN ALSO BIND TO THE LPS BOUND STATE OF THE COMPLEX STABILISING THE B-TACO AND CAPTURING A	

NON-PHYSIOLOGICAL COMPLEX OF LptDE. THANATIN BINDING HERE COULD CONCEIVABLY BLOCK INTERACTIONS WITH LptA AND THEREFORE INHIBIT LPS TRANSPORT BY BREAKING THE FORMATION OF THE PERIPLASMIC BRIDGE.	266
FIGURE 5.28 ELECTRON DENSITY MAPS SHOWING DETERGENT MOLECULES WITHIN THE B-TACO. FIGURE ADAPTED FROM QIAO <i>ET AL.</i> ³²⁴ . ELECTRON DENSITY OF C ₈ E ₄ AND LDAO WITHIN THE HYDROPHOBIC GROVE OF THE LptD B-TACO.	269
FIGURE 5.29 COMPARISON OF OPEN B-BARREL LATERAL GATES. LEFT: BAM COMPLEX OPEN LATERAL GATE WITH THE OMP SUBSTRATE REMOVED FROM THE STRUCTURE FOR CLARITY (PDB: 6V05). RIGHT: FINAL OPEN <i>kplptDE</i> LATERAL GATE.	270
FIGURE 5.30 PIPELINE FOR PRODUCING DIFFERENTIAL SOLVENT PLOTS. (A) SOLVENT CONTACTS ARE CALCULATED FOR BOTH STATES OF THE PROTEIN (B) A DIFFERENCE MAP IS THEN GENERATED FOR EACH REPLICATE AND (C) AVERAGED WHICH CAN BE FILTERED USING A 95% CONFIDENCE INTERVAL TO PRODUCE (D) A RESIDUE-BY-RESIDUE DIFFERENTIAL SOLVENT PLOT. (E) ALTERNATIVELY, THE RESIDUES ARE THEN AVERAGED OVER THE EXPERIMENTALLY DETERMINED PEPTIDE MAP AND THEN SUBSEQUENTLY FILTERED BY A 95% CONFIDENCE INTERVAL TO PRODUCE (F) THE FINAL DIFFERENTIAL SOLVENT PLOT.	274
FIGURE 5.31 PEPTIDE COVERAGE. THE EXPERIMENTAL COVERAGE MAP IS SHOWN ABOVE FOR THE Re-LPS BOUND STATE, WHICH IS REPRESENTATIVE OVER ALL THE HDX COVERAGE MAPS FOR DIFFERENT STATES. THE MD COVERAGE MAP USED FOR PRODUCING THE DIFFERENTIAL COVERAGE MAPS ARE DRAWN ON TOP OF THE EXPERIMENTAL PEPTIDES. I CONSIDERED THE START AND END OF A PEPTIDE (CLEAVAGE SITES) WHEN THERE IS NO OVERLAP BETWEEN EXPERIMENTALLY DETERMINED PEPTIDES (MARKED WITH RED LINES).	275
FIGURE 5.32 DIFFERENTIAL SOLVENT WOODSPLOTS. SHOWN HERE IS AN EXAMPLE OF A 'WOODSPLOT' FOR THE DIFFERENTIAL SOLVENT PLOT OF THE Re-LPS BOUND OPEN STATE WITH THE CLOSED STATE OF THE PROTEIN ON AN INDIVIDUAL PEPTIDE (TOP) AND RESIDUE (BOTTOM) RESOLUTION.	276

TABLE 2.1 SUMMARY OF AT FORCEFIELDS. OVERVIEW OF AT FORCEFIELDS DISCUSSED IN THE MAIN TEXT WITH ASSOCIATED ADVANTAGES AND DISADVANTAGES.	74
TABLE 2.2 SUMMARY OF CG FORCEFIELDS. OVERVIEW OF CG FORCEFIELDS DISCUSSED IN THE MAIN TEXT WITH ASSOCIATED ADVANTAGES AND DISADVANTAGES.	78
TABLE 3.1 PDB IDs FOR LPT COMPLEX STRUCTURES. PROTEIN STRUCTURES USED DIRECTLY IN SIMULATIONS OR USED TO GENERATE THE HOMOLGY MODELS OF THE DIFFERENT SUBCOMPLEXES OF THE LPT SYSTEM.	150
TABLE 4.1 PDB IDs FOR LPT COMPLEX STRUCTURES. PROTEIN STRUCTURES USED DIRECTLY IN SIMULATIONS OR USED TO GENERATE THE HOMOLGY MODELS OF THE DIFFERENT SUBCOMPLEXES OF THE LPT SYSTEM.	214
TABLE 5.1 PDB IDs FOR LPT COMPLEX STRUCTURES. PROTEIN STRUCTURES USED DIRECTLY IN SIMULATIONS OR USED TO GENERATE THE HOMOLGY MODELS OF THE DIFFERENT SUBCOMPLEXES OF THE LPT SYSTEM.	268

CHAPTER 1 Introduction

1.1 Membrane and lipid biology

Proteins are large, complex macro-molecules which are comprised of long chains of smaller building blocks called amino acids. There are 20 different types of amino acid which make up proteins. The sequence of these amino acids determines the protein's particular three-dimensional structure and function¹.

Proteins are fundamental to crucial cellular pathways which regulate the viability of living organisms. They are intimately involved in all aspects of biological life such as: catalysis of biomolecular reactions, gene expression regulation, cellular signalling and transport of essential metabolic molecules². Commonly, proteins associate together to form larger complexes, where modulation of activity is controlled by small molecules binding (cofactors or substrates) or post-translational modifications (PTMs) which affects complex formation and conformation³⁻⁵. For example, small molecules such as zinc ions and ATP enhances the ability of the eye lens protein α -crystallin acting as molecular chaperone to maintain the transparency of the eye lens as they help stabilise the structure of the protein³.

Within a cell, membranes are composed of lipid bilayers which feature abundant and varied proteins, such as embedded 'integral' membrane proteins or 'peripheral' membrane proteins attached to the inner or outer leaflet of the membrane⁶. Integral and peripheral membrane proteins can be associated with the lipid bilayer in different ways. Integral membrane proteins extend across the bilayer, having regions which are

hydrophobic which pass through the membrane and interact with the lipid tails of the bilayer as well as regions that are hydrophilic and are exposed to water on the other side of the bilayer. Peripheral membrane proteins primarily in the soluble compartment of the cell but are anchored to bilayer by an amphipathic helix or by one or more covalently attached lipids. Membrane proteins stand out amongst the other classes of protein as they are prevalent on the outer surfaces of the cell and therefore have important responsibilities in maintaining cellular homeostasis control and cell integrity⁷ (Figure 1.1).

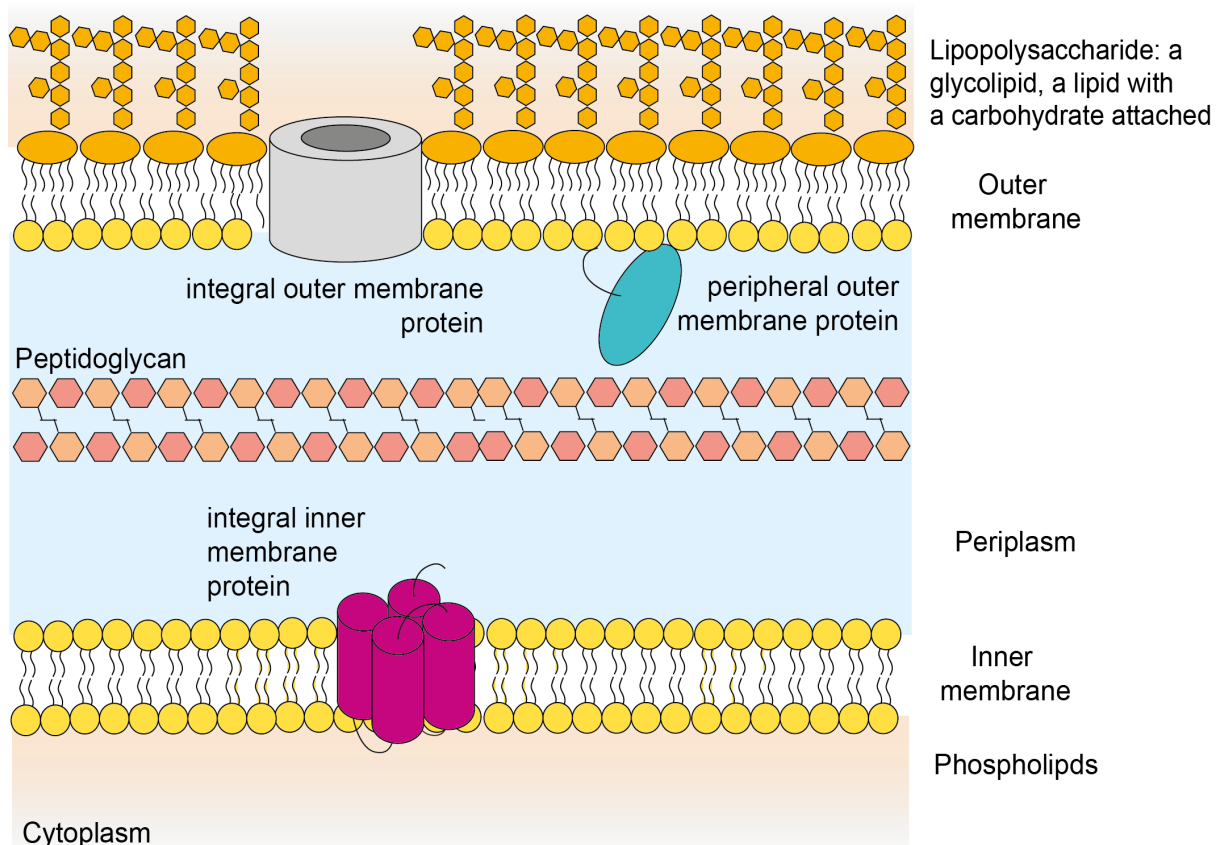


Figure 1.1 Schematic of the prokaryotic Gram-negative cell membrane. The membrane is a dynamic entity, comprising proteins interacting with different types of lipids

1.1.1 The significance and diversity of membrane proteins

In contrast to soluble proteins (and to a similar extent peripheral membrane proteins), integral membrane proteins feature surface hydrophobic residues which contact the surrounding lipid environment. These residues comprise 'transmembrane' domains or TMDs and possess either an α -helical or β -barrel type structure⁶. Soluble domains are often connected to these TMDs and can reside on the cytoplasmic or periplasmic sides of the lipid bilayers (if in bacteria) or out into extracellular space⁸. Depending on their biological function, membrane proteins adopt specific structures to carry out their roles which can either be described as: channels, receptors, enzymes, or transporters⁹ (**Figure 1.2**).

An example of membrane protein channels are those which mediate the passage of ions across cell membranes, which is essential for maintaining cellular homeostasis as well as cell signalling. They are classified accordingly as ligand-gated, voltage-gated or mechanosensitive channels¹⁰ and their activity can be modulated by their lipid environment¹¹⁻¹³. For example the mechanosensitive channel of large conductance (MscL) from *Escherichia coli* (which opens in response to membrane tension) was modulated by the direct interaction of the surrounding phospholipid environment as shown by fluorescent studies¹⁴.

In terms of receptors, the most prevalent of which exist in eukaryotic cells are represented by G protein-coupled receptors (GPCRs). These membrane proteins have a distinct structure featuring seven trans-membrane helices and react to a diverse set of stimulants including: light, other proteins, small molecules, and peptides.

Furthermore, work on membrane protein crystallography and GPCRs lead to at least two Nobel Prizes in chemistry in 1988 and 2012¹⁵.

Intramembrane proteases are enzymes embedded within the membrane which facilitate hydrolysis of peptides and proteins in the cell membrane itself. Furthermore, they are also involved in responding to egregious protein aggregation as well cellular differentiation, peptide signalling and lipid metabolism¹⁶.

Membrane protein transporters utilise either potential gradients, counter ions or ATP hydrolysis to drive the movement of molecules through the membrane. Substrates can vary widely from small molecules such as amino acids, carbohydrates and drugs^{17,18} to large lipid species such as lipopolysaccharides¹⁹⁻²¹.

Due to the vital roles membrane proteins play in microbial and human cells, more than 60% of current therapeutics target them^{22,23}. Regarding pathogenic bacteria and cancer cells, transporter membrane proteins called efflux pumps have evolved to expel drugs and therefore lie behind the prevalence of multidrug resistance development²⁴⁻²⁶. Furthermore, membrane proteins have been shown to be present in viral envelopes promoting infection²⁷. Other viruses, such as SARS-CoV-2 (which causes coronavirus disease 2019 or COVID-19), bind to eukaryotic membrane proteins using their spike glycoproteins to gain entry into the cell²⁸.

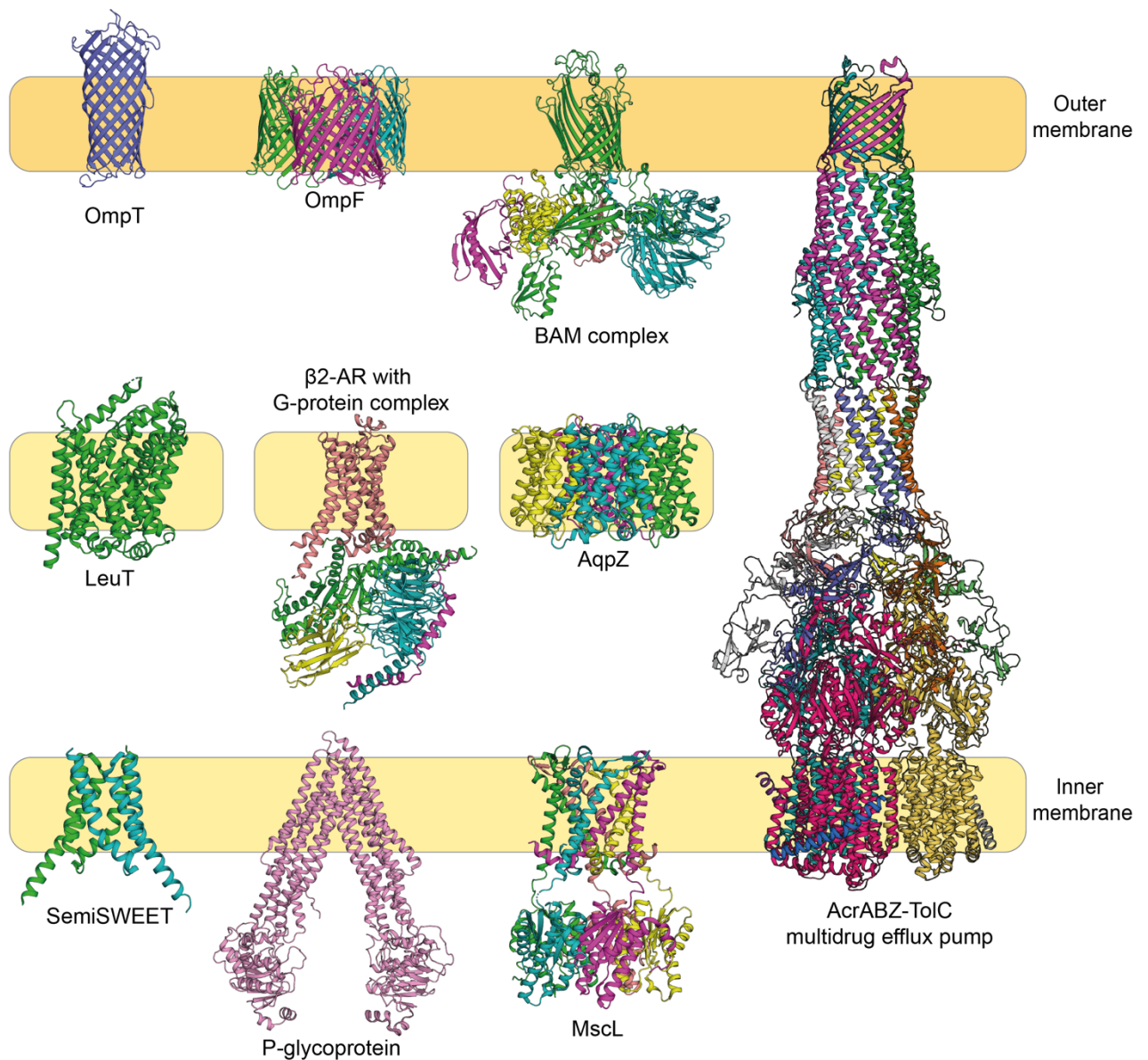


Figure 1.2 examples of membrane proteins. OmpT (PDB ID: 1I78), OmpF (PDB ID: 3K1B) and the BAM (PDB ID: 5LJO) complex are examples of β -barrel outer membrane proteins. AcrABZ-TolC (PDB ID: 5O66) as well as P-glycoprotein (PDB ID: 4KSC) is an multi drug efflux pump. LeuT (PDB ID: 3F3A) and SemiSWEET (PDB ID: 4X5M) are responsible for transporting leucine and sugars. MscL (PDB ID: 4Y7K) is a mechanosensitive channel and AqpZ (PDB ID: 2ABM) is a water channel. β 2-AR is an a GPCR protein in complex with its G protein (PDB ID: 3SN6).

All membrane proteins share a relationship with their lipidic environment, a relationship which itself dramatically varies depending on location and function of the membrane protein as well as the types of lipids within the environment. I have briefly outlined above the significance of membrane protein function, due to their many and diverse roles in cells and therefore the importance of studying their structure and function. However, due to their hydrophobic and often amphipathic nature, experimental and structural work is challenging, which I will describe in more detail below.

Over the last few decades, the primary tools for determining atomic resolution structures of membrane proteins have been NMR, X-Ray diffraction and advances in cryo-electron microscopy (cryo-EM) instrumentation and methodologies²⁹. The first membrane protein structure was obtained in 1975 of bacteriorhodopsin, through pioneering electron microscopy studies³⁰. In 1985, the first crystal structure of a membrane protein, the photosynthetic reaction centre from *Rhodospseudomonas viridis* was obtained by Michel and co-workers³¹. By 2000, there was 26 unique structures of integral membrane proteins in the PDB (most of them β -barrel structures). Here, studies were dependent on proteins found in natural high abundance and each representing years of effort by dedicated teams²⁹. In the late 2000s GPCR engineering work by Rosenbaum *et al.* constructed a β_2 -adrenergic receptor (β_2 AR) fusion protein in which T4 lysozyme (T4L, an abundant antimicrobial enzyme which crystallises very easily) replaces most of the third intracellular loop of the GPCR yielding a high-resolution structure of the receptor³². Over the last few decades, improvements in methods in integral membrane protein expression, solubilization, and stabilization as well as advances in cryo-EM, has led to a large

increase in the number of solved integral membrane protein structures. As of November 2021, there are 5891 integral membrane proteins structures found in the PDB³³.

Membrane proteins are often expressed in low yields compared to their soluble counterparts. This is mainly due to the limited expression space in the membrane itself (i.e. where membrane proteins folding occurs) compared to the soluble compartments within the cell³⁴. Low expression yields are a major limitation for membrane protein characterisation and structural work^{35,36}, especially in the structural techniques like X-ray crystallography which often requires several milligrams of protein material for the crystallisation process. As mentioned earlier, membrane proteins are inherently amphipathic which makes resolving certain structures via X-ray crystallography and cryogenic-electron microscopy (cryo-EM) techniques difficult³⁷. As briefly remarked upon, membrane proteins require a hydrophobic environment and will often aggregate in solution, therefore, an amphipathic system for solubilisation is needed to preserve their native fold³⁸⁻⁴⁰. Traditionally (and generally), detergents have been the method of choice to solubilise membrane proteins^{41,42}. Detergents have a hydrophobic chain and a hydrophilic headgroup and form micelles in solution due to their conical-like geometry, which are formed when the concentration of detergent reaches the critical micelle concentration (CMC) and is in equilibrium with individual detergent molecules. The tails of the detergent align with the hydrophobic domains of membrane proteins, thereby solubilising the protein by shielding it from solvent⁴³. It does this in a similar manner to the monolayer arrangement of detergents between the air-water interface, where the detergent molecules form a monolayer at the hydrophobic protein surface. However, due to the irregularity of the transmembrane segment of the membrane

protein coupled with the detergent's energetic drive to avoid hydrophilic-hydrophobic contacts with water, detergent molecules will form a 'wedge' shape in a ring around the hydrophobic membrane sector⁴⁴. Often, in this process endogenous lipids may be solubilised and captured (co-purify) with membrane proteins giving rise to the potential of probing lipid-membrane protein interactions^{40,45}. However, this must be caveated, as the detergent environment does not fully replicate the native bilayer environment as much of the dynamic native lipid annulus is removed and isolated by the detergent mimetic. Furthermore, (as described above) detergents forming a non-native wedge ring around hydrophobic sections of the membrane protein, which are not reflective of the physiochemical properties of a bilayer environment. Therefore, alternative solubilising agents may be required to study these interactions in more detail such as nanodiscs or liposomes, which employ small patches of lipid bilayers to capture membrane proteins for study^{46,47}.

1.1.2 Lipopolysaccharide and Bacterial membranes

Bacteria are traditionally classified according to the Gram staining technique as either Gram-negative or Gram-positive, which are differentiated by their cell wall features. Gram-negative bacteria possess two membranes, an inner membrane (IM) and an outer membrane (OM), with a thin peptidoglycan layer (seven to eight nanometres) between the two (**Figure 1.3**)⁴⁸. Gram-positive bacteria feature a cytoplasmic membrane with a much larger peptidoglycan wall of around 20 to 80 nanometres⁴⁸⁻⁵⁰.

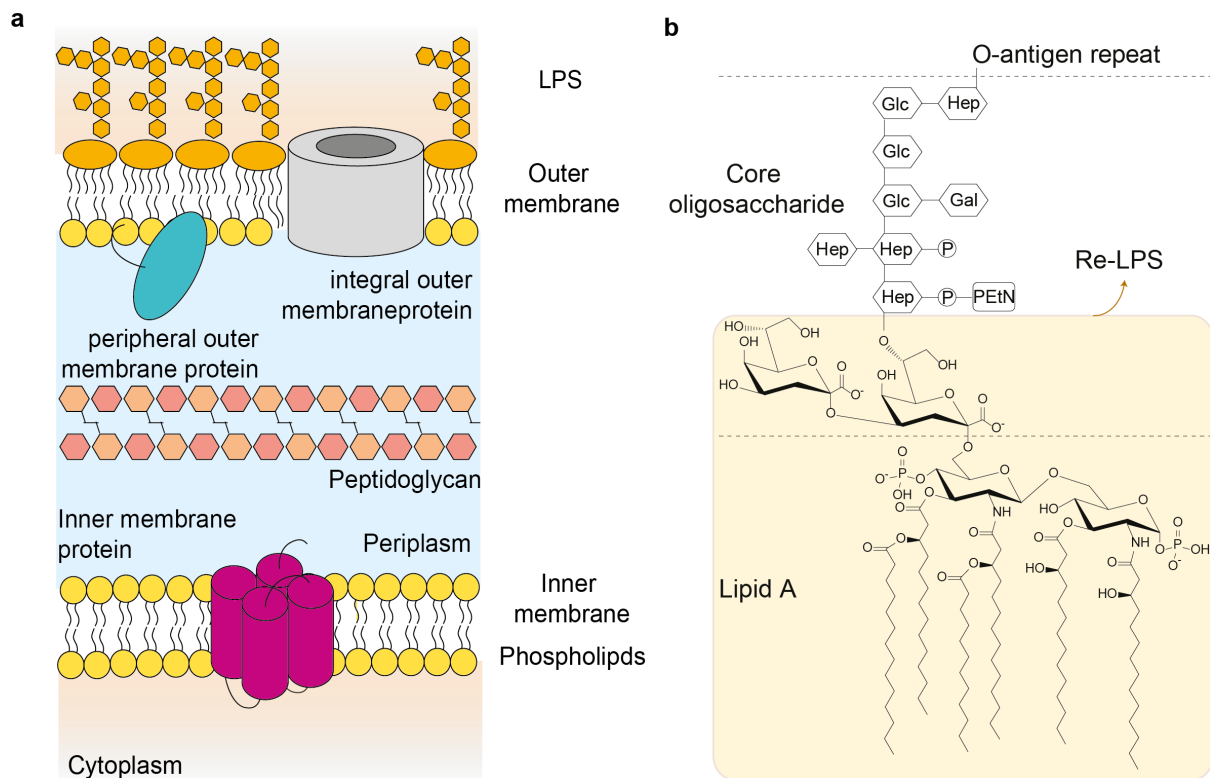


Figure 1.3 Schematic of Gram-negative bacterial membrane, figure modified from Florentino et al⁵¹. (a) Cartoon schematic of Gram-negative bacteria membrane depicting the inner membrane shielded by the peptidoglycan layer. The outer membrane's outer leaflet feature LPS exclusively. (b) A skeletal structure of LPS, organised in three parts: lipid A, lipid A glycosylated with two keto-deoxyoctulosonate moieties (lipid A-kdo₂ or Re-LPS) including the rest of the core-oligosaccharide and the O-antigen repeat unit.

The common model for microbiological investigation for Gram-negative species is *Escherichia coli* due to several factors including such as the diversity of its genome its metabolic flexibility, substrate breadth, growth rate in chemically defined media, and its ability to develop resistance against anti-biotics for the study of bacterial persistence⁵². The membrane properties of *E. coli* have been extensively studied. Here, the predominant lipid species within the membrane are phospholipids (PLs), of which, three types have been reported: the zwitterionic phosphatidylethanolamine (PE, ~75% of membrane lipids), and the anionic lipids phosphatidylglycerol (PG, ~20%

of membrane lipids) and ~5% cardiolipin (CDL)^{53,54}. It is important to note that these compositions are dynamic where lipids can be localised and sub compartmentalised depending on genetic or environmental factors such as cell wall curvature, lipid interactions with membrane proteins or the presence of antibodies interacting with the cell^{7,54}. Gram-negative bacteria have an additional and significant lipid type which is predominantly situated on the outer leaflet of the OM, a glycolipid known as lipopolysaccharide (LPS). In fact, ~75% of the cell surface is covered in LPS which is roughly 2×10^6 LPS molecules present in Gram-negative bacteria⁵⁵. The OM is also populated with beta-barrel OMPs, which can cluster together in supramolecular assemblies, as shown by studies conducted with fluorescent labelling⁵⁶. Furthermore, cells with OM defects or are exposed to harsh chemical environments accumulate PLs in the outer leaflet of the OM which break the OMP networks to form new domains in the outer leaflet. However, in healthy cells, the OM is a mosaic of phase-separated LPS-rich and OMP-rich regions⁵⁷. The membrane protein complex which handles the transport of LPS within the cell is the subject matter of this thesis. Below, I describe the structure, function and transportation of this essential glycolipid to many species of pathogenic bacteria (**Figure 1.3b**).

1.1.2.1. LPS structure

LPS is a glycolipid structured in three parts: lipid A, core oligosaccharide (OS) and the O-antigen. Lipid A is the hydrophobic portion of the molecule lining the OM and is an acylated β -1'-6-linked glucosamine disaccharide⁴⁹. In *E. coli* the glucosamines of lipid A are acylated at the 2, 3, 2' and 3' positions and phosphorylated at the 1 and 4' positions⁴⁹ (**Figure 1.3b**). Two secondary acyl chains are also present on the more distal glucosamine so that lipid A has six acyl tails in total⁴⁹. The core-oligosaccharide

portion of LPS is formed of non-repeating saccharide units bonded to the lipid A glucosamines, which are typically 3-deoxy-D-*manno*-oct-2-ulosonic acid (Kdo) residues, heptoses, and various hexoses^{49,58}. This core-oligosaccharide portion can be further modified with phosphates and phosphoethanolamines^{49,58-60}. The O-antigen is by far the largest portion of LPS and is composed of a repeating oligosaccharide unit composed of two to eight sugars⁶¹⁻⁶³.

LPS has a generally well conserved structure, though there are many subtleties and variations across species and strains^{49,58-60,64,65}. The lipid A section of LPS can undergo acyl chain chemical modifications under environmental conditions (such as temperature or the amount of cations and positively charged antimicrobials present) but is largely conserved at the species level whereas the core-oligosaccharide varies largely, even among strains themselves^{49,58-60,66-68}. Here, modifications include the addition of various sugar groups such as PtEN which is mediated via the enzyme EptB to the first Kdo sugar, or the enzyme WaaP for the addition to the first heptose of the inner core⁶⁹. The O-antigen is by far the most diverse portion of LPS^{58,62}, and in some cases not even synthesised by some Gram-negative bacteria^{58,70}. Without the O-antigen, LPS is known as 'rough' whereas 'smooth' LPS contains the O-antigen⁵¹.

1.1.2.2. LPS function

The main function of LPS is to form a permeability barrier against toxic molecules such as antibiotics and bile salts to protect the cell^{71,72}. As LPS is the primary point of contact with the host immune system it is also the main source of bacterial pathogenicity^{68,72}.

In terms of the relative sizes of the hydrophobic and hydrophilic portions (compared to other lipids species, like phospholipids, in the IM and OM) LPS is a strongly amphipathic; where the lipid A portion (with six acyl tails) residing in the OM provides a barrier to hydrophilic molecules and the core-oligosaccharide with the O-antigen (often 1000s saccharide units in length) provides a huge hydrophilic barrier to any hydrophobic compounds^{71,72}. Furthermore, LPS packs together very efficiently in a highly ordered manner, where X-ray diffractions studies of isolated LPS molecules reveal that the saturated tails are in a near crystalline arrangement (the diffraction of this packing show a sharp 4.25 Å reflection compared to more diffuse 4.5 Å reflection of liquid crystalline lipids)⁷³. The six acyl tails of lipid A hydrophobically associates with other LPS tails where extensive saturated interactions here cause low fluidity in the OM⁷¹. Packing of the negatively phosphorylated groups at the 1 and 4' positions are mediated by divalent cations such as Mg^{2+} which intercalates between lipids enhancing the overall LPS interactions^{71,72}.

LPS features on the surface of many bacterial pathogens, and therefore, the host-immune system has evolved to respond to its presence to such an extent that LPS is recognised as a pathogen associated molecular pattern (PAMP)^{68,72,74}. Here the immune system response may be so drastic as to be fatal to the host. Hence, LPS has been termed an 'endotoxin' due to its observed cell associated (or endo) toxicity. The immune system has developed to target the most conserved element of the molecule, lipid A, which it does primarily through the TLR4 (toll-like receptor 4)^{68,72,74}. As alluded to above, lipid A exhibits some level of diversity, whereas the O-antigen portion demonstrates a huge level of variety and therefore different LPS molecules have a varying ability to trigger an immune response⁶⁸. In fact, evolution of less immunogenic

lipid A, helps pathogens to avoid an immune response, such as *Yersinia pestis* (which caused the bubonic plague), reduced lipid A tail acylation at mammalian physiological temperatures⁷⁵. Alternatively, some bacteria use a highly variable O-antigen to evade the host-immune response of the more conserved lipid A motif by stimulating the production of antibodies for the O-antigen motif, where the length of the O-antigen chain extends to prevent the deposition of antibodies on the cell surface⁶⁹. Furthermore, possession of an O-antigen unit itself helps pathogens to avoid phagocytosis by immune cells^{58,76}. It is abundantly clear that understanding pathways involved in LPS biogenesis and system involved in LPS transport and translocations are of paramount importance for the development of novel antibacterials.

1.1.2.3. LPS biogenesis

The biosynthetic pathway for LPS is complex and is summarised here^{49,66,77}. Synthesis occurs in three distinct cellular compartments: the cytoplasm, IM and periplasm and involves the convergence of separate pathways in an orchestrated fashion, much like a factory assembly line. For the synthesis of lipid A with its core-oligosaccharide at the inner leaflet of the IM, three convergent biosynthetic pathways are required the: (i) lipid A portion (known as the Raetz pathway), (ii) the Kdo sugars and (iii) the rest of the core-oligosaccharide (**Figure 1.4**). The first stage of the Raetz pathway is the LpxA enzyme acylating the sugar nucleotide UDP-N-acetylglucosamine (UDP-GlcNAc) on the carbon atom 3 of the sugar, which is then followed by LpxC (a zinc metalloenzyme) mediated deacylation of the amide attached to the carbon atom 2. This deacylation stage in particular has been a target of novel antibiotics^{78,79}, such as the development of a substrate-analogue inhibitor TU-514 which potently complexes with its acyl tail with LpxC as revealed in NMR solution structure studies⁷⁸. Following this, the

successive action of acetyltransferase LpxD, pyrophosphatase LpxH and disaccharide synthase LpxB results in the production of a lipid IV_A moiety, which is the β -1-6 disaccharide portion connected to four acyl tails^{49,66,77}.

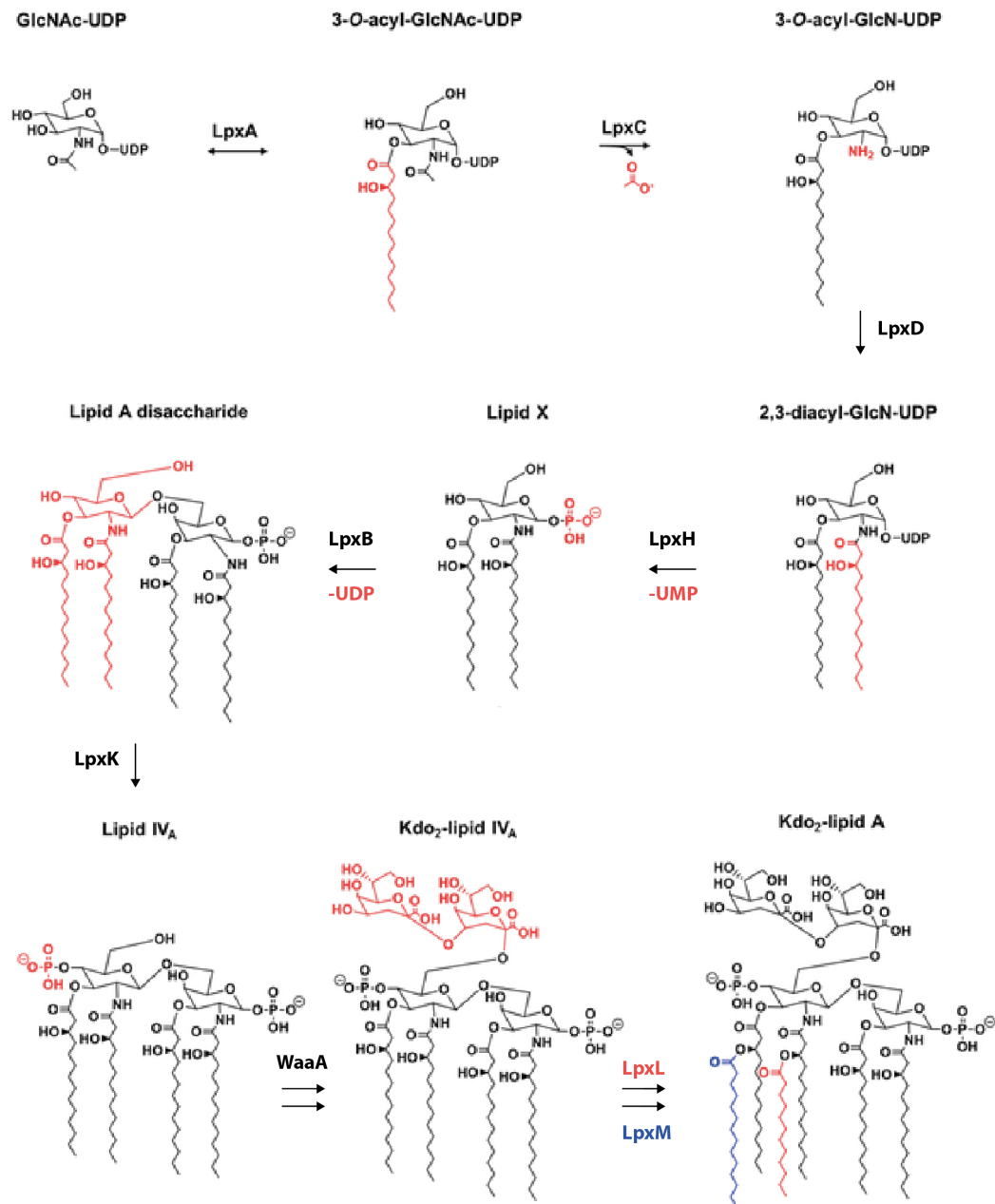


Figure 1.4 The Re-LPS biosynthesis pathway. The modifications to the proceeding structure are made in red. Modifications that are made by LpxL and LpxM are labelled in red and blue respectively. This figure is modified from Bertani *et al.*⁶⁹.

A second pathway generates the Kdo sugars, where the first reaction in the pathway is catalysed by KdsD, which has also been the target of novel antibacterials such as

the development of various 2-deoxy analogues of Kdo exhibiting inhibition of the pathway⁸⁰. The addition of the Kdo moieties to lipid IV_A is then mediated by the CMP-Kdo transferase⁸¹ to produce lipid IV_A-Kdo₂. Subsequently the hexa-acylated lipid A-Kdo₂ moiety is produced by the activity of LpxL and LpxM^{49,66,77}. The rest of the sugar units comprising the core-oligosaccharide are attached using glycosyl-transferases to produce the lipid A-core oligosaccharide portion of LPS⁵¹. This section is flipped across the IM from the cytoplasm to the periplasm by the ABC-transporter MsbA.

The components of the O-antigen unit are synthesised in the cytoplasm and are similarly flipped across the IM and are attached there by the lipid carrier undecaprenyl diphosphate. Here, subsequent extension of the O-polysaccharide chain can occur via *wzy*- or synthase-dependent pathways⁸¹. Finally, the WaaL ligase attaches the O-antigen unit to the lipid A-core oligosaccharide, resulting in a fully matured LPS molecule. From here, the Lpt system machinery transports LPS from the IM to the OM as detailed below.

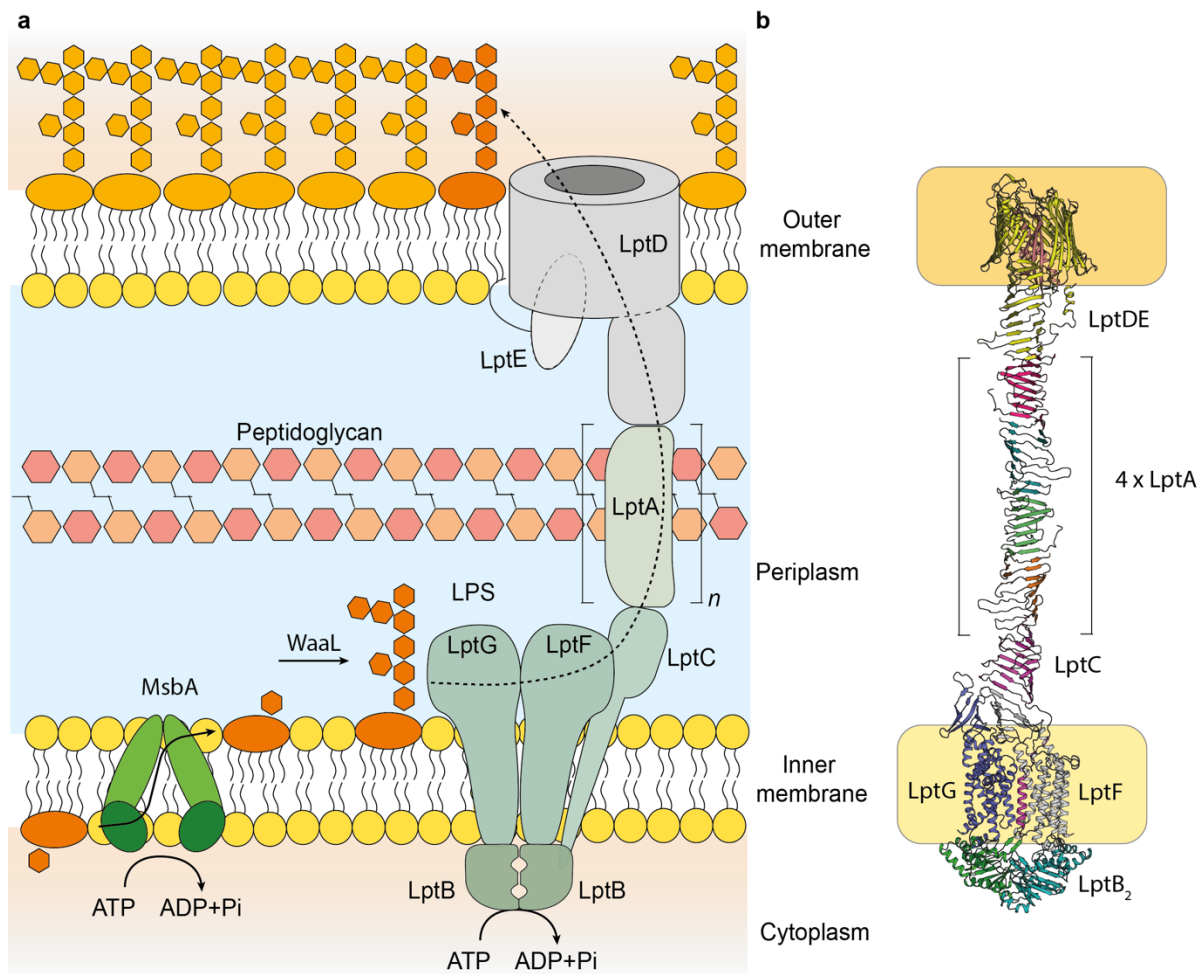


Figure 1.5 Lpt system schematic. (a) Cartoon depicting the Lpt system spanning two bilayers. LPS is initially synthesised on the inner leaflet of the IM and is translocated by the ABC transporter MsbA to the periplasmic side. The transport of LPS from the IM to the OM is handled by the Lpt system, consisting of LptA-G. LPS is extracted from the IM by LptB₂FG, handed to LptC which forms a periplasmic bridge with the soluble protein LptA and the N-terminal region of LptD. LPS moves through this bridge where it then reaches the OM and inserted by the translocon complex LptDE. **(b)** Crystal structure of the Lpt system comprising: *EcLptB₂FGC* homology model from *E. cloacae* (PDB ID: 6MIT) as well as the *EcLptDE* homology model of *KpLptDE* (PDB ID: 5IV9). LptA connects the two subcomplex (PDB ID: 2R19) (where four units are arbitrarily chosen here). The chute is represented as a cartoon and was generated via alignment of the bridge subunits in PyMOL.

1.1.3 The Lpt System

After final chemical modifications of LPS at the outer leaflet of the inner membrane, LPS requires transport across the periplasm to its final destination at the OM. On the face of it, this is an incredibly challenging prospect; physiologically this requires

selective extraction from the IM, encapsulation of the hydrophobic LPS tails, excursion through the periplasmic milieu and then final insertion into the OM. The system responsible for LPS transport is called the Lpt system and is also the subject of primary investigation in this thesis⁸².

The complex itself is composed of seven essential proteins, which span three different environments of the cell: LptB in the cytoplasm, LptF and LptG in the IM, LptA bridging the periplasm and LptD with LptE in the OM^{82,83} (**Figure 1.5**).

LptB, found in the cytoplasm, is the nucleotide binding domain (NBD) of the IM ABC-transporter, LptB₂FGC which hydrolyses ATP to power transport of LPS^{84,85} (**Figure 1.6**). Recent structural studies suggest that the TM domains of LptF and LptG house LPS within the transporter, where LptC inserts its TM helix (TM_c) between LptF/G^{86,87}. Here, a mechanism is proposed whereby: LPS (i) laterally enters the cavity (independent of ATP hydrolysis)⁸⁷, is (ii) recognised through transient interactions with LptF/G. Upon this, (iii) TM_c subsequently moves away from the TM helices of LptF/G to allow for association with LPS and (iv) ATP hydrolysis causes inward movement and orthogonal extrusion towards the periplasmic bridging units⁸⁶. The 'outward open' conformation of the LptB₂FGC transporter is similar to previous structures of many well studied and understood ABC-transporters^{18,88}, as the LptB (NBD) domains are separate and ready to accept ATP. It should be noted that TM_c is not required for transport, and in fact, functional studies suggest that the activity of the transporter increases when it is absent. Therefore, TM_c is suggested to have a regulatory role within the transporter to moderate ATP activity for efficient coupling with LPS transport⁸⁷, assuming one ATP molecule is used per cycle of LPS expulsion towards

the periplasmic chute⁸⁶. Though the precise role of TM_c needs to be understood, it is clear that the transporter extrudes LPS into the periplasmic domain of LptC^{83,85}. The IM transporter complex, LptB₂FGC, of the Lpt system in the main subject of **chapter 3** in this thesis. In the chapter I begin to characterise a new conformation of the transporter to address the potentially larger role of LPS within the IM.

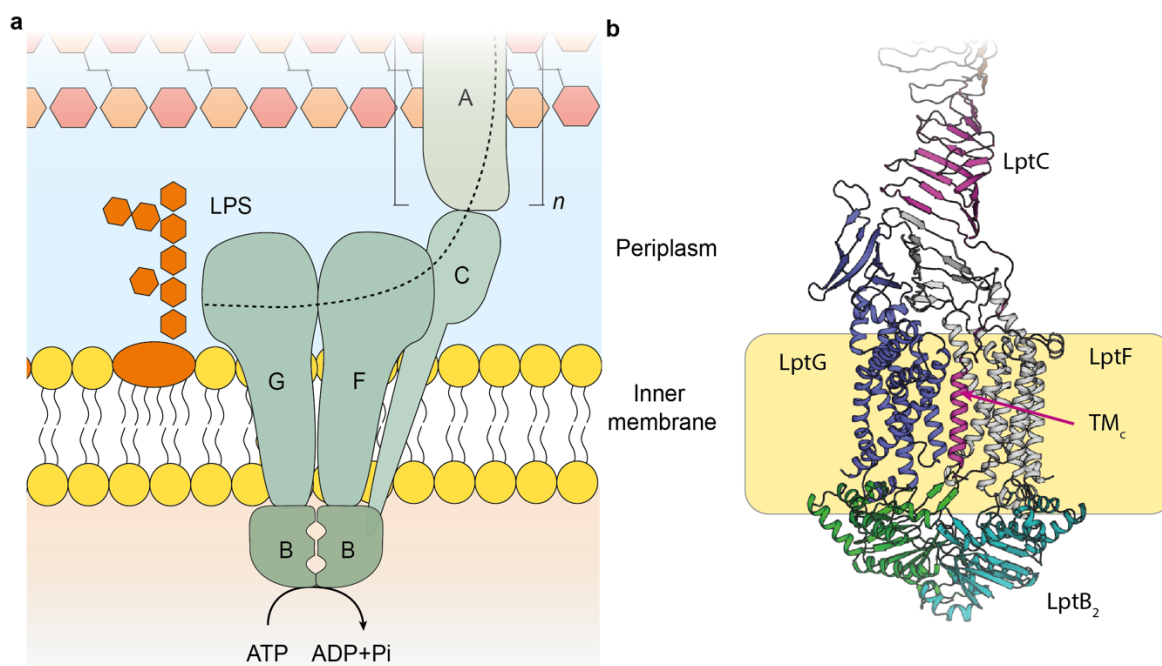


Figure 1.6 LptB₂FGC system schematic. (a) Cartoon depicting LptB₂FGC which extracts and pushes LPS up through the system (b) Crystal structure of EcLptB₂FGC homology model from *E. cloacae* (PDB ID: 6MIT).

The LptC domain itself has a characteristic β -taco fold (the first reference of this type of fold is from TamB of the translocation and assembly module, TAM⁸⁹), where the acyl tails of LPS nestle inside its hydrophobic groove with the saccharide units exposed to the periplasm^{90,91}. Subunits LptF, LptA and the periplasmic domain of LptD share the same β -taco fold, and in fact, shown to complex in a ‘head-to-tail’ manner to form a bridge across the periplasmic space^{21,91,92} (**Figure 1.7**). This bridge extends the hydrophobic groove from LptF to LptD allowing the tails of LPS to slide through to

the OM translocon complex LptDE. In support of this 'LptFCAD complex', studies have shown the LPS has been crosslinked to each member of the bridge, furthermore, each of the periplasmic components are required for the successful in vitro reconstitution of the bridge^{47,85,93}. The 'main' component of the bridge LptA, can associate with LptC and LptD, though it is also able to form 'head-to-tail' oligomeric complexes by itself^{91,94}. Therefore, it is very likely that more than one copy of the LptA subunit exists within the bridging complex, though, this exact number is currently unknown. Under normal *E. coli* growth conditions, it has been shown that a LptA variant deficient in homo-oligomerisation (alongside the deletion of the *lptA* gene from *E. coli*) is viable and therefore implies that only one LptA subunit within the complex is sufficient⁹⁵. Regardless, a study in *Salmonella* has demonstrated that modulating the length of Lpp (an OM lipoprotein which links the OM to the peptidoglycan layer) can change the width of the periplasm (the distance between the IM and the OM)⁹⁶. From these results, it seems very logical to suggest that the number of LptA units varies to span differently sized periplasmic compartments. **Chapter 4** of this thesis investigates the thermodynamic landscape LPS experiences on transit across the periplasm, as well as the effect of multiple LptA bridging units upon this landscape.

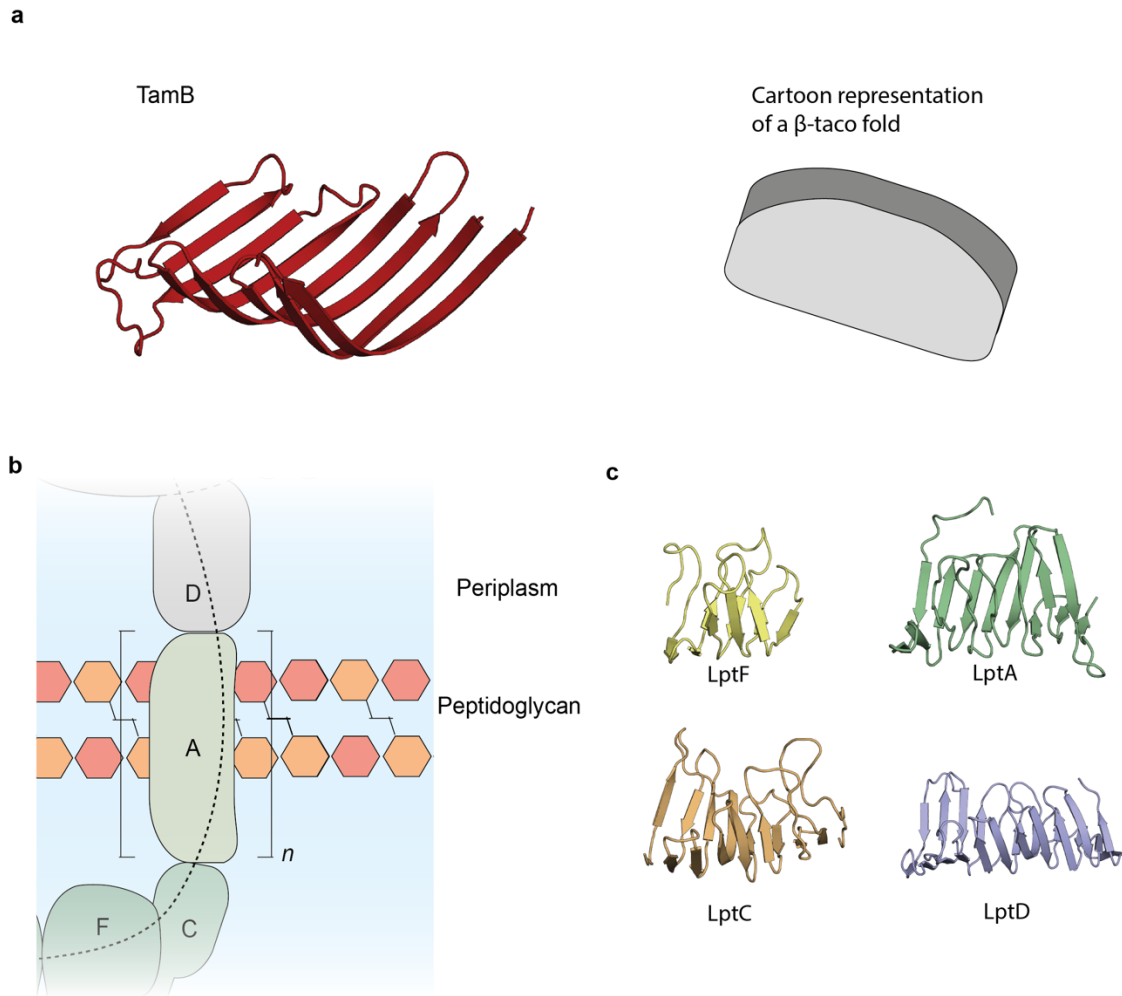


Figure 1.7 LptCFAXD system schematic. (a) Left: cartoon structure of TamB (PDB ID: 5VTG) the first recognised example of a ‘ β -taco’. Right: next to a cartoon representation (b) Cartoon depicting the Lpt bridging complex LptCFA_xD, which shuttles LPS across the periplasm (c) Structures of the β -taco protein bridging subunits: *Kp*LptF (PDB ID: 5L75), *Ec*LptC (PDB ID: 3MY2), *Ec*LptA (PDB ID: 2R19) and *Kp*LptD (PDB ID: 5IV9).

Once LPS reaches the OM, its installation is facilitated by LptD and LptE⁸³. LptD has a large 26 stranded β -barrel domain which sits in the OM alongside a periplasmic β -taco domain which associates with LptA (**Figure 1.8**). Between the β -strands 1 and 26 of the LptD β -barrel exists a lateral putative gate, mediated by an array of hydrogen bonds^{21,83}. LptE exists inside the lumen of the LptD β -barrel, to form an LptDE ‘plug-and-barrel’ complex⁹⁷. From structural and *in silico* modelling it has been suggested

that the LptD β -taco domain forms an intramembrane hydrophobic hole with the inner leaflet of the OM, allowing the LPS acyl tails a 'track' to be threaded into the membrane. Meanwhile, the LPS saccharide units are passed through the hydrophilic lumen of the β -barrel, which swings the lipid tails through the LptD putative exit gate to align itself within the OM²¹. The role of the lipoprotein LptE is less clear, however, its strong affinity for LPS implies that its presence helps to facilitate the movement of LPS through the lumen of the β -barrel^{98,99}. LptE is also critical for the viable assembly of LptDE in the OM^{97,100}. The OM translocon complex is the subject of **chapter 5**, where I seek to use MD to supplement HDX experimental work to verify and validate the mechanism of LPS installation.

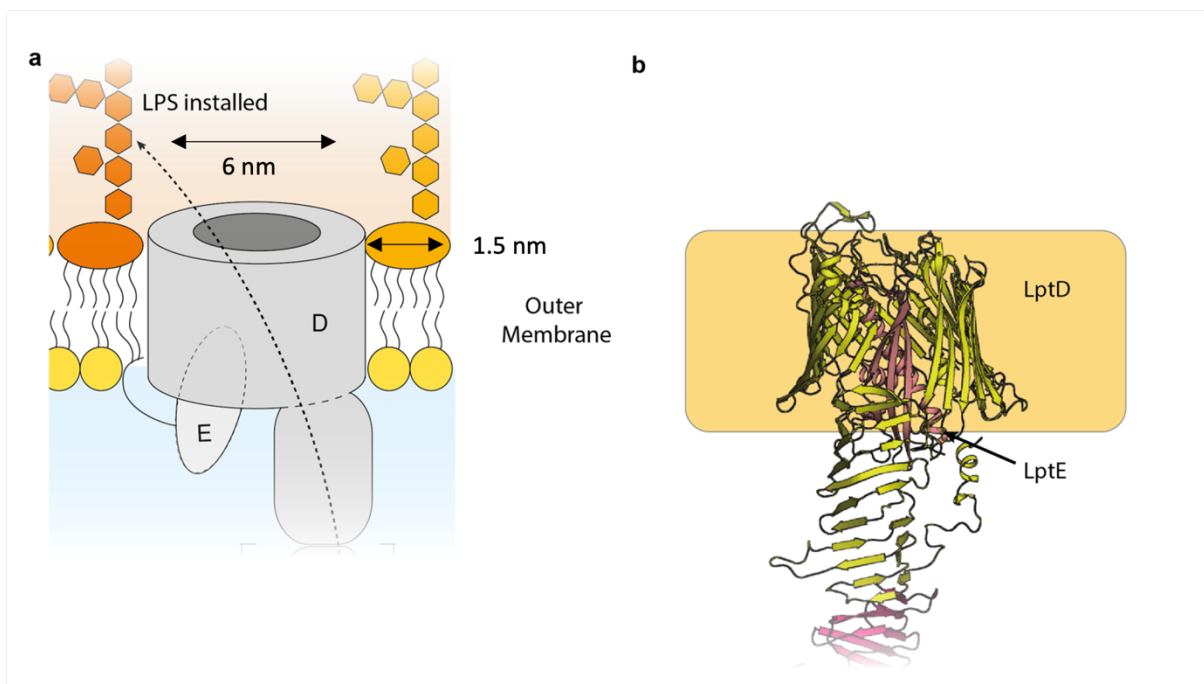


Figure 1.8 LptDE schematic. (a) Cartoon depicting the translocon complex LptDE which inserts LPS into the OM with pore diameter compared to the size of LPS (b) Structure of the *Ec*LptDE homology model of *Kp*LptDE (PDB ID: 5IV9).

This collection of studies has led to the currently proposed 'PEZ model' of LPS transport through the Lpt system⁸³. This model which draws comparison to the LPS being transported as the Austrian PEZ candies and the Lpt system itself to the PEZ

candy dispenser. The dispenser is spring loaded with a base platform which pushes a stack of PEZ candies towards the top of the dispenser so that a candy can be removed from the top and consumed. The model, therefore, proposes that the system is stacked full of LPS molecules such that the ABC-transporter acts as a spring to push the lipids forward through the bridge where the terminal LPS gets inserted in the OM via the LptDE complex. Hence, the LptFCAD bridge, primarily acts as a rigid infrastructure with the passive role of funnelling the lipids through the periplasm. The *in vitro* reconstitution of the Lpt system supports this PEZ model^{47,83}, where the transport occurs between separate liposomes containing LptB₂FGC and LptDE with LptA in solvent in an ATP dependent manner.

1.2 Molecular dynamics

1.2.1 Molecular dynamics and membrane proteins

As mentioned earlier, membrane proteins have diverse and essential roles⁹ and it will come as no surprise that membrane proteins interact with multiple lipid species. Generally, these protein-lipid interactions are vitally important for the structure and function for membrane proteins. Lipids can scaffold integral membrane proteins by oligomerising them in the bilayer like in the case of SemiSWEET (a bacterial sugar transporter) where its monomer-dimer equilibrium is tuned by the concentration of CDL, or by attaching peripheral proteins to the membrane such as with β -barrel assembly machinery (BAM) complex lipoproteins BamB,C,D,E¹⁰¹⁻¹⁰³. Functionally, they can assist via allosteric means by slowing the transition between states in the case of cholesterol with β 2 adrenergic receptor^{104,105} or can have a direct role as substrates with lipid scramblase proteins which help define the distribution of lipids within each leaflet of membranes^{106,107}. Therefore, the characterisation of these

protein-lipid interactions are subject to experimental work (such as mass spectrometry techniques which I cover in **section 1.3**) as well as computational studies, such as molecular dynamics (MD) simulations.

Experimental techniques used for structure elucidation such as X-ray crystallography and/or cryo-electron microscopy generally involve the use of detergents to solubilise membrane proteins which removes most of the surrounding native lipids (lipidome) from the protein itself. Some lipids may get caught up with the detergent during this process which affords their identification in the density maps produced by X-ray techniques¹⁰⁸⁻¹¹⁰. In order to catch some of these lipid interactions, nanodiscs¹¹¹ (a protein which encapsulates a small lipid bilayer patch) or an amphipathic polymer (SMALP) which also encloses a lipid bilayer¹¹² may be used in these structural techniques. The SMALP method uses this styrene maleic acid co-polymer (SMA) to extract membrane proteins directly from native membranes. In contrast to detergents, the SMA extracts the membrane protein with the surrounding lipids portion around it which preserves the native lipid environment¹¹². Whereas nanodiscs are generally formed after a membrane protein has been solubilised by detergent before they are exposed to a solution of phospholipids and an encircling amphipathic helical protein belt called the membrane scaffold protein (MSP)¹¹¹. However, the use of these alternative solubilisation methods is frequently difficult, and if lipid species are bound assignment of their densities can often be ambiguous (though developments in cryo-EM methodologies yielding better resolution structures may reduce this problem). Furthermore it is hard to extrapolate upon the prevalence of any of these interaction events as being representative of the protein's native environment¹¹³. This is because nanodiscs generally require detergent solubilised membrane proteins before

assembly and so lose the majority of the surrounding lipidome. SMALPs, on the other hand, necessitate the assumption that the SMA-extraction process has no bias or preference for a particular lipid¹¹⁴.

Computational methods¹¹⁵, and in particular, molecular dynamics (MD) simulations¹¹⁶ offer an extremely useful tool for identifying and characterising protein-lipid interactions. A helpful analogy of MD simulations is that they can be considered as a 'molecular microscope'^{117,118} allowing the user to 'zoom in' and observe membrane proteins at atomistic (AT) resolution interacting with water, ions or lipids. The user can also 'zoom out' and observe the system at lower resolutions known as 'coarse grained' (CG)^{119,120} representations if they wish to monitor the system over longer timescales (albeit with loss of atomic detail). This has successfully been used to model protein-lipid systems at the nanoscale^{116,121}. Further discussion of the theory and background of AT and CG simulations is covered in **chapter 2**.

Simulation work has been used to predict lipid interaction sites for a large number of integral membrane proteins, helping to elucidate the surrounding liposome such as those within aquaporins^{122,123}. Specific protein-lipid interactions such as with cardiolipin have been explored for a number of systems including not only the membrane-exposed surface of the cytochrome bc1¹²⁴ and the cytochrome c oxidase¹²⁵ but also with the rotor of metazoan ATP synthases¹²⁶ as well as with SecY, assisting protein secretion¹²⁷. In these studies, simulations were carried out (either CG or AT) to define the precise cardiolipin interaction sites for these systems. Furthermore, MD simulations have been successful in characterising PIP₂ binding sites for ion channels including Kir2.2¹²⁸ and Kv7.1¹²⁹ potassium channels. More

recently, a MD approach has been employed to study other PIP₂ lipid binding to Class A GPCR receptors¹³⁰ shown to help stabilise the active site of the GPCR as well as being selectively bound by TRP channels¹³¹. Simulations, especially CG simulations, have the benefit of being high-throughput and online data bases such as MemProtMD take advantage of this by simulating new structures of membrane protein in a bilayer for accessible lipid protein characterisation³³. Furthermore, MD simulations can give insight into the lipid annulus surrounding a given membrane protein. For instance, with the transient receptor potential (TRP) channel present in ciliary membranes of the kidney known as PC2, a phospholipid binding pocket was suggested in MD simulations, where free energy calculations implied that this site was selective for PIP₂ lipids. This suggestion helped identify low-resolution lipid-like density in a concurrent cryo-EM data¹³¹.

1.2.2 Molecular dynamics and free energy calculations

As illustrated above, MD simulations have been used to reliably predict membrane protein-lipid binding sites and interactions. Moreover, it is of importance to quantify the strength of these interactions especially due to the crowded and diverse nature of membranes¹³². This helps indicate the extent to which binding of a certain lipid is likely within its native environment rather than a detergent solubilised environment.

Potential of mean force calculations (PMFs) is a commonly used targeted approach from estimating the free energy of protein-lipid interactions^{125,133,134}. The free energy is calculated as a function of one-dimensional coordinate (or collective variable) which is usually the distance of the headgroup of the lipid away from the protein-lipid binding site. To measure the free energy along this coordinate a number of methods are

available such as ‘umbrella sampling’¹³⁵ or ‘replica exchange umbrella sampling’^{136,137}.

In this thesis I employ the use of umbrella sampling to calculate the free energy of binding of Lipid A to the periplasmic bridging units in **chapter 4**. The theory and background of PMF calculations are covered in more detail in **chapter 2**.

Other methods of estimating the free energy of protein-lipid interactions include free energy perturbation (FEP)¹³⁸ or absolute binding free energy (ABFE)¹³⁹ calculations as well as metadynamics. Briefly, for FEP or ABFE calculations the bound lipid molecule is ‘alchemically’ transformed from its original state (the target lipid) into a ‘generic lipid’¹⁴⁰. This process is then separately repeated with the lipid in the bilayer alone (without the protein) and the difference between these energy values corresponds the energy of binding of the target lipid to the protein. Metadynamics allows for more extensive sampling of any particular energy landscape over any well-defined CV. The more commonly used variant of metadynamics, known as well-tempered metadynamics (WTMetaD)¹⁴¹, samples the defined CV using a history dependent bias potential. This has the effect of reducing the time a system spends in a particular energy minimum which drastically improves the efficiency of sampling. WTMetaD can be used to sample a CV of a particular membrane patch around a protein to compute the free energies and binding poses of lipids to a membrane protein system. Recently it has been employed to replicate the PIP₂ binding sites with Kir2.2 as well as all three CDL sites on AAC¹⁴² in good agreement with PMF and FEP studies carried out on those same binding sites.

Much of this thesis is collaborative in nature and promotes the power of uniting experimental work with *in silico* studies. In particular, the power and utility of combining

MS techniques with molecular dynamics which has shown on numerous occasions to yield congruous and supplementary results^{130,143}. In order to best achieve this, it is important to briefly overview the theory and utility of these experimental techniques. I will only be reviewing the methodologies of the latest native mass spectrometry (MS) and hydrogen deuterium exchange (HDX) techniques used by my collaborators in **chapter 3** and **chapter 5**.

1.3 Mass spectrometry and structural biology

1.3.1 Native MS

The emergence of soft ionisation techniques as well as alterations and improvements to MS instrumentation has facilitated the study of non-covalent intact protein complexes in the gas-phase^{144,145}, including mega-Dalton sized complexes¹⁴⁶ such as intact viruses¹⁴⁷ and ribosomes¹⁴⁸. Further instrumental developments paved the way to observe the first membrane protein complex in the gas phase, BtuCD (an ABC-transporter from *E. coli* – also discussed in **chapter 3**)¹⁴⁹, where the protein was solubilised in detergent before it was stripped off in the instrument to allow for a well resolved spectrum. Modern instrumentation has taken advantage of state-of-the-art mass analysers allowing for the resolution of small molecules bound to protein complexes^{150,151}.

The most obvious use of native MS is the elucidation of subunit stoichiometry within a complex, which may be further interrogated by breaking apart the complex in the instrument and separately isolating the fragments for analysis¹⁵². Here, native MS has helped to determine ambiguous density with Cryo-EM or X-ray crystallography structures of complexes¹⁵²⁻¹⁵⁴. For instance, MS was able to identify RNaseK (a

hydrophobic ribonuclease protein) in the sample preparation suggesting its presence in the ambiguous Cryo-EM density of a rat brain vesicular-ATPase¹⁴⁸.

Furthermore, mass spectrometry assists in the assignment of ligand densities in structural methods^{155,156} (**Figure 1.9**). The importance of identifying lipid binding with membrane proteins cannot be understated, as studies have shown the lipids have roles in influencing oligomeric state¹⁰¹ as well as function¹⁵⁷. Native MS instrumentation may be coupled with other techniques such as liquid chromatography (LC), which can separate out lipid species within a sample mixture before mass analysis allowing for the precise identification of lipid species with mass (700 – 5,000 Da)¹⁵⁸. Furthermore, native MS can be coupled with HDX techniques allowing for the elucidation of conformation or structural dynamics.

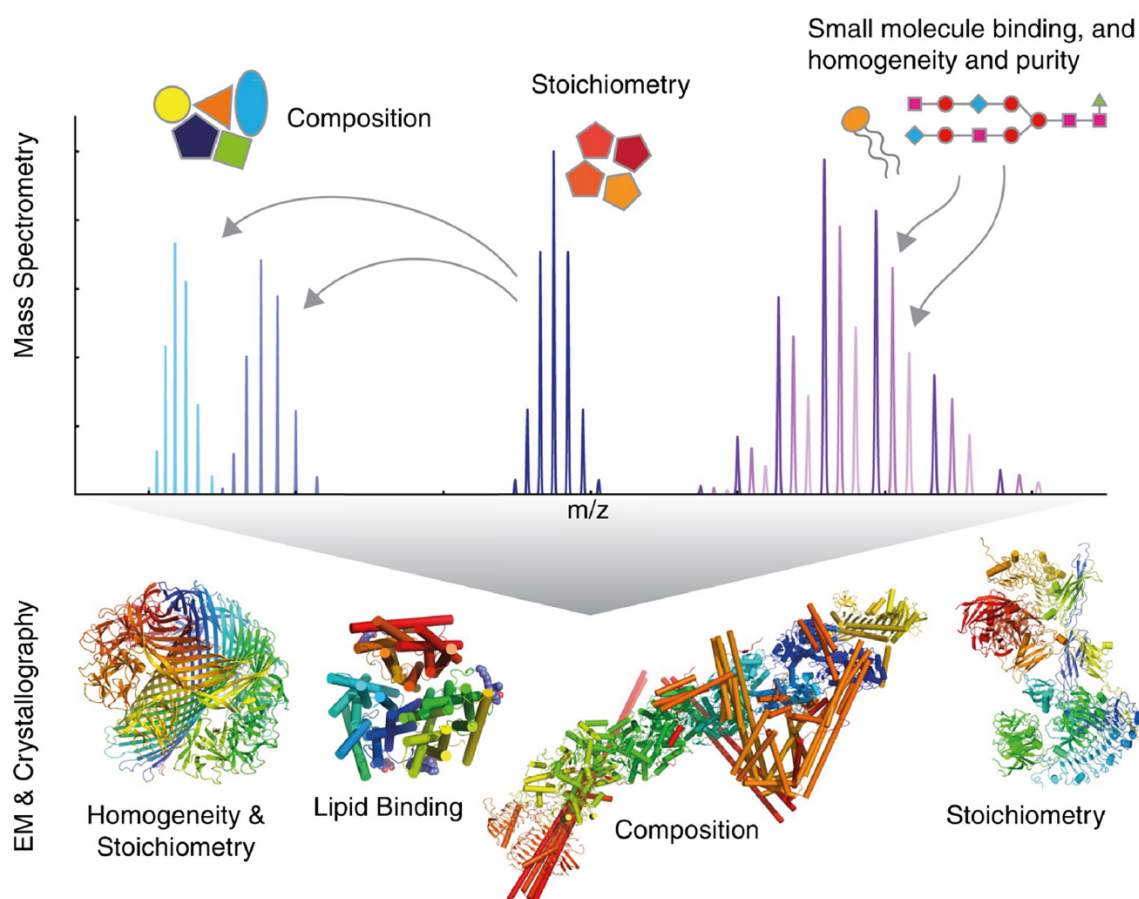


Figure 1.9 Native MS and structural biology. Native MS is a powerful analytical technique used to characterise complex stoichiometry and ligand or small molecule binding. It is often used to screen and optimise conditions for X-ray crystallography or Cryo-EM protocols. Figure obtained from Liko *et al*¹⁵⁹.

1.3.2 Native MS theory and background

In the most simple terms, mass spectrometry ionises species in a sample and separates them according to their ‘mass over charge’ ratio, m/z . The process can be best understood in three stages: (i) desolvation and ionisation of analytes from solution at the source, (ii) species are separated by their m/z values, and lastly, (iii) measured by the detector.

1.3.2.1. Electrospray ionisation

With native MS work the first stage in the process utilises electrospray ionisation (ESI), which is known as a ‘soft’ ionisation method so to preserve the native fold and integrity of biomolecular complexes (**Figure 1.10a**). ESI was a technological breakthrough for the analysis of bio-macromolecules and was pioneered by Professor John Fenn who won the chemistry Nobel prize in 2002¹⁵. In practise, nano-ESI which only uses microlitres of sample and typically expels droplets of less than 100 nm¹⁶⁰ (compared to normal ESI which uses millilitres of sample and has a typical droplet size of 1-2 micrometres¹⁶¹), is employed where more easily desolvated analyte solution is expelled into the instrument for analysis¹⁶². This technique requires a small gold coated capillary needle, which is filled with analyte solution with a voltage applied between the capillary and the cone (the entrance of the MS instrument)^{163,164}. The application of a voltage on the sample in the capillary needle generates a field which permeates the solution and results in the release of charged droplets. On transfer to the cone in the gas phase, these droplets evaporate, shrink and break apart further via electrified droplet fission which eventually results into gaseous ionic species from the droplet¹⁶⁵. The exact mechanism of ionisation is not precisely understood, however there are a number of proposed models to help account for the phenomenon, such as the ‘charge residue model’ or the ‘chain ejection model’^{166,167}.

1.3.2.2. Collisional activation

The purpose of collisional activation is to use an inert gas (like argon) at various stages through the MS instrument (namely the initial source and collision cell) to collide with the gaseous ions. This affords tuneable levels of energy activation (by altering the gas pressure as well as the potential difference across the cells to accelerate the ions

faster) which can be used to 'clean' the ionic biomolecular systems (from unwanted adducts), unfold them or dissociate them¹⁶⁸.

Collisional activation is very useful for 'cleaning' salt and solvent adhering to the protein sample¹⁶⁸ which helps to clarify the mass spectrum and reduce peak width for clearer analysis. This is particularly essential for removing detergent bound membrane protein complexes in the gas phase to obtain clearly resolved membrane protein spectra¹⁶⁹. At higher levels of activation, complexes can be unfolded or dissociated which can be probed by ion mobility (IM) techniques to comment on the stability of complex species¹⁷⁰.

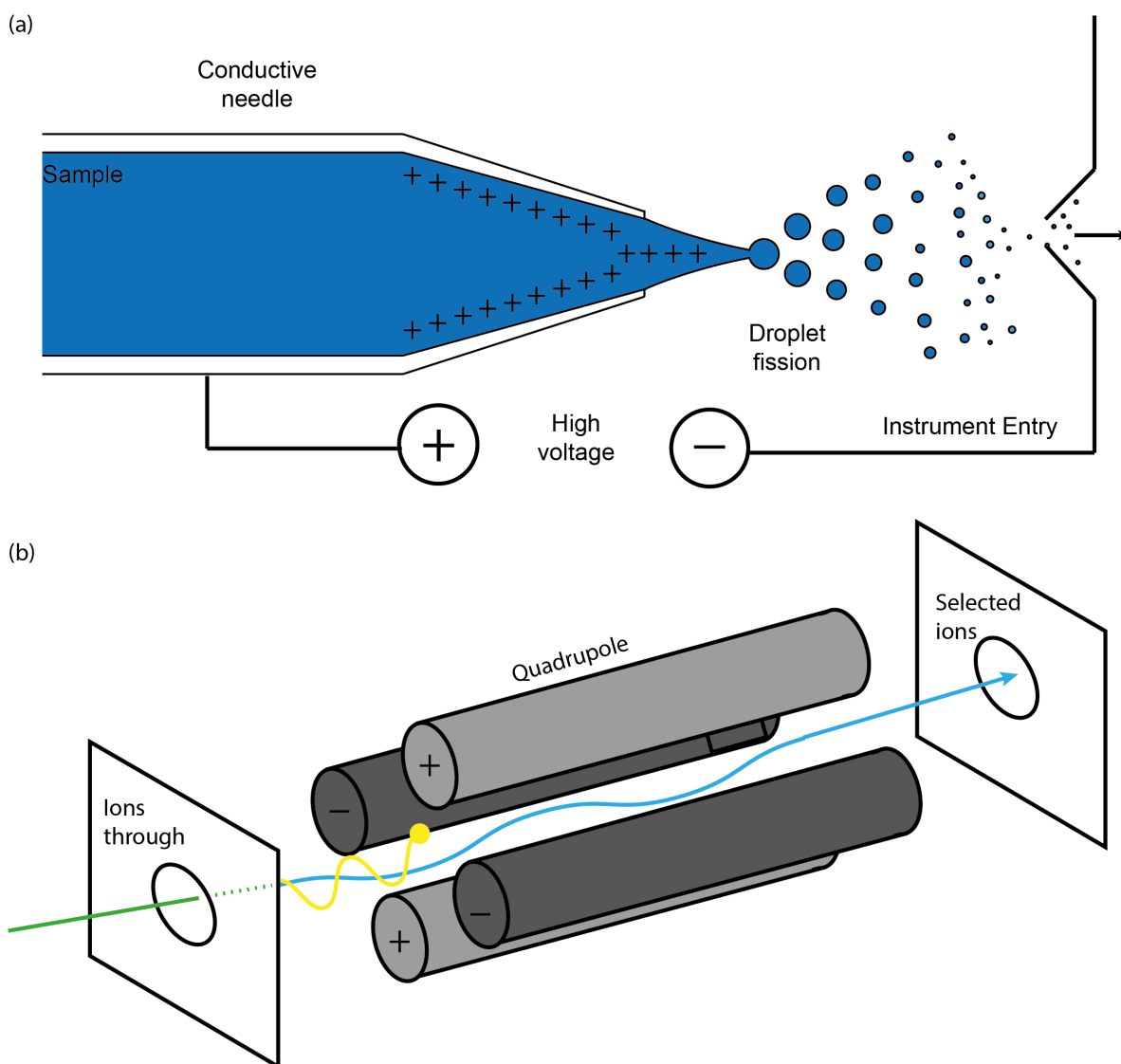


Figure 1.10 Figure 9: Electropray and quadrupole mass analysers. (a) An overview of the electropray process where the application of a voltage between the capillary and the entry of the instrument distorts the analyte to release droplets, which break up and result in the generation of gaseous ions for mass analysis. **(b)** These ions (green) enter a field produced by the quadrupole. Here, the yellow ions are unstable and crash with the electrodes however the blue ions can stably traverse the field and make it the next stage of the instrument.

1.3.2.3. Mass analysers

Mass analysers are used to determine the m/z values of the ionic species in the gas phase. The quadrupole mass analyser is ubiquitous in MS instruments and has been used since the first pioneering native MS experiments¹⁷¹, in modern instruments

however, they function mainly as mass filters, where they allow through only certain ions over a selected mass range (**Figure 1.10b**). The quadrupole is made up of four cylindrical electrodes aligned in parallel where a potential difference is applied to opposite electrodes and opposing charge to adjacent rods. This produces two superimposed fields where on uniform (DC) field which has strength U and other is an oscillating radio frequency (RF) field of amplitude V and frequency f . For every m/z value there exists an area of (U, V) space which allows safe passage of the ion without colliding with electrodes of the quadrupole¹⁷². Therefore, the quadrupole can be used as a mass filter, before the ions are measure by another analyser called an Orbitrap, in modern state of the art instruments. These experiments are known as 'MS/MS'.

Orbitrap mass analysers are the most powerful analysers on the market and have excellent resolution, even at high m/z values¹⁷³. Orbitraps trap ions in a spindle like electrode within a chamber with two electric fields: a radial field which bends the ion path towards the spindle and axial which pushes the ion towards the widest part of the chamber. The ions in the field oscillate around the spindle at frequency determined by its m/z value. The outer electrode of the chamber can track the axial oscillations of the ions and detect their image current, which is recorded to produce a mass spectrum by a Fourier transform¹⁷⁴. This orbitrap technology was the mass analyser used by collaborator's experiments for native MS work¹⁵¹.

1.3.3 Hydrogen deuterium exchange mass spectrometry

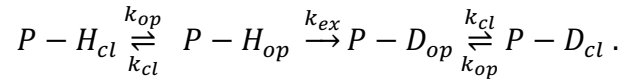
Hydrogen deuterium exchange mass spectrometry (HDX-MS) is a labelling technique in solution which takes advantage of the ability of hydrogens in the backbone amides of proteins to exchange with deuterium in deuterated buffer. After the protein is

labelled, it is digested into peptide fragments corresponding to different regions of the protein, which are measured for their increase in mass using MS¹⁷⁵. This technique measures the H→D exchange of back bone hydrogens however other hydrogens exchange with the solvent, such as those involved with side chain hydroxyl, sulfhydryl, amine, and amide groups of residues serine, cystine, lysine and glutamine etc... However, the exchange is so fast and dynamic that they back exchange on quenching and HPLC steps¹⁷⁶. Additionally, C-H bond do not undergo exchange with deuterium in any experimentally relevant timescales¹⁷⁷. Quenching is an extremely important part of the HDX-MS workflow as it prevents (or rather slows) the back exchange of back bone hydrogens from buffers used in subsequent steps in the HDX-MS procedure (detailed below). Typically, this involves lowering the temperature to 0°C and the pH of the sample to 2.5, which decreases the exchange rate by an order of 10,000 (where the average half-life of a deuterium label is about 1 h) in order to prevent back exchange of deuterium to hydrogen¹⁷⁸.

As hydrogens of the protein are involved in the formation of secondary structures, the H→D exchange rate is an indicator of the stability of such hydrogen bonds, and therefore representative of the conformational changes of the protein. This exchange can also provide rich information on the dynamic structure of protein-ligand complexes. Apart from proline and the first residue of the protein (or peptide) every residue has an amide N-H group amenable to exchange, and therefore HDX allows for the analysis of conformational dynamics for all protein domains.

In these types of experiments, the H→D exchange events are enabled by local fluctuations of the protein from folded to partially unfolded or open conformation. This

occurs when hydrogen bonds are weakened in the secondary structure of the protein allowing for the transiently unfolded/open conformations. This is described by the equilibrium reaction¹⁷⁹:



Here, $P - H_{cl/op}$ defines the unlabelled protein in both its closed and open state, whereas $P - D_{cl/op}$ is the deuterated protein in its closed and open state. k_{op} , k_{cl} and k_{ex} is the rate of opening, closing and H→D exchange respectively.

Generally, the most common exchange dynamics proteins experience is the ‘EX2 regime’, which is defined by the opening (k_{op}) and closing (k_{cl}) rates being much faster than the H→D exchange rate, k_{ex} ¹⁸⁰⁻¹⁸². Hence, the likelihood of exchange events occurring over the course of an isolated opening event is very low and therefore the repeated opening and closing events need to occur to for an overall exchange reaction. Eventually, this can be understood as an isotopic distribution slowly shifting from a low m/z to a high m/z value over time (**Figure 1.11**).

On the other hand, the ‘EX1 regime’ is defined by the opening (k_{op}) and closing (k_{cl}) rates being much slower than the H→D exchange rate, k_{ex} . The result of this is that the isotopic distribution becomes bimodal where the low mass part of the distribution characterises those residues which have not yet exchanged, whereas the high mass distribution represent residues which have already been exchanged (**Figure 1.11**). In this regime, the rate of opening is equal to the chemical exchange rate ($k_{op} = k_{ex}$),

and therefore means that the rate of conformational change as well as the half-life of a certain conformational states can be inferred. As alluded to earlier, most proteins followed EX2 behaviour however, an increasing number have been found to exhibit EX1 kinetics¹⁸³⁻¹⁸⁵. The regimes described above are two limiting situations, in fact the high- and low-mass populations can also overlap because of mixed EX2/EX1 behaviour of a particular protein segment. This situation is referred to as EXX kinetics¹⁸⁶.

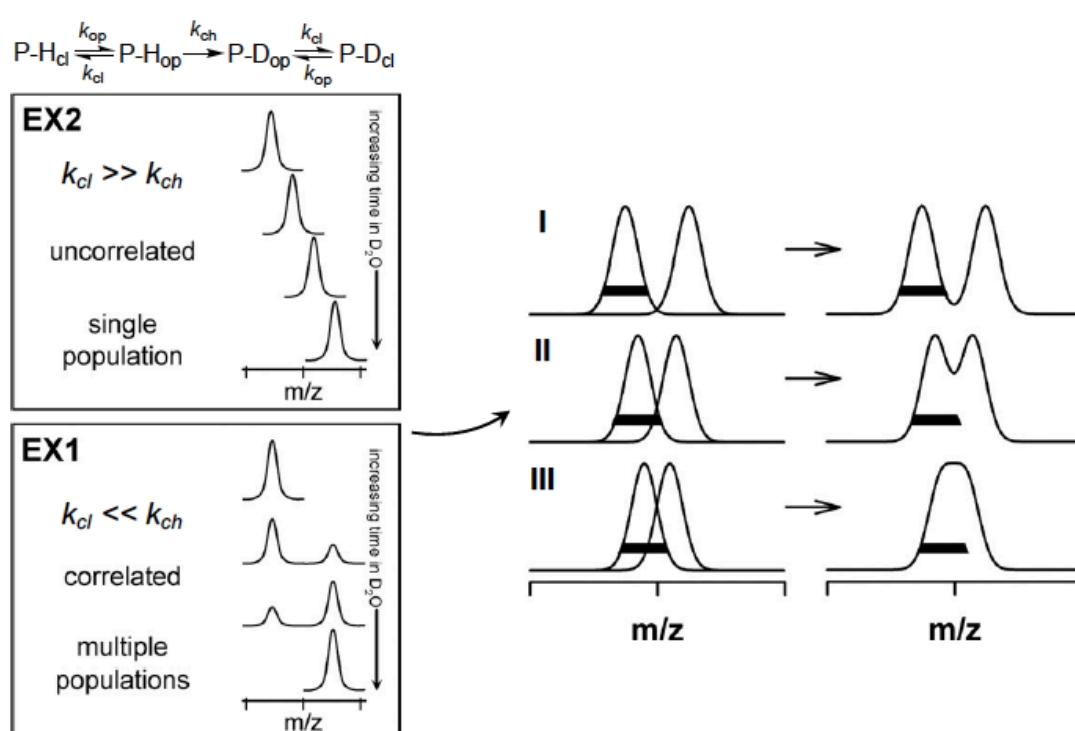


Figure 1.11 Different HDX kinetic regimes. The panel on the left depicts the difference between the EX1 and EX2 regimes. The most common regime, EX2, shows an isotopic population gradually increasing in m/z over time. For EX1, a bimodal distribution is present, with the intensity of the higher m/z population rising over time. The right side of the panel shows the EXX regime which is a mixture of both EX1 and EX2 isotopic mass distributions. Figure adapted from Wies *et al.*¹⁸².

1.3.3.1. HDX experimental workflow

The workflow for an HDX experiment begins with dilution of the protein in a deuterated buffer, followed by reaction quenching, proteolytic digestion, high-pressure liquid

chromatography (HPLC) separation and then finally, MS analysis (**Figure 1.12**). Collaborators performed continuous labelling experiments, whereby the reaction is quenched at different timepoints by lowering pH to 2.0 or to 2.3 - 2.5 and temperature to 0 °C^{187,188}. This range of pH is important as the base catalysed mechanism of exchange is dominant at physiological pH (7.0 - 8.0), whereas the acid catalysed mechanism becomes dominant below pH 3.0. Between pH 2.0 and 3.0 the acid and base catalysed mechanisms contribute equally to exchange such that the water catalysed mechanism becomes dominant. This is where the exchange rate is at its minimum¹⁸⁹. The most widely used protease for digestion is pepsin, as it is an acid protease and operates under quenched conditions. Every peptide has to be sequenced in order to identify the corresponding protein region, which is generally achieved by tandem MS techniques and using proteomic peptide ID software such as ProteinLynx Global SERVER (PLGS)^{175,176}. After digestion, peptides are subsequently desalted and separated out using reverse phase HPLC. The peptides are injected into the mass spectrometer with online ESI-MS or MALDI-MS^{181,190}.

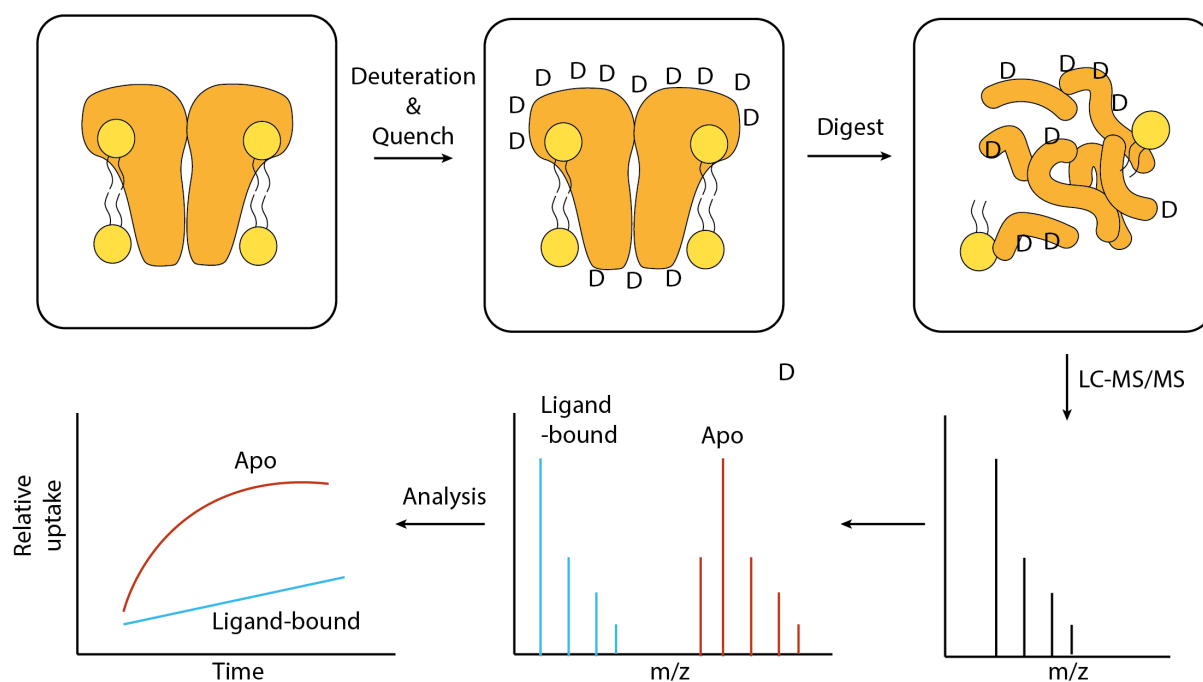


Figure 1.12 Experimental HDX workflow. The protein is incubated in a deuterated buffer for an exchange reaction, which is subsequently quenched by acid after a specific time interval. This process is repeated with the protein in complex with a ligand. The complex is then digested, and the peptides are separated and analysed by LC-MS. The protection level of each peptide is monitored at specific time intervals for both the apo and ligand bound state.

An important step of any HDX-MS experiment is to compare the deuterium uptake of a protein of interest in a different state i.e. a mutated state, ligand or lipid bound state, this is termed: differential HDX. This is critical as differences in deuterium uptake between states can be used to infer the effects of (for example) lipid binding on the conformation of the protein.

However, HDX experiments can only (strictly speaking) inform on the hydrogen bonding within a protein and therefore it is important to supplement findings with techniques such as MD to uncover molecular details underpinning any observed conformational changes^{46,143,185}.

1.4 Aims and objectives of this thesis

I have detailed the importance of studying membrane proteins above and outlined the significance of the lipopolysaccharide transportation pathway, the Lpt system, for developing new antibiotics. Therefore, the focus of this thesis is the Lpt system, which I study in three distinct parts, the IM complex LptB₂FGC (**chapter 3**), the periplasmic bridging protein complex LptFCA_xD (**chapter 4**) and the OM translocon LptDE (**chapter 5**).

Chapter 3 contextualises the role of LPS within the inner membrane and asks whether LPS is involved more widely with inner membrane proteins such as BtuCD and AmtB. The binding work here is related to the ABC-transporter complex LptB₂FGC which extracts LPS from the IM. Here, I examine how the influx of LPS affects the conformation of the transporter and I characterise in detail the binding of LPS within the hydrophobic cavity of LptB₂FGC as well as the effects on the periplasmic domains. I relate my *in silico* studies with native MS experimental analysis to shed light on how LPS functions within the IM, particularly with its transporter complex LptB₂FGC.

Next, I address the thermodynamics of lipid transport through the periplasmic units of the Lpt system, LptFCA_xD, in **chapter 4**. I use computational methods to calculate the binding of lipid A with the bridging subunits in isolation as well as collectively with larger section of the bridging complex. My aim here is to understand the energetics lipid A experiences along periplasmic bridge.

The final section of the Lpt system is examined in **chapter 5**, where my MD simulations help to demystify and rationalise experimental HDX data and Native MS work which

examines the effect of substrate and inhibitor binding to LptDE translocon complex. Here, I uncover in molecular detail the process of LPS insertion into the OM.

Overall, the aim of this work is to study each component of the Lpt system using MD simulations. In doing so, I have demonstrated the power of such *in silico* techniques to describe such important drug targets in detail especially emphasising the indispensability of computational techniques in tandem with experimental analysis.

CHAPTER 2 Theory and Methods

2.1 Molecular dynamics simulations

2.1.1 An overview of the principles of molecular dynamics

Molecular dynamics solves the equations of classical Newtonian mechanics iteratively, over time, to derive the motion of a system of interacting particles. However, for a precise and accurate account of the dynamics of a system, quantum mechanical considerations would need to be included which factor in orbitals and the electronic configuration of atoms and molecules. Models which support this are known as QM/MM (quantum mechanics/molecular mechanics) but can only be applied on small or restricted systems (such as an enzyme active site) and within short timescales (picoseconds) due to their incredibly high computational expense, though recently QM calculations can now be performed for larger systems than these via linear scaling DFT calculations¹⁹¹⁻¹⁹³. In order to address larger systems (i.e. relevant to biomolecular scales) such as protein complexes, lipid interactions, ions moving through channels¹⁹⁴, as well as longer timescales (microseconds to milliseconds) a number of approximations will need to take place. To achieve this, 'classical' MD employs molecular mechanics (MM) force fields to estimate the quantum mechanical potential energy surface (PES)¹⁹⁴.

MM forcefields are built on the integral Born-Oppenheimer (BO) approximation, which uses the fact that atomic nuclei are thousands of times greater than electrons by mass, to separately define the terms for nuclear and electron motion¹⁹⁵. Therefore, this means that modelling electronic positions within the molecular systems would not be required. On top of the BO approximation, MM forcefields are composed of terms describing the potential for distinct physical properties such as bond strength, angle, electrostatics and dispersion forces between atoms and molecules¹⁹⁴. Experimental data from infra-red¹⁹⁶ and NMR spectroscopy¹⁹⁷ and X-ray crystallography¹⁹⁸ as well as QM calculations¹⁹⁹ supply the values of parameters used in forcefields.

2.1.2 Molecular mechanics force field

Consider a system of N atoms, where the position of each atom is defined by a set of Cartesian coordinates such that:

$$\mathbf{r}_N = \{x_1, y_1, z_1, \dots, x_N, y_N, z_N\} \quad 2.1$$

The total energy of this system (the Hamiltonian) is the sum of both the potential and kinetic energy $K(p_N)$ (as a function of momenta), where the potential energy $U(r_N)$ is defined by:

$$U(r_N) = U_{bonded} + U_{non-bonded} \quad 2.2$$

which is simply the sum of the potential of the bonded and non-bonded components of the system.

2.1.2.1. The potential energy of the bonded interactions

The expression of the potential energy of the bonded component, i.e., covalent interactions (used generally, from simple organic compounds to large macromolecular proteins) is as follows:

$$\begin{aligned} U_{bonded} = & \frac{1}{2} \sum_{bonds} k_d (d - d_0)^2 \\ & + \frac{1}{2} \sum_{angles} k_\theta (\theta - \theta_0)^2 \\ & + \frac{1}{2} \sum_{dihedrals} k_\omega (1 + \cos(n\omega - \phi)) \\ & + \frac{1}{2} \sum_{improper} k_\psi (\psi - \psi_0)^2 \end{aligned} \quad \mathbf{2.3}$$

and is composed of terms to describe bond stretching, bending, and rotating respectively with an additional (final) term in **equation 2.3** which arises from stereochemical considerations such as improper torsions (**Figure 2.1**).

The first term in the expression is a classical harmonic potential which expresses atomic bonds as balls on springs, here k_d is the force constant of the bond and d is the stretched or compressed length of the bond itself with d_0 the equilibrium distance. k_d is often very high which represents the stiffness of the bond requiring a large amount energy to stretch or compress it. On small deviations from d_0 , the harmonic oscillator is a reasonable estimation of the potential¹⁹⁴, however on much larger deviations from equilibrium the Morse potential is a better estimator as the function tends to a dissociation energy with converging vibrational energy levels²⁰⁰:

$$U_{bonds} = D_e \{1 - \exp[-a(d - d_0)]\}^2 \quad 2.4$$

with D_e defining the depth of the energy well and the constant a relating to the prior force constant k_d – though within biomolecular systems such large bond length deviations would be rare. In comparison, the value of k_θ is much smaller than k_d which reflects the relative ease in which bonds may be bent from its reference state θ_0 versus the energy required to compress or stretch a bond. To improve the accuracy of the forcefield higher order terms may be added in as part of a Taylor expansion for particularly strained molecules (such as higher order derivatives of k_θ with respect to angle deviation, θ), though these are uncommon within standard forcefields¹⁹⁴.

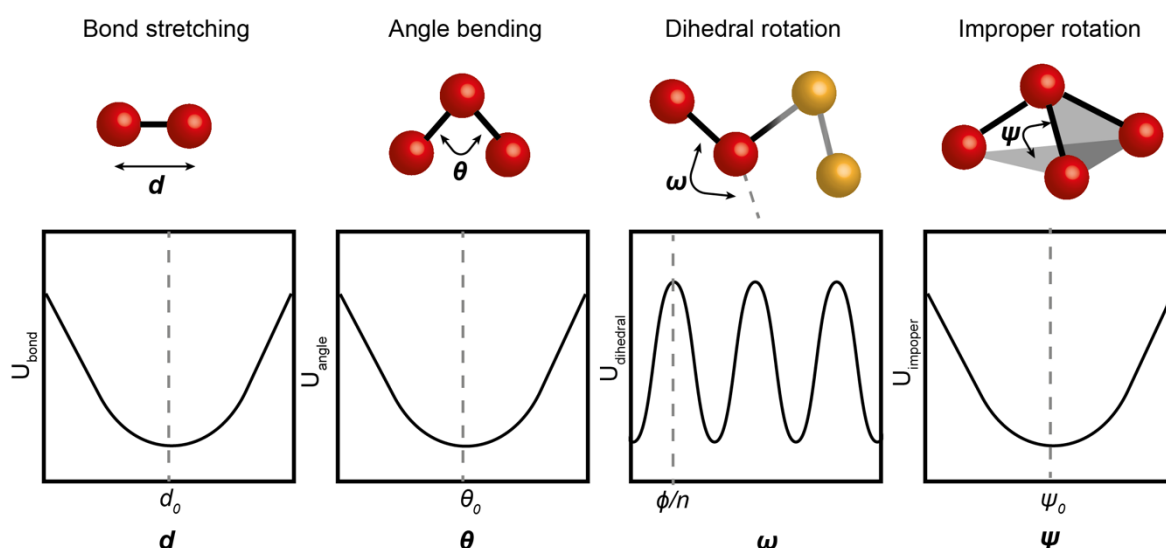


Figure 2.1 The potential of bonded interactions. The four main types of potential through bonded interactions are represented as a schematic

The third term in the expression for bonded potential describes the dihedral component which includes the dihedral angle ω which describes the interaction made between atoms separated by three bonds (i.e., the position 1-4 relative to each other). Similar to the physical description of the previous force constants with the expression; the magnitude of the torsional constant k_ω defines the barrier to rotation. Here also, n

(termed multiplicity) refers to the number of minima in the energy function where ϕ determines the position of these minima. In a molecule, there is often more than one term for each dihedral with ultimately results in a truncated cosine series¹⁹⁴.

The importance of the fourth term is to address 'out of plane' bending, to preserve the planarity of groups with flat geometry such as aromatics or those with chiral restrictions^{194,201}.

As mentioned before, the separation of these potential energy terms is not strictly justified by a QM theory, however it is certainly convenient and allows for a level of helpful physical interpretation of the results of the model¹⁹⁴. Though, in certain cases, these approximations are not appropriate for addressing particular properties of a system, such as the accurate prediction of vibrational spectra. Cross terms can be introduced to remedy this, to factor in the interdependence of bond stretching, bending and rotating which can help to predict structural properties. One such example of a cross term is the Urey-Bradley potential where the distance between (1,3) atoms determine the variation of energy required for angle bending¹⁹⁴. Due to the convenience of separating out potential energy terms as well as the non-existence of a unique solution to translate QM effect into classical MM, the potential energy will continue to be considered as a summation of the respective energy terms as detailed above.

2.1.2.2. The potential energy of non-bonded interactions

Non-bonded interactions are considered between all pairs of atoms, i and j which are separated by three or more intramolecular bonds as well as atoms from other molecules. The interaction between these atoms is characterised by two types (i) the

van der Waals force and (ii) electrostatic components which can be written together as:

$$U_{non-bonded} = \sum_{i=1}^{N-1} \sum_{j=i+1}^N \left\{ 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right\}. \quad 2.5$$

Here the van der Waals forces U_{vdW} are expressed as:

$$U_{vdW} = \sum_{i=1}^{N-1} \sum_{j=i+1}^N 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad 2.6$$

which is known as the Lennard-Jones potential. σ_{ij} refers to the distance at which the energy is zero between the pair of atoms and ε_{ij} corresponds to the interaction strength. The potential described contains both an attractive term and a repulsive term, the former is derived physically from London dispersion forces arising from instantaneous dipoles between bonds and the latter comes from the Pauli exclusion principle¹⁹⁴. This principle captures the strong repulsion between atoms at very short distances, typically 0.3 nm – a quantum mechanical phenomenon which states that no two fermions can occupy the same space. More correctly the repulsive term should be written as:

$$U(r_{ij}) = \exp\left(-\frac{2kr_{ij}}{a_0}\right) \quad 2.7$$

where a_0 is the Bohr radius, however, calculating an exponential can be computationally expensive and so the form of the repulsive term to the power of twelve is a more efficient approximation¹⁹⁴. With increasing r_{ij} the Lennard jones potential rapidly approaches zero, which is truncated after a certain distance (**Figure 2.2**)¹⁹⁴.

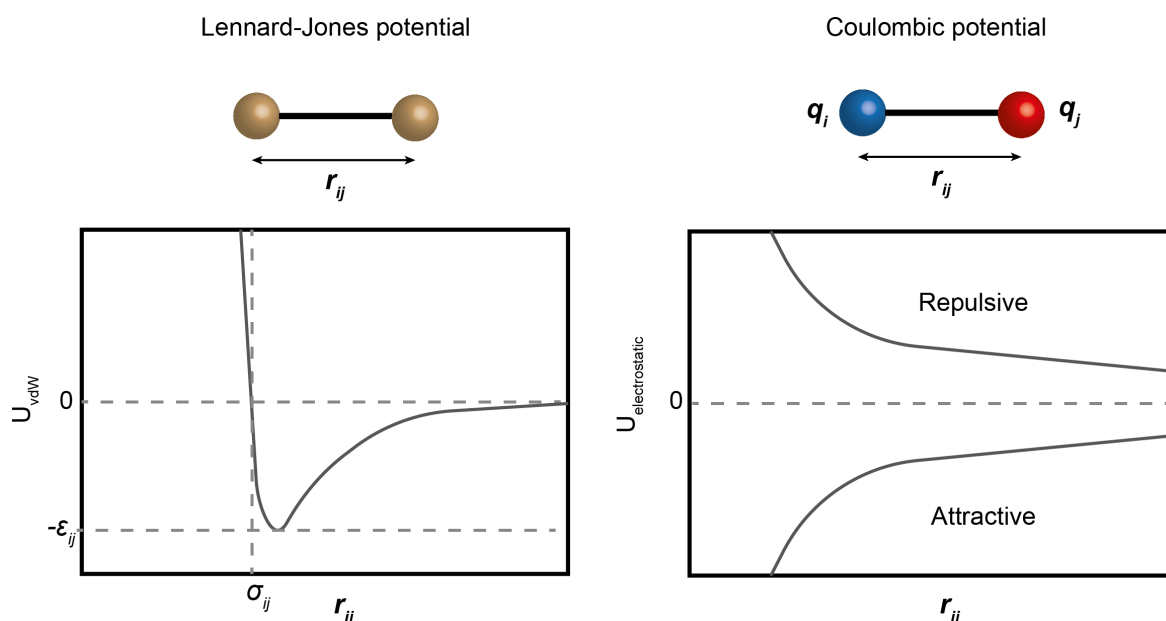


Figure 2.2 The potential of non-bonded interactions. The two main types of potential through non-bonded interactions between nuclei i and j are represented above, Left describes the Lennard Jones potential, Right: the coulombic potential.

The second term in the expression for the potential of non-bonded interactions is:

$$U_{electrostatic} = \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad 2.8$$

which factors in the electrostatics between atoms and is based on Coulomb's law. This is simply a sum over the point charges of the system, with r_{ij} being the distance between nuclei i and j with partial charge $q_i q_j$, where ϵ_0 denotes vacuum permittivity. The decay of the electrostatic potential in the inverse distance between point charges r_{ij} , however a more considered treatment is required for long range interactions within MD²⁰².

2.1.3 The particle mesh Ewald method

To calculate long range electrostatic interactions the particle mesh Ewald (PME) is typically used^{203,204}. It assumes that the periodic cell is indefinitely replicated and uses

a Fourier based approach to solve for the Poisson equation in the repeating system. PME is an interpolation of the reciprocal-space Ewald sum²⁰⁵, which is method originally used for calculating electrostatics in crystal lattices. Within the periodic array, all interactions are summed where the electrostatic interactions here are separated into short-range and long-range interactions. Both these types of interactions converge rapidly in both real and Fourier space, where the rates of convergence are dependent on the width of the Gaussians which cancel when added to the nuclear point charges. Regarding the real space term, a cut-off distance is applied along with a limiting number of modes to the Fourier space term. By implementing the PME algorithm, the computational expense is reduced as the number of electrostatic interactions have been brought down from N^2 to $N\log(N)$, with N being the number of atoms in the system²⁰³.

2.2 Force fields within MD

2.2.1 Atomistic (AT) Forcefields

As detailed above, a forcefield is an expression factoring in the potential of the bonded and non-bonded interaction between all the particles within a system. In practise the most commonly used ones are the Chemistry at Harvard Macromolecular Mechanics (CHARMM)^{206,207}, Groningen Molecular Simulation (GROMOS)^{208,209}, Assisted Model Building with Energy Refinement (AMBER)^{210,211} and the Optimised Potentials for Liquid Simulations (OPLS)²¹² force fields. On the one hand AMBER, CHARMM and OPLS explicitly model every atom including the hydrogens and are called 'all-atom' forcefields, whereas GROMOS approximates an aliphatic carbon and hydrogen into one interaction site. The CHARMM force field²¹³ derives its parameters from both OPLS and AMBER which have been able to reproduce experimental results for phospholipid bilayers. These force fields have been shown to reliably produce the

experimental density of lipids (such as dipalmitoylphosphatidylcholine) by modifying the Lennard-Jones parameters of the hydrocarbon tails²¹⁴.

Table 2.1 Summary of AT forcefields. Overview of AT forcefields discussed in the main text with associated advantages and disadvantages.

AT Forcefield	Advantages	Disadvantages
CHARMM	Large number of compatible lipid models, with parameters for proteins, nucleic acids, some small molecules, glycans, and lipids ²¹⁵⁻²¹⁷ . Furthermore, the online CHARMM-GUI is a helpful resource for constructing simulations for a wide range of systems ²¹⁵ .	Despite deriving its parameters from OPLS and AMBER, significant differences were observed in the configurations of two peptides studied with CHARMM and AMBER ²¹⁸ .
AMBER	Array of compatible lipid models, but the forcefield is more frequently used for protein, nucleic acid, and druglike molecules.	Not as widely available lipid parameters compared to CHARMM.
OPLS	One of the first models used in which the parameters were extensively optimised to reproduce the liquid state thermodynamic properties for a variety of small organic	Compared to other forcefields in this table it has been found to severely underestimate the magnitude of hydrogen bonding interactions ²¹⁹ .

	molecules. OPLS Can accurately describe coulombic and van der Waals interactions of molecular complexes compared to <i>ab initio</i> calculations ²¹⁹ .	
GROMOS	Many parameters for proteins and a good range of lipids, as well as other molecule types ^{220,221} .	GROMOS approximates an aliphatic carbon and hydrogen into one interaction site and therefore has a lower resolution than the other forcefields in this table.

2.2.2 Coarse Grain (CG) force fields

In a coarse grained forcefield, heavy atoms are grouped into sets to produce single interaction sites which drastically decreases the computational expense of simulations²²². This is a direct consequence of (i) reducing the number of interactions within a timestep of MD simulation and (ii) reducing the degrees of freedom, which increases the timescale of the trajectory to up to microseconds or milliseconds.

The resolution of coarse-grained systems can vary by the number of atoms mapped to a single interaction site, which can be found in the range of 3-6 atoms per CG bead. However, there is a trade-off between the extent of coarse-graining (which decreases computational expense) and accurately reproducing the physical properties of the system.

2.2.2.1. Parameterisation of a CG-forcefield

The first step in parameterisation of a CG force field is how to define the sets of atoms from the AT system which would make up the set of CG beads²²³. To do this, one needs to consider (i) system mapping, which involves the number, types and connectivity of the sites, and (ii) the coordinate mapping which specifies the beads with respect to the AT model²²⁴.

Once this has been established the forcefield can be parametrised using a top-down or a bottom-up approach^{223,225}. A top-down approach prioritises a replication of the physicochemical properties of the system such as heat of vaporisation or density. However, despite very good reproducibility of physical parameters, structural fidelity can be lost in the process. On the other hand, for a bottom-up approach, parameters are derived from structural data banks or atomistic molecular dynamic simulations in order to reproduce structural properties. Adopting this method produces accurate coarse-grain models, however, they are restricted to the reference state in which the parameters were based on which makes interpreting the results (from a physical chemistry perspective) challenging. Therefore, in practice, both approaches are employed. Though, with all CG methodologies, a compromise is always made in the accuracy of the system owing to the reduction of the degrees of freedom on coarse-graining.

2.2.2.2. CG forcefields used for systems with biomembranes

The application of coarse-grained approaches to systems with biomembranes has been highly successful²²⁶⁻²²⁸. The most commonly adopted CG forcefield is known as the Martini forcefield, supported by a large community of users and includes

parameters for proteins^{229,230}, lipids²³¹, glycolipids²³², carbohydrates²³³ and nucleic acids²³⁴. Martini also benefits from being a modular and extensive forcefield, with a myriad number of tools for analysis²³⁵, system construction²³⁶ and transformation back to AT resolution²³⁷. It has been successfully used to describe membrane protein-lipid interactions²³⁸ as well as simulations of an entire virus where CG MD simulation of the AT dengue virial envelope was used describe properties such as area per lipid values, bilayer thickness, and lipid diffusion coefficients²³⁹. This Martini forcefield is used for all the CG work presented in this thesis (**Figure 2.3**).

Other CG force-fields utilise a similar level of resolution to that of the Martini forcefield such as the Shinoda-DeVane-Klein (SDK) models which utilises a 3:1 mapping²⁴⁰. This model has successfully been used to investigate phase behaviour and the fusion of vesicles¹²⁰, however, the number of parametrised lipids is small with no proteins having yet been parameterised either. The SPICA forcefield, which has been more recently developed, has the same level of 3:1 resolution with a larger range of parametrised biomolecules²⁴¹. Similarly, ELBA forcefield uses a 3:1 mapping scheme but also features a water model which uses soft, dipole moments²⁴². Furthermore, ELBA uses dipoles within lipids to improve the accuracy of the electrostatics, albeit only a few lipids have been modelled so far. The main work carried out using ELBA has been to model drug permeations through bilayers.

An example of a top-down forcefield is SIRAH^{243,244}, which uses a similar mapping resolution as Martini as well as explicit treatment of water and ions, with a range of parameterised biomolecules²⁴⁵. Within this cast of biomolecules there are differences in resolution depending on the molecule i.e., polar and aromatic groups use a higher level of resolution compared to hydrophobic groups. As parametrisation uses

structural data no levels of restraints are required to hold together the secondary structure which are required with Martini models.

Table 2.2 Summary of CG forcefields. Overview of CG forcefields discussed in the main text with associated advantages and disadvantages.

CG Forcefield	Advantages	Disadvantages
SDK	Has been used to accurately predict the behaviour or self-assembling macromolecular systems such as liposomes ¹²⁰ .	Small number of parameterised lipids and no parameterised proteins.
SPICA	Has a large range of parametrised biomolecules and can accurately predict domain formation within various lipids in membranes ²⁴¹ .	Has only recently been used to include parameterised proteins ²⁴⁶ .
ELBA	Includes a polarisable water model which uses soft, dipole moments ²⁴² . ELBA also uses dipoles within lipids to improve the accuracy of the electrostatics.	Only a very few lipids have been parameterised so far.
SIRAH	Has a good range of parameterised biomolecules	There are inconsistent levels of resolution depending on the

	currently including parameters for lipids, proteins, nucleotides, and aqueous solvent ²⁴⁵ .	molecule. There are higher resolutions with polar or aromatic groups compared with hydrophobic ones which will cost computational power.
Martini	Largest set of parameterised lipids and biomolecules and has as successfully been applied to detailed studies of individual membrane proteins and their interactions with lipids. The forefield has large support from the community with a huge range of analysis tools.	Can overestimate the aggregation propensity of molecules such as glycans and proteins.

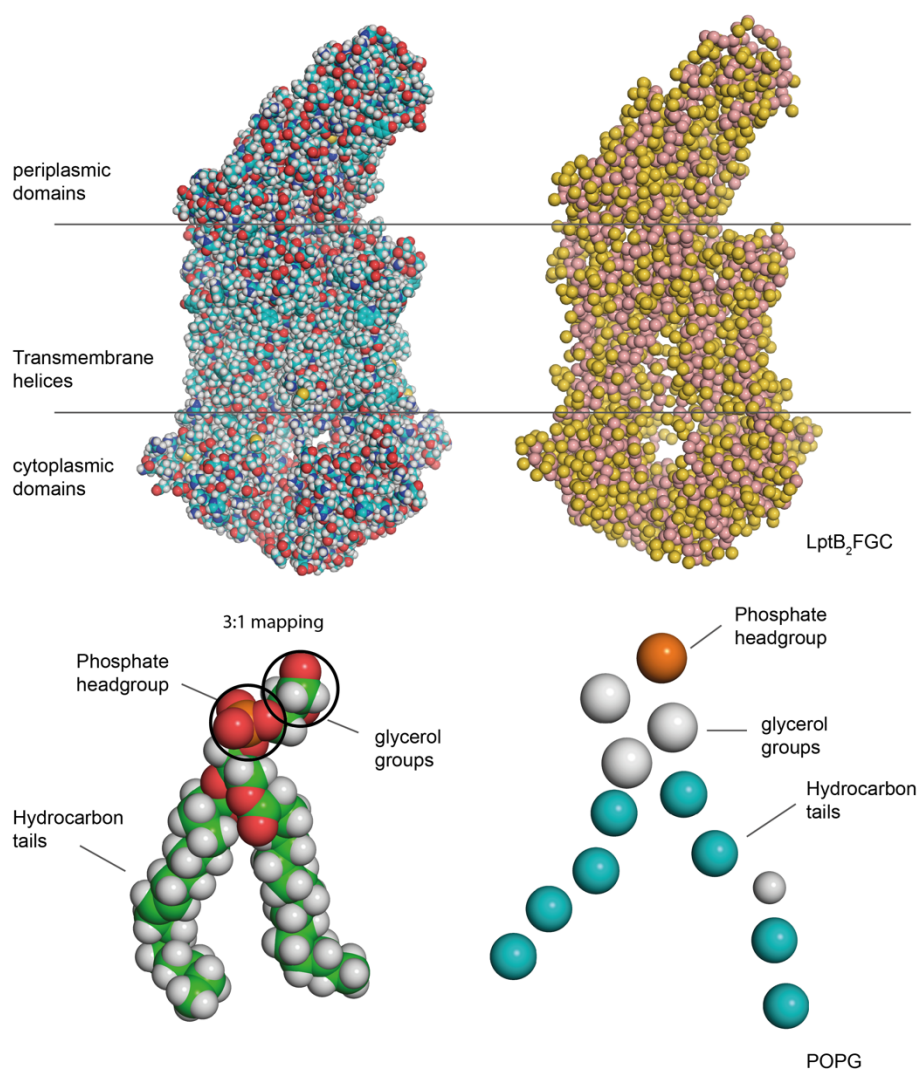


Figure 2.3 AT and CG representations. Top: Comparison of the LptC-skewed *pa*LptB₂FGC protein in both its atomistic representation (CHARMM forcefield) and its coarse-grained representation (Martini forcefield). Bottom: POPG lipid in the atomistic and CG versions of the structure with CHARMM and Martini forcefields respectively.

2.2.2.3. The Martini Forcefield

The forcefield was mainly developed with lipid interactions within a membrane in mind²⁴⁷. Initially, all particles were considered to have the same mass which was equivalent to the mass of four water molecules, set at 72 Da. These particles could then only interact through (i) harmonic bond stretching and (ii) harmonic angle bending for bonded interaction potentials and (i) coulombic and (ii) Lennard-Jones interactions

for non-bonded potentials. For the bonded interactions, only a single bond length was used for the harmonic bond stretching potential and two bond angles with the harmonic angle bending, where only one value for the force constant was used respectively. For every CG bead the value of σ_{ij} of the Lennard-Jones potential was always 4.7 Å, whereas for the coulombic term, the potential moves to zero from 0 Å to 12 Å. Furthermore, the non-bonded electrostatic potential has no long-range electrostatic potential and includes a value for ϵ_r in the expression:

$$U_{electrostatic} = \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}} \quad 2.9$$

In the original iteration of the forcefield, only four bead types were used:

- (i) polar (**P**); for partial charges and hydrophilic atoms,
- (ii) non-polar (**N**); for neutral and hydrophilic atoms,
- (iii) apolar (**C**); for atoms with a mix of polar and apolar properties
- (iv) and charged (**Q**) for ions.

The non-polar (**N**) and charged (**Q**) beads, were further classified into their respective hydrogen bonding abilities i.e. donor (*d*), acceptor (*a*), donor and acceptor (*da*), and no hydrogen bonding ability (*0*). Once these CG beads have been defined, distinct interaction strengths were set between them, which was factored into the Lennard-Jones potential respectively. These Lennard-Jones force field parameter values originated from an alkane-water system to replicate the experimental densities, as well as the solubility and diffusion of the pure system²⁴⁷.

From this initial base set of assumptions, the Martini forcefield has been refined further by new bead types and by improving bonding and non-bonded interactions^{229,231}. This,

in addition to its modular build and simple interaction potentials has increased the Martini forcefield popularity in the community.

2.2.2.4. Martini forcefield details

As mentioned, the number of bead types have increased and the interactions between them have been refined since the original iteration of the forcefield. Bead types can now come in different sizes: normal (**N**), small (**S**) or tiny (**T**) which are used to justify differences in the 4:1 mapping scheme i.e. in cases where three heavy atoms constitute a bead, it may be replaced by a small bead (**S**) instead of a normal bead (**N**). Furthermore, they can also be employed to account for atoms in rings with reduced availability such that three heavy atoms could be described as tiny (**T**) instead of small (**S**). Bead size modifications in the system vary the σ_{ij} and the interaction strength ε_{ij} values of the Lennard-Jones non-bonded potential between them.

A major limitation of the Martini model is the lack of directionality of intermolecular interactions which is especially acute when attempting to accurately reproduce hydrogen bonds interactions. Hydrogens are vital in maintaining secondary structure of proteins, hence, to account for this, an ‘elastic network’ is used to hold the protein structure together which itself originates from the input structure, which is derived from harmonic potentials between C α beads within a defined cut-off²³³. The use of such a network, unfortunately, eliminates the accurate prediction of large-scale conformational changes altogether. Another limitation is that Martini beads may feature an overall charge which does not account for the partial charges from the heavy atoms. This misrepresents the electrostatics in certain cases such as for a deprotonated carboxylate for example. Also, it should be noted that for uncharged

beads, non-bonded interactions are simply determined by the Lennard-Jones potential.

In the Martini 2 forcefield, nine different interactions levels are accounted for which only the sigma value of the van der Waals potential altered with bead size. With Martini 3, there are twenty distinct interactions possible. Interactions are rated on a scale of 0 (hyper attractive) to 19 (super repulsive), where every permutation of bead type and size has a particular interaction level reflective to the atomic forces in play. This increase in interaction levels from Martini 2 to Martini 3 allows for more fine-tuning in the reproduction of experimental solubilities. As with atomistic forcefields, the CG potentials are modelled with the usual expression of the harmonic oscillator including angle terms as well as the dihedral potential. Due to the non-existence of hydrogen bond directionality within a Martini CG system, proper dihedrals are present to help enforce the secondary structure of the protein, whereas the improper dihedrals help to impose planarity within certain molecules. The parametrisation of bonded interactions comes mainly from atomistic simulations, and the non-bonded interactions are based on experimental thermodynamic data.

2.2.2.5. The Martini 'stickiness' phenomena

For several forcefields (both CG and AT), there is noted a susceptibility for proteins and glycans to aggregate more than they 'should'²⁴⁸⁻²⁵². It is generally considered that this arises from inaccuracies of the water model utilised, which has effects on the solute-solute, water-solute, and water-water interactions^{249,251}. It is also suggested that inaccuracies in the AT forcefield are compounded on translation to CG by parameterisation and therefore produces an overcompensation in protein-protein interactions²⁵².

This particular issue of ‘stickiness’ has often been associated with the Martini forcefield, not just from a protein-protein perspective,^{253,254} but with glycan interactions as well²⁵². This stickiness consideration does have an influence on protein-LPS interactions due to the associated saccharide units. On a practical note, this may manifest in an overestimation of the LPS-protein binding energy, this was one of the reasons why lipid A (LPS without the core saccharide and O-antigen units) was used for the thermodynamic binding studies in **chapter 4**. One such solution has been to use a scaling factor to the non-bonded interactions, which has been applied to both glycans²⁵² and proteins alike^{253,254}. Though, this does not address the issue of stickiness in the Martini forcefield, it is more just a sophisticated ‘work around’.

2.3 Energy minimisation

In order to produce a stable system before performing an equilibration step for a MD production run, it is vitally important to reduce initial steric clashes within the system. Energy minimisation is the process of minimising these clashes, which is carried out by a choice of two methods: minimising potential energy by steepest descents or by conjugate gradient algorithm which both utilise a change in particle coordinates along a gradient to produce a lower energy conformation of the system.

The steepest descents method uses an iterative line search along the negative gradient of the energy to find the energy minima at sequential positions. On the other hand, the conjugate gradient method uses previous vectors to derive future search directions and therefore reduces the need to repeat minimisation over those same vectors²⁰⁰.

2.4 Integrator algorithms in Molecular Dynamics

Molecular dynamics is a ‘deterministic’ method, whereby, the future state of the system can be calculated by the current state. MD evaluates the evolution of the system over time using integrator algorithms.

Within a system of interacting particles: the force F_i any particular atom, i , at position r_i experiences is the negative derivative of the potential energy function:

$$F_i = - \frac{\partial U(r_i)}{\partial r_i}, \quad 2.10$$

which can also be written as:

$$F_i = m_i \ddot{r}_i \quad 2.11$$

where m_i is the mass of atom i with $\ddot{r}_i$ denoting the acceleration (the second derivative of position r_i). Therefore, within three dimensions, 3N set of second order differential equations need to be solved. However, the force on each atom is dependent on all the interaction partners within the system, and so an analytical solution to these differential equations cannot be obtained. Instead, these equations need to be solved numerically using a finite difference method. The integration is broken down into small time steps δt , where the prevailing force upon an atom, at time t , is calculated from the vector sum of the surrounding interacting atoms. The acceleration of each atom can be derived from the overall force upon that atom to calculate the new position and velocity at time $(t + \delta t)$, assuming that the force remains constant during δt . At time $(t + \delta t)$, with new coordinates for the atoms, the forces upon them are reevaluated to produce the new positions and velocities at time $(t + 2\delta t)$. This process is continually reiterated to produce a ‘trajectory’ which describes how the dynamic variables of the system

evolve over time. At the start of the MD simulation, the initial velocities are randomised according to the Maxwell Boltzmann distribution. It is also common practice to re-run the simulation over a series of replicates in which the initial trajectories are re-set and randomised again²⁰⁰. Within MD, an algorithm is required to determine positions of particles from their velocities and accelerations. Within the GROMACS^{255,256} package, this algorithm is known as the ‘leap-frog algorithm’²⁵⁷ which is derived from the Verlet algorithm based on two Taylor expansions²⁵⁸ of $r_i(t + \Delta t)$ and $r_i(t - \Delta t)$ about time, t :

$$r_i(t + \Delta t) = 2r_i(t) - r_i(t - \Delta t) + \frac{1}{m_i}F_i(t)(\Delta t)^2. \quad 2.12$$

In order to calculate a new position at time $(t + \Delta t)$, the Verlet algorithm uses the position at time $(t - \Delta t)$ as well as at time t . However, this requires knowledge of time $(- \Delta t)$ when $t = 0$, which is solved by performing a Taylor expansion about $r_i(t)$.

Instead, the leapfrog algorithm uses the position and forces at time t and $(t - \frac{\Delta t}{2})$ to update the position and velocities, using a half time step:

$$v_i\left(t + \frac{\Delta t}{2}\right) = v_i\left(t - \frac{\Delta t}{2}\right) + \frac{1}{m_i}F_i(t)\Delta t, \quad 2.13$$

$$r_i(t + \Delta t) = r_i(t) + v_i\left(t + \frac{\Delta t}{2}\right)\Delta t \quad 2.14$$

It is termed the ‘leapfrog’ algorithm to reflect the fact that the velocity ‘jumps over’ the coordinate to give the next half-step value of the velocity.

2.5 Settings and parameters of MD simulations

2.5.1 Temperature and pressure

Integrating the Newton's equations of motion directly yields an NVE ensemble where particle number, N , the volume of the system, V , and the energy, E is conserved. A separate methodology is often adopted for simulations of biomolecular systems known as the NPT ensemble. This ensemble defines the system in more physically useful parameters (i.e. more relatable to experiment) such as pressure, P , and temperature, T . In this ensemble, both pressure and temperature are kept constant at the appropriate physiological values. To do this, requires barostat and thermostat algorithms which, in this present study, are known as the Parrinello-Rahman barostat²⁵⁹ and the velocity scaling thermostat respectively²⁶⁰.

2.5.1.1. Velocity re-scaling thermostat

The temperature of a molecular dynamics system is a statistical thermodynamic quantity defined by the position and momenta of all particles and can be determined from the kinetic energy as:

$$KE_{NVT} = \sum_{i=1}^N \frac{1}{2} m_i v_i^2 = \frac{3}{2} N k_B T \quad 2.15$$

with k_B as the Boltzmann constant. The thermostat scales any particle velocities which fluctuate from a fixed specified temperature defined by an external heat bath, where these fluctuations decay exponentially with respect to time as:

$$\frac{dT(t)}{dt} = \frac{T_0 - T(t)}{\tau_T}, \quad 2.16$$

where τ_t is a time constant. This velocity re-scaling temperature coupling thermostat is similar to the weak coupling Berendsen algorithm²⁶¹, where the kinetic energy of the system, KE , is corrected by this modification:

$$dKE = (KE_0 - KE) \frac{dt}{\tau_T} + 2 \sqrt{\frac{KE \cdot KE_0}{N_f}} \frac{dW}{\tau_T}. \quad 2.17$$

Where N_f is the number of degrees of freedom and dW is a Wiener process, which is defined as ‘Brownian motion’ where the change of the position of the particle occurs over iterative fixed timespans and its displacement within a particular timespan obeys a normal distribution function²⁵⁶.

2.5.1.2. Parrinello-Rahman barostat

For an ideal gas, pressure is inversely related to volume (according to Boyle’s law) which can be modified to maintain a constant pressure. Like the velocity re-scaling barostat, the system is coupled to an external pressure bath which is re-scaled accordingly to a reference, P_0 :

$$\frac{dP(t)}{dt} = \frac{P_0 - P(t)}{\tau_P} \quad 2.18$$

where, like before, τ_P is a pressure constant. Typically, for simulations containing a lipid bilayer pressure, coupling along x, y and z (isotropic coupling) would squeeze the membrane, therefore, pressure along the membrane x, y axes vary independently to the pressure along z (called semi-isotropic coupling).

2.6 Periodic boundary conditions

Within an 'open' system, particles can move freely without any limitations, however this is problematic, as particle density is often required to be constant. On the other hand, a 'closed' system may give rise to surface effects which could influence the physical behaviour of the system¹⁹². To avoid these problems, 'periodic boundary conditions' (PBC) are frequently employed to replicate the image of the simulation box indefinitely along x , y and z . Here, when a particle leaves the edges of the box it will enter through the opposite face of the adjacent image. Within this, only the forces within the original simulation box are considered¹⁹².

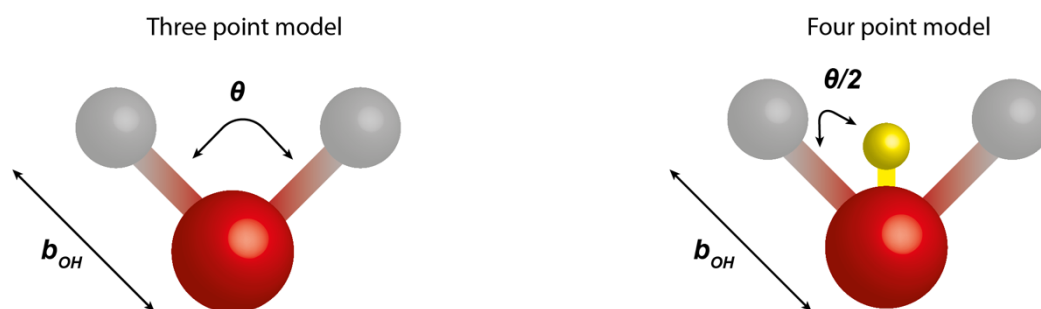
For systems with bilayers and proteins embedded in them, PBCs allow for the extension of the membrane patch along the axes of the membrane, however this gives rise to a number of issues when defining the characteristics of the membrane. The main issues here are non-physiological undulation of the membrane due to lowered amplitudes at the edges of the PBC as well as the reduction in thermal fluctuations of soft-condensed matter, when small cells are used²⁰². Therefore, with simulations conducted in this project which include predominantly membrane embedded proteins, the minimum image convention is used to approximate short-ranged forces and the PME method used to account for long range electrostatic interactions.

2.7 Water models

The simplest way to simulate water within MD trajectories is to use rigid, non-polarisable water molecules which are approximated as a set of point charges²⁶²⁻²⁶⁴. The two types of models which use these simplifications are known as the simple point charge (SPC)²⁶⁵ and transferable intermolecular potential (TIP)²⁶⁶ model.

Respectively, each model has been refined to include further physical features such as three-point charges which are known as the SPC/E²⁶⁶ and TIP3P²⁶⁷ models. These three-point charges are arranged to reflect the three atoms of an H₂O molecule. Further to this, a fourth point charge to mimic the water quadrupole moment was added to the TIP3P model to produce the TIP4P^{267,268} water model (**Figure 2.4**). This fourth interaction site was added to the bisector plane of the molecule and improves the charge distribution modelled. A updated version of this four-point model has slightly modified the positions of point charges from the hydrogens and the oxygens of the molecule, which improves the accuracy of the electrostatics and is known as the OPC model²⁶⁹. Further to this, models with greater number of point charges have also been made including the TIP5P (with five point charges) which replaces the charge of the quadrupole moment from the TIP4P model with two point charges which mimic the lone pair of electrons on the oxygen²⁷⁰.

Atomistic



Coarse-grain

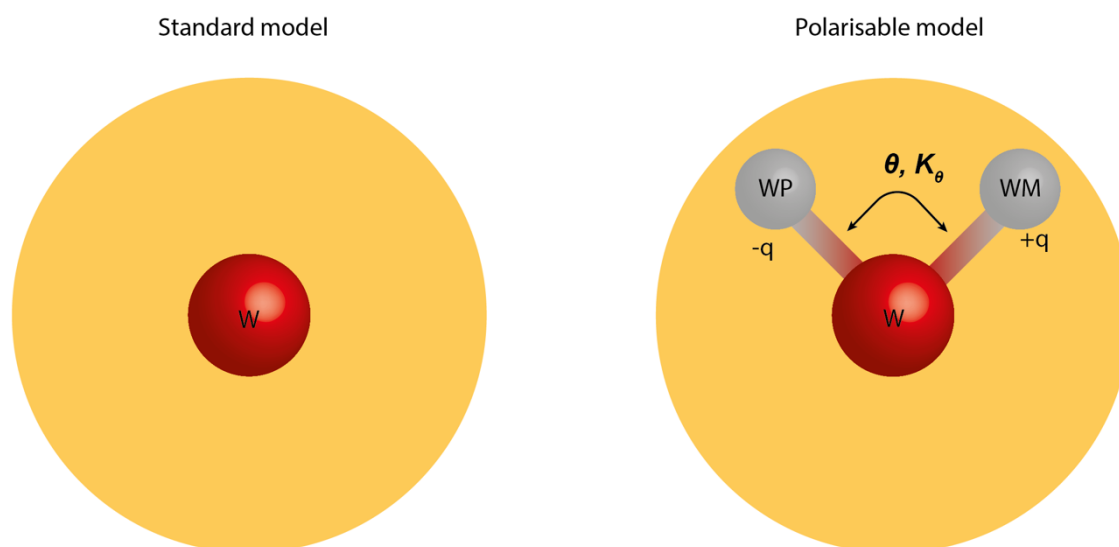


Figure 2.4 AT and CG water models. Top: shows the schematic of a three-point and four-point AT water model, where the oxygen is coloured red, and the hydrogens coloured white. For a three-point model charges are located on the atoms respectively whereas the four-point model included a virtual site between the hydrogens. Bottom: CG model of martini water in both its standard model and polarizable model

For the CG Martini forcefield, water is modelled as a P_4 particle, where four water molecules are treated as a course-grained bead and the water is treated explicitly. Unfortunately, martini water freezes at a temperature in the range of 280-300 K, though this problem surfaces mainly when there are nucleation sites within the system, which irreversibly freezes the water within short timescales. To prevent this, antifreeze particles can be introduced, labelled BP_4 ('Big P_4 ') which are used to disturb the lattice packing of the molecules²³¹. Within the Martini forcefield, this water model has been further adapted to include a polarisable version using a three-bead model to represent

four water molecules, where there are electrostatic interactions between their charged particles WP and WM, with associated angle parameters between the beads, θ and K_θ (**Figure 2.4**)²⁷¹.

Within this thesis, the two water models used are the martini water models for CG simulations and TIP3P for all-atom simulations.

2.8 Potential of Mean Force (PMF) calculations

The potential of mean force (PMF) is a very convenient method of measuring the force of pulling a particle (or substrate) of interest along a coordinate, especially a well-defined one-dimensional coordinate such as the axis along the periplasmic chute as discussed in **chapter 4**. The PMF $W(\xi)$ along a determined coordinate ξ , is a fundamental concept in statistical mechanics and is defined from the average distribution function, as $\langle \rho(\xi) \rangle$:

$$W(\xi) = W(\xi^*) - k_B T \ln \left[\frac{\langle \rho(\xi) \rangle}{\langle \rho(\xi^*) \rangle} \right] \quad 2.19$$

where ξ^* and $W(\xi^*)$ are arbitrary constants, such that the reference state is set at zero free energy^{272,273}. The average distribution function is derived from the Boltzmann weighted average which is expressed:

$$\langle \rho(\xi) \rangle = \frac{\int d\mathbf{R} \delta(\xi'[\mathbf{R}] - \xi) e^{-U(\mathbf{R})/k_B T}}{\int d\mathbf{R} e^{-U(\mathbf{R})/k_B T}} \quad 2.20$$

where the total energy of the system is $U(\mathbf{R})$, which is a function of the coordinates. $\mathbf{R}$ and $\xi'[\mathbf{R}]$ represents a function which depends on a particular collective variable of the system.²⁷³ Generally, it is impractical to calculate the PMF $W(\xi)$, or rather the

distribution function $\langle \rho(\xi) \rangle$ from a MD simulation directly as large energy barriers along ξ will prevent accurate sampling of the configurational space (especially within the small timescales of a MD simulation). Therefore, in order to circumvent this problem, biased sampling techniques like umbrella sampling can restrain the system to allow computation of high energy configurations for an accurate calculation of the PMF along ξ .

Umbrella sampling uses a harmonic potential $w_i(\xi)$ to bias sampling in certain regions along ξ where a number of simulations, i , called ‘windows’ are run. Here:

$$w_i(\xi) = \frac{1}{2} k_\xi (\xi - \xi_i)^2 \quad 2.21$$

defines the harmonic potential imposed on the coordinate ξ in the particular region of ξ_i with the force constant k_ξ to force sampling at ξ_i . Imposing an umbrella potential in small discrete overlapping intervals over ξ results in an accurate computation over the energy barriers along the collective variable.

From these window simulations the individual probability distributions, $\langle \rho_i(\xi) \rangle$, can be combined to produce the overall $\langle \rho(\xi) \rangle$ which must factor in the correction for the bias, $w_i(\xi)$. To do this, the weighted histogram analysis method (WHAM)²⁷⁴ is often employed as a function within GROMACS²⁷⁵ upon the set of window simulations. The overall probability distribution is calculated by iteratively solving a set of coupled non-linear equations:

$$\langle \rho(\xi) \rangle = \frac{\sum_i n_i \langle \rho_i(\xi) \rangle}{\sum_j n_j \exp \left(-\frac{w_j(\xi) - F_j}{k_B T} \right)} \quad 2.22$$

$$\exp\left(-\frac{F_i}{k_B T}\right) = \int d\xi \langle \rho(\xi) \rangle \exp\left(-\frac{w_i(\xi)}{k_B T}\right) \quad \mathbf{2.23}$$

where F_i defines the free energy introduced courtesy of the biasing potential and n_i is the number of independent points to construct $\langle \rho(\xi) \rangle^{273}$.

CHAPTER 3 Characterising LPS binding with inner membrane proteins

3.1 The role of LPS within the Inner Membrane

The inner membranes (IM) of Gram-negative bacteria contain primarily phospholipids and proteins. In comparison, the outer membrane (OM) has an asymmetric composition where the outer leaflet is lipopolysaccharides (LPS) and the inner leaflet is predominantly phospholipids (PL)⁷⁷.

LPS is fundamental to the OM as its amphiphilic chemistry aids protection from both bactericidal hydrophobic and polar compounds like antibiotics and its presence maintains the structure of the membrane^{71,276}.

The glycolipid LPS is made up of three parts: lipid A, core oligosaccharide and O-antigen polysaccharide. LPS, when fully mature, is laterally extracted from the IM to be shuttled across the periplasmic space to be inserted into the outer leaflet of the OM by seven essential proteins LptA-G, a complex known as the Lpt system^{83,277}.

Despite the fact that the IM contains a prevalence of PL, over the last few years an increasing number of IMPs have been shown to co-purify with LPS such as: MscL²⁷⁸, AcrB²⁷⁹ and BtuCD²⁸⁰.

At first, this observation was mainly dismissed as an artefact of MS purification as in the process of purification of membrane proteins, both the OM and IM are disrupted mechanically and subsequently solubilised in detergent. Therefore, the lipidome and proteins of both the OM/IM can homogenise during this procedure (which also raises the question as to whether the lipids identified experimentally on MS are representative of the native environment).

However, it is also clear that in native-MS work carried out by collaborators (Leonhard H. Urner (LHU), Francesco Fiorentino (FF), Denis Shutin (DS), Mark Agasid (MA), Tarick J. El-Baba (THE), Jani Reddy Bolla (JRB), Carol V. Robinson (CVR)) that the protein-lipid interactions identified depend on the properties of both the detergents used to solubilise the protein as well as the purification protocol. Furthermore, membrane protein delipidation depends on the properties of detergent micelles in solution²⁸¹ and can be tuned by the properties of the detergent micelles such as their overall hydrophobicity and packing density²⁸².

From Native MS observation, two distinct types of IMP binding lipids have been observed by collaborators (LHU, FF, DS, MA, TRE, JRB, CVR): 'PL binding' and 'LPS binding' which can easily be distinguished by their detergent chemistry²⁸³. For 'PL binding' IMPs have their binding to PL's starkly attenuated against detergent hydrophobicity and packing density. However, for LPS binding IMPs delipidation is not significantly affected by changing the hydrophobicity and packing density of detergent micelles when using low detergent concentrations, at two times their critical aggregation concentration ($2 \times c_{ac}$, i.e. the concentration at which any detergent aggregate forms as opposed to critical micelle concentration (CMC) which is specific

to the formation of micelles). Instead, gradual delipidation of protein-LPS complexes occurs when applying different column volumes (CVs) of IMAC wash buffer containing a high concentration of detergent (50xcac). Here, protein-LPS interactions appear to be governed by far stronger hydrophobic and electrostatic interactions when compared to 'PL binding' IMPs.

Due to clear observation of two different classes of IMPs emerging from an apparently homogenous mixture of IM/OM lipids (due to cell lysing on purification), it's certainly plausible that LPS may have an important role when binding to other IMPs.

Therefore, we need to understand whether LPS binding to IMPs is physiologically relevant or just an artefact of MS purification and whether we can comment upon LPS having a larger role to play within the IM than originally understood. If this is true, how then does LPS differ in its interactions with its specific inner membrane transporter protein LptB₂FGC versus other IMPs? However, it is important to mention that there is no noteworthy data on IMP-LPS interactions (other than the IMP proteins which constitute the LPS biosynthetic pathway) outside native-MS experimental observations.

In order to address the first question, I will consider two case studies, one with a PL-binding protein AmtB and the other considering an LPS-binding protein BtuCD. Later I will compare these case studies with work carried out with the IM LPS transporter LptB₂FGC to understand the behaviour of LPS

3.1.1 Case study 1: PL-binding protein AmtB

The trimeric ammonia channel (AmtB) was chosen by collaborators (LHU, FF, DS, MA, TRE, JRB, CVR) as a model system to study lipid binding as it is an IMP of *E. coli* and its PL-binding properties have been established before in a previous native-MS study²⁸⁴. The binding properties of AmtB was compared with those of an LPS-binding, sensor histidine kinase, AcrB²⁷⁹.

Collaborators (LHU, FF, DS, MA, TRE, JRB, CVR) fully delipidated AmtB and AcrB and titrated equimolar amounts of lipids PG, PE, Cardiolipin, lipid A and Re-LPS (a derivative of LPS with two ketodeoxyoctonic acids (Kdo) residues belonging to the core oligosaccharide bound to lipid A) and observed lipid binding intensities relative the apo-MS state using native-MS²⁸³. Despite the nature of the protein, whether it was 'PL-binding' (AmtB) or 'LPS-binding' (AcrB), there was not apparent discrimination for lipid binding from either. Collaborators (LHU, FF, DS, MA, TRE, JRB, CVR) found, that for both proteins, higher relative intensities of protein-CDL, protein-lipid A and protein-Re-LPS binding compared to lower relative intensities of protein-PE/PG binding

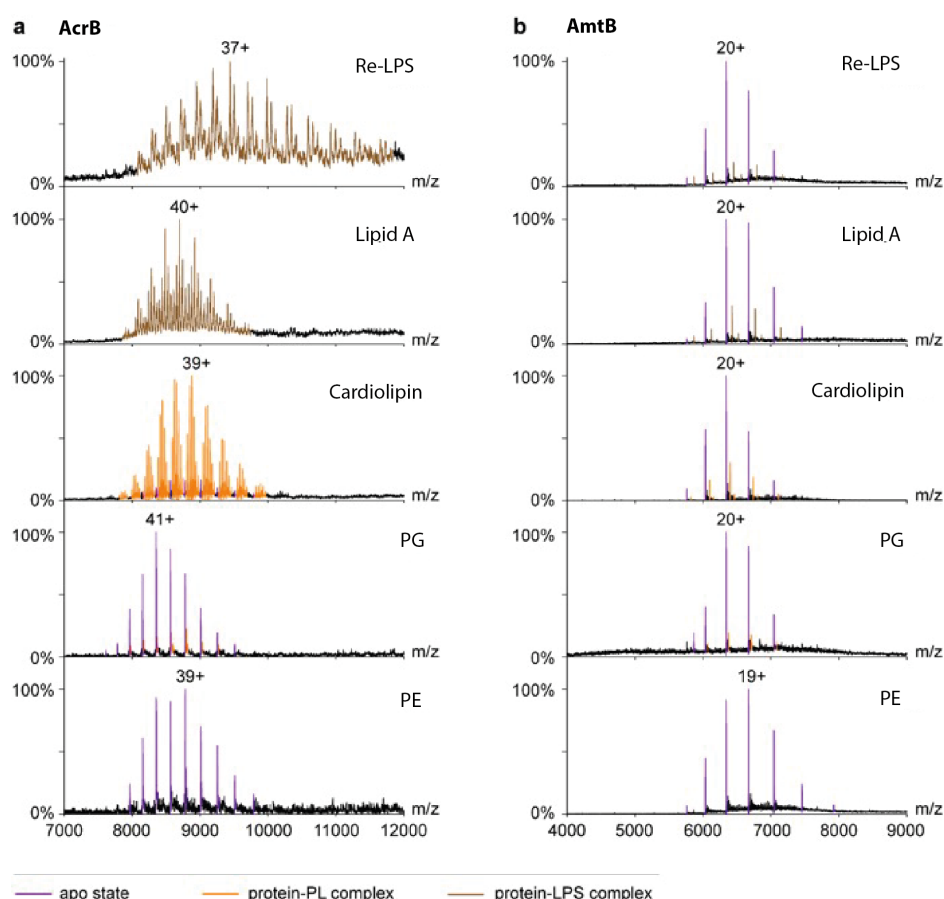


Figure 3.1 Observing AcrB and AmtB lipid interactions with native-MS. Figure modified from Urner *et al*²⁸³, all work here was conducted by collaborators (LHU, FF, DS, MA, TRE, JRB, CVR) Fully delipidated AcrB and AmtB were mixed with equimolar amounts of Re-LPS, lipid A, cardiolipin, POPE and POPG under the same conditions and their protein-lipid interactions were monitored using native-MS. Both AmtB and AcrB demonstrated higher relative protein-lipid intensities for Re-LPS, lipid A and cardiolipin compared to protein-PL intensities. This suggests that co-purified lipids for either AmtB or AcrB are indicative of their local lipid environments within the native membrane.

To help verify these experimental results obtained by collaborators (LHU, FF, DS, MA, TRE, JRB, CVR), I can compare these results with *in silico* studies of AmtB, where I placed the protein in a membrane of 83 % POPE, 5 % cardiolipin, 10 % POPG and 2 % of either Lipid A or Re-LPS. Clearly, both lipid A and Re-LPS can bind to AmtB if given the opportunity to do so (**Figure 3.2**) and demonstrate higher binding occupancy than with POPG. However, I was not able to observe greater amounts of occupancy

contacts with cardiolipin (compared to POPG), though this would most likely be due to competition between the lipids in mixed bilayer.

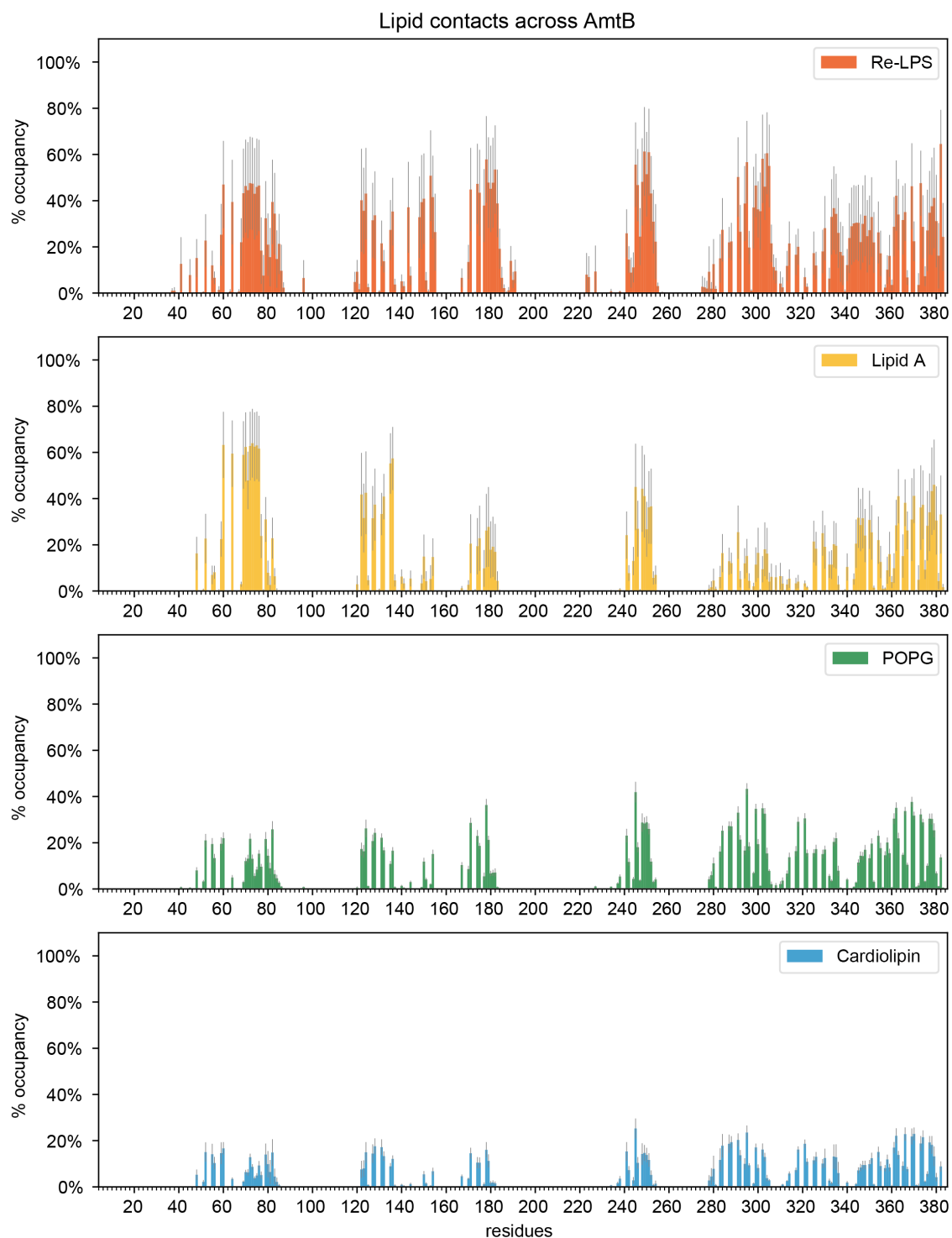


Figure 3.2 % lipid occupancy plots for AmtB. Average % lipid contacts for Re-LPS, lipid A, cardiolipin and POPG were monitored for AmtB in a bilayer of composition 83 % POPE, 5 % cardiolipin, 10 % POPG and 2 % of either lipid A or Re-LPS. As with the experimental results, contacts was observed for lipid A or Re-LPS despite AmtB

being described as having PL-binding properties. If given the opportunity LPS derivatives will bind to AmtB if these lipids are proximal within the environment. LPS is found within the IM before transport, but it is unknown whether they freely interact with IMPs in the cell itself.

The standard purification procedure performed by collaborators (LHU, FF, DS, MA, TRE, JRB, CVR) consistently yield AmtB bound with PL's and AcrB bound to LPS. However, after delipidation both AmtB and AcrB form complexes with LPS (as also revealed in the *in silico* binding studies), therefore this suggests that co-purified lipids are representative of the lipidome of the proteins in their native state. Despite the possibility that OM/IM lipids can mix during the purification process, it seems that the experimental and MD simulations suggest that LPS derivatives will bind to AmtB if these lipids are proximal within the environment. Therefore, this implies that OM/IM 'mixing' during purification is far less homogenous than otherwise assumed and hence co-purified lipids with IMPs are representative of either their PL or LPS rich local lipid environments. Furthermore, this continues to suggest that LPS has a greater and more promiscuous role within the IM than originally appreciated. It will be useful now to consider an LPS-binding protein, BtuCD to shed light on the potential role LPS could play with other proteins in the inner membrane.

3.1.2 Case study 2: LPS-binding protein BtuCD

BtuCD is an ABC transporter, a class of important drug targets as they mediate the transport of vital molecules, such as vitamin B₁₂ (in the case of BtuCD) as well as biosynthetic precursors and trace metals through membranes. As such, BtuCD forms a heterotetramer that consists of two transmembrane subunits (BtuC) that non-covalently interact with two cytoplasmic ATPase domains (BtuD)²⁸⁵.

Generally, it has been shown by collaborators (LHU, FF, DS, MA, TRE, JRB, CVR) that LPS-binding IMPs require a number of different column washes of IMAC buffer at high detergent concentrations to delipidate the protein. BtuCD is an unusual example because this is not the case. It is important to mention that BtuCD was overexpressed in purification. A greater prevalence of any MP in the membrane will likely bias for the local lipid environment it would favour natively (much like a fingerprint¹⁰⁵), therefore we could be studying an overcompensation of LPS. Collaborators were not fully able to remove LPS when purifying BtuCD from C43(DE3) *E. coli* cells using conditions previously optimized for another LPS-binding protein AcrB (even after 40 column washes) (**Figure 3.3a**). Complete delipidation could only be achieved when purifying BtuCD from ClearColi® *E. coli* cells whose cell wall contains no LPS but lipid IV_A - a LPS precursor with no additional glycan head (**Figure 3.3c**). ClearColi® *E. coli* cells have seven genetic deletions which remove two secondary acyl chains from LPS to produce lipid IV_A where one additional compensating mutation enables cell viability with lipid IV_A²⁸⁶.

To understand what underpins these markedly strong BtuCD-LPS interactions shown by experiments performed by collaborators (LHU, FF, DS, MA, TRE, JRB, CVR) where LPS was not fully removed from BtuCD after 40 column washes in detergent, I performed coarse grained molecular dynamics (MD) simulations in an IM bilayer containing 2% of either core-LPS, Re-LPS or lipid A. I observe contacts between LPS (considering all the atoms of the lipid, i.e. the core-saccharide headgroup and acyl tails) with the BtuCD monomer as well as contacts common across all the LPS derivatives important for LPS binding (**Figure 3.3b and Figure 3.4**). The amino acid

sequences corresponding to lipid A, Re-LPS and core-LPS involved in LPS-binding are repetitive in nature and can be regarded as ‘fingerprints’ (**Figure 3.4**). *In silico* work from studies conducted with outer membrane proteins reveal a similar pattern which therefore implies that protein-LPS interactions for both the IM and OM are driven by comparable intermolecular forces²⁸⁷.

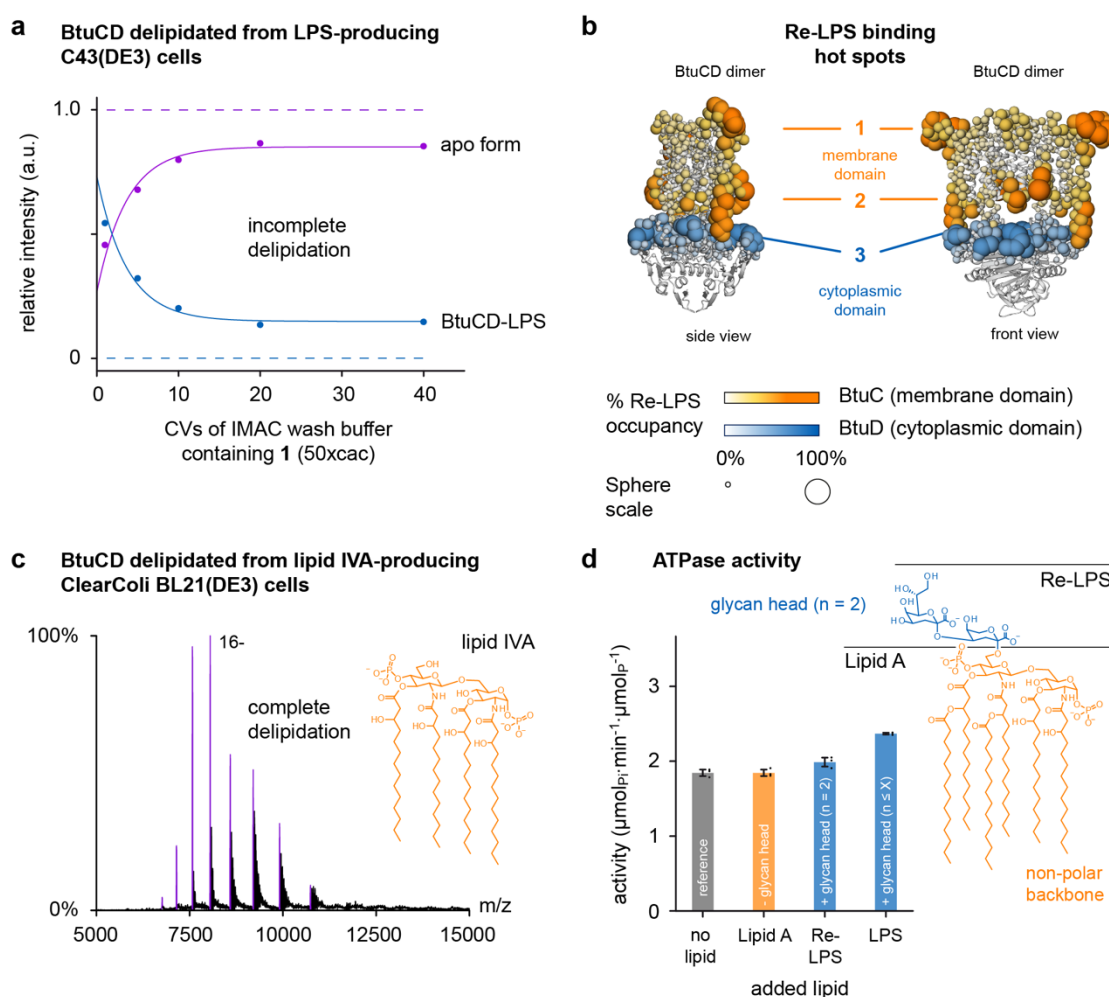


Figure 3.3 Functional and structural behaviour of BtuCD with LPS interaction. Figure modified from Urner *et al*²⁸³, all work here for (a),(c-d) was conducted by collaborators (LHU, FF, DS, MA, TRE, JRB, CVR) whereas (b) was work carried out by myself. (a) Normal delipidation methods failed to completely remove LPS from BtuCD derived from ‘core’ LPS-producing *E. coli* strains. (b) % Re-LPS occupancy contact sites for LPS binding with BtuCD for both the transmembrane domain BtuC (orange) and cytoplasmic domain BtuD (blue) respectively, where the C α of the back backbone are represented as spheres and are scaled by % lipid contact. (c) BtuCD was expressed and purified from ClearColi®, a construct which features lipid IVA instead of LPS in its cell wall, in order to fully delipidate the protein. (d) BtuCD ATPase activity was quantified with LPS derivatives with increasing length of the glycan

headgroup: lipid A ($n = 0$) < Re-LPS ($n = 2$) < 'core' LPS ($2 < n \leq 10$). The ATPase activity of BtuCD increases as the number of glycan groups increase.

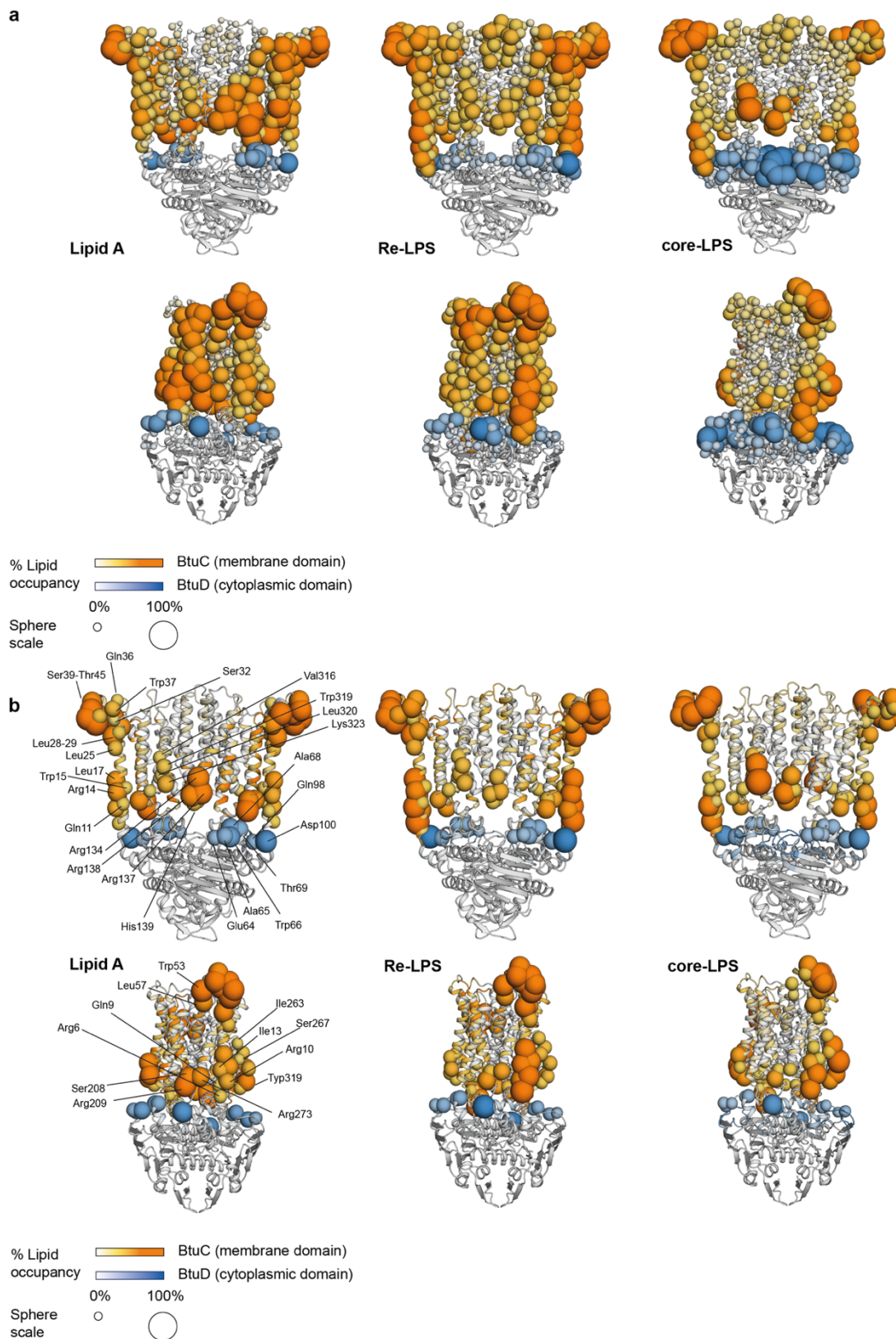


Figure 3.4 Lipid binding across BtuCD. (a) % Lipid occupancy highlighted by colour orange and blue for transmembrane and cytoplasmic domains respectively. Ca

spheres are scaled by lipid occupancy from 0 to 100%. **(b)** Residues highlighted show only residues which are both above 25% occupancy (arbitrary cut-off) and in common between each lipid species (lipid A, Re-LPS and core-LPS). Amino acid sequences involved in binding to 'LPS can be described as 'fingerprints' as they follow a repetitive trend.

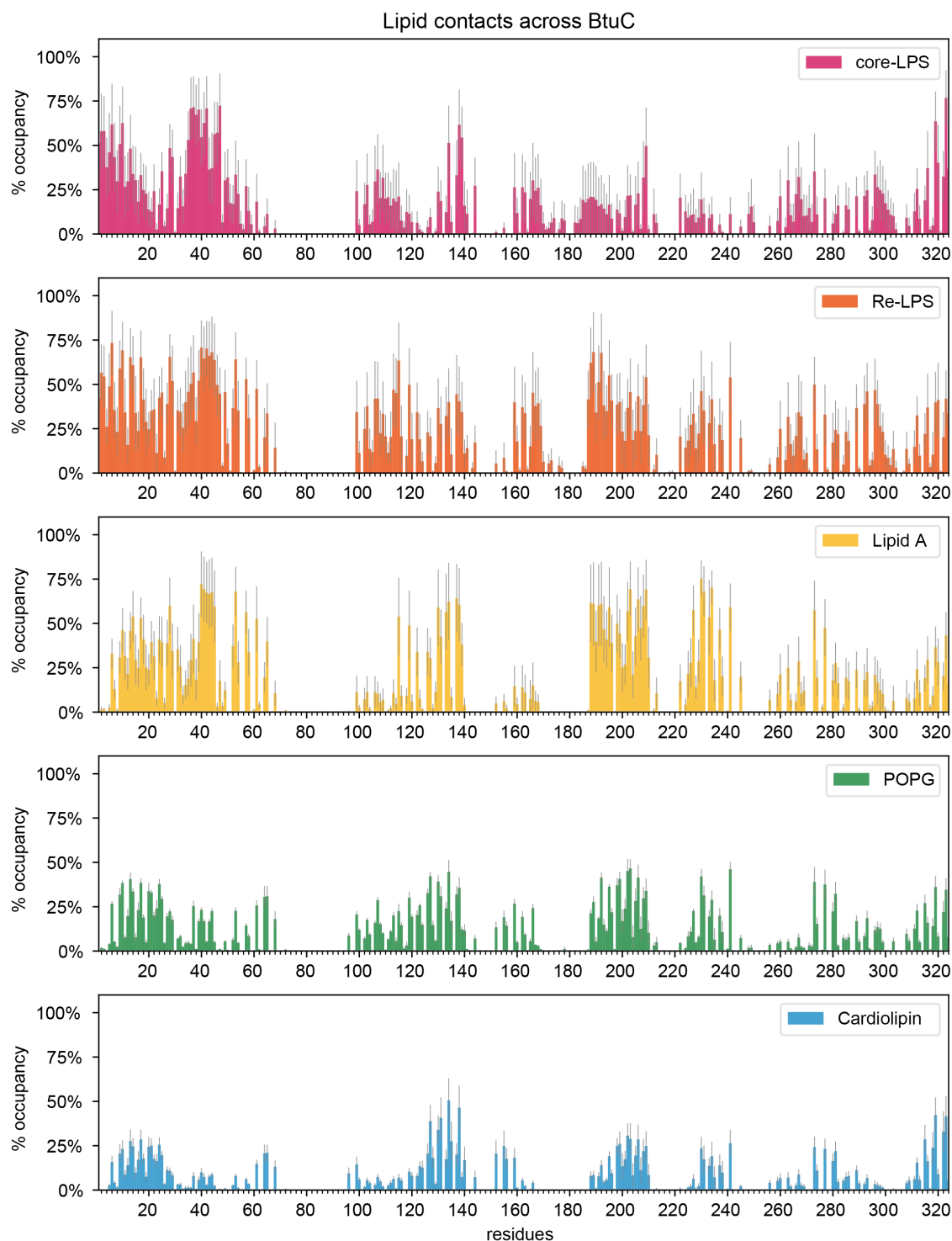


Figure 3.5 Quantifying LPS contacts across the transmembrane domain of BtuCD (BtuC). Figure adapted from Uner *et al*²⁸³, all work here completed by myself. % occupancy plots of lipid A (n = 0), Re-LPS (n = 2), 'core' LPS (2 < n ≤ 10)

contacts with BtuC. Error bars shown in grey represent the standard deviation between five replicates. Cardiolipin and POPG % occupancy plots shown here as a control.

3.1.2.1. Soluble domains of the ABC-transporter with LPS

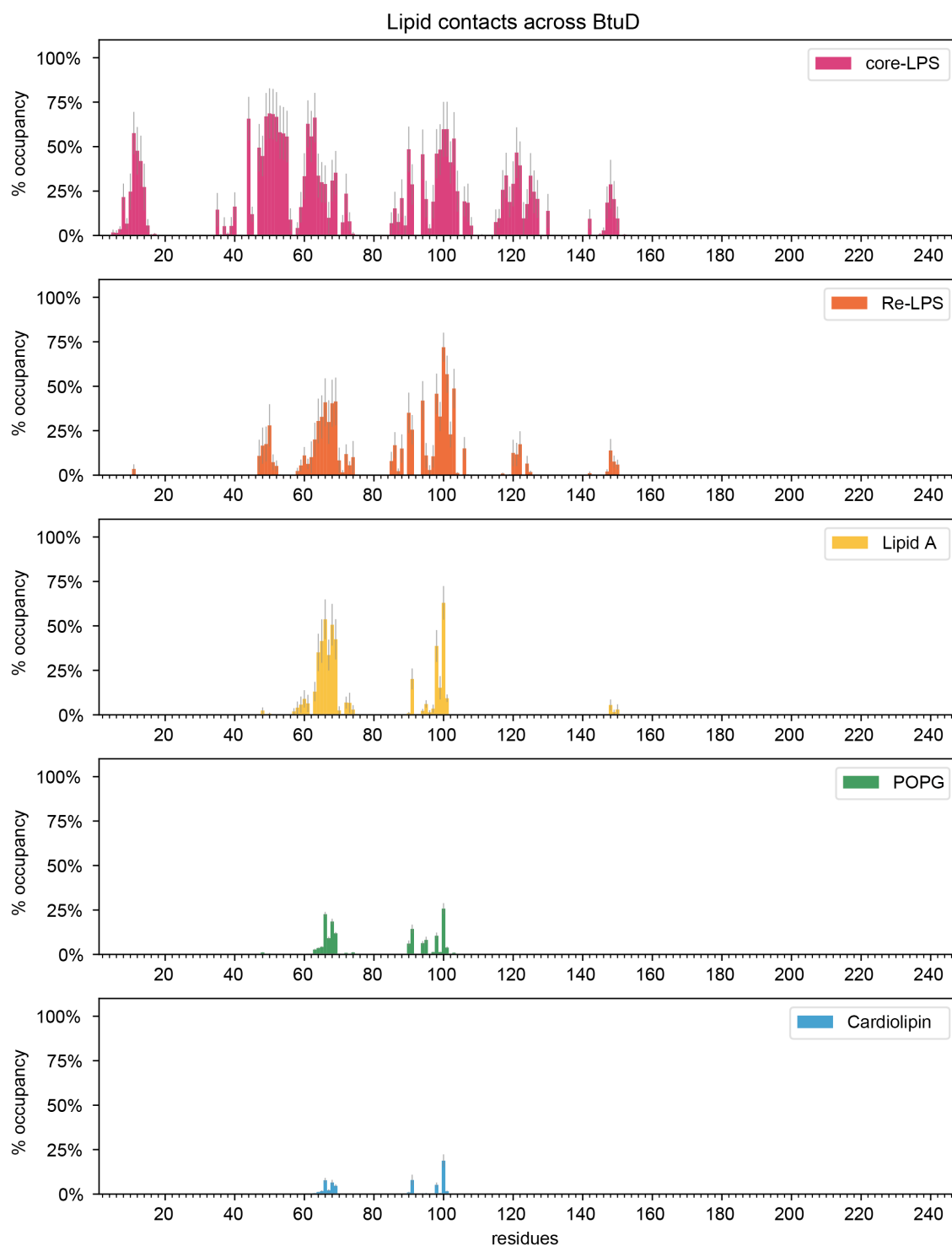


Figure 3.6 Quantifying LPS contacts across the soluble domain of BtuCD (BtuD). Figure adapted from Uner *et al*²⁸³, all work here completed by myself. % Occupancy of lipid contacts across BtuD and error bars shown in grey represent the standard deviation between five replicates. Contacts increase with the number of glycan units (n) attached to lipid A in the following order: lipid A (n = 0), Re-LPS (n = 2), 'core' LPS (2 < n ≤ 10).

Unsurprisingly, we found that the relative abundance of lipid contacts at the cytoplasmic region BtuD increases with the saccharide units (n) attached to lipid A core of LPS: lipid A (n = 0) < Re-LPS (n = 2) < 'core' LPS (n = 10) (**Figure 3.6**)

It is clear from the MD simulations that the glycan units in 'core' LPS increase the abundance of contacts with the cytoplasmic domain BtuD. This implies that the saccharide units contribute towards the strength of BtuCD binding and helps to explain why 'core'-LPS cannot be delipidated by collaborators (LHU, FF, DS, MA, TRE, JRB, CVR) using their standard protocol.

3.1.2.2. LPS interacts with ABC cassette of BtuCD near the binding site of ATP.

From experimental work carried out by collaborators (LHU, FF, DS, MA, TRE, JRB, CVR), LPS has been shown to bind strongly to BtuCD, as it was not able to be delipidated at all from 40 column washes of excess detergent. Furthermore, my MD simulations reveal that the saccharide units of LPS contribute to the strength of LPS binding with the ABC-transporter. What has not been established, though, is whether LPS plays a functional role with the transporter and if it could modulate the ATPase activity of the protein. Collaborators (LHU, FF, DS, MA, TRE, JRB, CVR), measured the ATPase activity of apo-BtuCD in the presence of lipid A (n, number of saccharide units = 0), Re-LPS (n = 2), 'core' LPS (2 < n ≤ 10), and similar to the abundance of

cytoplasmic contacts made with the soluble domain revealed by MD simulations, ATPase activity was found to increase with increasing glycan units. Greater ATPase activity was observed on binding with (8% higher) Re-LPS and (20% higher) core-LPS (**Figure 3.3d**) significant to three repeat experiments. This suggests that the ATPase activity is stimulated in the presence of glycans.

To compare to experimental observations, my MD simulations reveal that the additional saccharide units in Re-LPS ($n = 2$) can form contacts with the ATP-binding pocket of BtuD (**Figure 3.7**). Simulations with 'core' LPS ($2 < n \leq 10$) also reveal that these glycan units increase the contacts with the ATP-binding pocket of the cytoplasmic domain (**Figure 3.7**). Therefore, the presence of these glycan units increases the interactions of the lipid with the ATP-binding pocket and therefore stimulate greater ATPase activity when bound to BtuCD.

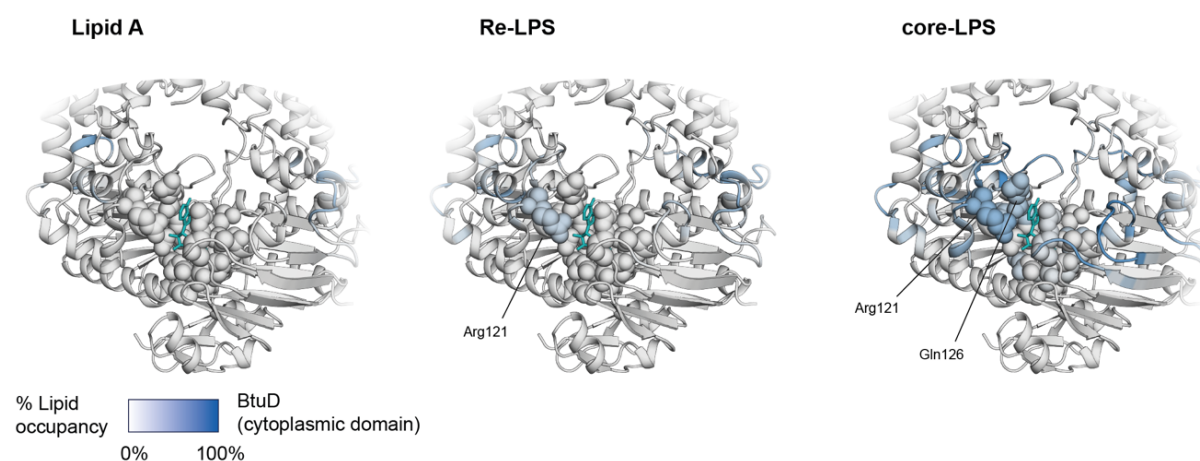


Figure 3.7 BtuCD cytoplasmic binding contacts between Re-LPS and core-LPS with the ATP-binding pocket. Figure adapted from Urner *et al*²⁸³, all work here completed by myself. The ATP-binding pocket in BtuD is highlighted in spheres and defined as those residues 4 Å away from the bound nucleotide which is drawn as sticks and coloured cyan. Residues which contact the saccharide units of LPS are coloured blue, where interactions with the ATP pocket increase as the number of glycan units (n) increase in the order: lipid A ($n = 0$) < Re-LPS ($n = 2$) < 'core' LPS ($2 < n \leq 10$). Residues Arg121 and Gln126 (of the ATP binding pocket) are shown to contact the saccharide units of LPS may be particularly relevant for the activity of the transporter.

3.1.3 LptB₂FGC soluble binding contacts with LPS

It is helpful, at this stage, to compare these observations, with an IMP definitively known to bind LPS within the IM – the LPS transporter protein itself, LptB₂FGC. This is clearly seen in LPS bound cryo-EM structures of the complex (**Figure 3.8**)⁸⁶.

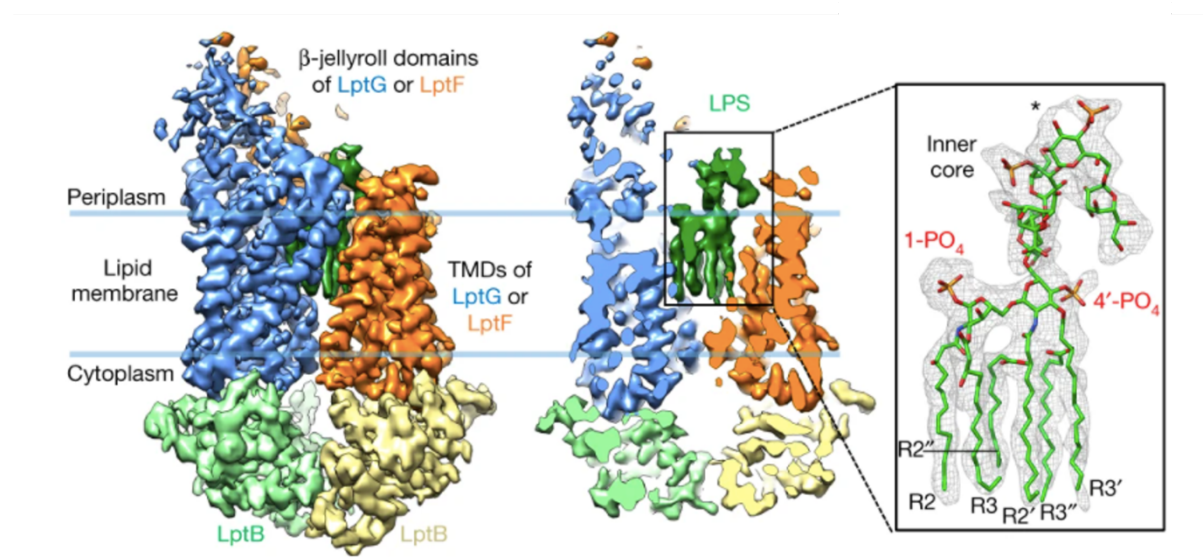


Figure 3.8 Interaction of liposaccharide with LptB₂FGC. Figure adapted from Li *et al.*⁸⁶. Surface and cross-sectional view of the LptB₂FG cryo-EM density with the LPS molecule inside the hydrophobic cavity of LptG (blue) and LptF (orange). Insert shows the atomic model of LPS superimposed with the cryo-EM density (grey).

The complex which resides in the inner membrane which extract LPS, is known as the ATP-binding cassette (ABC) transporter LptB₂FGC^{288,289}. The transporter orthogonally extrudes LPS to the soluble portion of the complex, LptC which forms a bridge with several LptA units across the periplasm and ultimately interacts with the β -taco domain of LptD^{91,290,291}, where the heterodimeric complex LptDE finally installs LPS within the OM (**Figure 3.9**).

The subcomplex LptB₂FGC itself is composed of 5 subunits and is known as a type VI transporter²⁹². It is an unusual ABC transporter in so far as it transports LPS directly from the bilayer compared to facilitating the passage of the substrate across the membrane like most ‘classical’ transporters²⁹². The nucleotide-binding domains (NBDs) of the transporter, LptB, hydrolyse ATP for the energy to transport LPS through the system and sit on the cytoplasmic side of the IM. LPS laterally diffuses into the hydrophobic cavity made between the transmembrane domains (TMDs) of LptF and LptG which both feature β -taco domains, where the β -taco pertaining to LptF complexes with LptC to form the periplasmic bridge. Between the hydrophobic cavity of the TMDs is a transmembrane α -helix (TM_C) which helps to regulate ATPase activity for more efficient LPS lateral extraction and subsequent transport^{86,87,293}.

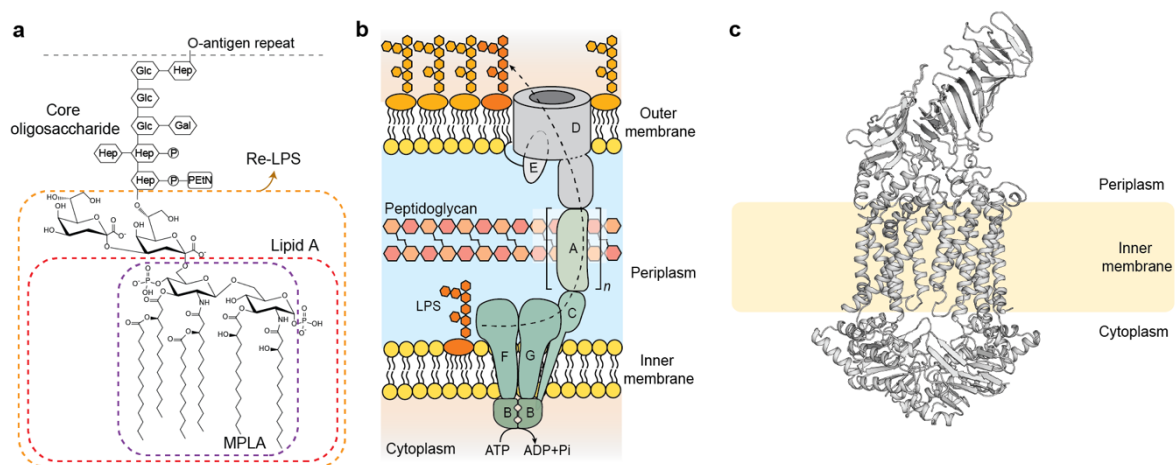


Figure 3.9 Cartoon schematic describing the LPS transport pathway. Figure adapted from Fiorentino *et al*⁵¹, all work here completed by myself and FF. (a) A skeletal representation of LPS where its tripartite structure is described: lipid A, core oligosaccharide and the O-antigen repeating unit. The orange box highlights a derivative of LPS, called ‘Re-LPS’ which consists of the lipid A unit connected to two ketodeoxyoctonic acid (Kdo) moieties. The red box shows the lipid A unit and the purple box outlines the mono-phosphate lipid A (MPLA) **(b)** cartoon schematic of the Lpt system: LPS is extracted from the inner membrane (IM) by the ABC transporter complex LptB₂FGC, where it is subsequently shuttled through the periplasm via a soluble bridge of LptA units to be finally installed in the outer membrane (OM) by LptDE. **(c)** Recently obtained Cryo-EM structure by collaborators (FF, JRB, CVR) of LptB₂FGC from *Pseudomonas aeruginosa*. The soluble β -taco domain of LptC is known to interact with LptA to form the periplasmic bridge. The transmembrane

domains LptF, LptG and TM_c (which I modelled in by using the resolved structure of *E. cloacae* as a template) form a hydrophobic cavity which laterally extracts LPS from the IM to be extruded through the periplasmic bridge. The transport of LPS is powered by ATP which is hydrolysed by the ABC transporter cassette protein LptB.

As with BtuCD, it is interesting to report that collaborators (FF, JRB, CVR) could only remove LPS adducts from LptB₂FGC only after expression in ClearColi® - which implies the strength of the glycan units binding to the transporter. I performed coarse grained molecular dynamics (MD) simulations of LptB₂FGC in an IM bilayer containing 2% of either Re-LPS, lipid A or MPLA. In a similar manner to what was observed with BtuCD, increased contacts are apparent between lipids with glycans and the soluble domain of the ABC-transporter LptB₂FGC. More contacts are clearly observed between Re-LPS (n = 2) and LptB when compared to lipid A (n = 0) and LptB (**Figure 3.10**). Here also, shows the effect of removing one of the phosphate groups from lipid A (MPLA), which arguably and unsurprisingly decreases the contacts with the soluble domain.

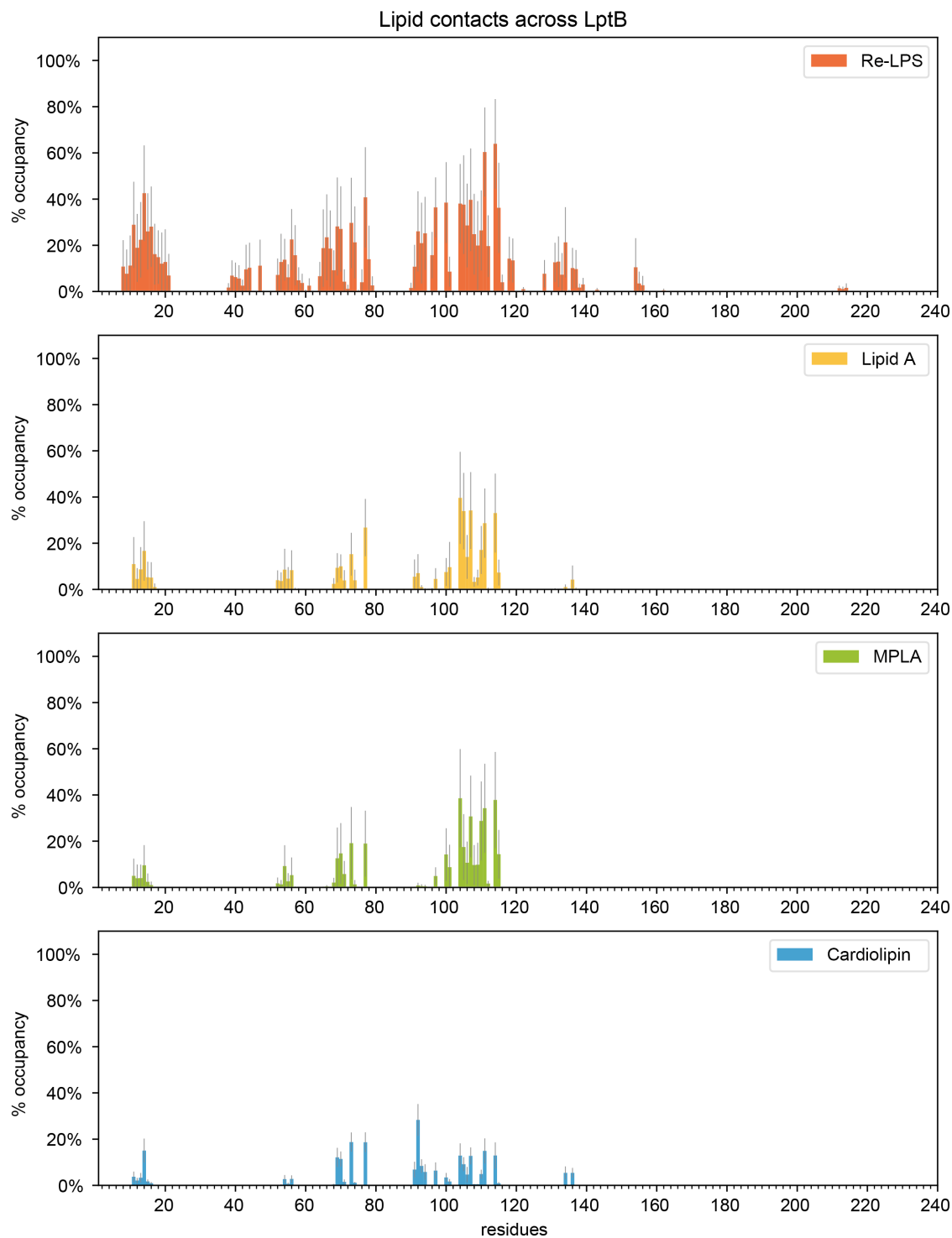


Figure 3.10 Investigating lipid contacts across the cytoplasmic domain of LptB₂FGC. % Occupancy of lipid contacts across LptB and error bars shown in grey represent the standard deviation between five replicates. Contacts increase with the number of glycan units (n) attached to lipid A in the following order: lipid A (n = 0) and Re-LPS (n = 2). Removing one of the phosphate groups also reduces the number of contacts of the lipid with the soluble domain.

Native MS experiments conducted by collaborators (FF, JRB, CVR) have shown an increase in relative intensities between LPS derivatives and LptB₂FGC in the order: monophosphate lipid A (MPLA), lipid A and Re-LPS. Therefore, glycan units contribute to the overall binding strength between LPS and LptB₂FGC as with BtuCD.

3.1.3.1. LPS interacts with the ABC cassette of LptB₂FGC

Important residues for the nucleotide and LPS binding has been characterised in recent cryo-EM and X-ray crystal structures which have also helped to describe the regulatory role of LptC within the complex itself^{86,87,293}. Further to this, and again, in a similar manner to BtuC, MD simulations reveal that Re-LPS (n = 2) interacts with the ABC cassette of LptB₂FGC near the binding site of ATP (**Figure 3.11**).

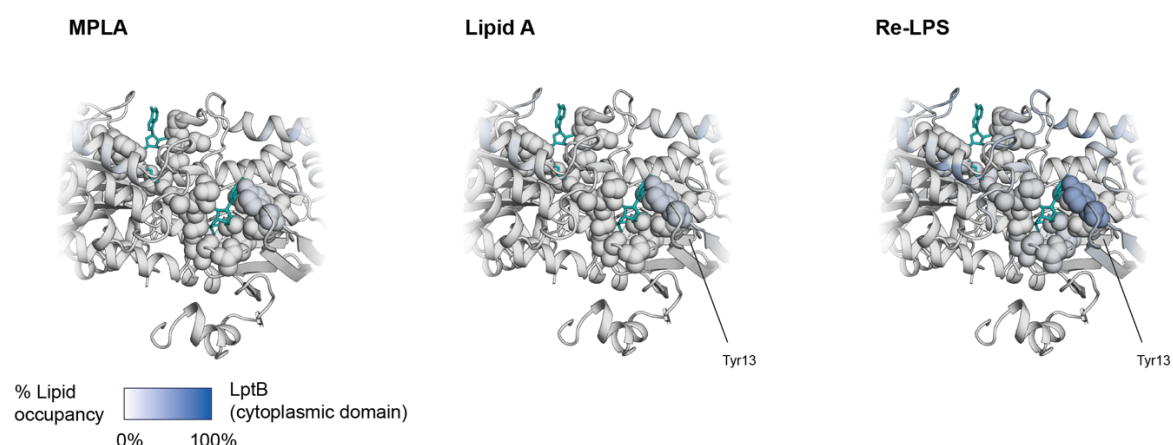


Figure 3.11 Interactions between LPS and the ATP-binding pocket of LptB₂FGC. The ATP-binding pocket in LptB is highlighted in spheres and defined as those residues 4 Å away from the bound ATP-PNP which is drawn as sticks and coloured cyan. Residues which contact the saccharide units of LPS are coloured blue, where interactions with the ATP pocket increase as the number of glycan units (n) increase in the order: MPLA (n = 0) < lipid A (n = 0) < Re-LPS (n = 2). Residue Tyr13 (of the ATP binding pocket) are shown to contact the saccharide units of LPS may be particularly relevant for the activity of the transporter.

Taken together, the experimental data obtained by collaborators (LHU, FF, DS, MA, TRE, JRB, CVR) as well as my MD simulations show that the number of saccharide

units in 'core' LPS not only contribute to the strength of protein-LPS interactions but can also stimulate protein function to a significant degree. Comparing observations from IMPs such as LPS-binding BtuCD with LptB₂FGC, which is known to transport LPS, and drawing similarities between their function implies that the role of core-LPS within the inner membrane is perhaps far larger than originally thought.

3.2 Characterising LPS interactions with LptB₂FGC

Despite recent structures helping to elucidate key interactions of LptB₂FGC between LPS and ATP^{86,87,293}, mechanistic details of transport remain elusive. It is still not yet clear how the transporter discriminates between lipids in the inner membrane to selectively extract LPS for transport. Due to the diversity of lipids within the IM, there also presents the possibility of other lipid species competing for or regulating the lateral recognition and diffusion of LPS. Similarly, is it still not clear how the LptC transmembrane helix works to regulate LPS transport and interacts with its substrate. These questions are essential to understanding the mechanism of transport but also are critical for the development of new antibacterials which target the biogenesis of LPS.

Currently, there are two published structures which fully describes all units of the ABC-transporter with the important LptC transmembrane helix and soluble domain from *V. cholera* (PDB: 6MJP) and *E. cloacae* (PDB: 6MIT)⁸⁷. Here, the emergence and comparison of different conformations of the protein can help provide much clarity between the stages of the mechanism of LPS transport. To this end, collaborators (FF, JRB, CVR, Chih-Chia Su (CS) and Edward W Yu (EWY)) describe a Cryo-EM structure of LptB₂FGC derived from *Pseudomonas aeruginosa* which features a

different conformation of the β -taco domain compared to previously described structures. As of February 2022, this structure has not been deposited in the PDB or published.

Further to this, structural and functional understanding of the LPS biogenesis pathway from *Pseudomonas aeruginosa*, in particular, will assist in antibiotic development. The pathogen itself is known to cause serious infection in patients already diagnosed with immunodeficiency diseases such as AIDS²⁹⁴. Moreover, LPS from *P. aeruginosa* cooperates with the type III secretion system to facilitate toxin injection into eukaryotic cells²⁹⁵.

In silico experiments can be employed here to help explain the functional and structural difference between these two structures and consequently shed light on the LPS mechanism of transport.

3.2.1 Structural studies of LptB₂FGC with its substrate

In year 2021, the new Cryo-EM structure of LptB₂FGC of *P. aeruginosa* reveals a skewed LptC β -taco domain compared with models of LptB₂FGC from *V. cholera* and *E. cloacae* with an upright β -taco domain. The structure with a LptC-skewed β -taco domain however, does not resolve the LptC TM helix which sits between LptF and LptG. To rectify this, I modelled in the TM_C (using MODELLER²⁹⁶) from *P. aeruginosa* using the resolved structure of *E. cloacae* as a template. In addition, I created a homology model of LptB₂FGC from *P. aeruginosa* in the β -taco domain in the upright position using SWISS Homology modeller²⁹⁷), using again the structure derived from *E. cloacae* as a template (**Figure 3.12**).

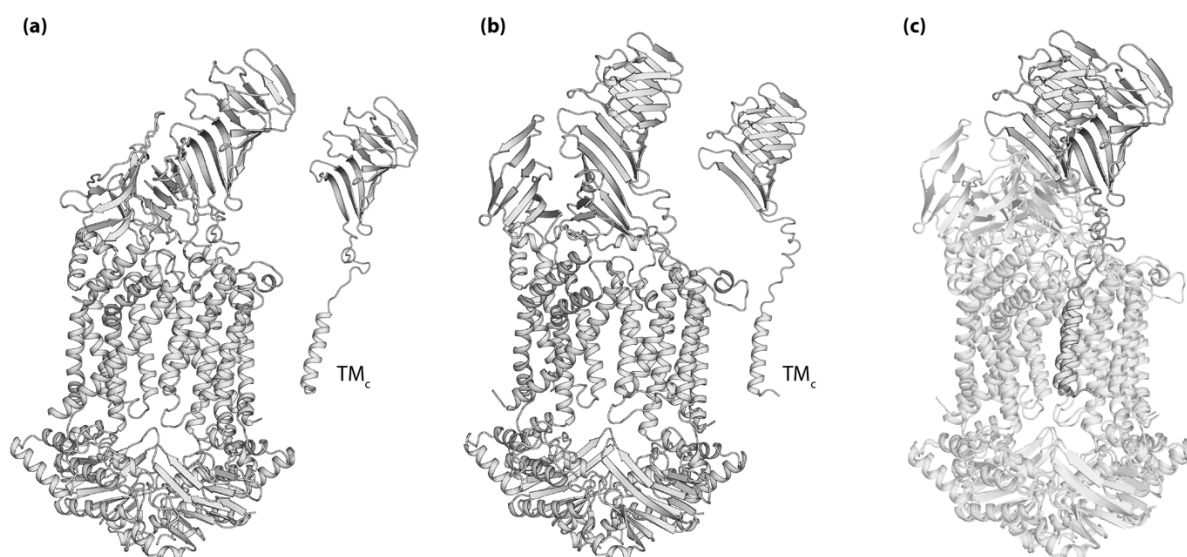


Figure 3.12 Structures of the 'LptC-skewed' and 'LptC-upright' conformation of LptB₂FGC transporter. (a) Cryo-EM structure of the 'LptC-skewed' conformation of LptB₂FGC where the TM_c helix has been modelled in using the resolved structure of *E. cloacae* as a template. (b) PaLptB₂FGC homology model of 'LptC-upright' conformation (c) 'LptC-skewed' and 'LptC-upright' conformations aligned on top of each other.

Alongside the new Cryo-EM structure, collaborators (FF, JRB, CVR, CS and EWY) performed native-MS binding experiments with the LptB₂FGC complex and observed interesting competitive binding effects between Re-LPS and cardiolipin. They found that adding increasing concentrations of cardiolipin (0.5 to 10 μ M) to a previously incubated solution of LptB₂FGC (1.5 μ M) and Re-LPS (1 μ M) significantly reduced the total amount of Re-LPS bound to the protein in a concentration-dependent manner. Atomic force microscopy (AFM) studies show that that LPS enriched, OMP-depleted regions may facilitate insertion of new membrane components⁵⁷, where fluorescence experiment suggests that new OMPs move out to the poles²⁹⁸. This could imply that the LPS complexes accumulate at the plane of division away from poles where cardiolipin localises such that the LptB₂FGC can freely transport LPS to the OM in the creation of new membrane.

In both structures of the complex with distinct positions of the soluble LptC β -taco domain, 'LptC-skewed' or 'LptC-upright', I wanted to compare and contrast the binding of its substrate within the cavity of the transporter for both conformations. In doing so, I could directly quantify the binding interactions between the substrate and the hydrophobic cavity with TM_c alongside the apparently competitive lipid cardiolipin. I can also compare these direct interactions with previous LPS bound LptB₂FG-TM_c structures from *Shigella flexneri*²⁹³. These LPS bound structures also show the lipid having shifted from being bound to the LptG side with TM_c present (SfLptB₂FG-TM_c) to the LptF side without TM_c (SfLptB₂FG). It was noted that the presence of TM_c enlarges the cavity and therefore reduces LPS interactions inside the transporter, without TM_c present the cavity reduces in size and forces LPS to shift position²⁹³. Both the LptC-skewed and LptC-upright structures of *P. aeruginosa* feature TM_c which affords direct observation of LPS behaviour between comparably sized cavities so that we can comment on the effect of LPS position within the transporter.

3.2.2 Lipid interactions with LptB₂FGC

3.2.2.1. Characterising lipid interactions with LptC

I performed three repeats of 200 ns atomistic simulation with Re-LPS, lipid A, MPLA and cardiolipin (as a control) bound within the cavity as well as the apo state of both the LptC-skewed and LptC-upright structures of *P. aeruginosa*. For each lipid the initial pose was templated and aligned from the bound LPS structure of SfLptB₂FG-TM_c (PDB ID: 6S8N)²⁹³. I subsequently calculated the lipid species occupancy for both conformations (**Figure 3.13, 3.14**) and mapped these contacts on the structures as B-factors (**Figure 3.15, 3.16**).

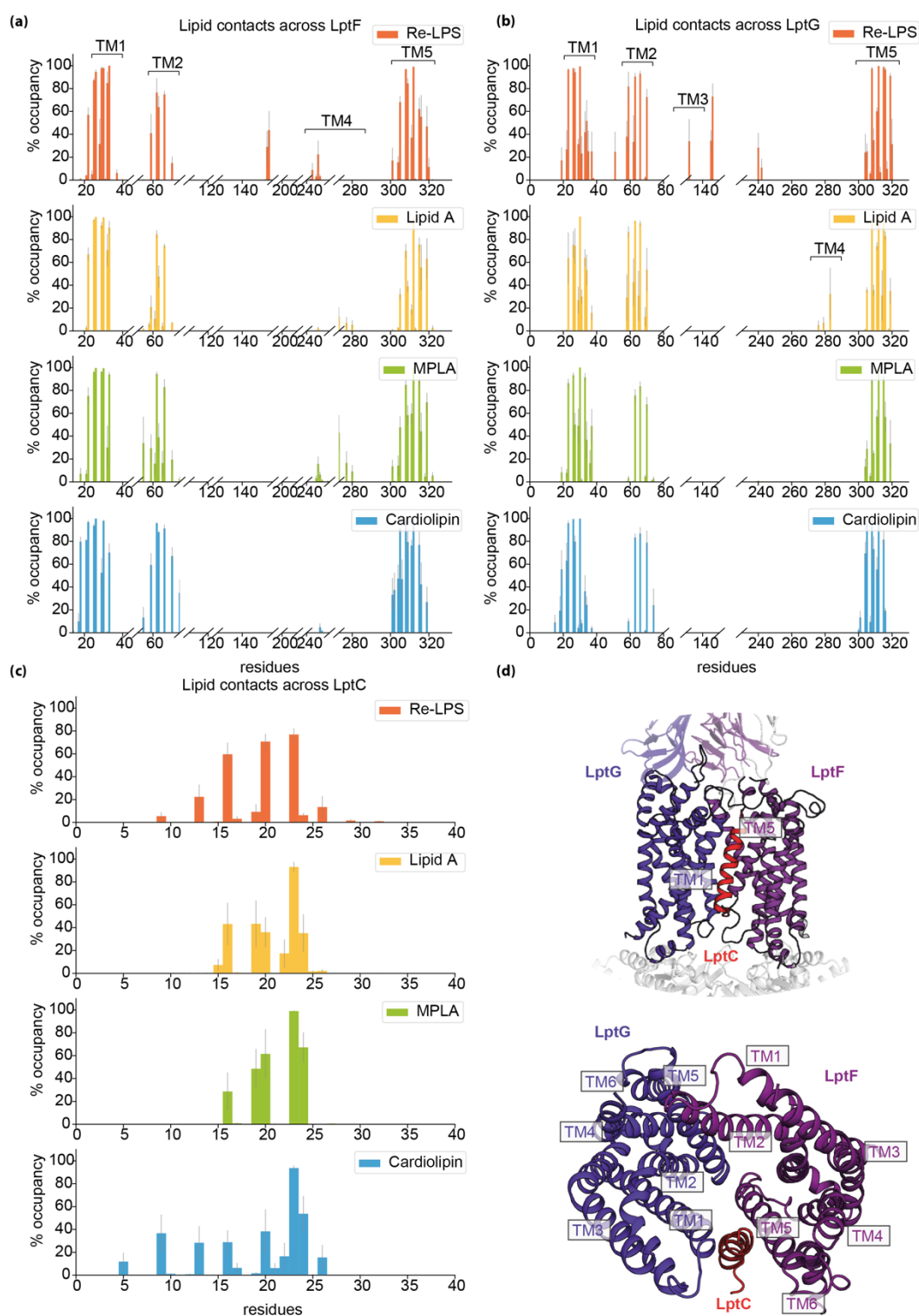


Figure 3.13 Lipid contacts with the hydrophobic cavity of the skewed-LptC conformation. Average % Occupancy of residue contacts between the LptC-skewed conformation of LptB₂FGC with Re-LPS, lipid A, MPLA and cardiolipin within the hydrophobic cavity. Average % occupancy contacts with lipids: Re-LPS, lipid A, MPLA and cardiolipin made between (a) LptF (b) LptG and (c) LptC. Error bars represent standard deviation over the three replicates. (d) TM helices in contact with lipids indicated on contact maps and labelled onto structure.

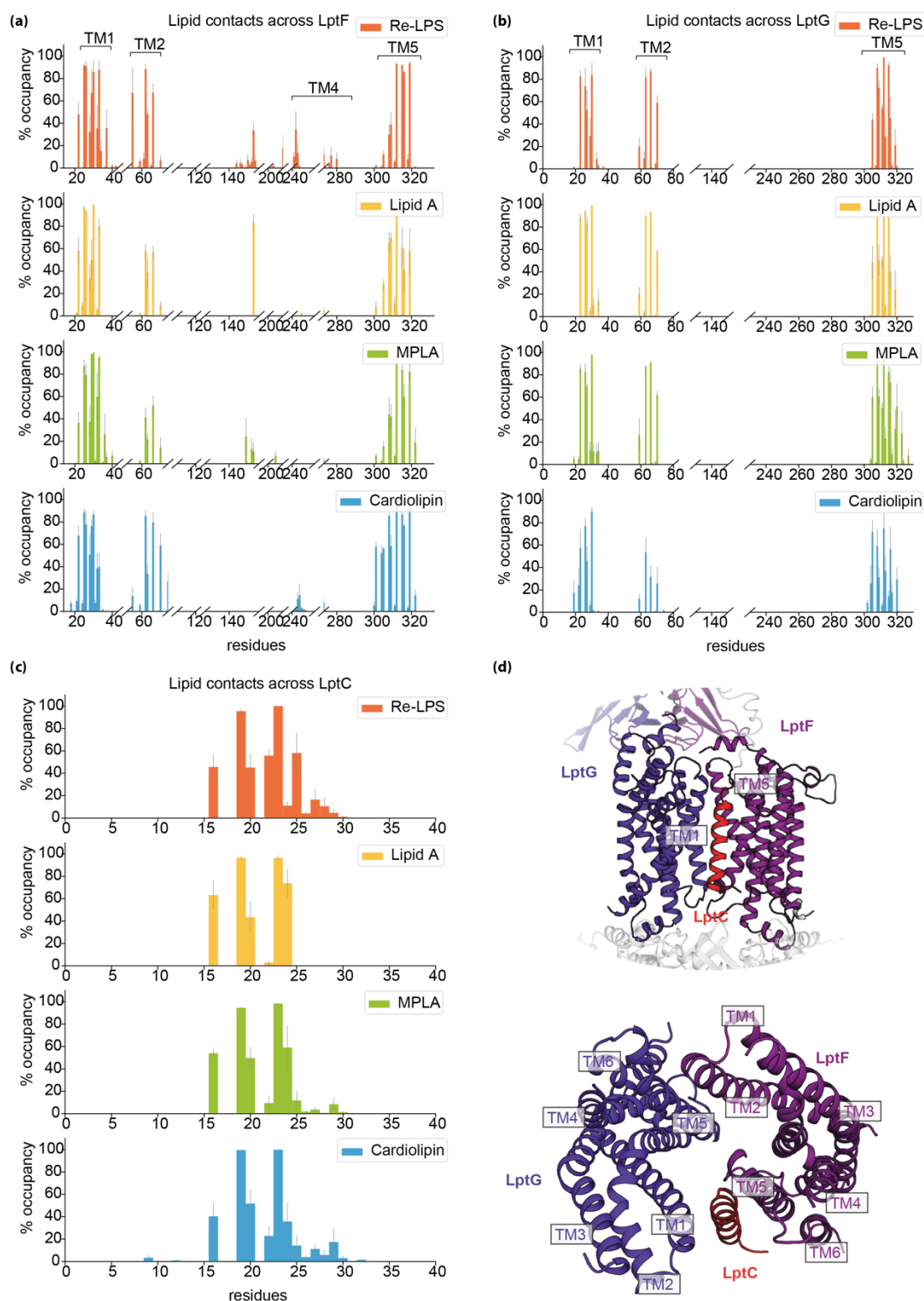


Figure 3.14 Lipid contacts with the hydrophobic cavity of the upright-LptC conformation. Average % occupancy of residue contacts between the LptC-upright conformation of LptB₂FGC with Re-LPS, lipid A, MPLA and cardiolipin within the hydrophobic cavity. Average % occupancy contacts with lipids: Re-LPS, lipid A, MPLA and cardiolipin made between (a) LptF (b) LptG and (c) LptC. Error bars represent standard deviation over the three replicates. (d) TM helices in contact with lipids indicated on contact maps and labelled onto structure.

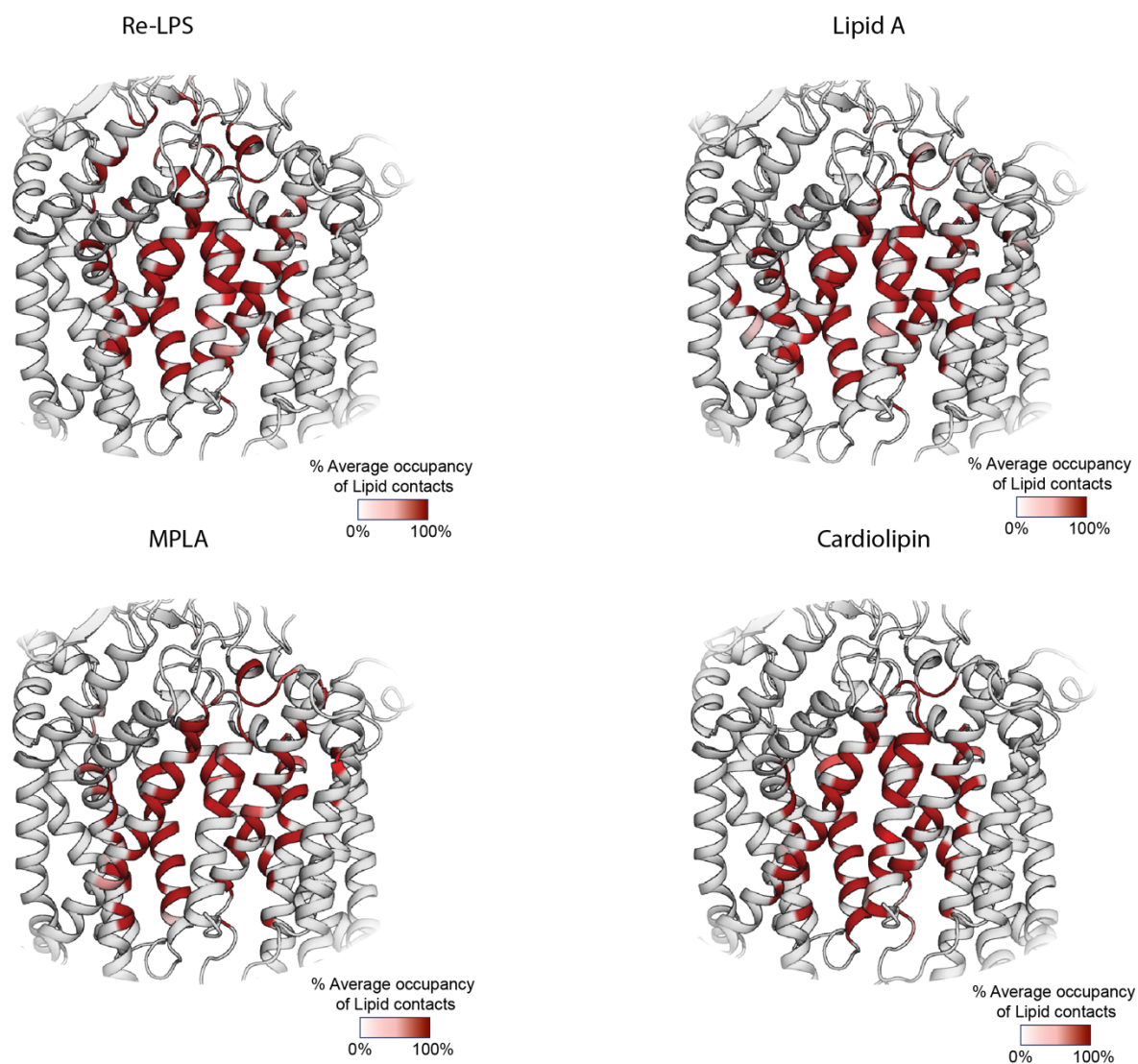


Figure 3.15 Contact map of lipid interaction with the hydrophobic cavity of the skewed-LptC conformation. Lipid contacts mapped onto the LptC-skewed conformation of LptB₂FGC for Top left: Re-LPS, Top right: lipid A, bottom left: MPLA and bottom right: cardiolipin where greater lipid contacts are highlighted in darker shades of red.

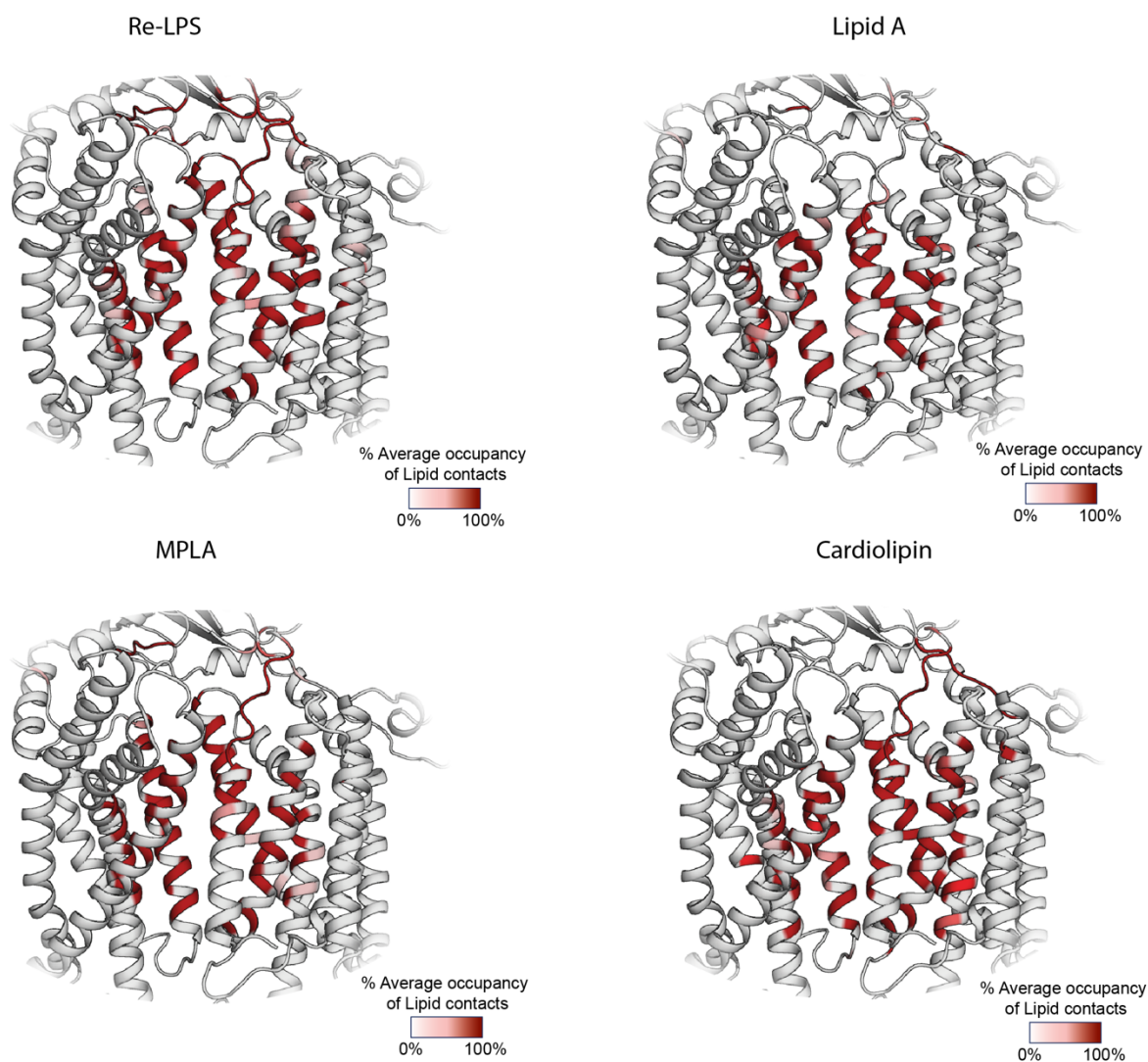


Figure 3.16 Contact map of lipid interaction with the hydrophobic cavity of the upright-LptC conformation. Lipid contacts mapped onto the LptC-upright conformation of LptB₂FGC for Top left: Re-LPS, Top right: lipid A, bottom left: MPLA and bottom right: cardiolipin where greater lipid contacts are highlighted in darker shades of red.

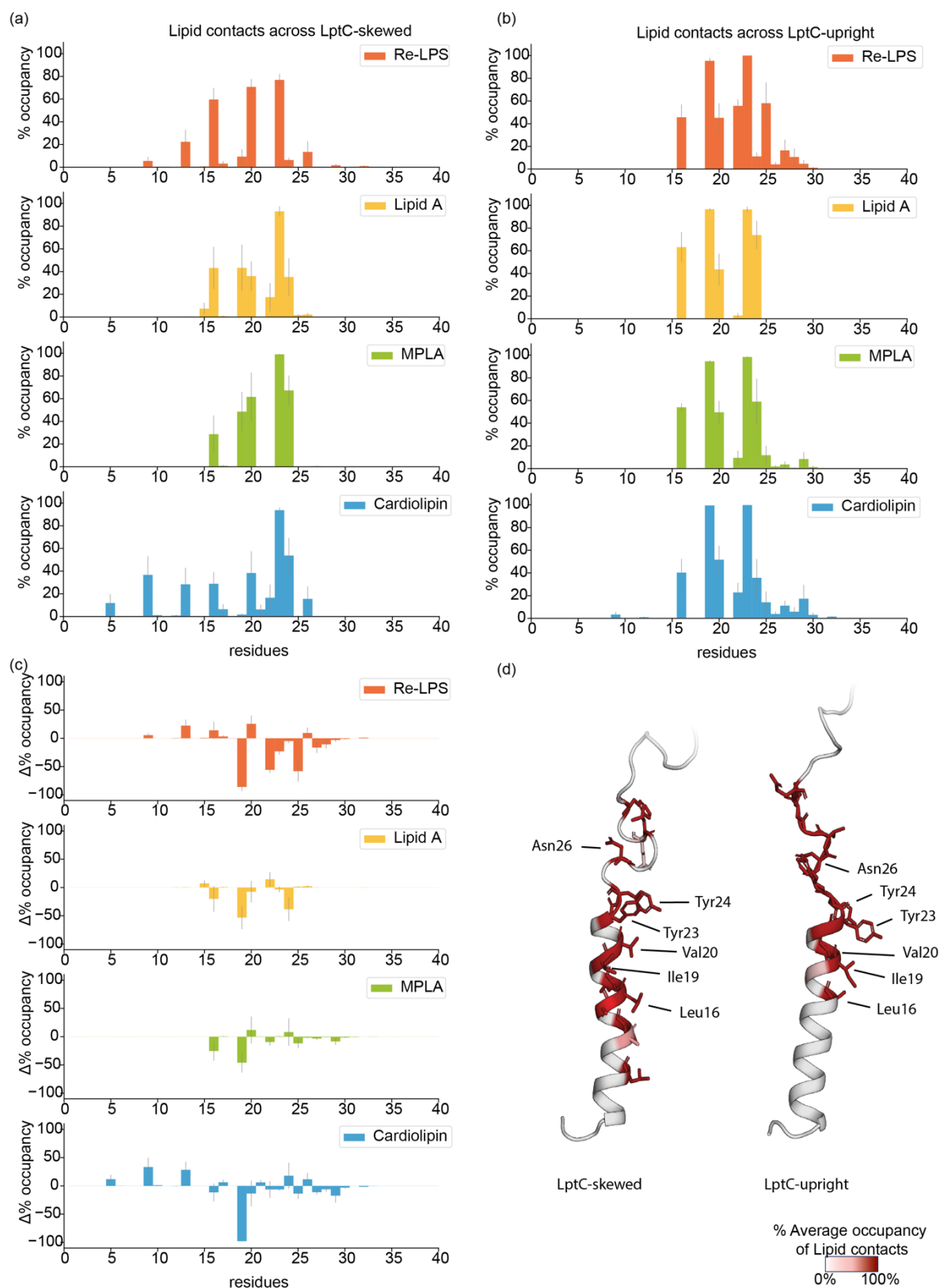


Figure 3.17 TMC interactions with lipids in the cavity of the LptB₂FGC transporter. % Occupancy contacts of the LptC transmembrane helix interactions with Re-LPS, lipid A, MPLA and cardiolipin within the cavity of the **(a)** LptC-skewed and **(b)** LptC-upright conformation of LptB₂FGC. Error bars represent standard

deviation over the three replicates. **(c)** Difference plot of % lipid contacts moving from LptC-skewed and LptC-upright conformations. **(d)** TM_c of the LptC-skewed and LptC-upright conformations where residues 16, 19-20, 23-24, 26 common to both conformations and are highlighted.

Lipid contacts with TM_c helix are very similar for Re-LPS and cardiolipin for both the skewed and the upright conformation (**Figure 3.17**). However, the distribution of the lipid contacts shifts from residues 9, 13, 16-17 to 25-34 when moving from the LptC-skewed to the LptC-upright conformation as Re-LPS is pushed further from the alpha helix towards the loops of LptC. In either case, residues 16, 19-20, 23-24, 26 are common to both conformations and are highlighted on **Figure 3.17** as important binding contacts for Re-LPS recognition. For the LptC-skewed conformation we see agreement with previously identified TM_c hydrophobic contacts Leu17 and Met24 of S/LptB₂FG-TM_c with LPS²⁹³.

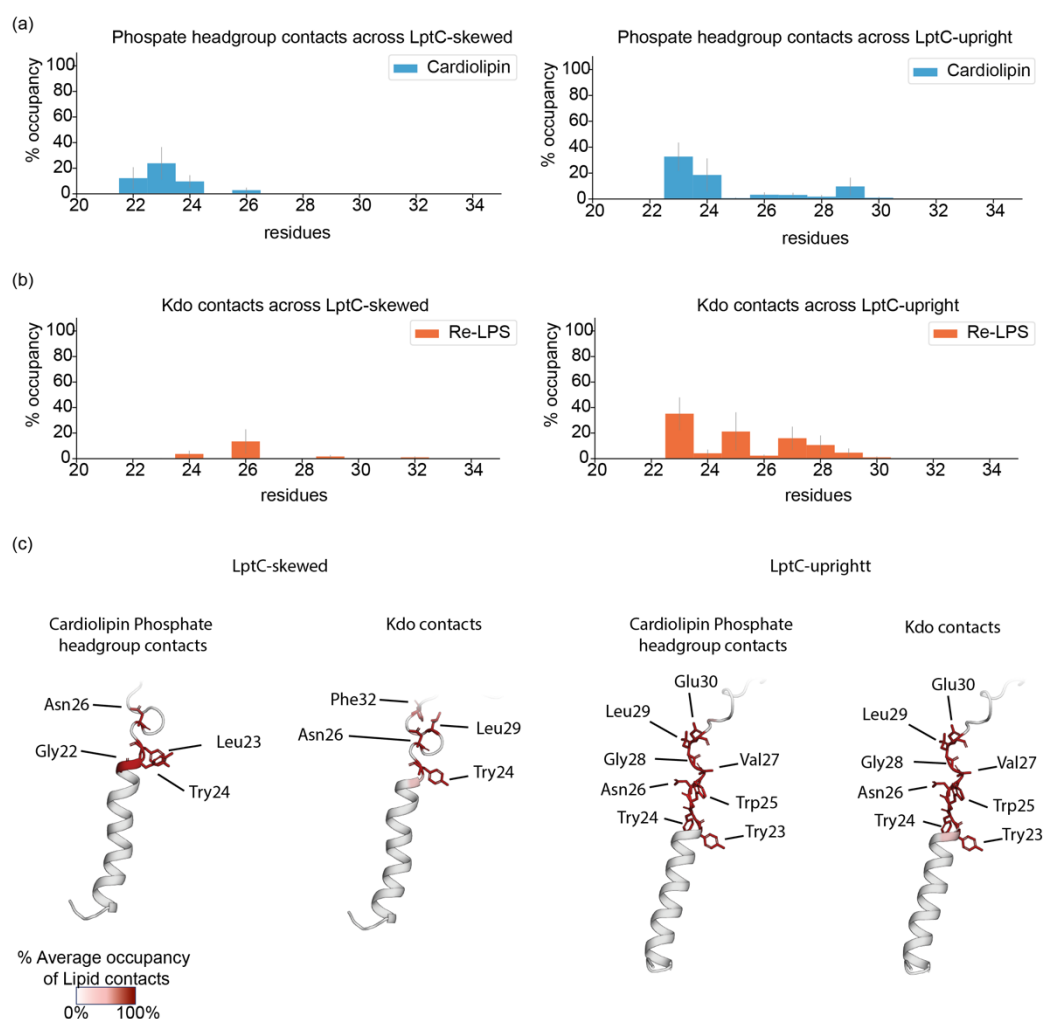


Figure 3.18 Headgroup interactions of cardiolipin versus Re-LPS with TM_c. (a) LptC TM helix % contacts of both the LptC-skewed and LptC-upright conformation with phosphate head group of cardiolipin compared to (b) the Kdo sugars of Re-LPS, where error bars represent standard deviation over the three replicates. (c) Rendered images of residues which contact the TM_c helix for both cardiolipin phosphate headgroup and the Kdo sugars of Re-LPS.

Across the TM_c, I also observe a very similar binding pattern of Re-LPS and cardiolipin contacting the same residues for both the LptC-skewed and LptC-upright conformation. Potentially then, as observed by the experimental collaborators (FF, JRB, CVR, CS and EWY), cardiolipin competes with LPS for the same recognition sites of the protein.

LptC has significant contacts with the phosphate headgroups of cardiolipin in both LptC-skewed and LptC-upright conformations, though no phosphate group contacts were observed for lipid A, MPLA or Re-LPS (**Figure 3.18**). However, I observed up to 30% occupancy contact with Re-LPS Kdo saccharide units with TM_c in both conformations: residues 24 and 26 for the LptC-skewed conformation and residues 23-29 for the LptC-upright conformation. Here these residues align with those that contact the cardiolipin phosphate headgroups, suggesting therefore, that the core saccharide units of LPS contribute to substrate recognition.

3.2.2.2. *Characterising lipid interactions with LptG*

Simply, by calculating % average occupancy difference plots between the LptC-skewed and the LptC-upright conformation I can observe where within the lipids reside most within the cavity. For LptG, the difference plots reveal that Re-LPS, lipid A and cardiolipin appears to have a greater concentration of contacts on LptG in the LptC-skewed conformation compared to the LptC-upright conformation (**Figure 3.19**).

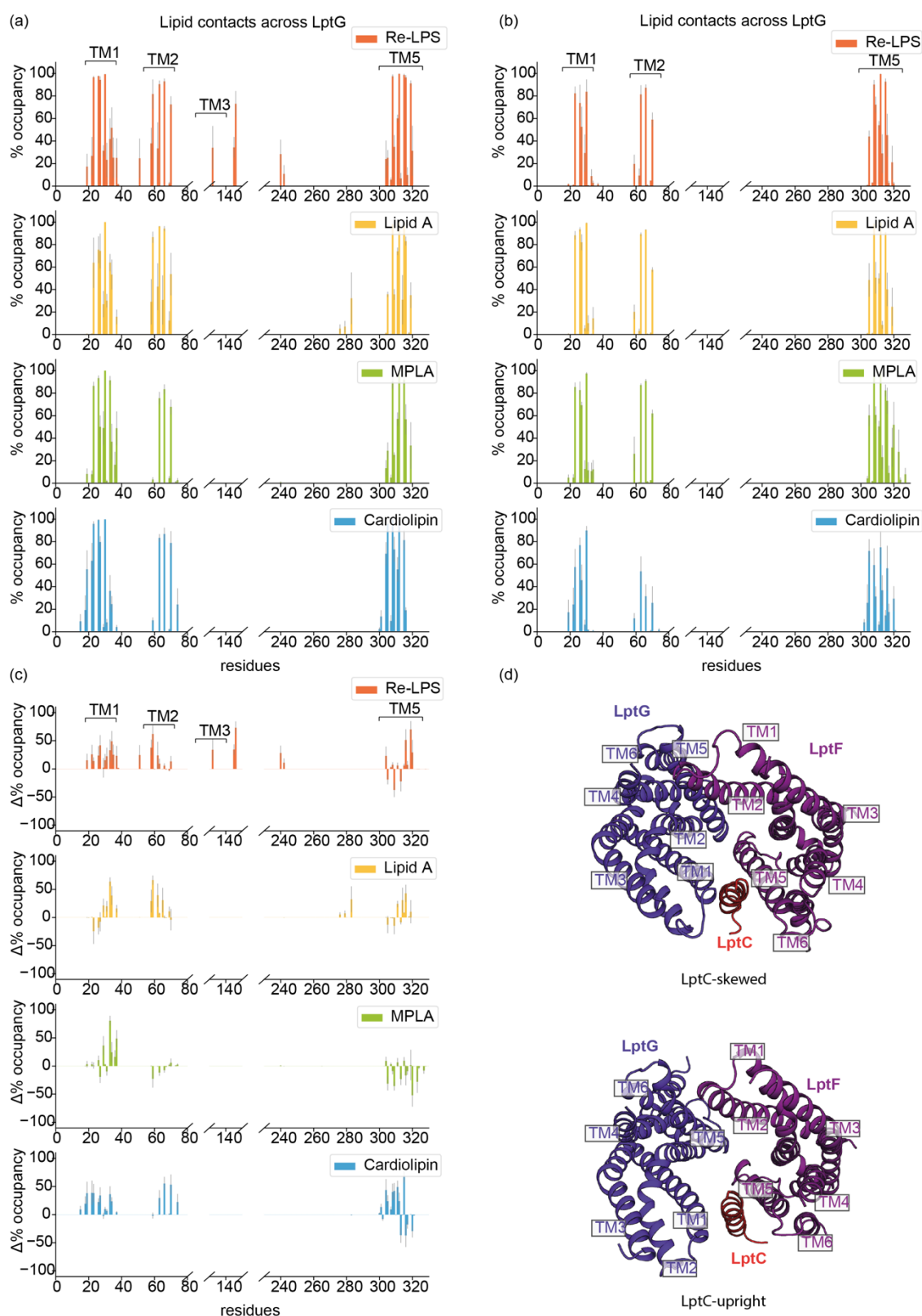


Figure 3.19 LptG interactions with lipids in the cavity of the LptB₂FGC transporter. % Occupancy contacts of LptG interactions with Re-LPS, lipid A, MPLA and cardiolipin within the cavity of the (a) LptC-skewed and (b) LptC-upright conformation of LptB₂FGC. Error bars represent standard deviation over the three replicates. (c) Difference plot of % lipid contacts moving from LptC-skewed and LptC-upright conformations, where +ve % contacts represent more interaction with the

LptC-skewed conformation and -ve % contacts represent more interaction with the LptC-upright conformation. **(d)** TM helices in contact with lipids indicated on contact maps and labelled onto structure.

Here, the % average occupancy difference contacts made between Re-LPS, lipid A and cardiolipin LptG seem relatively comparable between the conformational states of the transporter. All three lipids here (Re-LPS, lipid A and cardiolipin) have more explicit interactions with TM1, TM2, and TM5 of LptG in the LptC-skewed conformation. Again, it is important to note the similarity of the binding contacts of cardiolipin with Re-LPS which helps to corroborate experimental observation of competitive kinetics between the lipid species. MPLA, however, does not follow this trend and only forms more contacts with TM1 of LptG in the LptC-skewed conformation and forms greater contacts with TM5 in the LptC-upright conformation. This implies that the phosphate groups of the substrate form key interactions within the cavity which facilitate substrate recognition and mobility between the different conformations.

3.2.2.3. *Characterising lipid interactions with LptF*

I observe a similar (but less clear and one-sided) trend for the LptF occupancy difference plots, where there are slightly more contacts with TM1, TM2 and TM5 in the LptC-skewed conformation for each lipid respectively (apart from lipid A which appears to make only more contacts with just TM1 and TM2 (**Figure 3.20**)). Overall, greater interaction of the lipids occur within the LptC-skewed hydrophobic cavity where most contacts are seen with LptG which are most likely driven by phosphate headgroup contacts. The greater amount of lipid contacts observed on LptF and LptG with the LptC-skewed conformation implies a tighter hydrophobic cavity as TM1, 2 and 5 are closer to the centre than they are in the LptC-upright conformation.

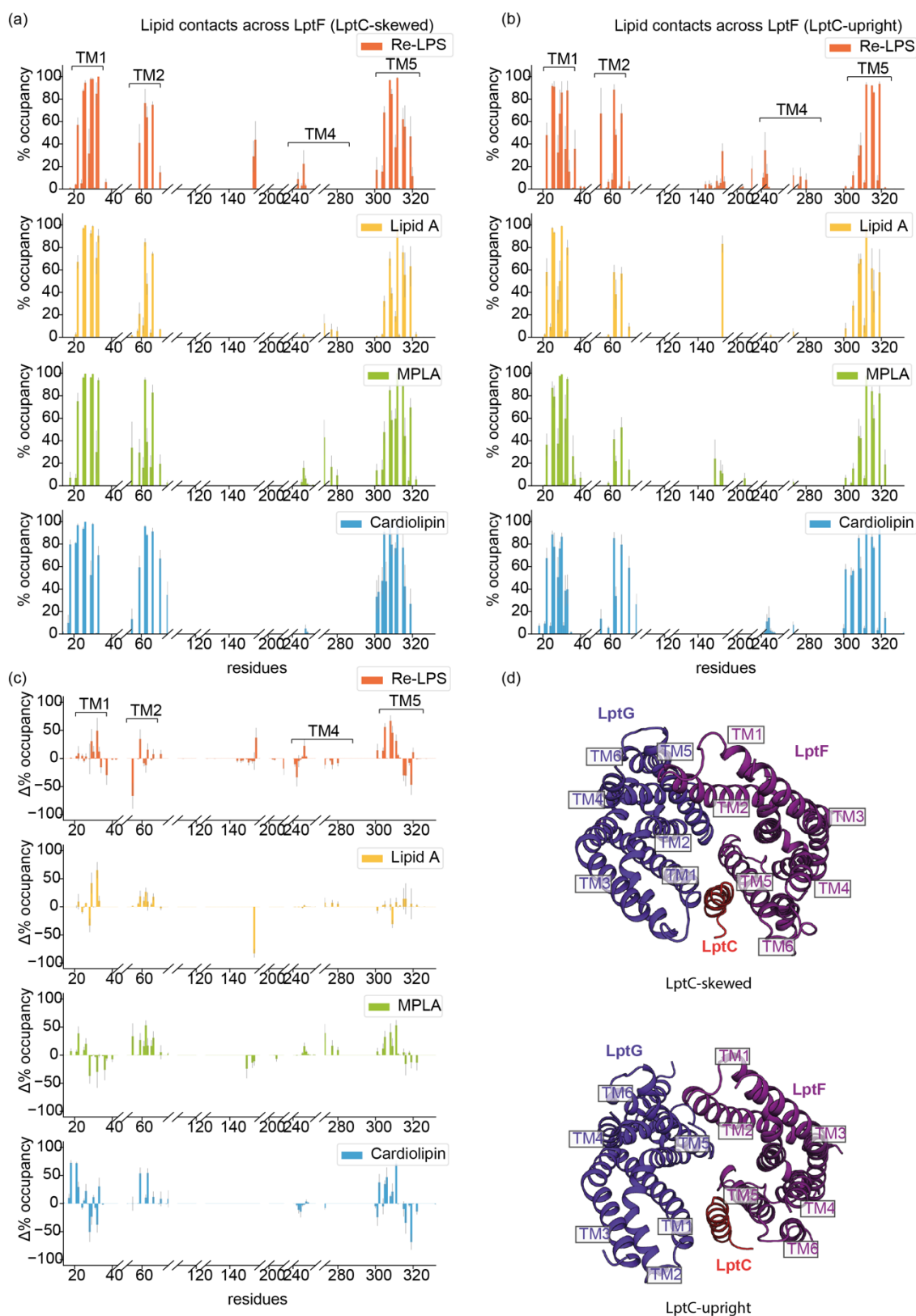


Figure 3.20 LptF interactions with lipids in the cavity of the LptB₂FGC transporter. % Occupancy contacts of LptF interactions with Re-LPS, lipid A, MPLA and cardiolipin within the cavity of the (a) LptC-skewed and (b) LptC-upright conformation of LptB₂FGC. Error bars represent standard deviation over the three replicates. (c) Difference plot of % lipid contacts moving from LptC-skewed and LptC-

upright conformations, where +ve % contacts represent more interaction with the LptC-skewed conformation and -ve % contacts represent more interaction with the LptC-upright conformation. **(d)** TM helices in contact with lipids.

3.2.2.4. Characterising head group interactions with *LptB₂FGC*

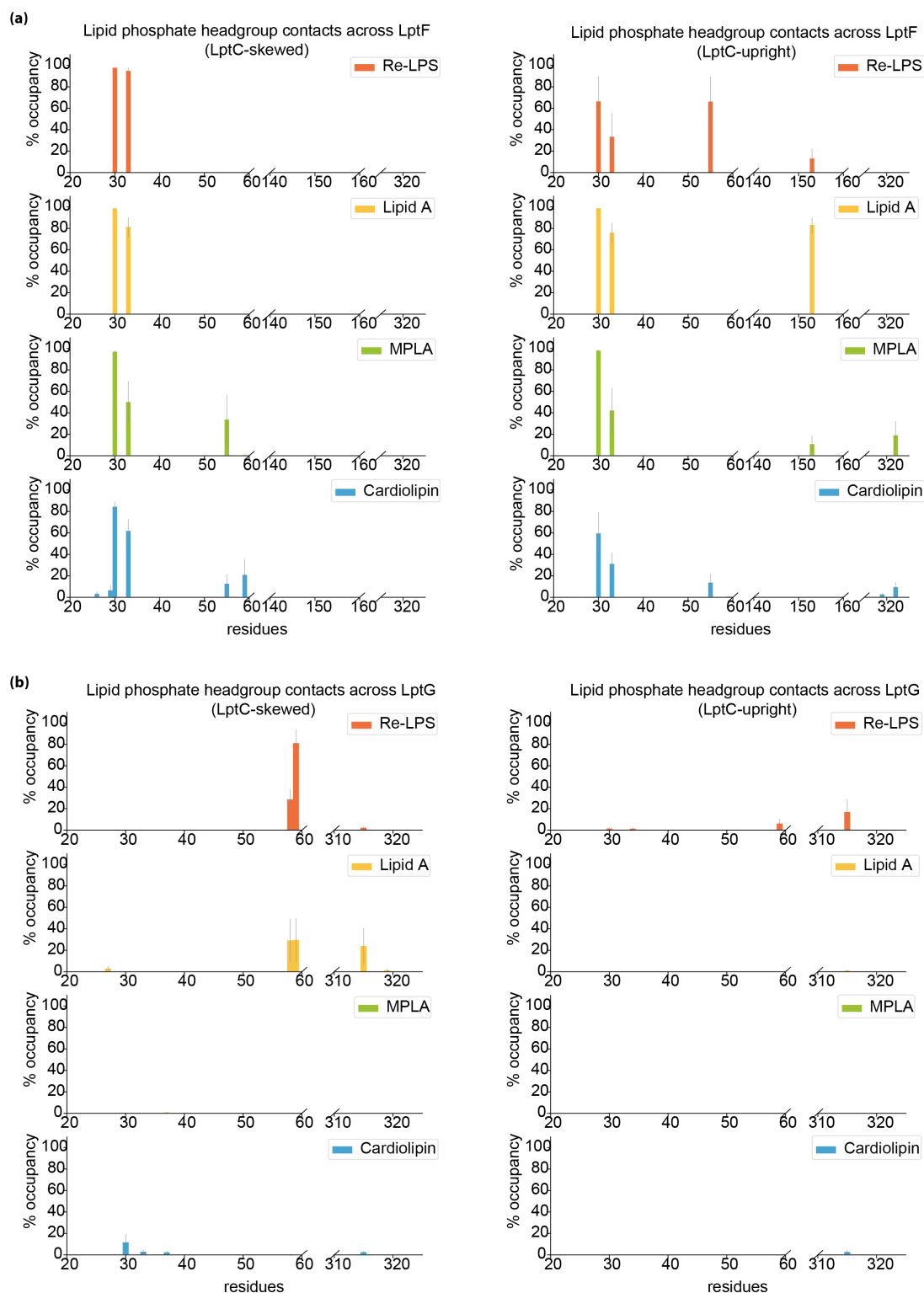


Figure 3.21 Phosphate headgroup contacts with the *LptB₂FGC* hydrophobic cavity. (a) % average occupancy phosphate headgroup contacts across LptF with Re-

LPS, lipid A, MPLA and cardiolipin within cavity for the LptC-skewed (left) and LptC-upright (right) conformations of LptB₂FGC. Error bars represent standard deviation over the three replicates. **(b)** % average occupancy of phosphate headgroup contacts across LptG with Re-LPS, lipid A, MPLA and cardiolipin within cavity for the LptC-skewed (left) and LptC-upright conformations of LptB₂FGC (right).

As noted above, when comparing the lipid contact profiles across the hydrophobic cavity for Re-LPS, lipid A, cardiolipin and MPLA, it was clear that MPLA (out of all the lipids tested) behaved differently because of its singular phosphate head group. Therefore, I sought to quantify the phosphate group contacts between the conformations as it seemed to be a key interaction underpinning substrate recognition and movement within the cavity.

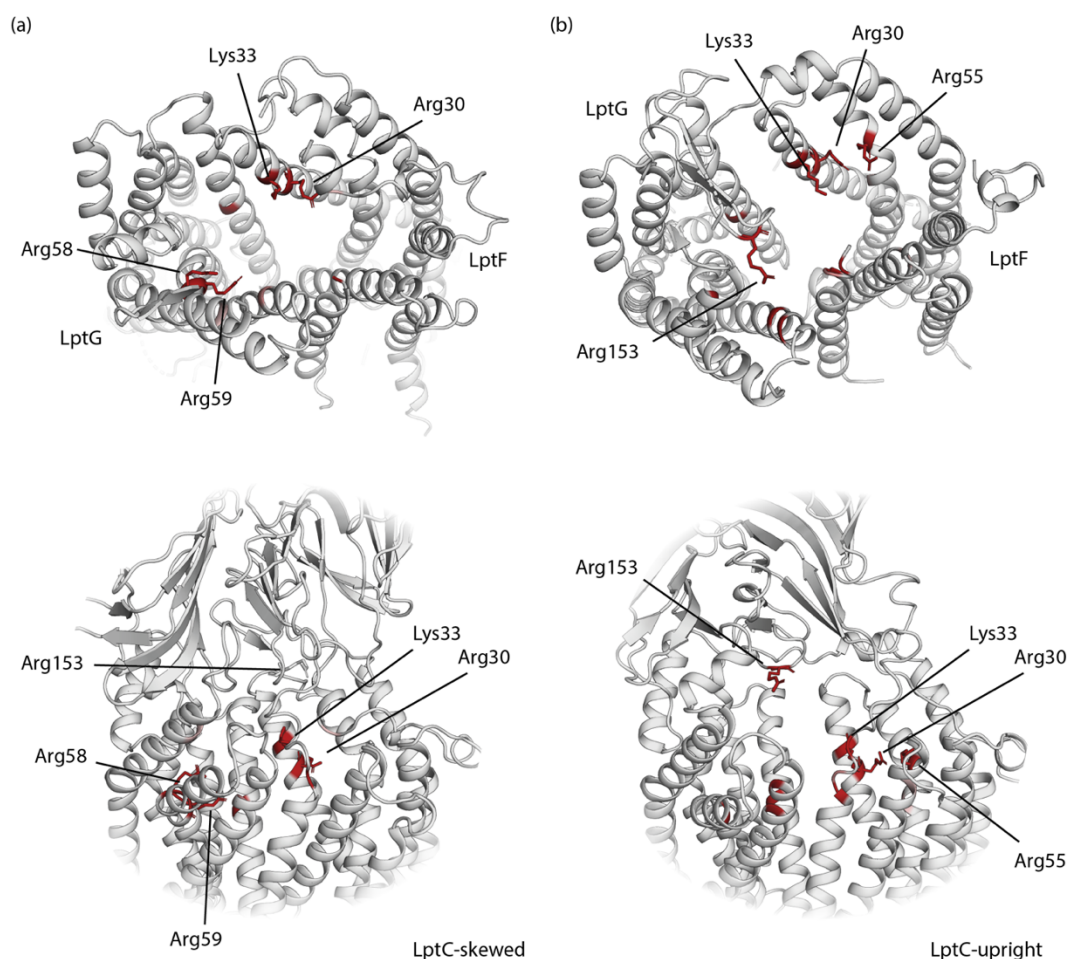


Figure 3.22 Key binding contacts of the LptB₂FGC hydrophobic cavity with the phosphate headgroup of Re-LPS. (a) Residues which contact phosphate groups of Re-LPS within cavity are depicted as sticks in both the LptC-skewed and **(b)** LptC-upright conformations of LptB₂FGC.

The phosphate headgroups of Re-LPS, lipid A, MPLA and cardiolipin consistently bind to LptF in the same place for both conformations for residue Arg33 and Lys33. Again, implicating cardiolipin's competitive tendencies when comparing its binding profile against Re-LPS (**Figure 3.21**). There are more LptG phosphate group contacts with the LptC-skewed structure than the LptC-upright. These contacts appear to have been transferred from LptG (Arg58 and Arg59) to LptF (Arg55 and Arg153) on moving from the LptC-skewed to the LptC-upright structure. Curiously, Arg153 belongs to one of the loops of the LptF β -taco domain and helps the transition of the LPS substrate from LptF to LptG for transport (**Figure 3.22**). Therefore, phosphate headgroups interactions with the hydrophobic cavity are key to substrate recognition and movement between conformations of LptB₂FGC.

The trend observed for the phosphate group contacts within the cavity is further exemplified when I consider only the Kdo sugar contact of Re-LPS with the transporter. I quantified the sites of where Re-LPS Kdo sugars interact with (**Figure 3.23**) which are markedly different from the LptC-skewed compared to the LptC-upright conformation. For the LptC-skewed structure Kdo contacts appear to be concentrated on LptG whereas they are primarily concentrated on LptF for the LptC-upright structure.

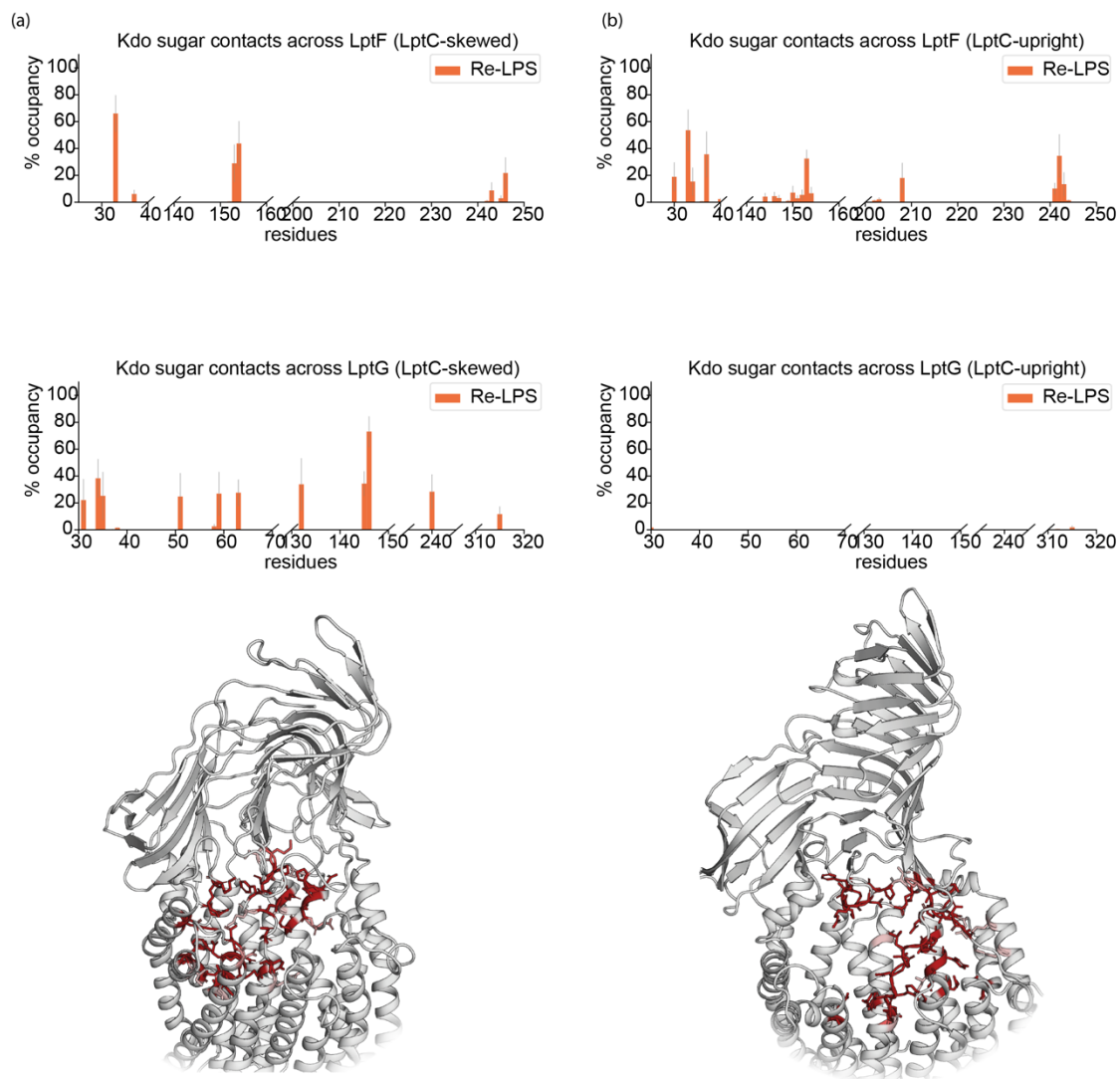


Figure 3.23 Key binding contacts of the LptB₂FGC hydrophobic cavity with the Kdo₂ glycan units. (a) % average occupancy Kdo₂ sugar contacts of Re-LPS within cavity for the LptC-skewed. Error bars represent standard deviation over the three replicates. (b) LptC-upright conformations of LptB₂FGC. Kdo₂ sugar contacts are depicted as sticks for both the LptC-skewed and LptC-upright conformations of LptB₂FGC.

Potentially then, this shift in conformation from the LptC-skewed state to the more dynamic LptC-upright structure triggers the inner core sugar and phosphate group to shift from LptG to LptF to prepare the substrate for transportation up the periplasmic chute.

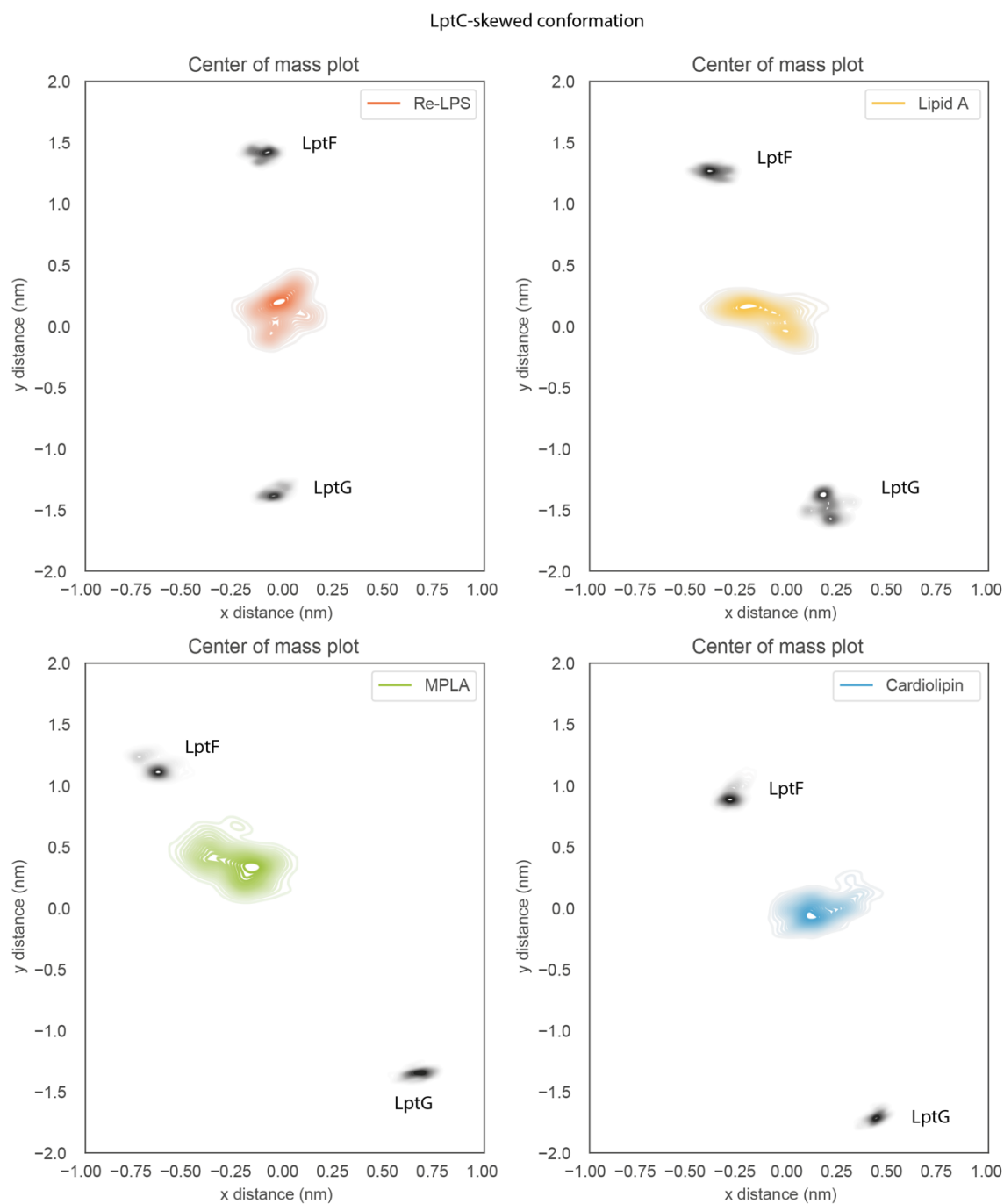


Figure 3.24 The COM of lipids in the cavity for the LptC-skewed structure. 2D kernel density estimates plotting the centre of mass (COM) position of lipid species within the hydrophobic cavity of the LptC-skewed structure relative to the COM positions of LptF and LptG. Top left: Re-LPS, top right: lipid A, bottom left: MPLA, bottom right: cardiolipin.

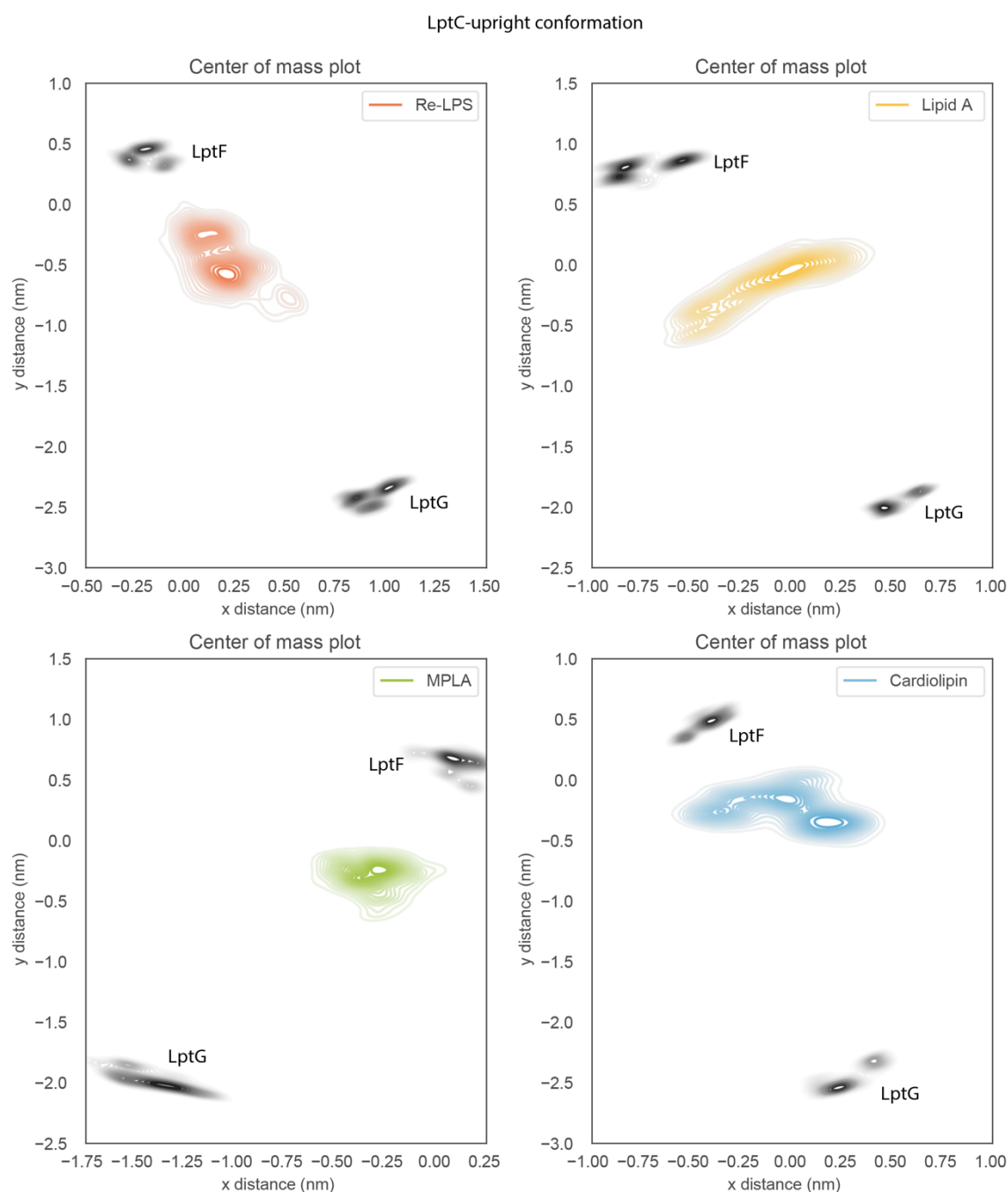


Figure 3.25 The COM of lipids in the cavity for the LptC-upright structure. 2D kernel density estimates plotting the centre of mass (COM) position of lipid species within the hydrophobic cavity of the LptC-upright structure relative to the COM positions of LptF and LptG. Top left: Re-LPS, top right: lipid A, bottom left: MPLA, bottom right: cardiolipin.

To verify such a shift in substrate position, I quantified the centre of mass (COM) of all the lipid species and plotted this relative to the COM of LptF and LptG units comprising the hydrophobic cavity. For the LptC-skewed conformation (**Figure 3.24**), the lipids

species Re-LPS, lipid A and cardiolipin are all centrally disposed, virtually equidistant between the COM of the LptF and LptG TM subunits. This accords nicely with observations made with the lipid contact analysis, where I note similar behaviour between cardiolipin and the LPS derivatives Re-LPS and lipid A. MPLA acts differently compared to the other lipids, which can again be accounted for by the singular phosphate headgroup which causes the lipid's COM to move (~0.5 nm) closer to the LptF subunit.

Strikingly, for the LptC-upright conformation (**Figure 3.25**), the lipid species which were equidistant between subunits in the LptC-skewed conformation, shift towards LptF. The COM of Re-LPS, lipid A and cardiolipin each move ~0.5 nm towards the LptF subunit. The COM of MPLA, however, appears to be consistent between the two conformations and does not shift its position unlike the other lipid species. Considering the lipid contacts in conjunction with the COM analysis of the lipids within the hydrophobic cavity the substrate shifts its position towards LptF which is driven by phosphate headgroup and sugar contacts. Cardiolipin shares similar contacts with the LPS derivatives and therefore can be considered a competitive lipid species to the transporter's substrate. It also seems reasonable to conclude that the LptC-skewed state of LptB₂FGC acts a resting state of the transporter, where the lipid substrate is housed centrally within the cavity. The conformational shift of the β -taco towards the upright position more easily facilitates contacts with the periplasmic bridge and triggers a shift of the LPS substrate driven by phosphate group and sugar contacts toward LptF. Here, LPS is closer to LptC and ready to be transported through the periplasm.

3.2.3 The conformational dynamics of the LptB₂FGC β -taco domains

In adding lipids to the central cavity, I was able to monitor their position and binding contacts which helped to describe differences between the LptC-skewed and LptC-upright conformations. From lipid contact studies, the TM_c plays a role in substrate recognition, but what is not known is how the substrate affects the position of the LptC β -taco domain which interfaces directly with the LptF β -taco domain. In order to understand this, I monitored the position of the vector running through the LptF β -taco as well as the LptG β -taco domain to comment upon conformational changes on binding to lipids within the hydrophobic cavity (**Figure 3.26**).

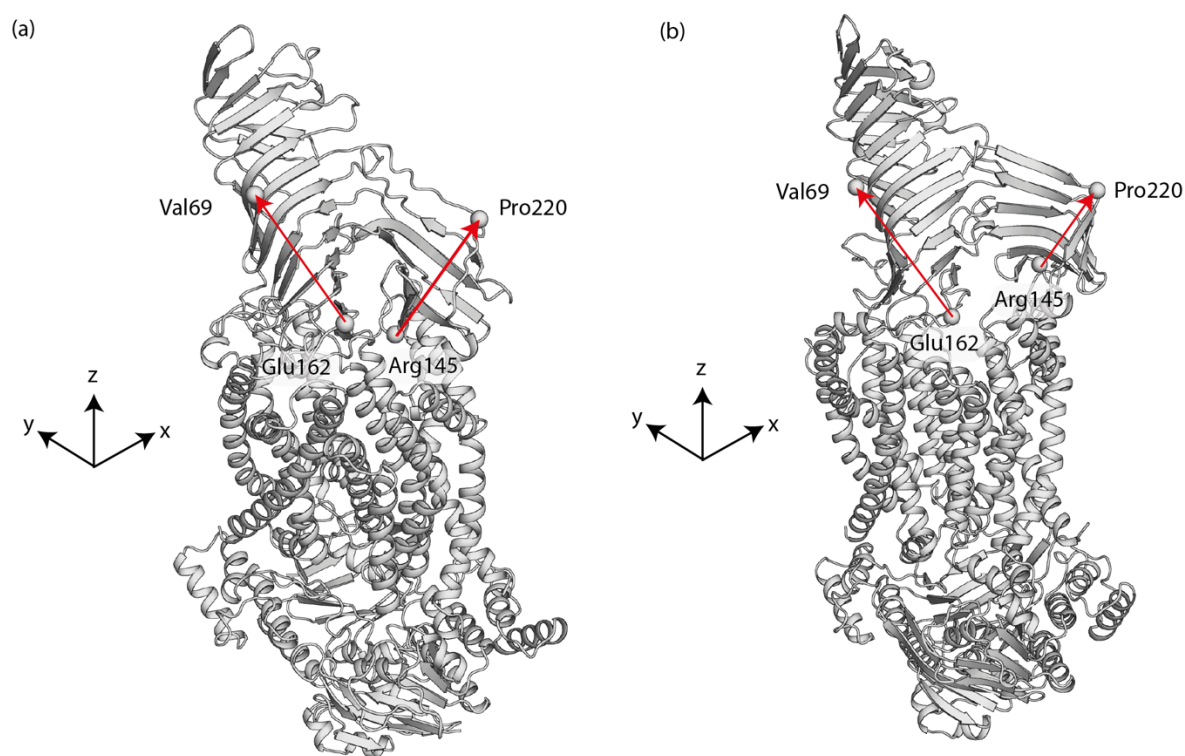


Figure 3.26 The LptF and LptG β -taco domain vector. Schematic illustrating the vector across LptF and LptG β -taco domain and the principal x, y, z axis of the LptC-skewed and LptC-upright conformation.

I monitored the x, y, z angle between the vector running through LptF and LptG β -taco and the principal x, y, z axis of the LptC-skewed conformation when bound to the lipid

species: Re-LPS, lipid A, MPLA and cardiolipin and compared this to the apo state of the protein (**Figure 3.27**).

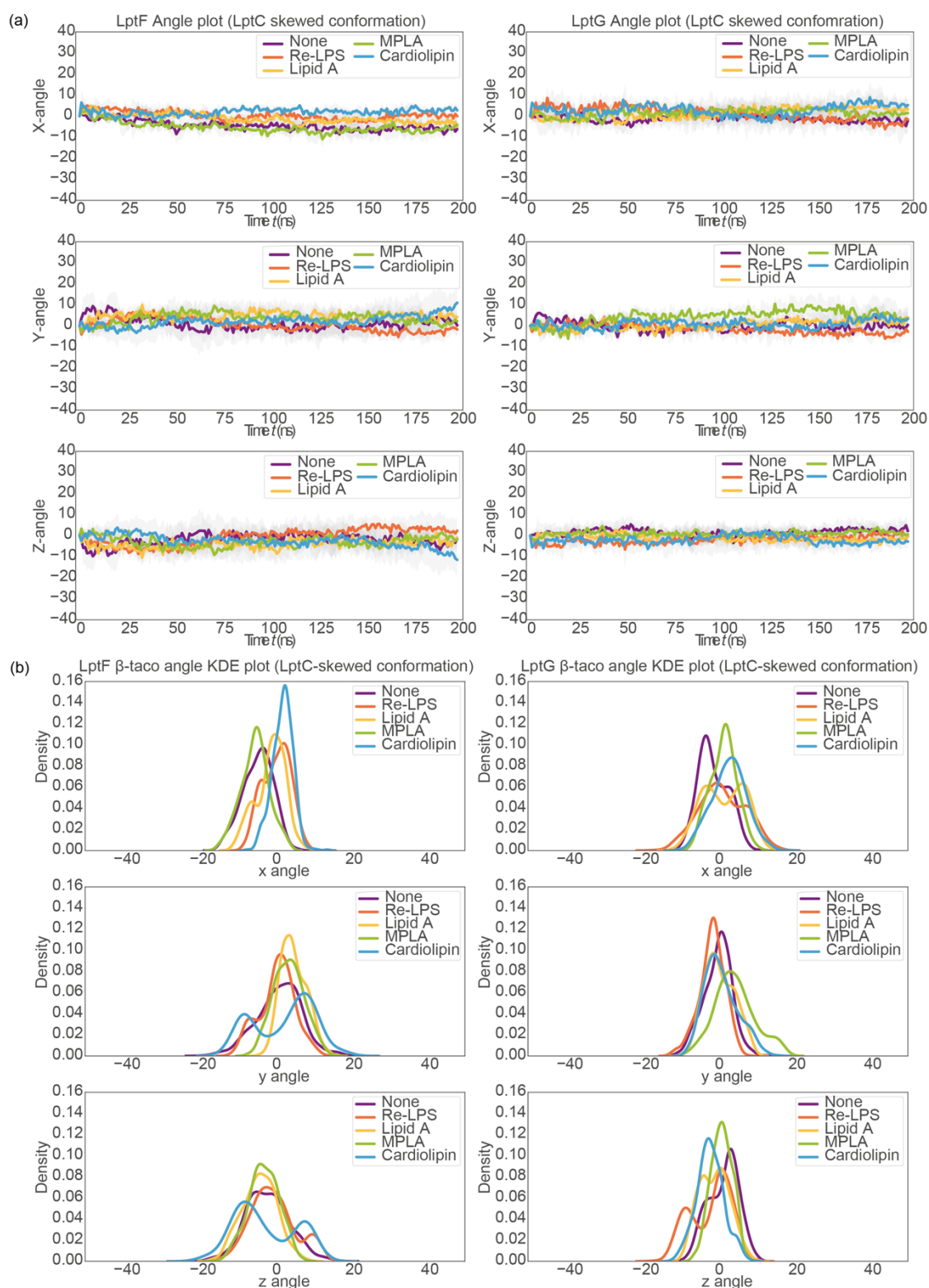


Figure 3.27 LptF and LptG β -taco domains angle plot in the LptC-skewed conformation. (a) Average x, y, z angle plot over time for the β -taco domains of LptF (left) and LptG (right) with no lipid present or Re-LPS, lipid A, MPLA and cardiolipin

present in the cavity. Shaded grey area represents the standard deviation over three replicates **(b)** The same x, y, z angle data for the β -taco domains of LptF (left) and LptG (right) plotted as a kernel density estimate.

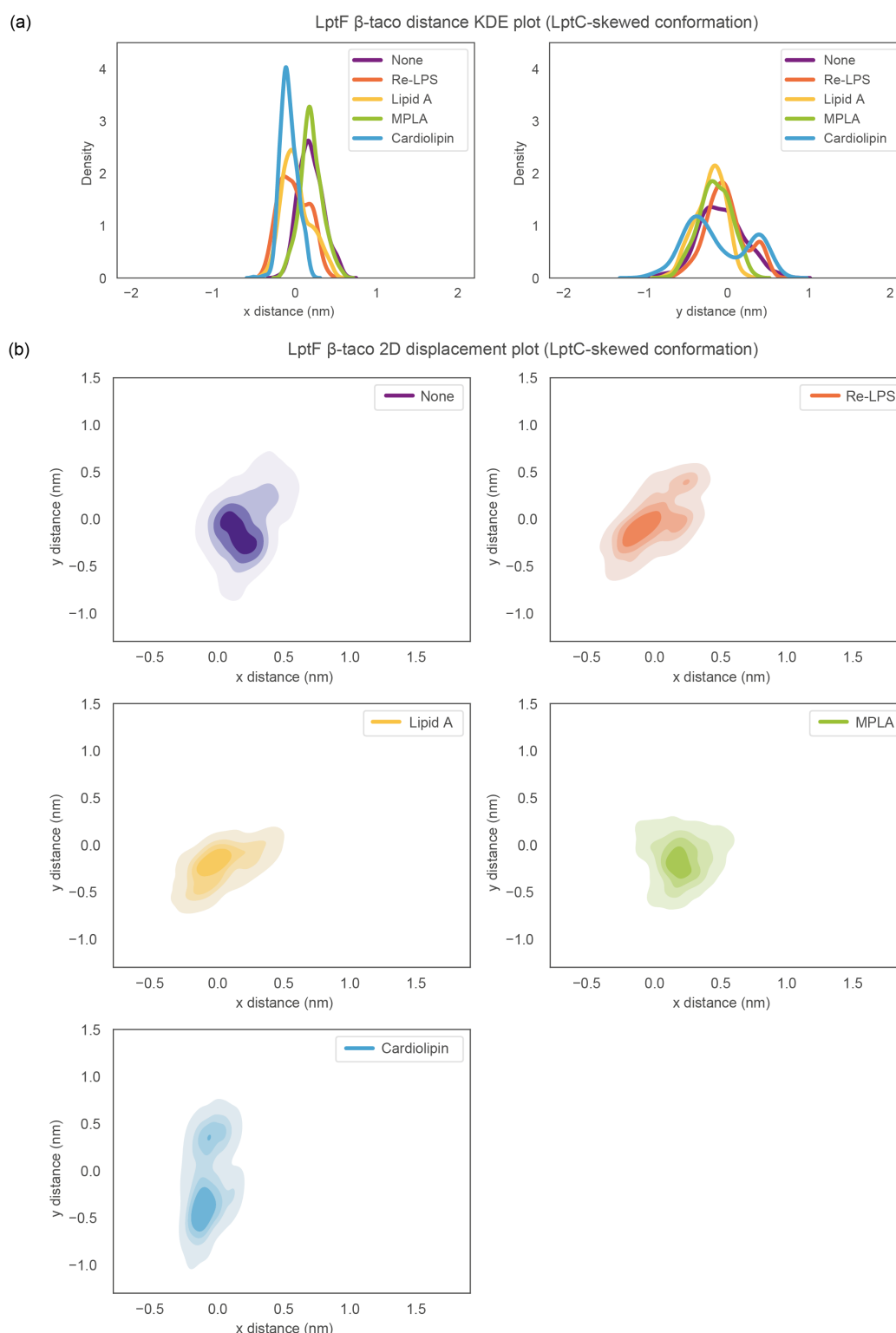


Figure 3.28 Displacement of LptF β -taco domain in the LptC-skewed conformation. (a) 1D x, y displacement of the LptF β -taco domain in the LptC-skewed conformation plotted as a kernel density estimate (KDE) with no lipid present or Re-LPS, lipid A, MPLA and cardiolipin present in the cavity. Where the positions $x = 0$ nm

and $y = 0$ nm are set as the initial position of the β -taco domain in each of the states respectively. **(b)** 2D KDE plot of x , y displacement across LptF β -taco with: Top left: no lipid present, Top right: Re-LPS, Middle left: lipid A, Middle right: MPLA and Bottom left: cardiolipin present in the cavity.

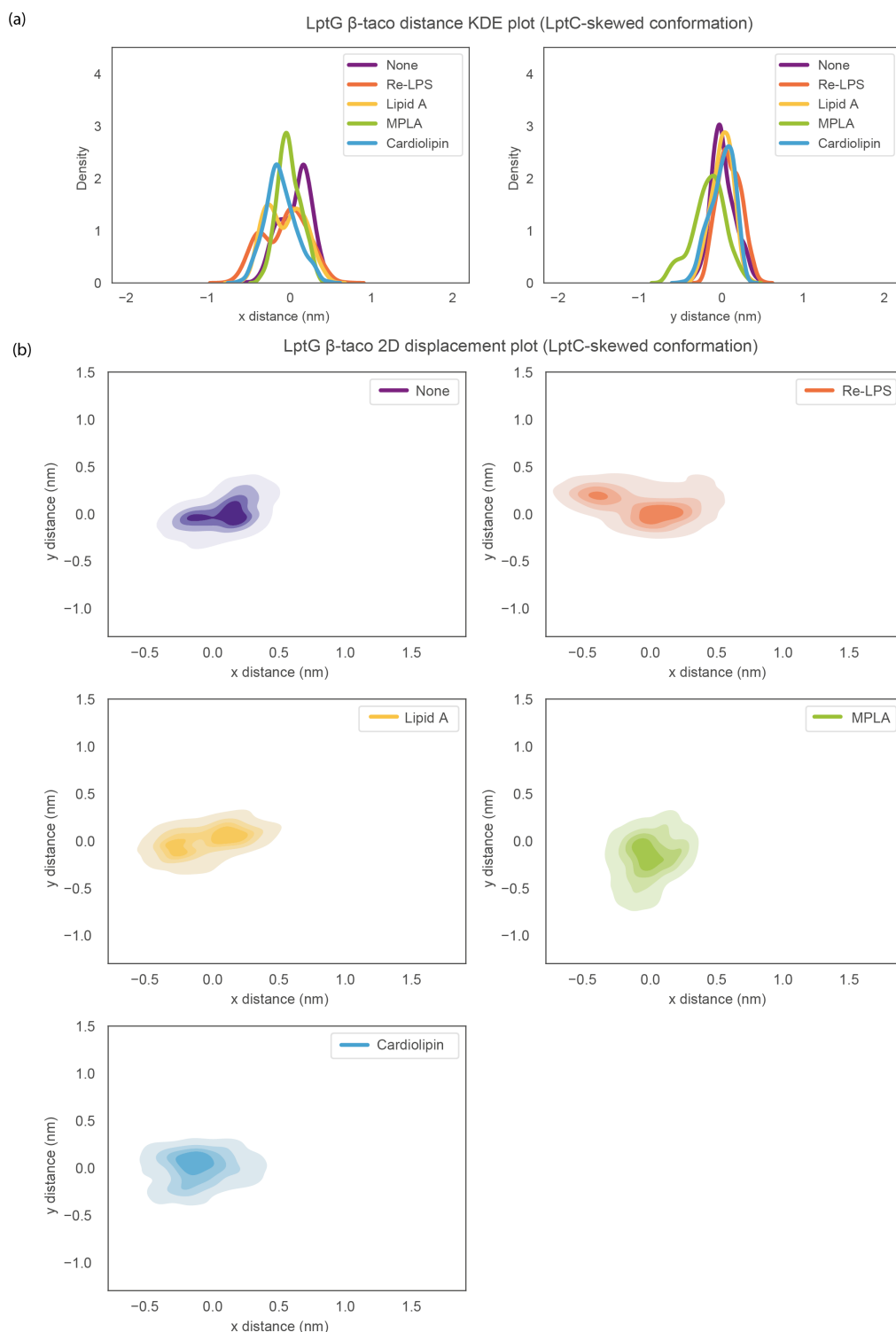


Figure 3.29 Displacement of LptG β -taco domain in the LptC-skewed conformation. (a) 1D x , y displacement of the LptG β -taco domain in the LptC-skewed conformation plotted as a kernel density estimate (KDE) with no lipid present or Re-LPS, lipid A, MPLA and cardiolipin present in the cavity. Where the position $x = 0$ nm

and $y = 0$ nm is set as the initial position of the β -taco domain in each of the states respectively. **(b)** 2D KDE plot of x , y displacement across LptG β -taco with: Top left: no lipid present, Top right: Re-LPS, Middle left: lipid A, Middle right: MPLA and Bottom left: cardiolipin present in the cavity.

I can more simply visualise the positional movement of the LptF and LptG β -taco domains by plotting the x and y displacement of the vector aligned to the β -taco's, which can be plotted as 1D and 2D KDE plots (**Figure 3.28** and **Figure 3.29**). Over the course of the simulation time, the x , y , z angle stays relatively stable for all the lipid bound states of the transporter compared to the apo state of the LptC-skewed conformation. Here, each x , y , z angle tends to deviate by around 5 degrees, regardless of whether a lipid was present or the type of lipid present in cavity.

As shown in **Figure 3.28** and **Figure 3.29**, the x and y displacement of the LptF and LptG β -taco soluble domains are minimal for the LptC-skewed conformation. These domains appear to be unaffected by the inclusion of particular lipids inside the transporter's cavity. This alludes to the possibility then that the LptC-skewed conformation pertains to a more 'rested' state of the transporter before the periplasmic bridge is assembled, where the β -taco domains remain conformationally inert. In contrast to the LptC-skewed conformation, the x , y , z angle of the LptF and LptG β -taco domain changes drastically for the LptC-upright conformation on binding to different lipids within the cavity **Figure 3.30**. Notably, with no lipids present within the cavity, the angle deviation is small and remains stable. However, during the course of the simulation, there are clear angle deviations observed when particular lipids reside within the cavity.

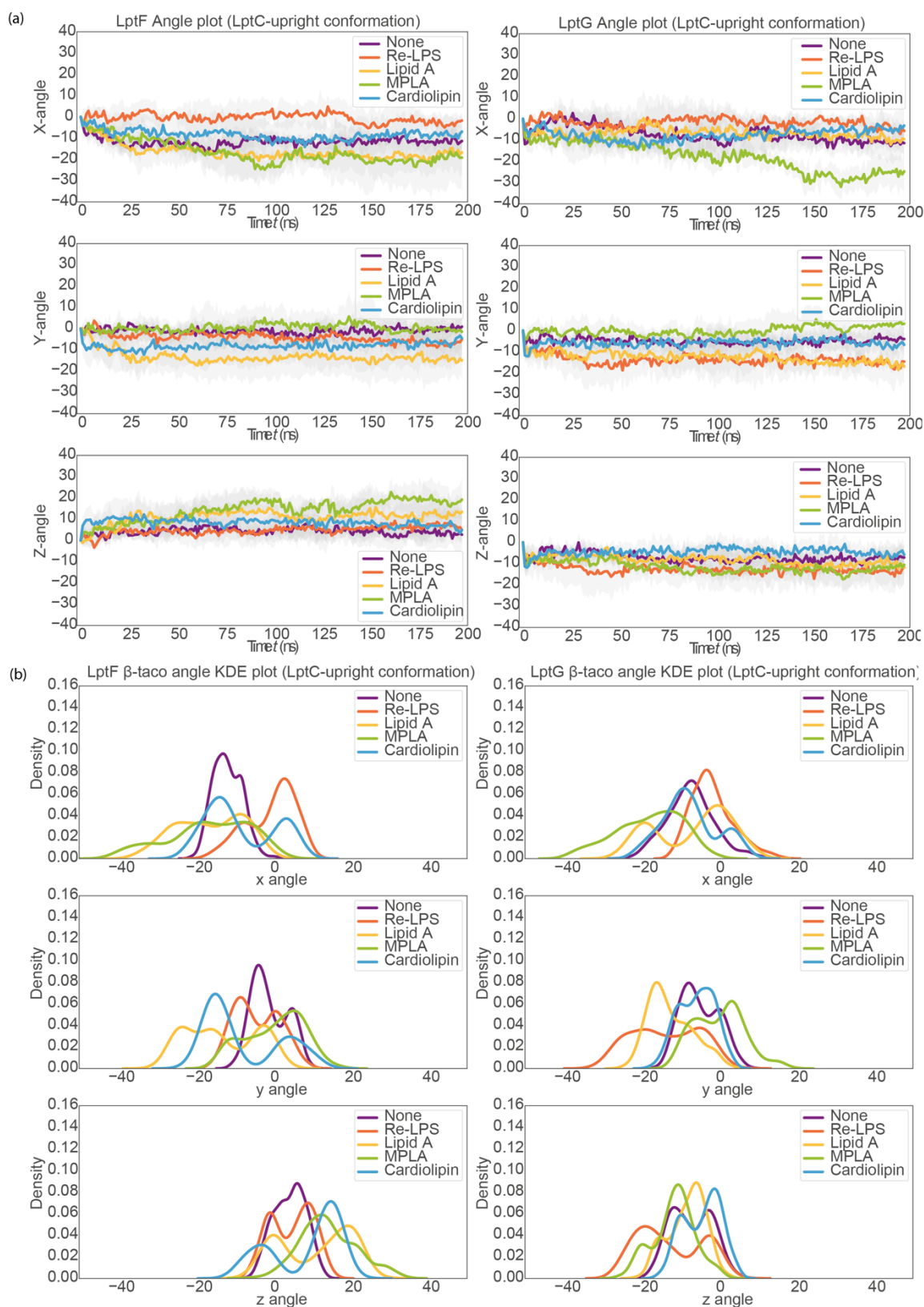


Figure 3.30 LptF and LptG β -taco domains angle plot in the LptC-upright conformation. (a) Average x, y, z angle plot over time for the β -taco domains of LptF (left) and LptG (right) with no lipid present or Re-LPS, lipid A, MPLA and cardiolipin present in the cavity. Shaded areas represent the standard deviation over three

replicates **(b)** The same x, y, z angle data for the β -taco domains of LptF and LptG plotted as a kernel density estimate.

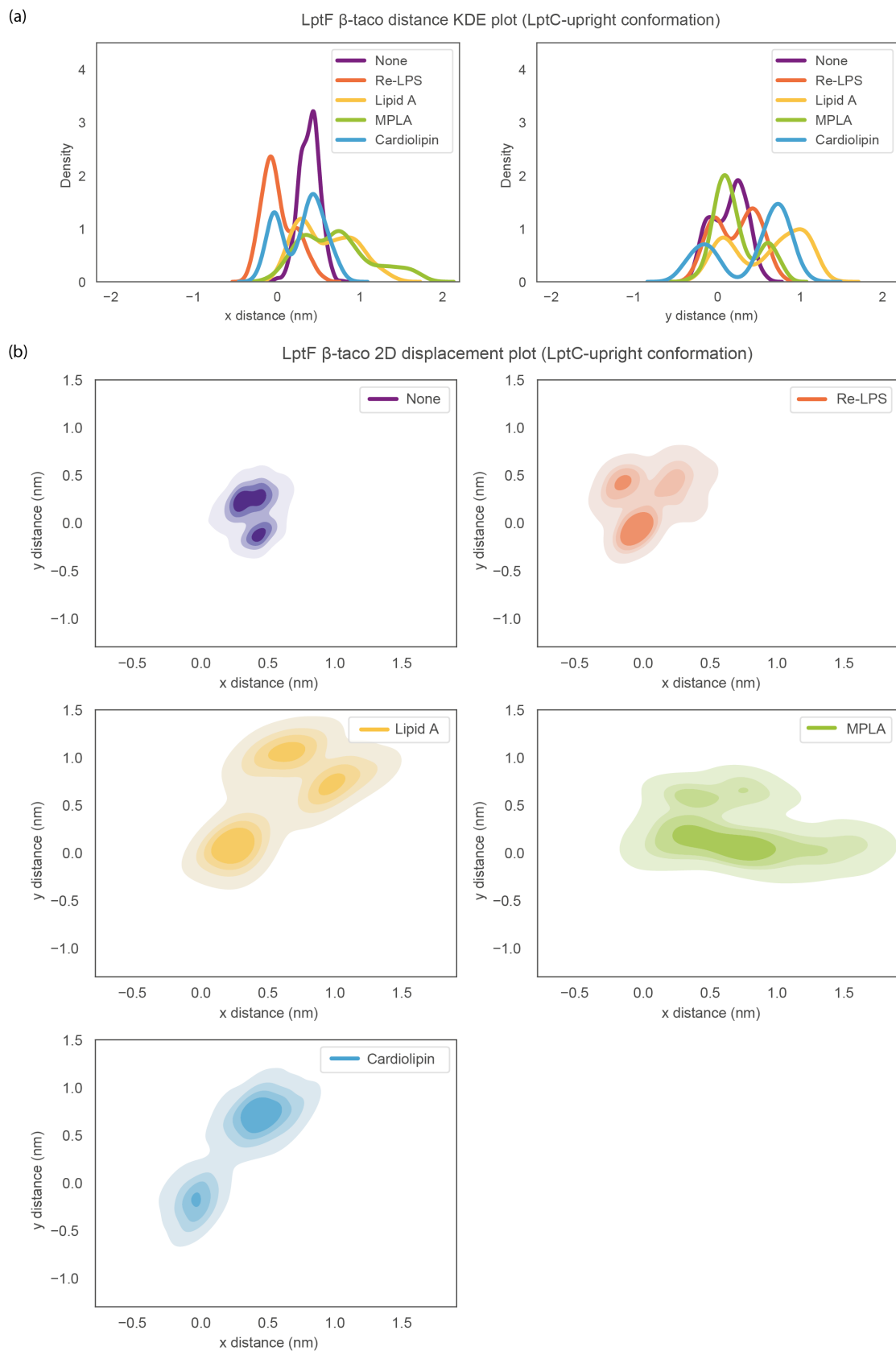


Figure 3.31 Displacement of LptF β -taco domain in the LptC-upright conformation. (a) 1D x, y displacement of the LptF β -taco domain in the LptC-upright conformation plotted as a kernel density estimate (KDE) with no lipid present or Re-LPS, lipid A, MPLA and cardiolipin present in the cavity. Where the positions $x = 0$ nm and $y = 0$ nm are set as the initial position of the β -taco domain in each of the states respectively. (b) 2D KDE plot of x, y displacement across LptF β -taco with: Top left: no lipid present, Top right: Re-LPS, Middle left: lipid A, Middle right: MPLA and Bottom left: cardiolipin present in the cavity.

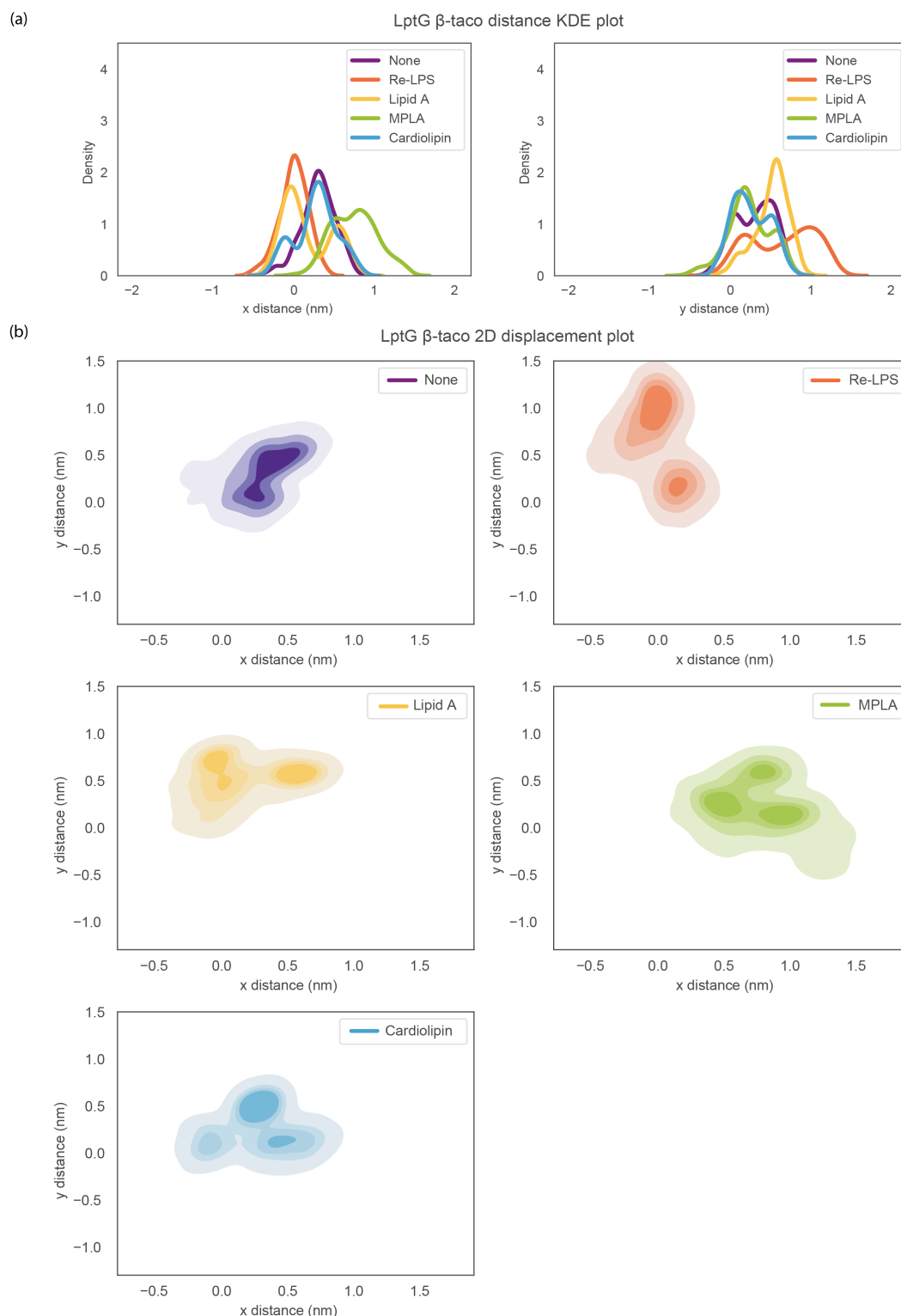


Figure 3.32 Displacement of LptG β -taco domain in the LptC-upright conformation. (a) 1D x, y displacement of the LptG β -taco domain in the LptC-skewed conformation plotted as a kernel density estimate (KDE) with no lipid present or Re-LPS, lipid A, MPLA and cardiolipin present in the cavity. Where the position $x = 0$ nm and $y = 0$ nm are set as the initial position of the β -taco domain in each of the states respectively. (b) 2D KDE plot of x, y displacement across LptG β -taco with: Top left: no lipid present, Top right: Re-LPS, Middle left: lipid A, Middle right: MPLA and Bottom left: cardiolipin present in the cavity.

For LptF, the general trend is a deviation of the x-angle by approximately -20 degrees and deviation of the z-angle by +15 degrees. The y-angle fluctuates deviates by 10-15 degrees around the origin. The LptG β -taco angle deviation follows a similar trend for the x and y angles, however the z angle deviates markedly in the opposite direction to that of LptF β -taco domain. This is an interesting observation as it implies correlation between the movement between the β -taco domains of the transporter, especially when compared to the comparatively minimal deviation observed in the LptC-skewed conformation.

Again, 1D and 2D KDE displacement plots are plotted for the LptF and LptG β -taco domains respectively (**Figure 3.31** and **Figure 3.32**). In both cases, the soluble domains are positively displaced for both the x and y distances when lipids are bound inside the cavity. The β -taco moves in roughly the same direction independent of the lipid type in a positive X, Y direction by approximately 1.3 nm. As alluded to above, because the same movement is observed LptG as for the LptF β -taco it suggests that the conformational adjustments of the β -taco domains of the LptB₂FGC transporter is highly correlated and coordinated. Notably however, the displacement of the β -taco domains changes the least when Re-LPS is present within the binding site. Therefore, it seems likely that the LptF β -taco domain (which directly coordinates with LptC) is stable and recognised by its key phosphate headgroup and saccharide moieties,

which facilitates facile coordination with the periplasmic protein LptA for bridge formation and subsequent transport.

3.3 *Pa*LptC interaction with *Pa*LptA

Due to the unusual conformational position of LptC in the LptC-skewed conformation of LptB₂FGC I wanted to understand how LptC of the skewed β -taco interacts with LptA, and whether this is in accordance with recent crosslinking studies performed by collaborators (FF, JRB, CVR, CS and EWY). They found that crosslinkers DSSO (disuccinimidyl sulfoxide, also Bis-(propionic acid NHS ester)-sulfoxide, Bis(2,5-dioxopyrrolidin-1-yl) 3,3'-sulfinyldipropionate) and DSBU (disuccinimidyl dibutyric urea, also known as ureido-4,4'-dibutyric acid bis(hydroxysuccinimide) ester) with maximum distance between C α 's being 2.35 nm and 2.69 nm respectively, suggest that Lys22 on LptA crosslinks with Lys131 and Lys176 of LptC (**Figure 3.33**).

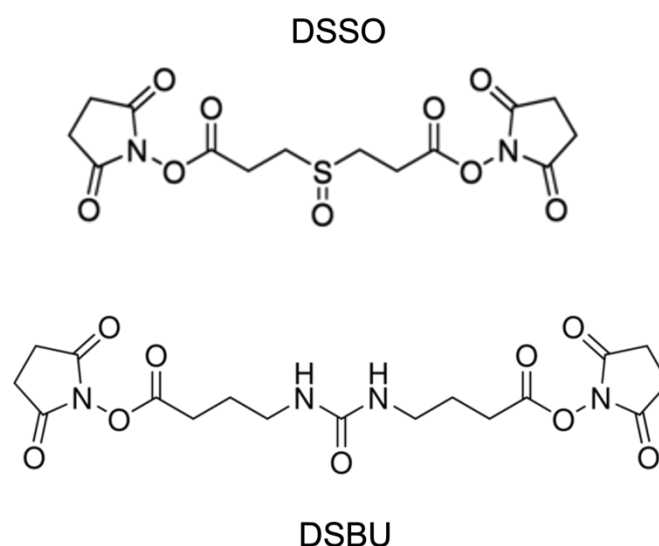


Figure 3.33 Chemical structures of crosslinkers. Illustration of disuccinimidyl sulfoxide (DSSO) and disuccinimidyl dibutyric urea (DSBU).

This means that DSSO crosslinked lysines can have a maximum distance of 2.35 nm between them where DSSO has an amine-reactive N-hydroxysuccinimide (NHS) ester at each end of a 7-carbon spacer arm. DSBU crosslinked lysines can have a maximum distance of 2.69 nm between them where DSBU possess an NHS ester at each end of an 11-atom spacer arm.

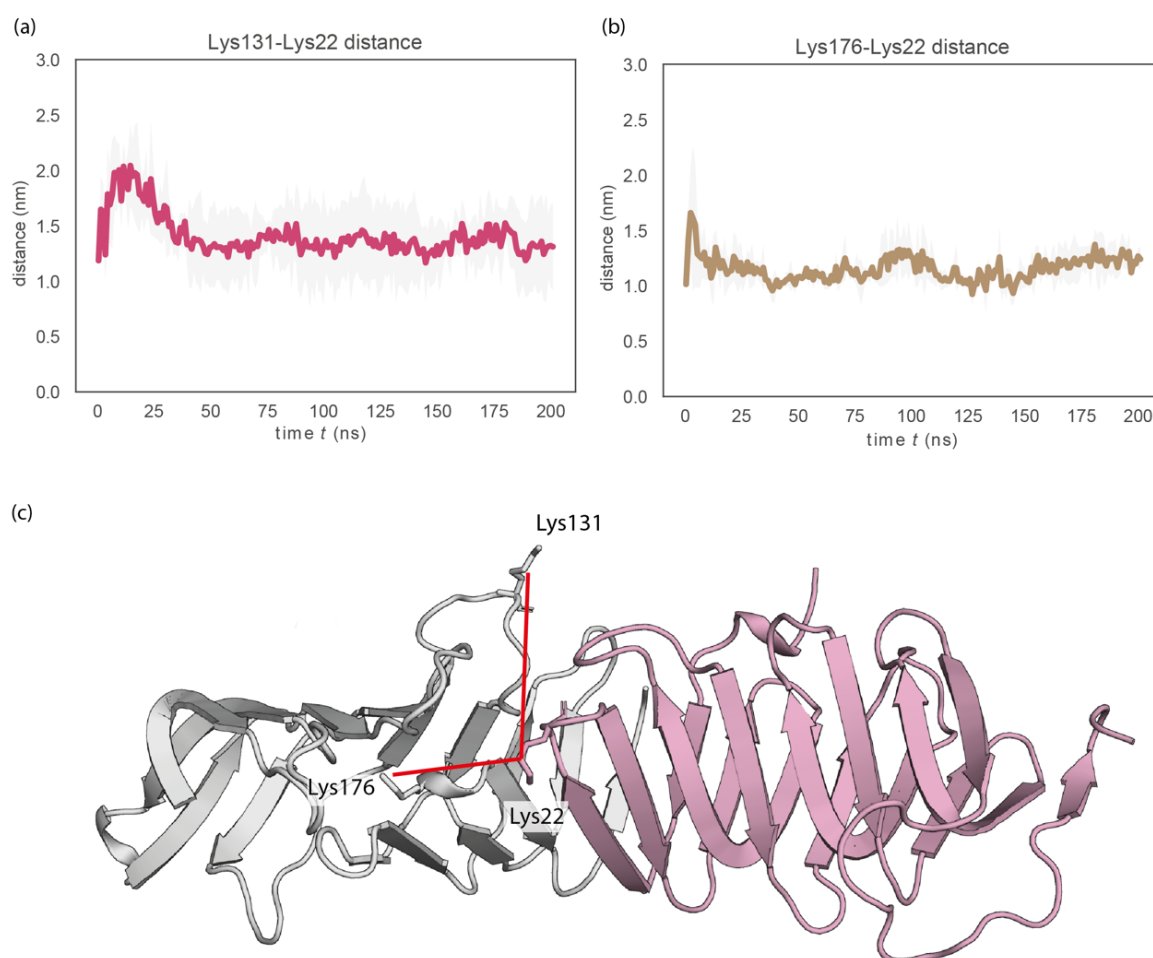


Figure 3.34 Lysine distance plot. (a) Average Ca-Ca distances measured between lysine 131 of LptC with lysine 22 of LptA as well as (b) the average distance between lysine 176 of LptC with lysine 22 of LptA. The shaded grey area represents the standard deviation over three replicates. (c) Structure of *PaLptAC* complex with the relevant lysines labelled.

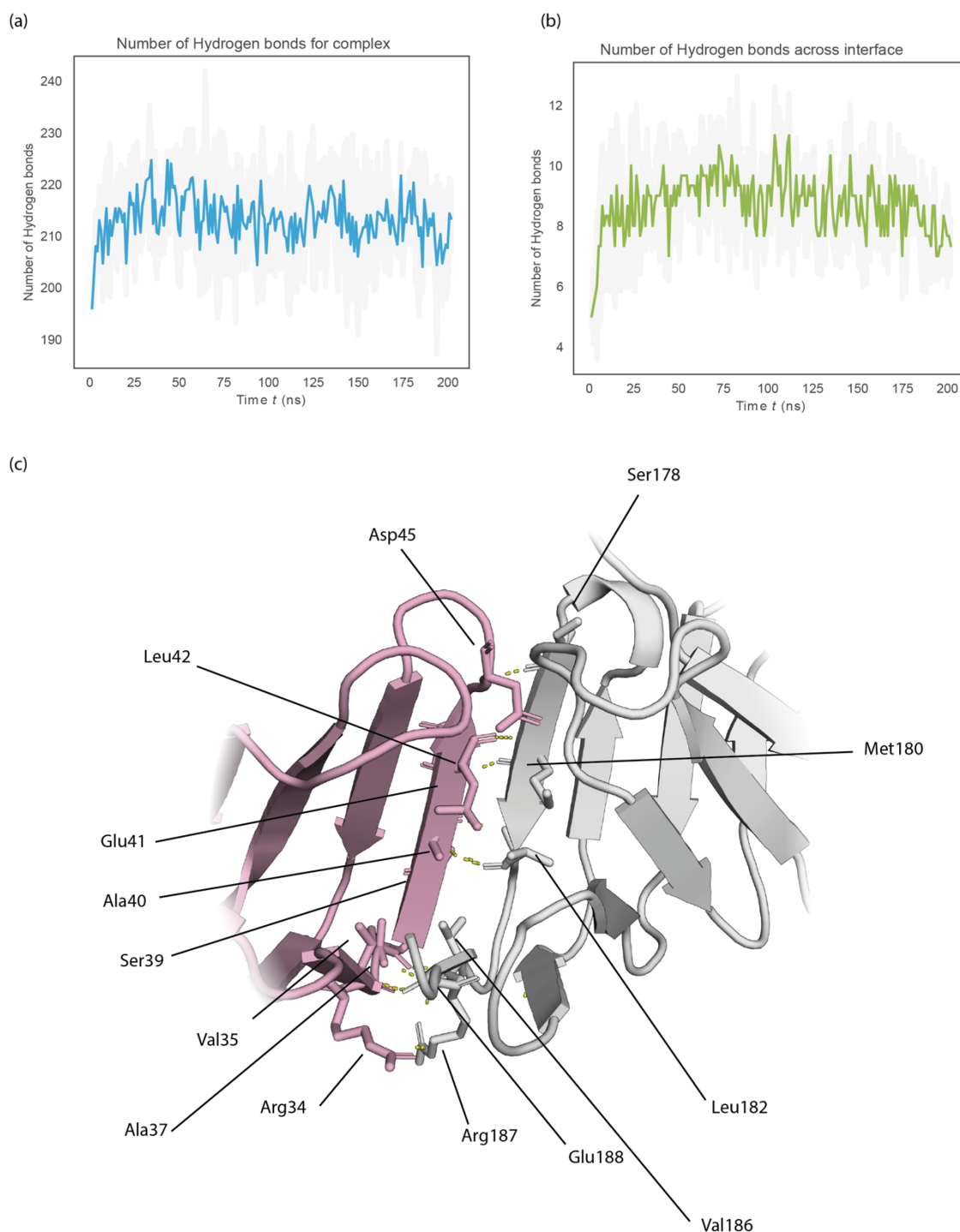


Figure 3.35 *PaLptAC* hydrogen bond count. (a) Average total hydrogen bonds across the *PaLptAC* complex over the simulation and (b) average total hydrogen bonds across the LptAC interface over the simulation time. The shaded grey area represents the standard deviation over three replicates. (c) The key residues which feature in the hydrogen bond network between LptA and LptC are represented as sticks. Hydrogen bonds are drawn on with dashed lines.

I generated the *PaLptA* homology model and docked this to *PaLptC* β -taco domain of LptC by aligning these domains. After obtaining this model I ran an AT simulation of

the complex for 200 ns over three replicates in a solvent box. I then measured the distance between these pairs of residues which were calculated to be ~ 1.05 nm and ~ 1.2 nm respectively which are within the experimental crosslinking range of 2.35 – 2.69 nm (**Figure 3.34**).

In order to verify the stability of the complex, I monitored the hydrogen bond count during the simulation time across the entire *Pa*LptAC complex which was consistent between 210-220 (**Figure 3.35a**), thereby implying that the model has formed a stable heterodimer. Furthermore, between the interface of LptA and LptC, the hydrogen bond number becomes consistent between 8-10 after 10 ns, suggesting that established coordination between proteins has been rapidly achieved (**Figure 3.35b**). The model produced accords with data obtained by collaborators and reveals the key residue interactions are those which constitute hydrogen bonds between the monomers of LptC (Arg34, Val35, Ala37, Ser39, Ala40, Glu41, Leu42 and Asp45) and LptA (Ser178, Met180, Leu182, Val186, Arg187 and Glu188) (**Figure 3.35c**).

3.4 Discussion

Detergent chemistry determines two different classes of IMP: PL binding and LPS binding. I can compare lipid binding native-MS work carried out by collaborators (LHU, FF, DS, MA, TRE, JRB, CVR)²⁸³ with *in silico* lipid binding experiments. For both a PL-binding IMP like AmtB and an LPS-binding IMP such as BtuCD, I demonstrate LPS binding to both protein classes as also observed in experiment (**Figure 3.2, 3.3**). This suggests that co-purified lipids for either AmtB or BtuCD are indicative of their local lipid environments within the native membrane and are not just an artefact of detergent purification. Therefore, this implies that core-LPS has a greater role within the IM which

needs to be fully explored by further MD and functional experiments. I quantified the LPS binding to the ATPase domains of BtuCD and found saccharide contacts near the ATP binding pocket. This is further supported by experimental data which suggests that the ATPase activity of BtuCD increases with the number of saccharide units attached to lipid A (**Figure 3.6, 3.7**). Here also, I could observe similar saccharide contacts with the ATPase domains of LptB₂FGC (**Figure 3.10, 3.11**), which specifically handles LPS as its substrate. My MD simulations help to explain that the glycan units contribute to the overall binding strength between LPS and BtuCD or LptB₂FGC and could also play a role in stimulating transporter activity.

The role of core-LPS within the IM is far larger than currently understood and may be involved in the native lipidome of IMPs such as BtuCD and AcrB²⁸³ (amongst others) which bind LPS. Therefore, this highlights the importance of targeting the LPS biogenesis pathway for the development of new antimicrobials.

In the context of how LPS functions in the IM promiscuously, it is important to assess then the details of how LPS interacts specifically with its transporter protein LptB₂FGC. By docking Re-LPS within the transporter in two distinct conformations of LptC (LptC-upright and LptC-skewed), I can begin to infer the functional role of the different states and compare this to lipid binding studies carried out by collaborators (FF, JRB and CVR). The position of Re-LPS within the hydrophobic cavity of the transporter shifts dramatically from one conformation (LptC-skewed) to the other, LptC-upright (**Figure 3.19, 3.20**). Initially, it is housed equidistant between LptF and LptG in the LptC-skewed conformation before moving towards the LptF in the LptC-upright state of the transporter. By quantifying the lipid contacts, I can observe that this shift is driven by

phosphate group contacts on Re-LPS as well as Kdo₂ interactions between the TM helices of LptF/G (**Figure 3.21 - 3.23**). Furthermore, cardiolipin within the cavity behaves in a very similar way to Re-LPS which helps justify the competitive kinetics between these lipids as observed by native-MS experiments as well as to highlight the key residues of interaction for substrate recognition. Perhaps then, the presence of cardiolipin within the local lipid environment of LptB₂FGC could help regulate LPS being extracted in the cavity.

The position of the LptF and LptG β -taco domains in the LptC-skewed conformation remain rigid and inflexible when different lipids (Re-LPS, lipid A, MPLA and cardiolipin) are accommodated within the hydrophobic cavity (**Figure 3.27 - 3.29**). This contrasts with the LptC-upright state where LptF/G rigidity is only observed when Re-LPS is docked within the cavity (**Figure 3.30 - 3.32**). This implies that the transporter is conformationally selective to LPS and must contain both phosphate groups and Kdo₂ saccharide units for it to be recognised. I suggest that the LptC-skewed conformation of the transporter acts as a 'resting state' before LPS is docked inside and shifts position from LptG to LptF to be ready for transport. This, in turn, causes the LptC domain to remain upright for facile binding to LptA units before LPS can be transported through the periplasm to the OM. At this juncture, the most important binding contacts between LptA and LptC are those which form hydrogen bonds between the interface. The complex between these β -taco domains is stable within my simulations where distances between experimentally linked lysines are shown to be easily within crosslinking range (**Figure 3.34, 3.35**).

3.5 Methods

3.5.1 Modification of Lpt structures used

Here is a table referencing all the structures taken from the PDB used to model different sections of LptB₂FGCA:

Table 3.1 PDB IDs for Lpt complex structures. Protein structures used directly in simulations or used to generate the homology models of the different subcomplexes of the Lpt system.

Protein	Organism	PDB code	Resolution (Å)
LptB ₂ FG ²⁹⁹	<i>P. aeruginosa</i>	5X5Y	3.46
LptB ₂ FG ³⁰⁰	<i>K. pneumoniae</i>	5L75	3.70
LptB ₂ FG ²⁹³	<i>S. flexneri</i>	6S8H	3.70
LptB ₂ FG ²⁹³	<i>S. flexneri</i>	6S8G	3.50
LptB ₂ FGC ⁸⁷	<i>E. cloacae</i>	6MIT	3.20
LptA ³⁰¹	<i>E. coli</i>	2R19	2.16
LptC ³⁰²	<i>E. coli</i>	3MY2	2.20

3.5.2 The IM sub complex LptB₂FGC

A number of crystal and cryo-EM structures exist for the LptB₂FGC complex derived from different species (*Shigella flexneri*, *Escherichia coli*, *Enterobacter cloacae*, *Vibrio cholerae*), with nucleotide bound and lipopolysaccharide bound states²⁹³. This also

includes the more recently resolved structures containing the β -taco domain of LptC from *E. cloacae* and *V. cholerae*⁸⁷ as well as from *Pseudomonas aeruginosa* with a skewed LptC β -taco domain which was used directly for AT/CG simulations.

3.5.2.1. Modelling the TM_c helix for *PaLptB₂FGC*

The *PaLptB₂FGC* structure does not possess a resolved LptC transmembrane alpha helix (TM_c), and so this was modelled in with MODELLER²⁹⁶ using the first 45 residues from LptC resolved from *E. cloacae* as a template. MODELLER uses an 'align2d' algorithm when you supply the programme with a structure which constructs an alignment by a 'gap penalty function' which considers the structural information supplied by the template by placing gaps in proximally close positions in space, as well as solvent exposed and curved regions²⁹⁶. The *PaTM_c* helix produced with the lowest DOPE (discrete optimised protein energy) score which is a statistical potential assessment based on the probability density function of the atomic cartesian coordinates, the lowest score is suggestive of the most favourable homology model produced by MODELLER³⁰³. The *PaTM_c* helix was aligned using PyMOL onto the *E. cloacae* structure, which was itself aligned to *PaLptB₂FGC*, where the loop was then connected to the β -taco domain of *PaLptC*. The PyMOL sculpt tool³⁰⁴ was used to move the loop away from any steric clashes near the TM helices of *PaLptFG*.

3.5.2.2. Building the homology model for the LptC-upright *PaLptB₂FGC* structure

The amino acid sequence from *PaLptB₂FGC* was uploaded to the SWISS-MODEL²⁹⁷ server where the LptB₂FGC structure from *E. cloacae* was supplied as a template. I selected 'alignment mode' which generates a pair-wise alignment of the supplied sequence to the template to generate the homology model of LptC-upright

PaLptB₂FGC. This mode was selected as a template was supplied, otherwise the SWISS-MODEL would search the PDB database for structures which may not produce the 'upright' conformation of LptC.

3.5.2.3. *Generating the lipid bound structures within the hydrophobic cavity*

To assess Re-LPS contacts within the hydrophobic cavity of the LptC-upright and LptC-skewed structure, I needed to dock the substrate within the transporter. The initial LPS binding pose was taken from the cryo-EM structure of *SfLptB₂FG* in complex with LPS (PDB ID: 6S8H)²⁹³. Here, both structures of the LptC-upright and LptC-skewed states of the transporter were aligned to *SfLptB₂FG* using PyMOL and then converted to a coarse-grain structure using Martinize python script²²⁹. These structures were then embedded in a lipid bilayer using INSANE composed of cardiolipin (5 %), POPG (20 %) and POPE (75 %) and solvated in a box of Martini water and 0.15 M NaCl. INSANE (INSert membrANE) is a CG building tool that generates membranes by distributing lipids over a grid³⁰⁵. I extracted the liposaccharide coordinates from the aligned LPS bound *SfLptB₂FG* AT structure considering only those according to a 4:1 CG mapping to build a coarse grained Re-LPS, lipid A and mono phosphate lipid A (MPLA) representation alongside a CG cardiolipin (aligning the phosphate head groups and four acyl tails with those of the most proximal LPS tails). I included these coordinates within the CG structures of the membrane embedded LptC-upright and LptC-skewed states and energy minimised them by steepest descents methodology to a target force of 1,000 kJ mol⁻¹ nm⁻². Each system was equilibrated with 100 ns of simulations time with a timestep of 20 fs in a Martini 2.2 forcefield²²⁹ in GROMACS 2019.1²⁵⁶ for 100 ns in three separate replicates. A pressure of 1 bar with the Berendsen barostat²⁶¹ was used. The coupling constant

(τ_P) here was set at 4 ps at a constant temperature of 310 K at a compressibility of $5.0 \times 10^{-6} \text{ bar}^{-1} \text{ K}$ for steady velocity scaling³⁰⁶. Each of the three equilibrated structure for both the LptC-upright and LptC-skewed states of the transporter for each respective lipid species (Re-LPS, lipid A, MPLA and cardiolipin) were then converted back to AT using CG2AT python script²³⁷. For CG2AT the CHARMM36 July 2017 updated forcefield¹⁹⁷ was selected alongside with TIP3P water²⁶⁷. This then generated all the equilibrated AT structures used for later production simulation and analysis.

3.5.3 Constructing the *Pa*LptAC complex

LptB₂FG as well as the loop and TM_c helix of *Pa*LptC was deleted from the LptC-skewed *Pa*LptB₂FGC complex, leaving only the β -taco domain. *Pa*LptA was constructed using SWISS homology modeller by uploading the UniProtKB³⁰⁷ *Pseudomonas aeruginosa* sequences and using *Ec*LptA as the target (PDB: 2R19). LptA and LptC were aligned using PyMOL to make the final the complex used for hydrogen bond analysis.

3.6 Atomistic (AT) simulations

3.6.1 *Pa*LptB₂FGC AT simulations

The various lipid docked structures (Re-LPS, lipid A, MPLA, cardiolipin and apo) of the LptC-skewed and LptC-upright LptB₂FGC conformations were prepared as outlined above. After conversion to AT, a further 1 ns of equilibration was required where position restraints were applied to the protein in the CHARMM36 forcefield.¹⁹⁷ Production simulations were then run for 200 ns with a 2 fs timestep without any restrains in explicit TIP3P water. The V-rescale thermostat was set at 310 K with semi-isotropic Parrinello-Rahman pressure coupling at 1 atm.^{259,260} Three replicates were

run for lipid bound state of the LptB₂FGC with the velocities reset for each replicate in GROMACS 2019.1²⁵⁶.

3.6.2 *Pa*LptAC AT simulations

The *Pa*LptAC complex is soluble and therefore will not be embedded in a membrane and so there is no need to convert to CG for equilibration. The *Pa*LptAC structure was solvated in a cubic box with 0.15M NaCl and explicit TIP3P solvent. The system was energy minimised by steepest descents to a target force of 1,000 kJ mol⁻¹ nm⁻² before NVT equilibration for 100 ps to reach a target value of 300 K. NPT equilibration followed, for 100 ps with a reference pressure of 100 bar. Three replicates of production simulations were run for 150 ns without restraints with a 2 fs timestep where the V-rescale thermostat was set at 310 K with isotropic Parrinello-Rahman pressure coupling at 1 atm^{259,260}. For each replicate the velocities were reset and was run in GROMACS 2019.1²⁵⁶.

3.7 CG-MD simulations

3.7.1 IM proteins CG simulations

I used the MemProtMD pipeline to generate CG representations for AmtB (PDB: 1U7G)³⁰⁸, BtuCD (PDB: 4R9U)³⁰⁹ and the LptC-skewed LptB₂FGC structures. Using a CG 4:1 mapping scheme a CG representation was made for MPLA, lipid A, lipid A bound to two ketodeoxyoctonic acid (Kdo) moieties (Re-LPS) and Lipid A-core oligosaccharide (core-LPS)³⁰⁵. For each simulation the membrane composition was 83 % POPE, 5 % cardiolipin, 10 % POPG and 2 % of either MPLA, lipid A, Re-LPS, LPS or cardiolipin (as a control). In each case, the simulation boxes were solvated with Martini water and made electrostatically neutral with 0.15 M of NaCl. The simulations were run with the MARTINI 2.2 force field³¹⁰ which each ran for 5 μ s

applying a pressure of 1 bar with the Berendsen barostat with a coupling constant (τ_P) of 4 ps and a compressibility of $5.0 \times 10^{-6} \text{ bar}^{-1}$ where the temperature remained constant at 310 K for steady velocity scaling, with a coupling constant (τ_T) of 1 ps²⁶¹. Five replicates were performed for each IM protein and for each lipid type for contact analysis.

3.8 Lipid contact analysis

CG Lipid contacts between IM proteins: AmtB, BtuCD and LptB₂FGC with lipid A, Re-LPS, and LPS were quantified with a local python script³³ using MDanalysis³¹¹ tools. Here, each residue of the protein was scored between 0 and 1 which is defined by the percentage of frames in the simulation where the residue was 6 Å away from a lipid bead for the lipid type under analysis, derived from the van der Waal radii of two Martini beads. Therefore 100% occupancy would mean that the residue in question was 6 Å away from a lipid bead for the entire duration of the simulation. These values can be extracted as beta-factors and may be represented on the structure in PyMOL when colouring the structure. In the case of LptDE AT simulations (**Chapter 5**), residues 4 Å away from Re-LPS were considered a 'contact' to be scored.

Contact values over the five separate simulations were averaged over repeating units in the ABC-transporters and plotted separately to increase sampling, i.e. BtuCD contacts were averaged over the dimer, AmtB over the trimer and LptB was averaged over the dimer.

3.9 Author contributions

LHU, FF, DS, MA, TRE, JRB, CVR carried out the native MS work with AmtB, AcrB and BtuCD and made **Figure 3.1, Figure 3.3a, c and d**. CS and EWY obtained the structure of LptC-skewed LptB₂FGC. FF, JRB, CVR carried out the native MS work with LptB₂FGC and cardiolipin, MPLA, lipid A and Re-LPS as well as the experimental cross-linking data. All other work including MD studies and analysis was carried out by JBS.

CHAPTER 4 The thermodynamics of lipid A transport through the Lpt system

4.1 Background to the periplasmic LPS chute

LptA is found connecting the IM ABC transporter LptB₂FGC and the OM heterodimer LptDE subcomplexes together, bridging the periplasm and forming the Lpt system^{92,312}. To do this, the N-terminal domain of LptA interacts with LptC, whereas the C-terminal domain interacts with LptD⁹². Here also, the N-terminal of LptC itself makes contact with the periplasmic domain of LptF. Both of these subunits, LptC and LptA, alongside the soluble domains of LptF and LptD have a very similar and distinctive LptD ‘ β -taco’ protein fold^{299,301,302,313} where the anti-parallel β -sheets form a ‘V shape’ with a hydrophobic pocket within the fold. The homologous fold between these subunits obviously implies the interactivity between the units to form a continuous bridging complex across the periplasm, much like LptA or LptC units crystallising end-to-end with their hydrophobic grooves aligned^{301,302}. Along the fold of each of the β -taco subunits there is a 90° twist, and when oligomerised together form a twisting groove along the periplasmic bridge, much like a ‘helter-skelter’. LptC, LptA and LptD have been shown to bind LPS, however very little is known about how these

units interact with the peptidoglycan layer^{51,85,90,314,315}. This bridge distinctively characterises the purported 'PEZ' system of transport whereby LPS is continuously streamed across (driven by ATPase activity) into the OM, assuming also that the bridge is completely stacked with lipid molecules (**Figure 4.1**).

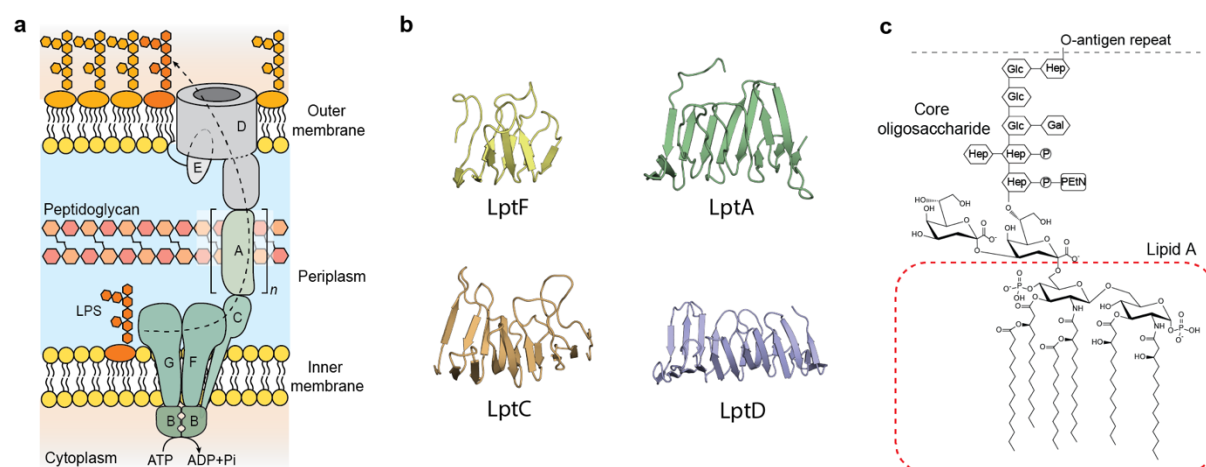


Figure 4.1 Cartoon schematic describing the LPS transport pathway. Figure adapted from Fiorentino *et al*,⁵¹ all work here completed by myself and FF. (a) cartoon schematic of the Lpt system: LPS is extracted from the inner membrane (IM) by the ABC transporter complex LptB₂FGC, where it is subsequently shuttled through the periplasm via a soluble bridge of LptA units to be finally installed in the outer membrane (OM) by LptDE. **(b)** Structures of the 'β-taco' protein bridging subunits: LptF, LptC, LptA and LptD **(c)** A skeletal representation of LPS where its tripartite structure is described: lipid A, core oligosaccharide and the O-antigen repeating unit where the red box highlights the lipid A unit.

As suggested by hydrogen deuterium exchange mass spectrometry (HDX-MS) data⁵¹ (**chapter 5**) and *in vivo* photo-crosslinking experiments⁸⁵, the tail of LPS are held within the hydrophobic pocket of the β-taco subunits where the hydrophilic saccharide unit is facing outward into solvent. NMR studies have shown interactions between LPS with the LptAC complex as well as with the monomeric LptA and LptC subunit⁹⁵. It was once suggested that the (ATP independent) driving force for LPS transport from LptC to LptA along the bridge was due to a stronger binding affinity of LPS with LptA, which was understood from a study in solution which revealed LptA displacing LptC-LPS

binding⁹⁰. Though further studies have suggested that this may not be the case, with K_d values of LPS with LptA ranging from 7 to 35 μM with values of LptC ranging from 11 to 28 μM ^{315,316}.

LptA, itself, forms head to tail oligomers in the presence of LPS (and therefore required mutation in the NMR study⁹⁵ to bind a singular unit) as well as concentration dependent oligomers in solvent (up to 22-mers)^{94,317}. Though it is still not yet known the number of LptA units which make up the LPS bridge, whether this is indeed variable or fixed in anyway in the system. The interactions between subunits are well characterised (as well as the arrangement of these comprising the bridge) but it is not understood how the residues of the periplasmic chute contribute to the transport of LPS through the system. Furthermore, it is also unclear the extent to which LPS binds with each of the bridging subunits and what contributes to the strength of binding interactions with LPS along the hydrophobic groove of the periplasmic bridge.

4.2 Opening up the β -taco subunits

Each of the β -taco subunits is crystallised in a closed conformation, with the internal hydrophobic groove interacting with itself^{51,85,90,314,315}. The subunits were opened by applying an alchemed²³⁶ methodology which grows in the lipids within the subunits to pry them open.

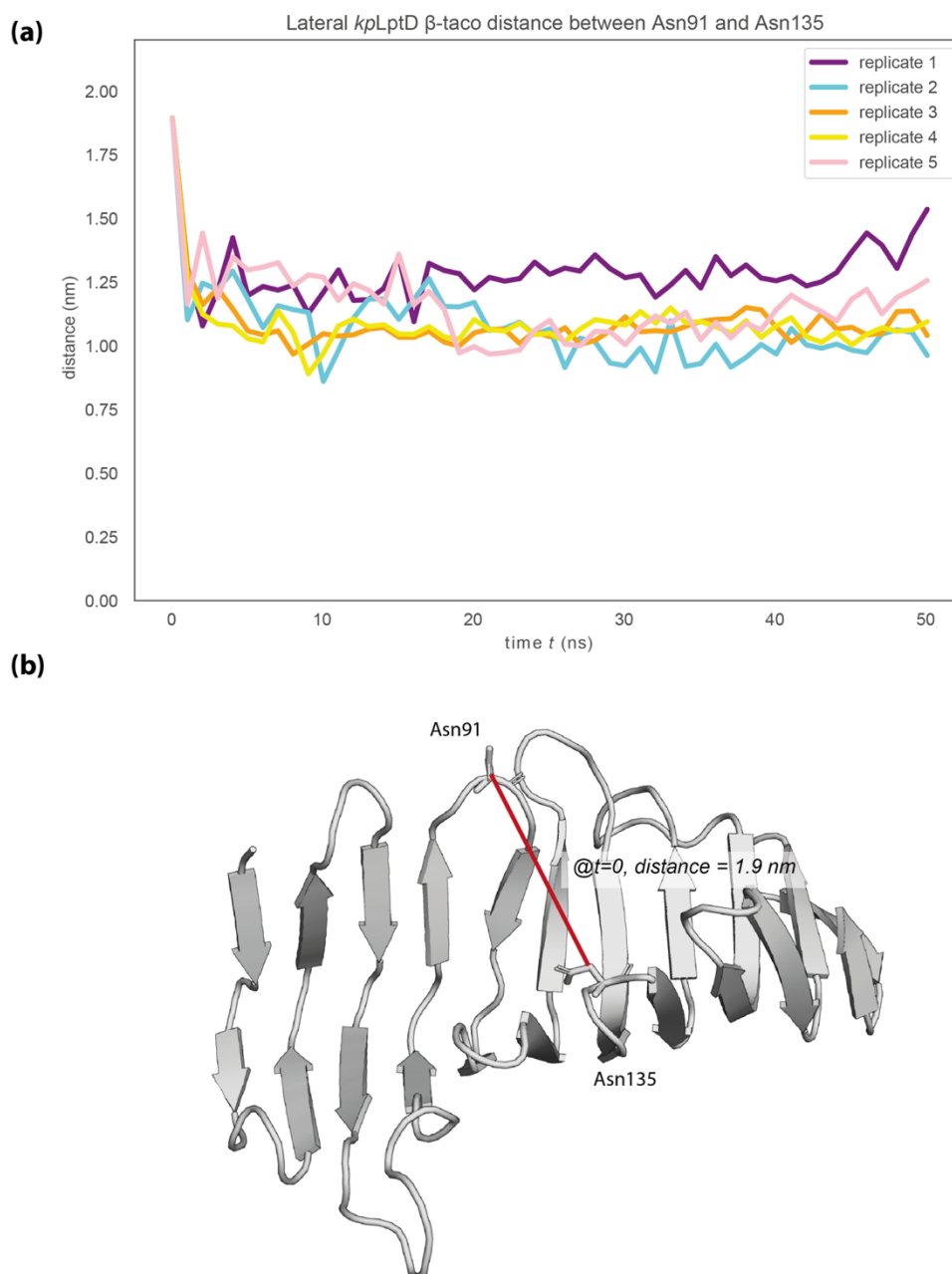


Figure 4.2 Distance plot of the sides of the *KpLptD* β -taco. (a) Average distance between backbone atoms of the two residues (Asn91, Asn135) on the either side of the LptD β -taco over five replicates. β -taco closure occurs within 1 ns of simulation time. **(b)** Residues depicted on schematic of *KpLptD*.

All atom *in silico* studies with the *KpLptD* β -taco subunit demonstrate rapid (within 1 ns) β -taco closure (~ 1.1 nm, as measure by the distance between Asn91 and Asn135) when the domain is fully open (1.9 nm), and the lipid is removed (**Figure 4.2**). This demonstrates the importance of hydrophobic contacts in the β -taco interacting with

itself when no lipid is bound and emphasises how solvent accessibility within the domain is highly unfavourable.

4.3 The binding energy of lipid A within each of the isolated subunits of the periplasmic chute is variable

With each β -taco having been opened (by growing lipid A from inside the domains), I wanted to calculate the binding energy of the lipid within each of the distinct units in isolation. This would give an indication of the strength of binding interactions which the lipid experiences through the chute. I performed the simulations necessary to calculate the energetics in coarse grain to significantly decrease the computational cost. To understand the binding energy, lipid A was used as the substrate (in all calculations) to simplify and speed up the simulations. This is because the tails of the LPS interact predominantly with the hydrophobic groove of the periplasmic chute, whereas the sugar units are assumed to be pointed towards the solvent and therefore not strictly necessary to factor into the binding calculation. However, this is an oversimplification as sugars may have a role interacting with the soluble loops of the periplasmic subunits and will interact with the hydrophobic cavity of the LptB₂FGC transporter (as shown in **Chapter 3**) as well as the lateral gate of LptDE (discussed in **Chapter 5**).

To quantify the binding energy of lipid A from the β -taco units which comprise the periplasmic bridge I used potential of mean force calculations (PMFs). Here, I employed steered MD to pull lipid A across the subunits from within the β -taco out into solvent in a 'card-swipe' like fashion. This motion occurs in both the forward (in the physiological direction of transport towards the OM) and reverse direction (towards the IM) to sample the entire protein groove (**Figure 4.3**).

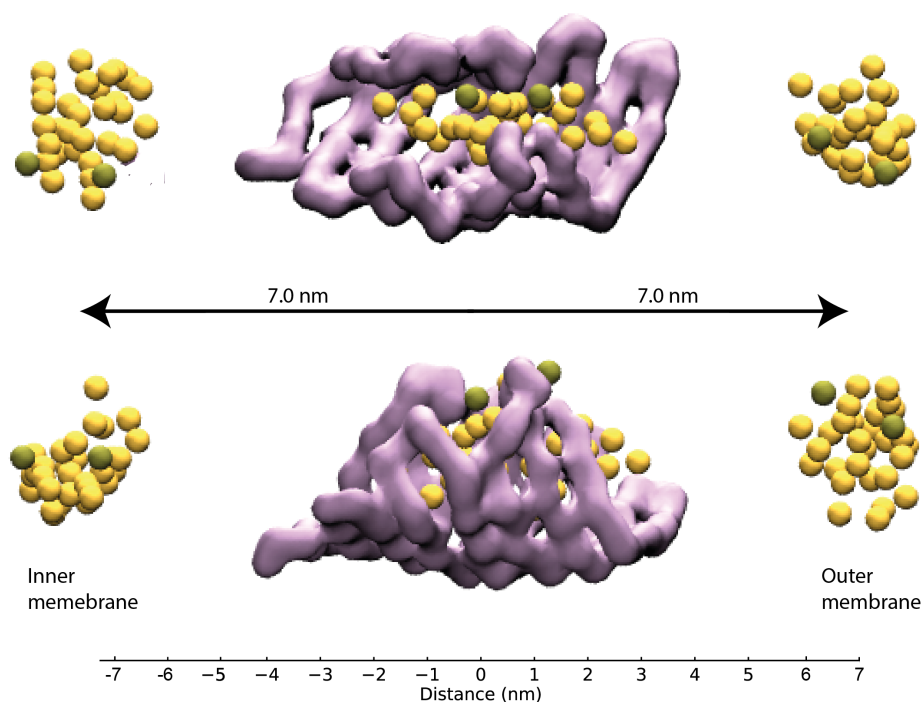


Figure 4.3 Schematic of steered MD. Lipid A being pulled out from the LptC β -taco in both the forward (towards the OM) and reverse direction (towards the IM).

The reason for using two PMF's from two 'forward' and 'reverse' pull-codes like this is a practical one, as it simplifies the movement of lipid A being threaded accurately through the β -taco. Subsequently, I stitch both PMF profiles together to realise the strength of the lipid A binding pocket through each of the β -taco domains in turn.

The binding profile obtained (**Figure 4.4a**), reveals the energy of lipid A from outside the β -taco in bulk solvent, where lipid A moves from experiencing unfavourable energetics to thermodynamic relaxation in the hydrophobic groove, before moving back out into solvent (which is set to zero). To ensure sampling is sufficient across the collective variable, specific positions which were undersampled, were detected and revisited (**Figure 4.4b,c**)¹⁴². There the lipid was constrained to achieve satisfactory overlap between histograms which totalled 385 for the entire PMF calculation (**Figure 4.4d**). This number of separate windows, was required to sample the entire length of

the collective variable from solvent, into the hydrophobic groove of LptF, and back out into solvent again.

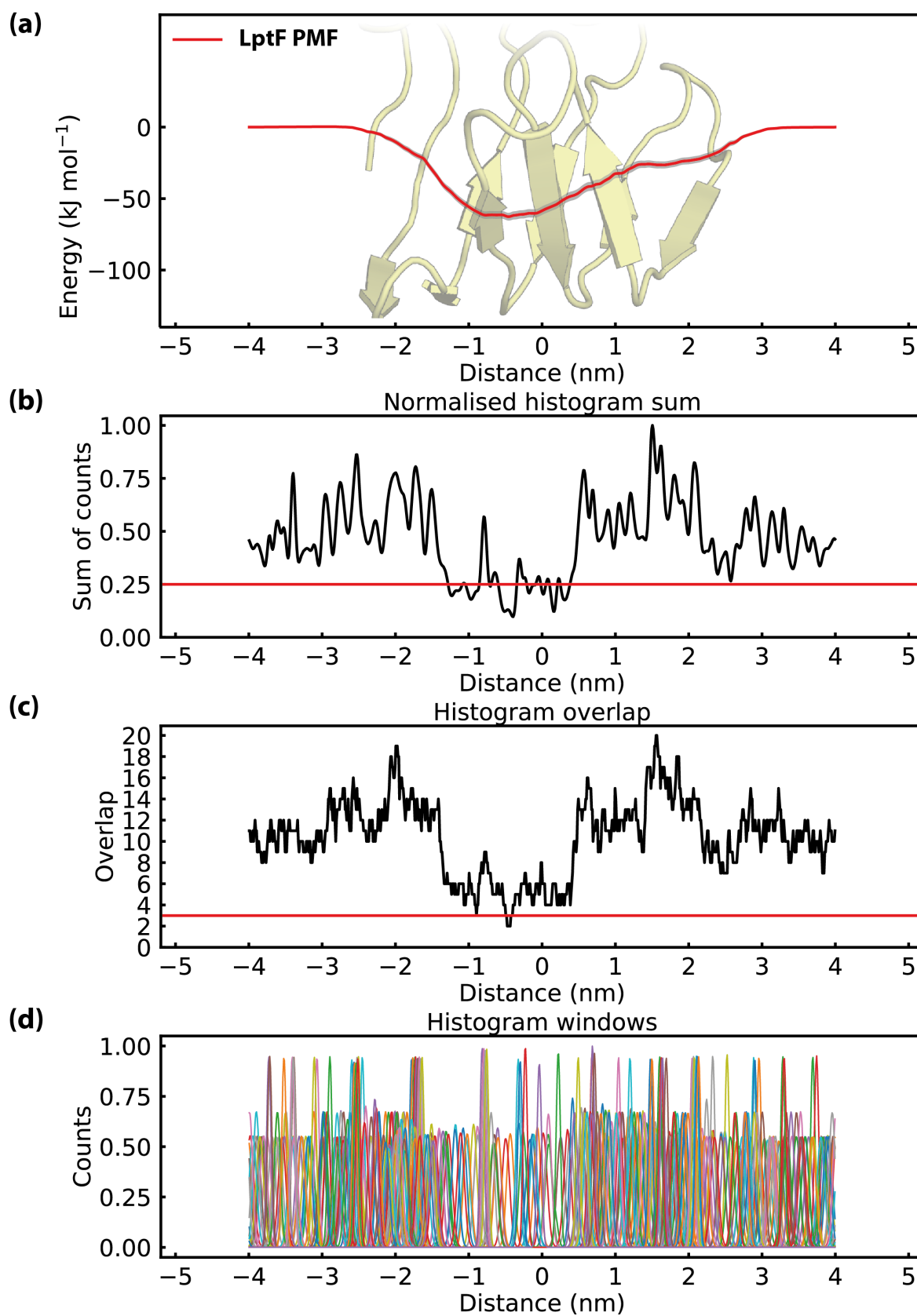
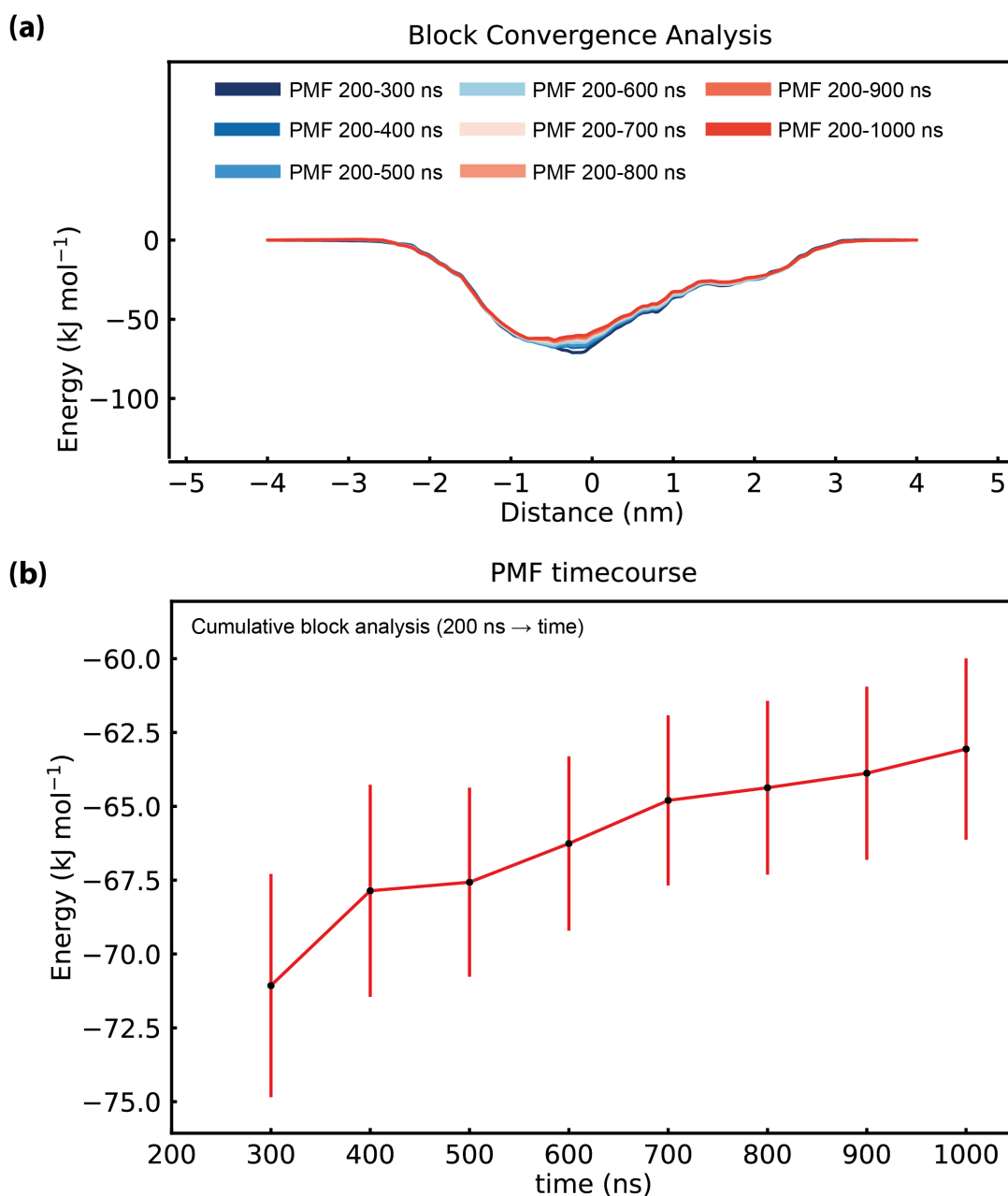


Figure 4.4 Binding energy of lipid A with the LptF β -taco. (a) PMF profile of lipid A moving through LptF, areas shaded in grey indicates the error and is obtained by bootstrapping the profile 200 times. Cartoon representation of the LptF subunit aligned with the energy profile in the background for reference. (b) Total counts summed across the collective variable of lipid A being pulled from the β -taco into solvent from the raw histograms, where the dataset is normalised to 1. The red line is a 'rule of thumb' marker for acceptable sampling positioned at 0.2 of total counts. (c) The number of overlapping histograms from the umbrella sampling across the collective variable. A count is considered when the overlap is $>10\%$ of the total histogram height. The red line corresponds to a 'rule of thumb' for sufficient overlap set to 3. (d) The histograms from the WHAM analysis, where every histogram corresponds to a separate sampling position of lipid A.



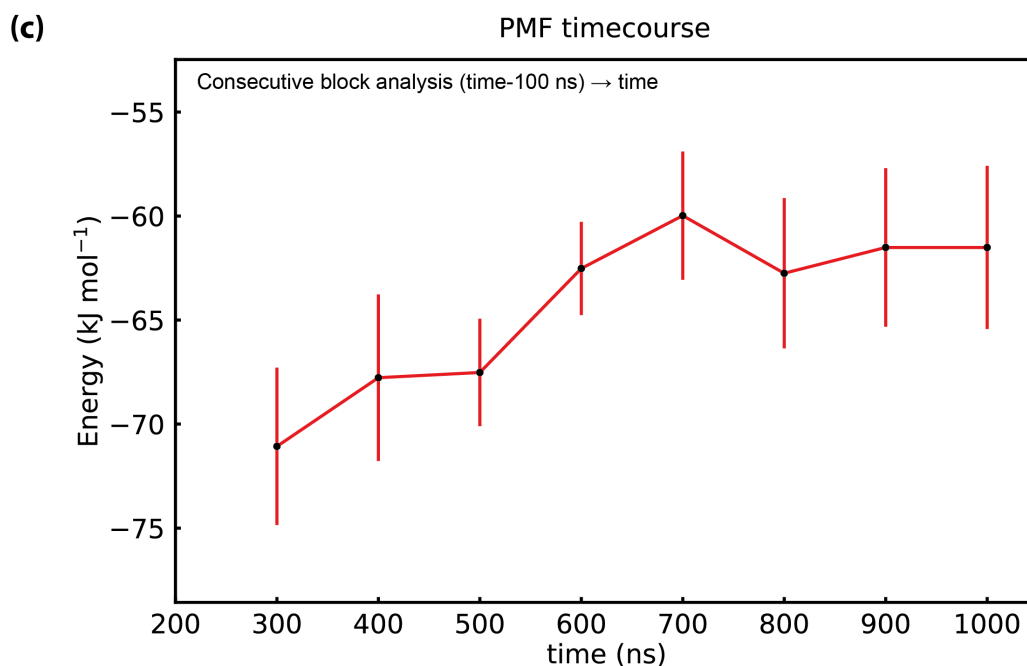


Figure 4.5 LptF PMF convergence analysis. (a) PMF profiles of increasing blocks simulation time by 100 ns for up to one microsecond. Colour scale from blue to red illustrates how the profile evolves through time. The first 200 ns was removed from each window to ensure that the system had fully equilibrated first. (b) The binding values (the lowest energy value) of each PMF profile over the course of the simulation time for increasing blocks of 100 ns. Errors calculated by bootstrapping each PMF profile 10 times. For instance, 200 – 300 ns, 200 – 400 ns, ..., 200 – 1000 ns. (c) The binding values (the lowest energy value) of each PMF profile over the course of the simulation time for consecutive blocks of 100 ns. For instance, 200 – 300 ns, 300 – 400 ns, ..., 900 – 1000 ns.

At this stage, it is important to assess how the binding energy changes over the course of the sampling time to ensure that the energy value has converged. I calculated the binding energy between increasing blocks of simulation time (e.g., 200-300 ns, 200-400 ns, 200-500 ns etc) and bootstrapped ($n=10$) each block to calculate the statistical error. I removed the first 200 ns from each window in the calculation to ensure the system had fully equilibrated first. The resulting PMF profiles are plotted on top of each other to illustrate how the landscape evolves through increasing sampling time (for up to one microsecond) (**Figure 4.5a**). The resulting binding values from 700 ns onwards are within error of each other, so I suggest the profile reaches a converged state after

700 ns (**Figure 4.5b**). The binding energy of the LptF β -taco is calculated to be $-62 \pm 3 \text{ kJ mol}^{-1}$ after one microsecond of simulation time. LptF is almost half the size of the other β -taco bridging units and consequently the binding energy for the other units would be assumed higher because of the larger hydrophobic pocket to accommodate lipid A.

The binding energy for the LptC β -taco is calculated to be $-130 \pm 2 \text{ kJ mol}^{-1}$, greater than twice the binding energy of LptF where 385 histograms were required for sufficient sampling here (**Figure 4.6**).

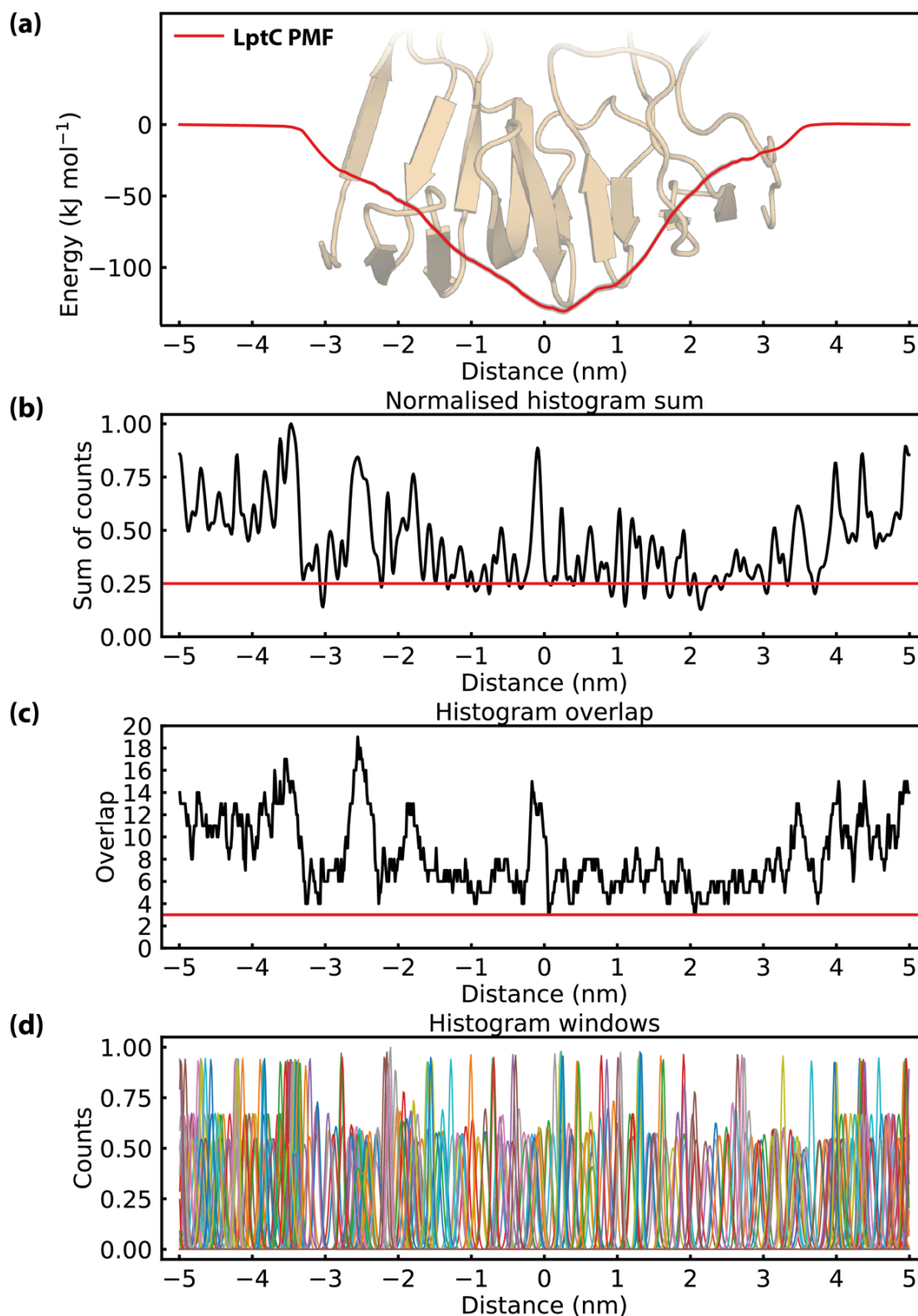
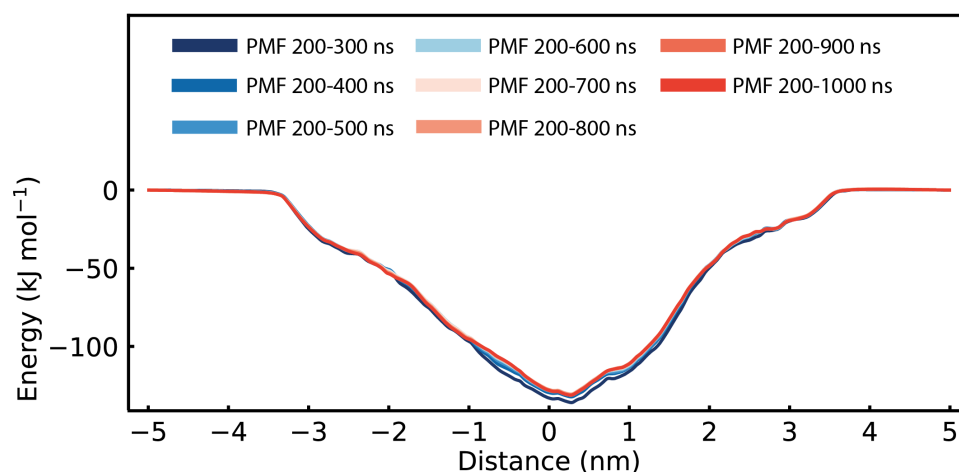


Figure 4.6 Binding energy of lipid A with the LptC β -taco. (a) PMF profile of lipid A moving through LptC, areas shaded in grey indicates the error and is obtained by bootstrapping the profile 200 times. Cartoon representation of the LptC subunit aligned with the energy profile in the background for reference. (b) Total counts summed across the collective variable of lipid A being pulled from the β -taco into solvent from the raw histograms, where the dataset is normalised to 1. The red line is a 'rule of

thumb' marker for acceptable sampling positioned at 0.2 of total counts. **(c)** The number of overlapping histograms from the umbrella sampling across the collective variable. A count is considered when the overlap is >10% of the total histogram height. The red line corresponds to a 'rule of thumb' for sufficient overlap set to 3. **(d)** The histograms from the WHAM analysis, where every histogram corresponds to a separate sampling position of lipid A.

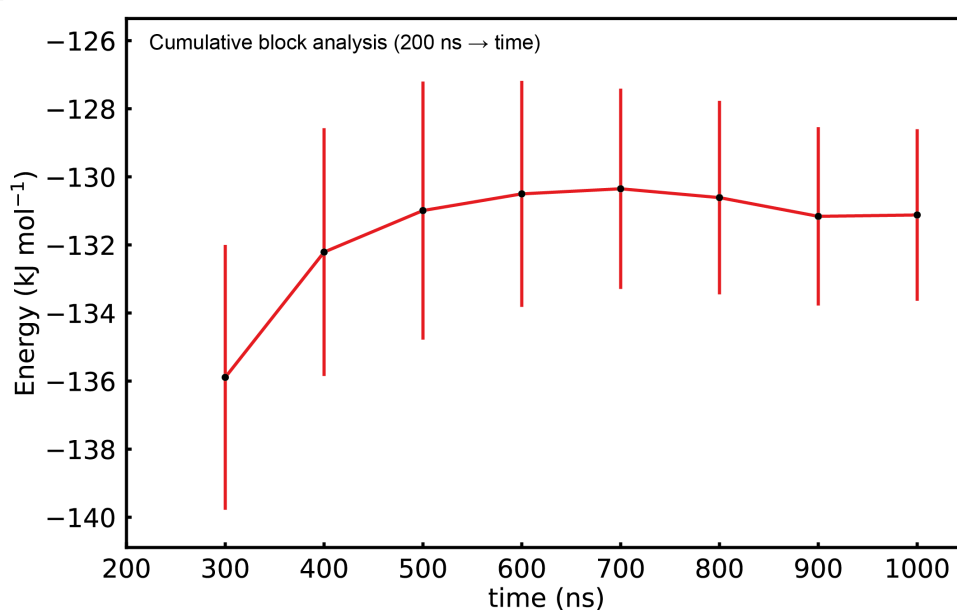
(a)

Block Convergence Analysis



(b)

PMF timecourse



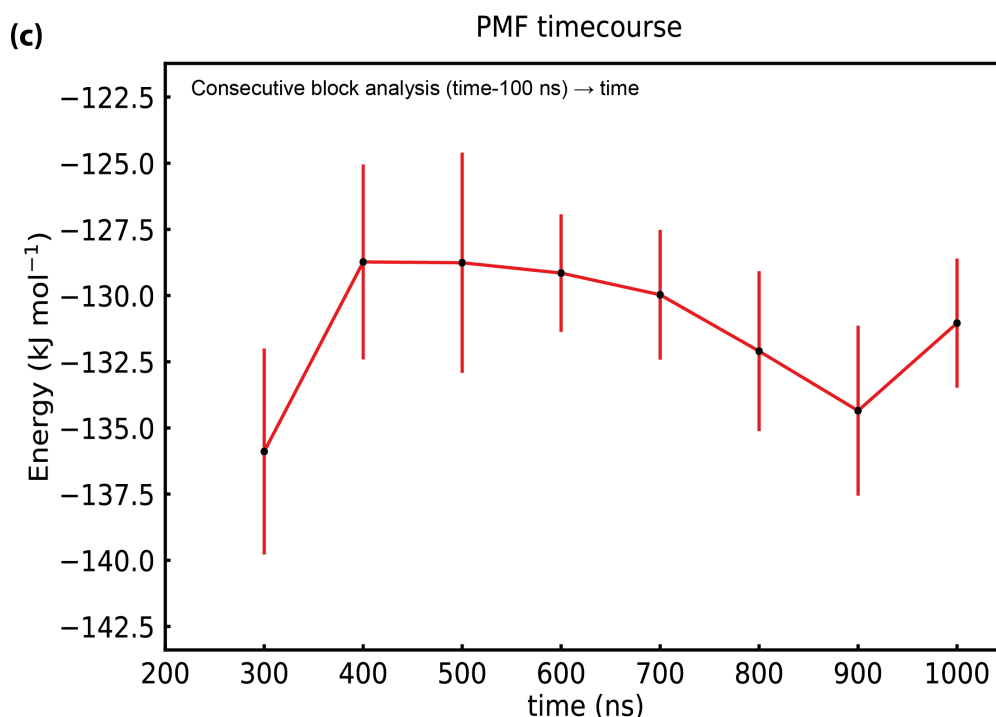


Figure 4.7 LptC PMF convergence analysis. (a) PMF profiles of increasing blocks simulation time by 100 ns for up to one microsecond. Colour scale from blue to red illustrates how the profile evolves through time. The first 200 ns was removed from each window to ensure that the system had fully equilibrated first. (b) The binding values (the lowest energy value) of each PMF profile over the course of the simulation time for increasing blocks of 100 ns. Errors calculated by bootstrapping each PMF profile 10 times. For instance, 200 – 300 ns, 200 – 400 ns, ..., 200 – 1000 ns. (c) The binding values (the lowest energy value) of each PMF profile over the course of the simulation time for consecutive blocks of 100 ns. For instance, 200 – 300 ns, 300 – 400 ns, ..., 900 – 1000 ns.

As before, I measure the convergence of the energy landscape in blocks of increasing simulation time (of 100 ns) and found the energy converged after 600 ns (**Figure 4.7**).

The PMF profile follows the same trend as before, where lipid A experiences high energy in solvent and thermodynamic relaxation through the hydrophobic groove of LptC. However, it is curious that the lipid does not experience ‘immediate’ relaxation into the groove, but rather a gradual sink in binding energy until the midpoint of the LptC β -taco i.e. a ‘V-shaped’ profile rather than a ‘U-shaped’ energy profile. This suggest that the lipid must gradually de-solvate first before reaching the centre point

of the domain before then gradually re-solvating as it moves out of the β -taco (**Figure 4.8**).

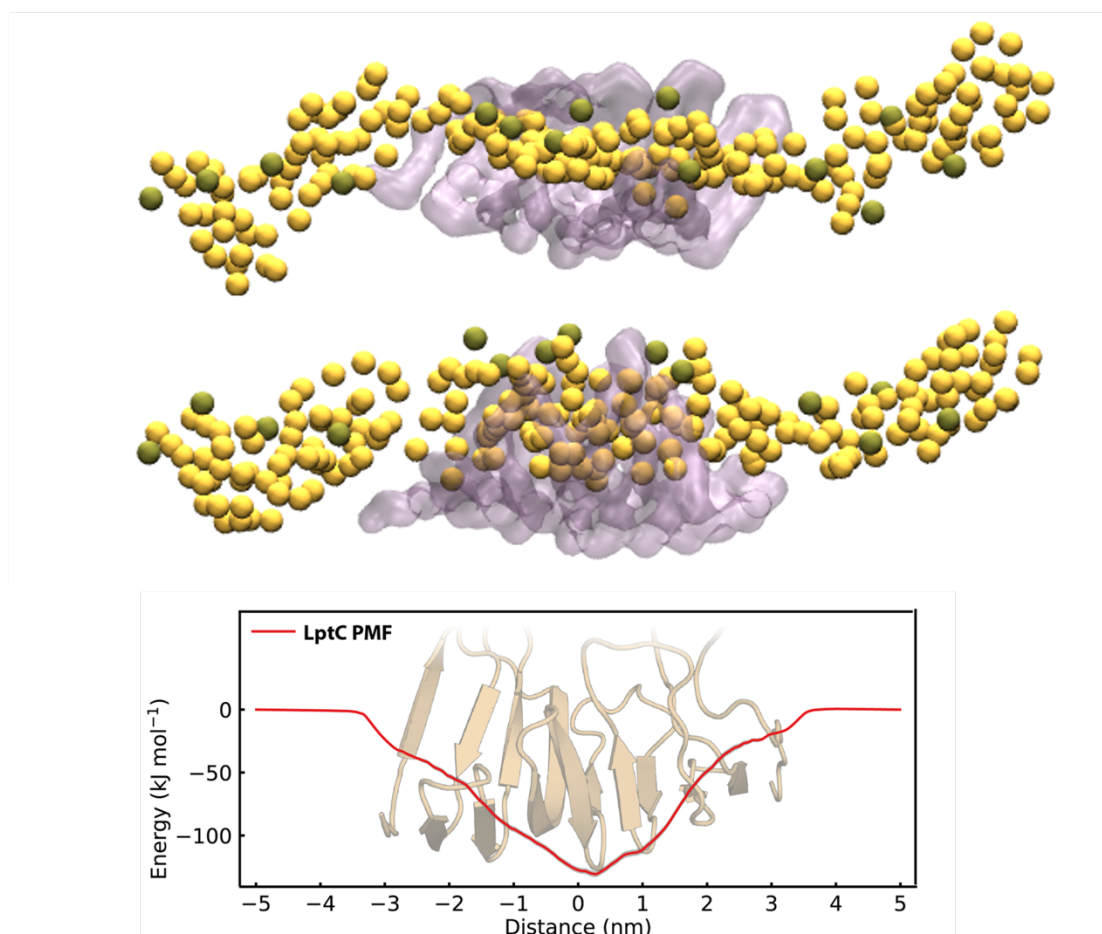


Figure 4.8: Snapshots of lipid A along the reaction coordinate. Top: looking down on the LptC β -taco. Middle: side view of the LptC β -taco as lipid A is pulled through the subunit. In both perspectives the β -taco is made transparent for clarity. Bottom: PMF profile overlayed for reference

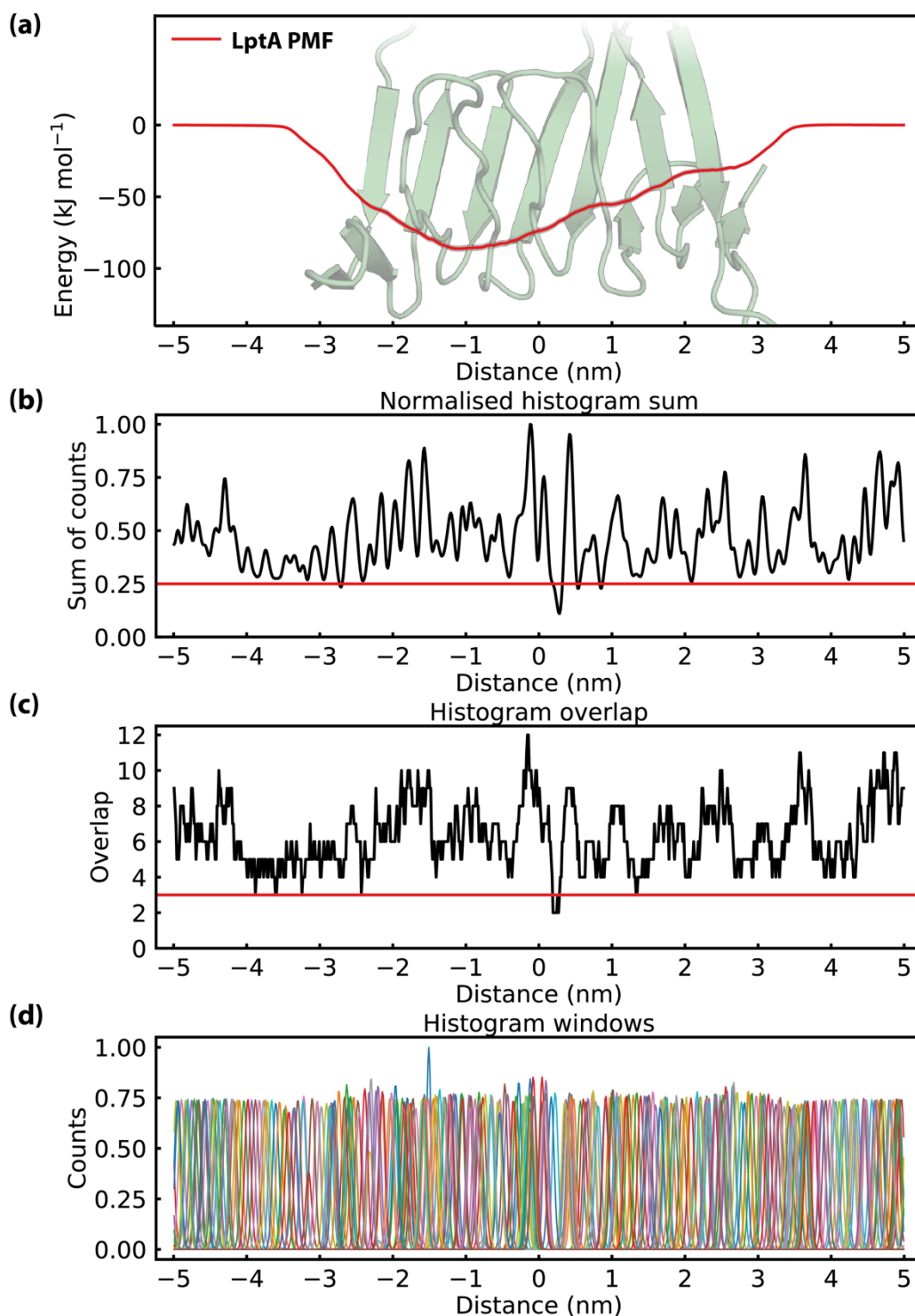
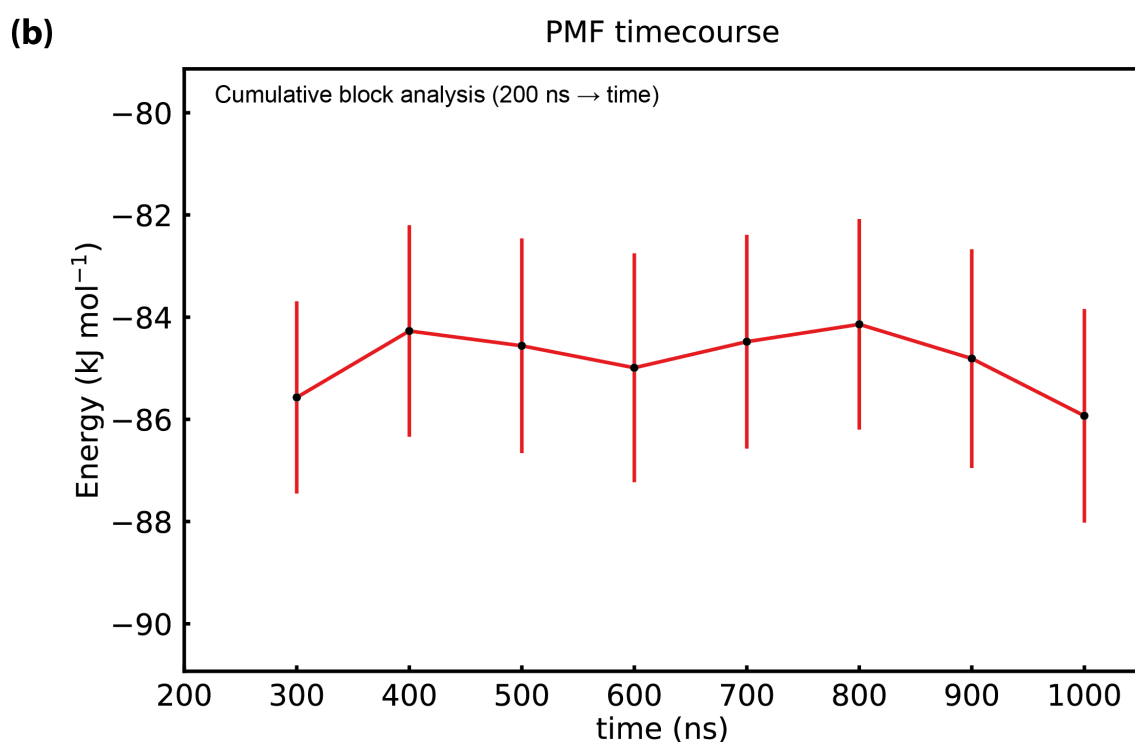
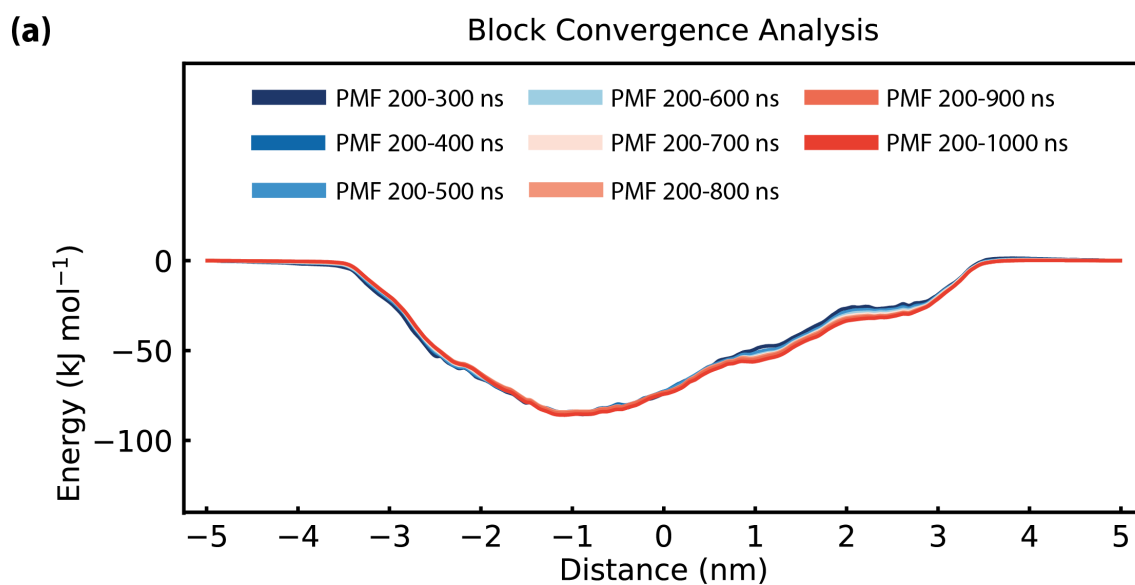


Figure 4.9 Binding energy of lipid A with the LptA β -taco. (a) PMF profile of lipid A moving through LptD, areas shaded in grey indicates the error and is obtained by bootstrapping the profile 200 times. Cartoon representation of the LptA subunit aligned with the energy profile in the background for reference. (b) Total counts summed across the collective variable of lipid A being pulled from the β -taco into solvent from the raw histograms, where the dataset is normalised to 1. The red line is a 'rule of thumb' marker for acceptable sampling positioned at 0.2 of total counts. (c) The number of overlapping histograms from the umbrella sampling across the collective variable. A count is considered when the overlap is $>10\%$ of the total histogram height.

The red line corresponds to a ‘rule of thumb’ for sufficient overlap set to 3. **(d)** The histograms from the WHAM analysis, where every histogram corresponds to a separate sampling position of lipid A.



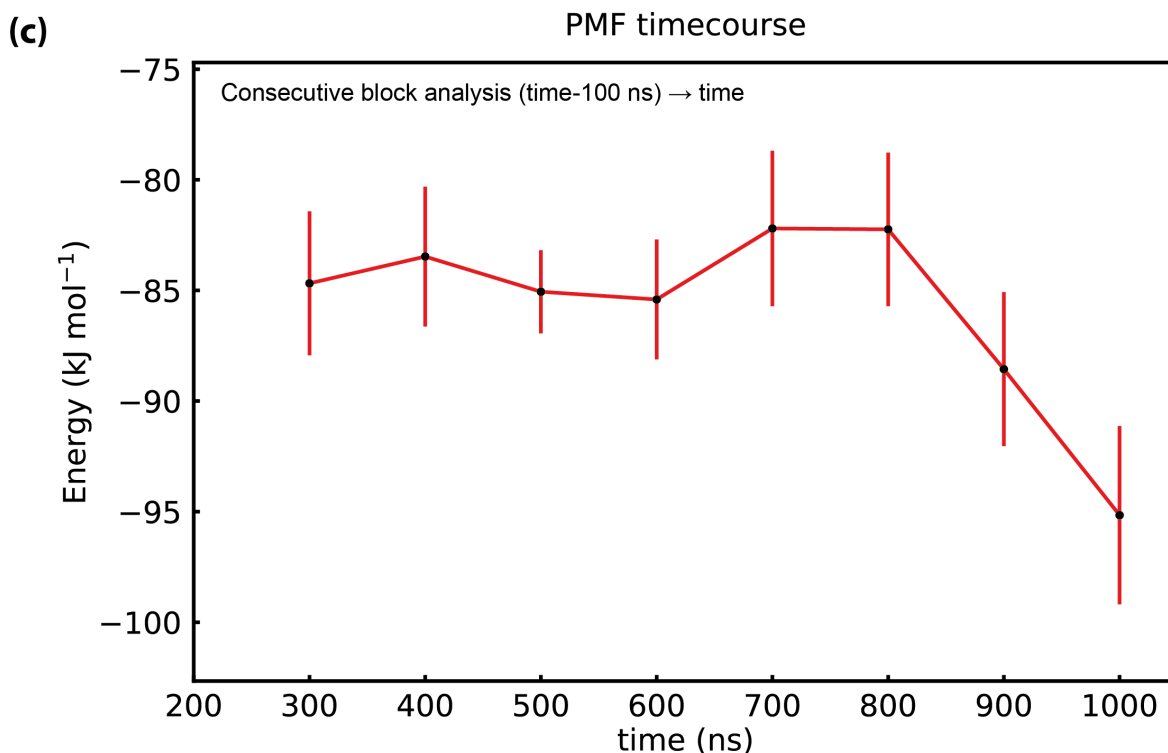


Figure 4.10 LptA PMF convergence analysis. (a) PMF profiles of increasing blocks simulation time by 100 ns for up to one microsecond. Colour scale from blue to red illustrates how the profile evolves through time. The first 200 ns was removed from each window to ensure that the system had fully equilibrated first. (b) The binding values (the lowest energy value) of each PMF profile over the course of the simulation time for increasing blocks of 100 ns. Errors calculated by bootstrapping each PMF profile 10 times. For instance, 200 – 300 ns, 200 – 400 ns, ..., 200 – 1000 ns. (c) The binding values (the lowest energy value) of each PMF profile over the course of the simulation time for consecutive blocks of 100 ns. For instance, 200 – 300 ns, 300 – 400 ns, ..., 900 – 1000 ns.

The main bridging unit of the periplasmic chute is LptA. In isolation, the calculated binding energy for LptA is $-86 \pm 2 \text{ kJ mol}^{-1}$, where 318 histograms were required for sufficient sampling (**Figure 4.9b, c, d**), where a converged plot was comfortably reached after 600 ns (**Figure 4.10**). Compared to the previous two PMF profiles, (LptF and LptC) the most favourable energy state of lipid A within the domain is 1 nm away from the centre of LptA. Therefore, the binding energy sampled by the lipid cannot ‘just’ be accounted for by slow de-solvation until the midpoint, followed by re-solvation.

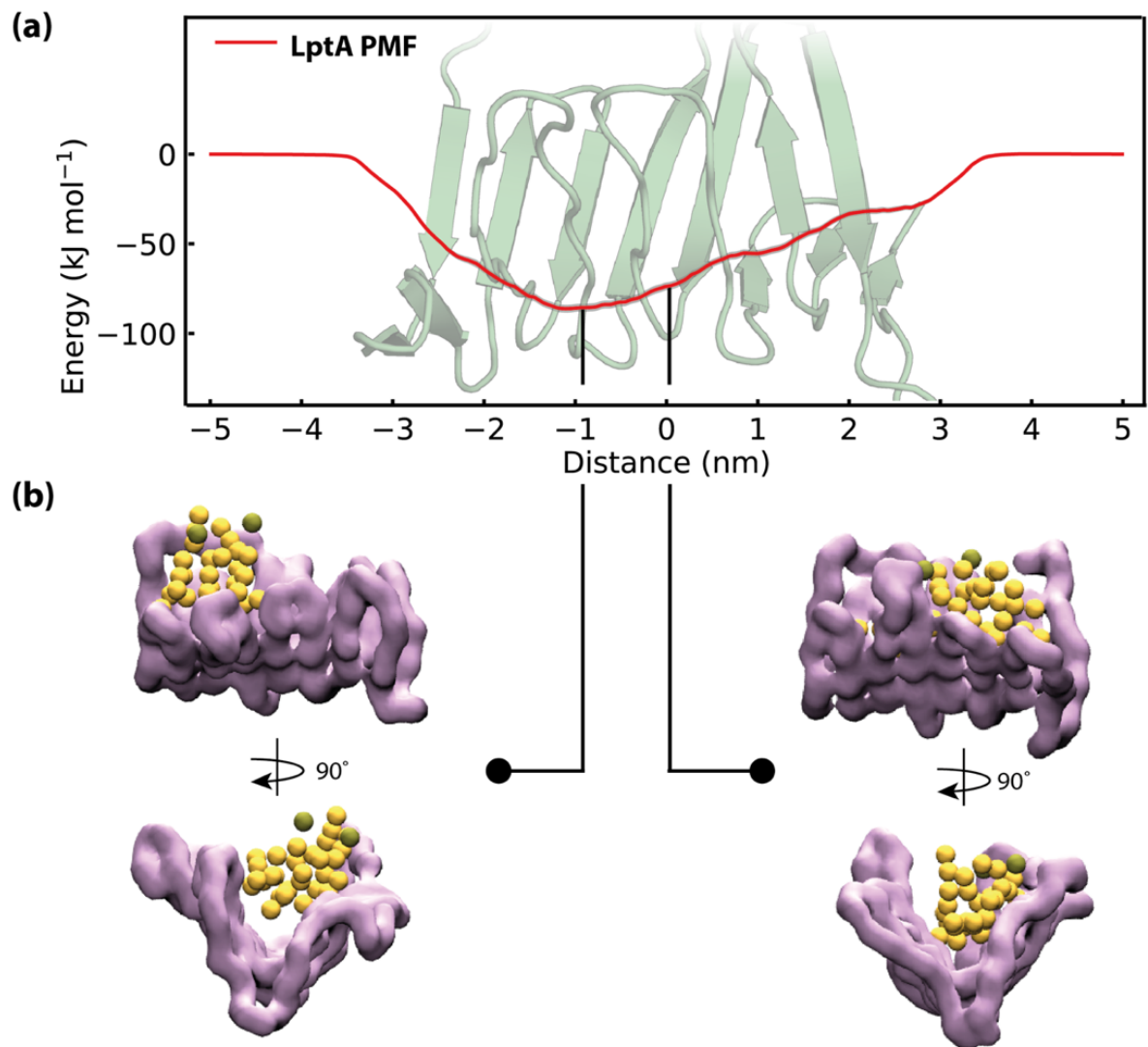


Figure 4.11 Lipid A binding conformation within the LptA β -taco. (a) PMF profile of lipid A moving through LptA, areas shaded in grey indicates the error and is obtained by bootstrapping the profile 200 times. Cartoon representation of the LptA subunit aligned with the energy profile in the background for reference. (b) Lipid A binding pose at -1 nm and at 0 nm labelled onto the LptA PMF binding profile.

The binding pose at -1 nm away from the centre of LptA reveals that all the lipid A tails are neatly aligned and tucked into the hydrophobic groove. At 0 nm, two of the lipid tails are slightly facing out of the β -taco which is clearly a more disfavoured lipid conformation within the domain (**Figure 4.11**). In comparison, the binding profile of lipid A through LptD is markedly different from the other β -taco domains (**Figure 4.12**).

To achieve this profile, 282 separate windows were required for sufficient sampling across the collective variable where convergence was reached after 600 ns (**Figure 4.13**). When considering the energy landscape of LptD, the COM of lipid A is -2 nm away from the centre of the subunit, the lipid sinks into a low energy state ($-87 \pm 5 \text{ kJ mol}^{-1}$) before rapidly approaching a highly unfavourable state as it moves across LptD, until it reaches another energy sink at +1 nm from the LptD centre. This high energy state of the lipid is caused by a salt bridge restriction (between residues Arg145 and Asp222) forcing the lipid tails out into solvent space (**Figure 4.14**).

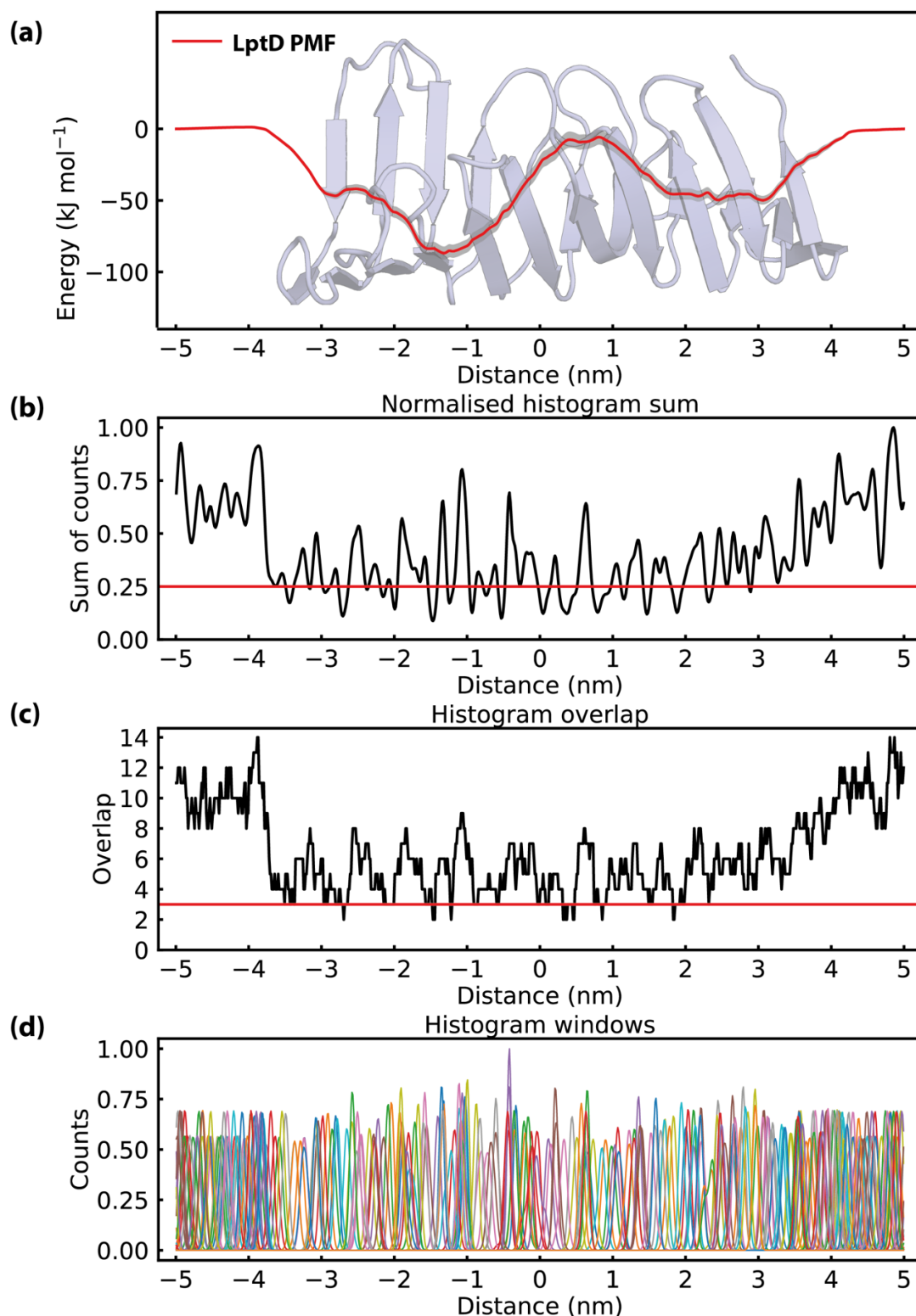
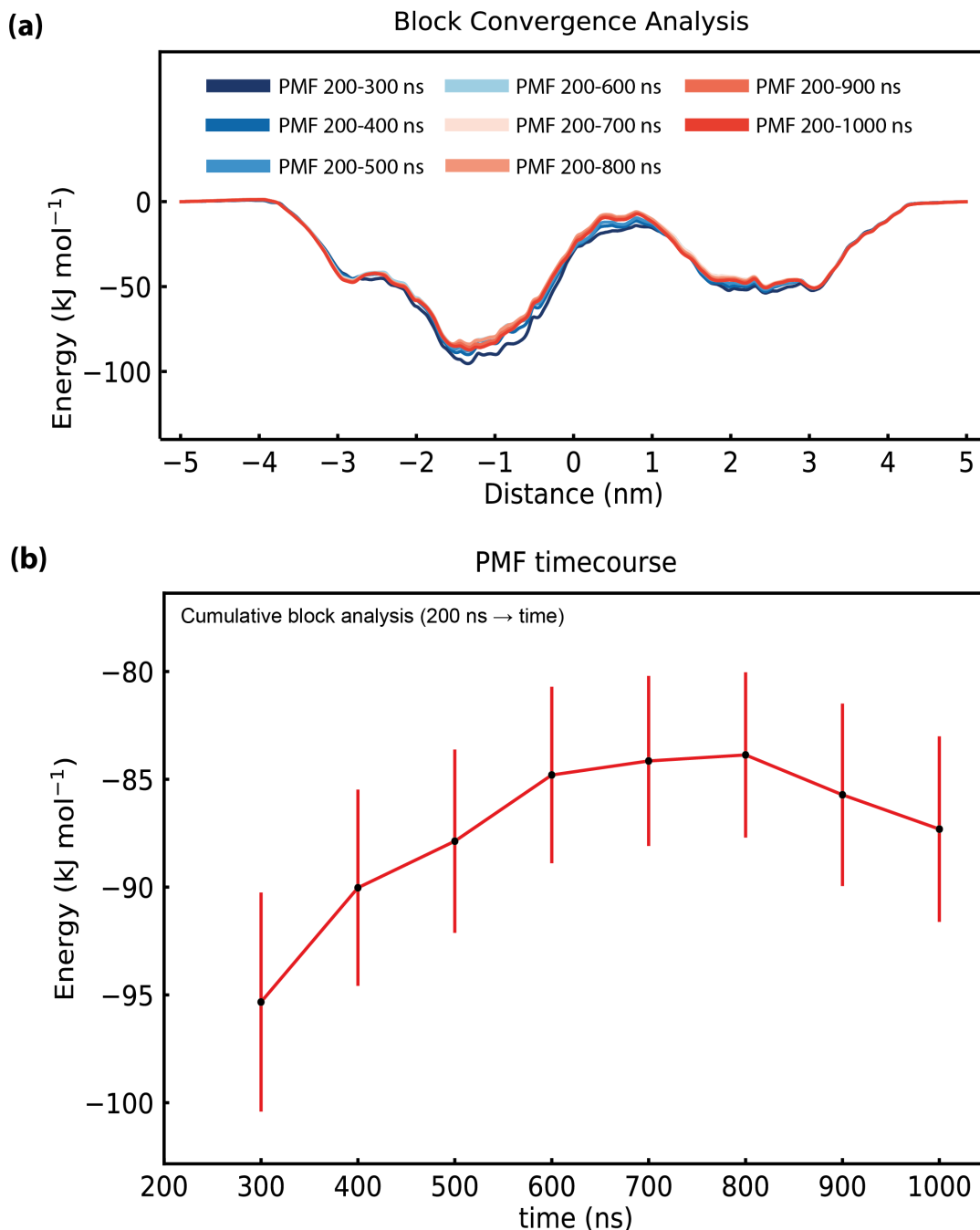


Figure 4.12 Binding energy of lipid A with the LptD β -taco. (a) PMF profile of lipid A moving through LptD, areas shaded in grey indicates the error and is obtained by bootstrapping the profile 200 times. Cartoon representation of the LptD subunit aligned with the energy profile in the background for reference. (b) Total counts summed across the collective variable of lipid A being pulled from the β -taco into solvent from the raw histograms, where the dataset is normalised to 1. The red line is a 'rule of thumb' marker for acceptable sampling positioned at 0.2 of total counts. (c) The

number of overlapping histograms from the umbrella sampling across the collective variable. A count is considered when the overlap is $>10\%$ of the total histogram height. The red line corresponds to a 'rule of thumb' for sufficient overlap set to 3. **(d)** The histograms from the WHAM analysis, where every histogram corresponds to a separate sampling position of lipid A.



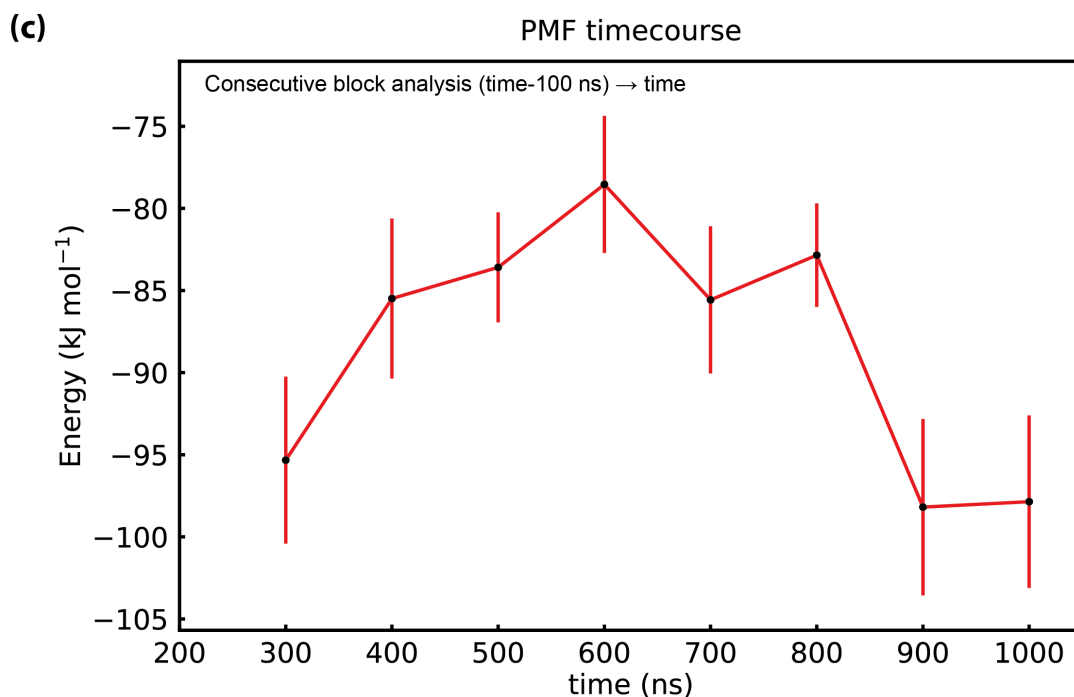


Figure 4.13 LptD PMF convergence analysis. (a) PMF profiles of increasing blocks simulation time by 100 ns for up to one microsecond. Colour scale from blue to red illustrates how the profile evolves through time. The first 200 ns was removed from each window to ensure that the system had fully equilibrated first. (b) The binding values (the lowest energy value) of each PMF profile over the course of the simulation time for increasing blocks of 100 ns. Errors calculated by bootstrapping each PMF profile 10 times. For instance, 200 – 300 ns, 200 – 400 ns, ..., 200 – 1000 ns. (c) The binding values (the lowest energy value) of each PMF profile over the course of the simulation time for consecutive blocks of 100 ns. For instance, 200 – 300 ns, 300 – 400 ns, ..., 900 – 1000 ns.

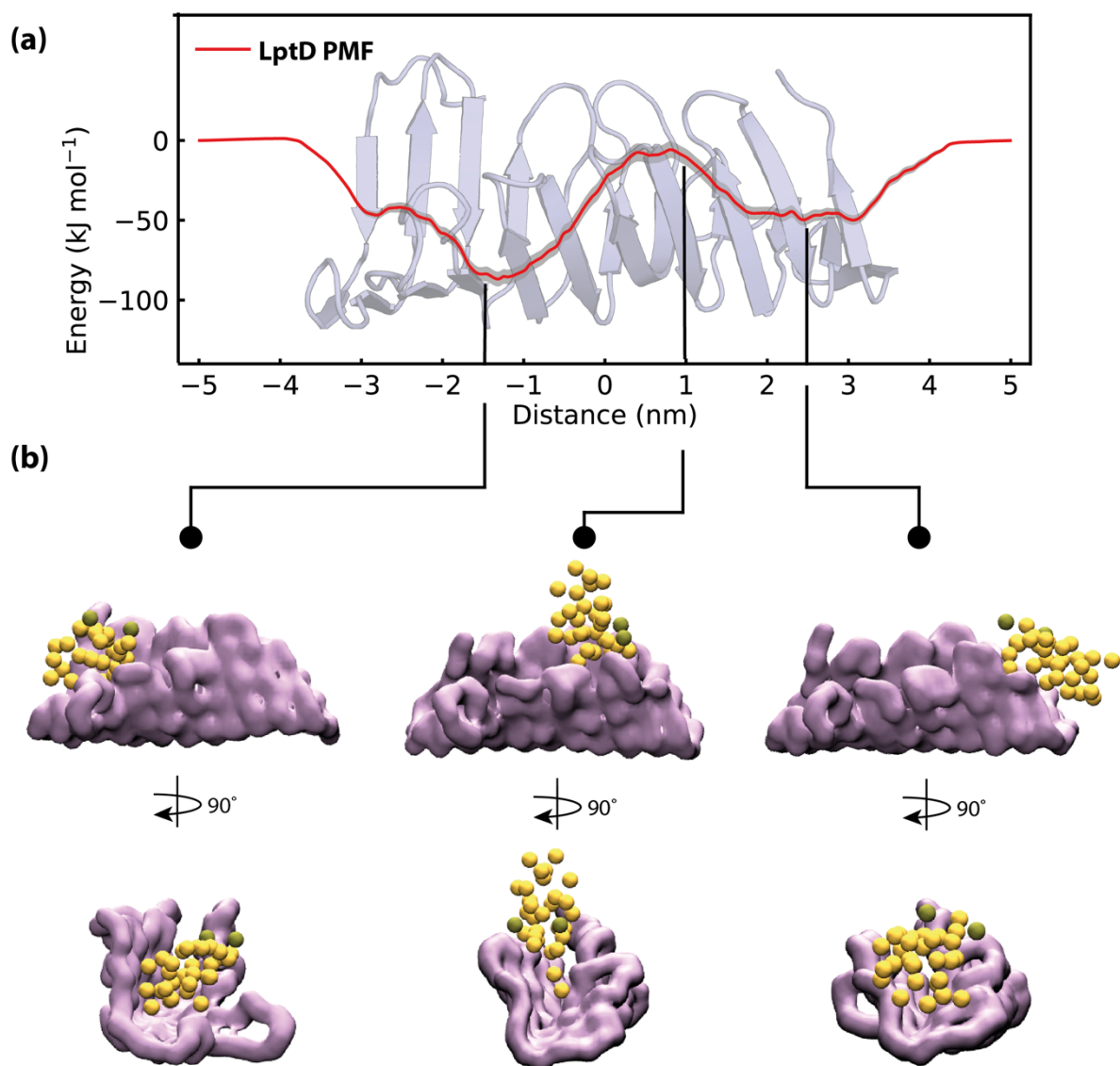


Figure 4.14 Lipid A binding conformation within the LptD β -taco. (a) PMF profile of lipid A moving through LptD, areas shaded in grey indicates the error and is obtained by bootstrapping the profile 200 times. Cartoon representation of the LptA subunit aligned with the energy profile in the background for reference. (b) Lipid A binding pose at -1.5 nm, +1 nm and at +2.5 nm labelled onto the LptA PMF binding profile.

Such a salt bridge may potentially act a gate or valve for OM LPS insertion, regardless, it is certainly clear that the chute must be open at least 2 nm between the β -strands of the β -taco for facile lipid transport (**Figure 4.15**). Therefore, the opening of the hydrophobic groove and the breaking of salt bridges is essential for LPS transport across the periplasm.

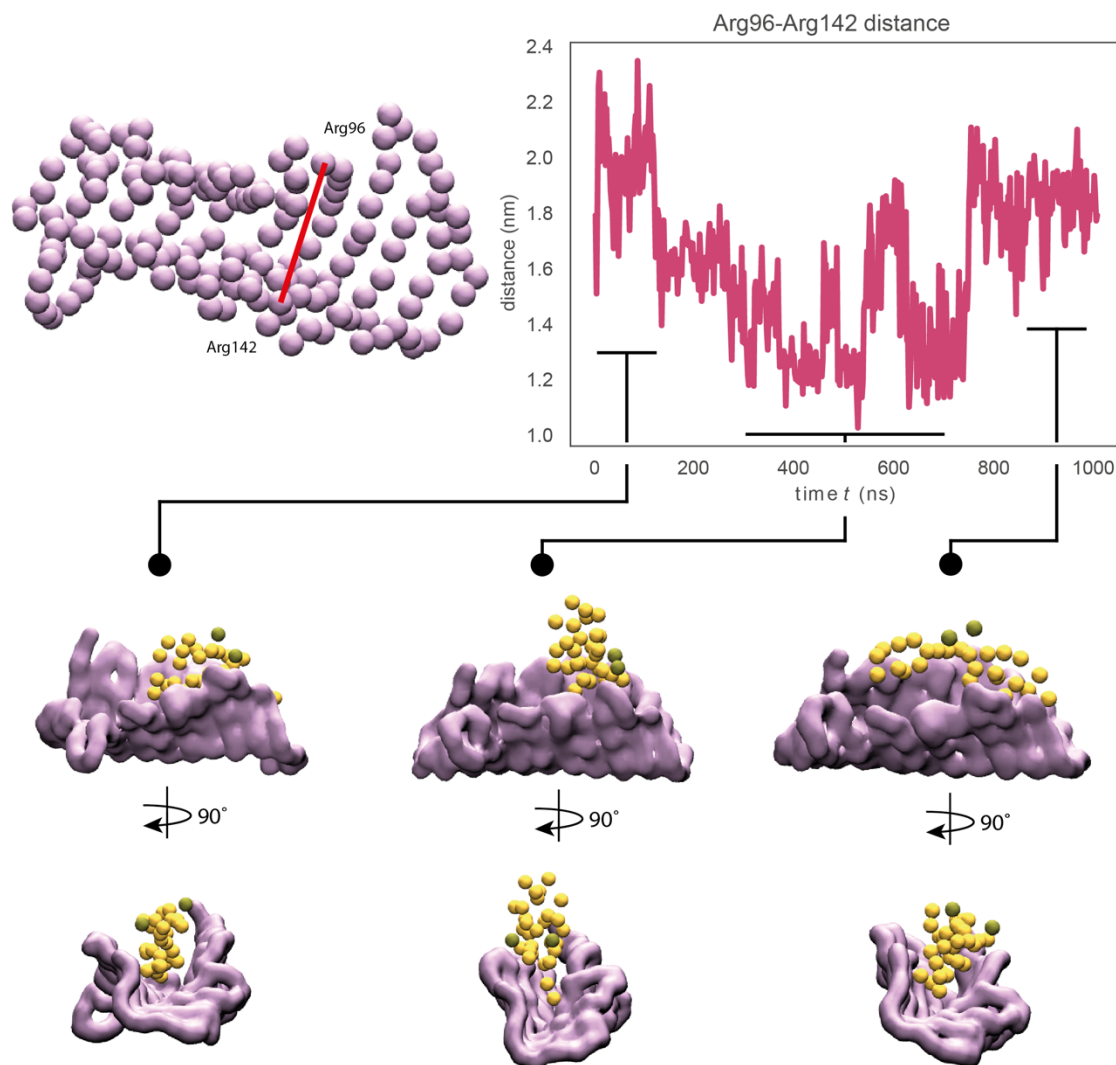


Figure 4.15 Width of LptD subunit during the umbrella sampling. Top: Distance over the β -taco measured between Arg96 and Arg142 over the 1 μ s sampling time for the window sampling at +1 nm. Bottom: representative snapshots of lipid A sampling of the 0-100 ns, 200-700 ns and 900-1000 ns sections of the simulation.

During the sampling window at +1 nm, I observe the salt bridge attempting to reform. This implies that more lipid A molecules must be present within the units to fully open the domains. LptD is ~ 5.1 nm in length which is large enough to accommodate a second lipid A molecule within its hydrophobic groove. If this was present, it would be likely that the salt bridge would not reform and therefore not push out the lipid A molecule observed in this instance.

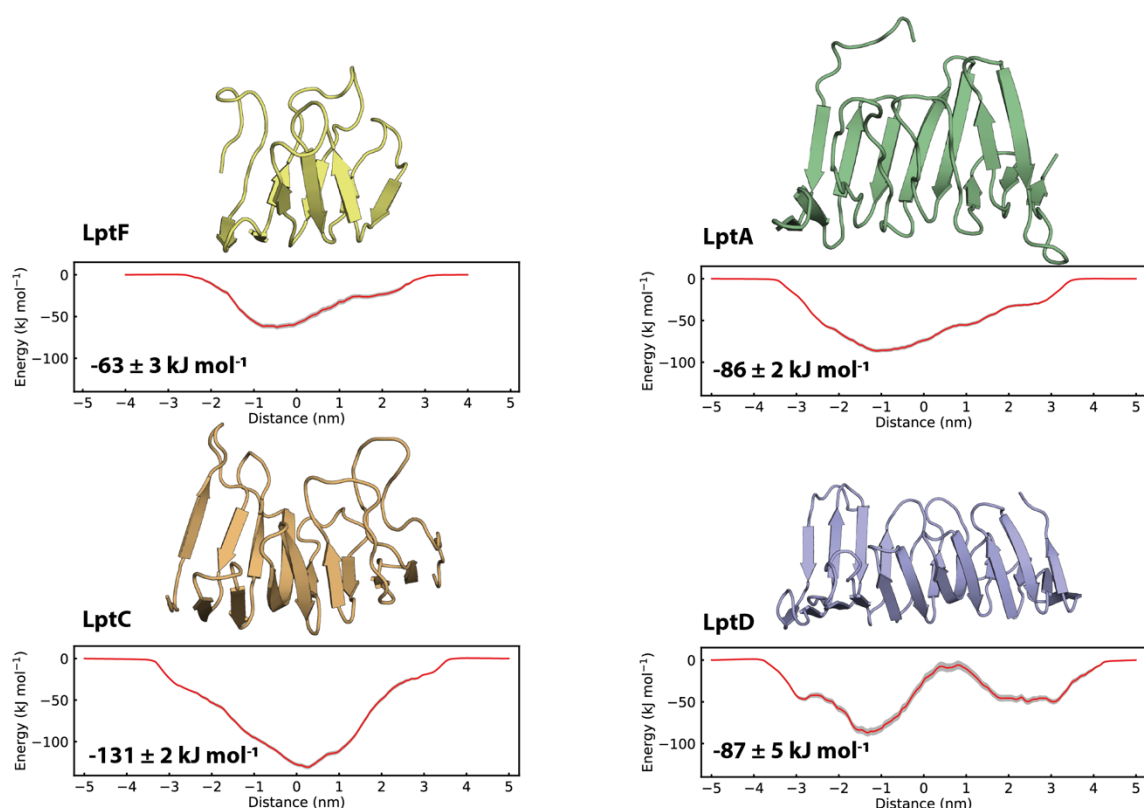


Figure 4.16 PMF profiles for the bridging subunits. PMF calculations from the lateral ‘card swiping’ through each of the β -taco units (a) LptF, (b) LptC, (c) LptA and (d) LptD. Areas shaded in grey indicates the error and is obtained by bootstrapping each profile respectively.

Between these lateral PMF pulls, I can begin to piece together what LPS may experience on transit through the periplasm (**Figure 4.16**). These profiles reveal a variable energy series: whereby lipid A experiences the highest energy within LptF, followed by a large energy sink into LptC, before entering a similar energetic environment within LptA and LptD.

However, there are several issues with this approach to understanding the binding energy of lipid A with each of the β -taco units. The most glaring of which is the variable amount of solvation lipid A experiences with each β -taco domain. For LptF, lipid A is barely housed inside the protein cleft and the lipid tails are free to contact the

surrounding solvent. However, for larger units like LptC, it is unclear whether the binding minimum across the β -taco is due to more favourable contact with the residues at that position or whether it is the point at which the lipid experiences the least solvation. For comparatively sized β -taco units such as LptC and LptA (length ~5.2 nm with 8 number of β -strands) it makes sense to compare binding energies in isolation. What I notice here is that LptC has a stronger binding interaction than with LptA, which lends itself to enabling a large thermodynamic sink for LPS to be drawn towards.

Another issue with this method, is the extent to which the β -taco domains are open, the LptD domain would require further opening for lipid A to be threaded through smoothly. Incidentally, however, this does demonstrate the importance of salt bridges across the β -taco and reveals, quite nicely, the minimum distance across the domain for lipid A transport, which is defined as 2 nm.

4.4 Binding energy of lipid A across multiple LptA subunits

To gain better insight into the thermodynamics of lipid A transport through the periplasm, I want to understand how the energy profile would change when the lipid has a chance to sample a larger section of the chute. The lipid here would experience less solvation effects when travelling through a bigger portion of the chute and should help reveal the extent to which solvation effects are influencing the PMF calculations. Additionally, I would also gain insight in how neighbouring β -taco units affect the binding energy of the lipid with the chute.

The number of LptA units which make up the bridge is not truly known³¹⁷, however in measuring the distance across the periplasm (~25 nm)³¹⁸ an educated guess of 3 or 4 units would be required, where a single LptA unit spans 3.5 nm. Thus, for this section three bridging units of LptA were assembled to form an LptA trimer. The hydrophobic groove of the LptA trimer was opened in a similar fashion to the monomers where lipid A was grown inside the units to open the complex in a uniform manner. Like the previous lateral pulls, lipid A was pulled from the centre of the LptA₃ out into solvent in both the forward (towards the OM) and reverse (towards the IM) direction.

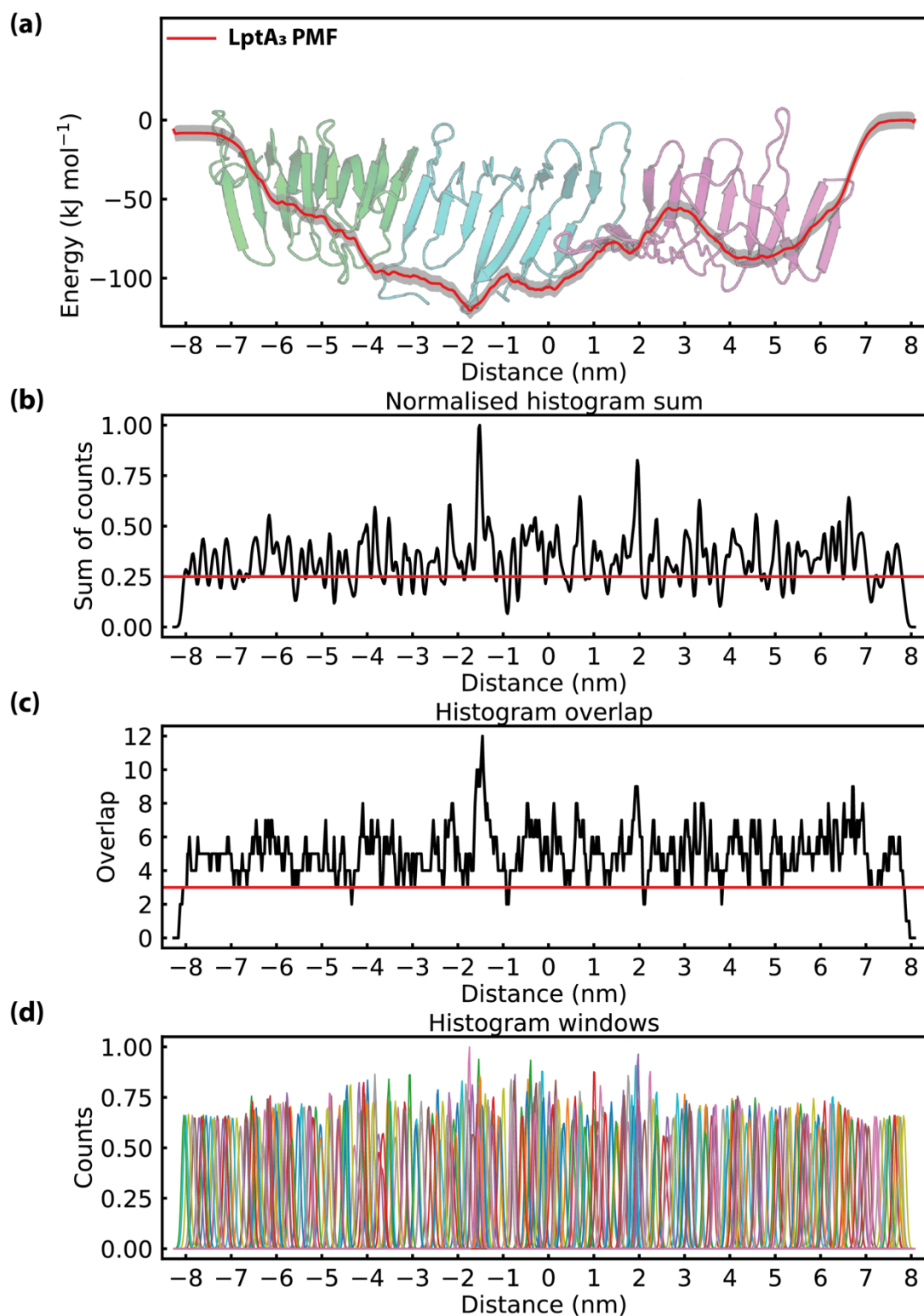
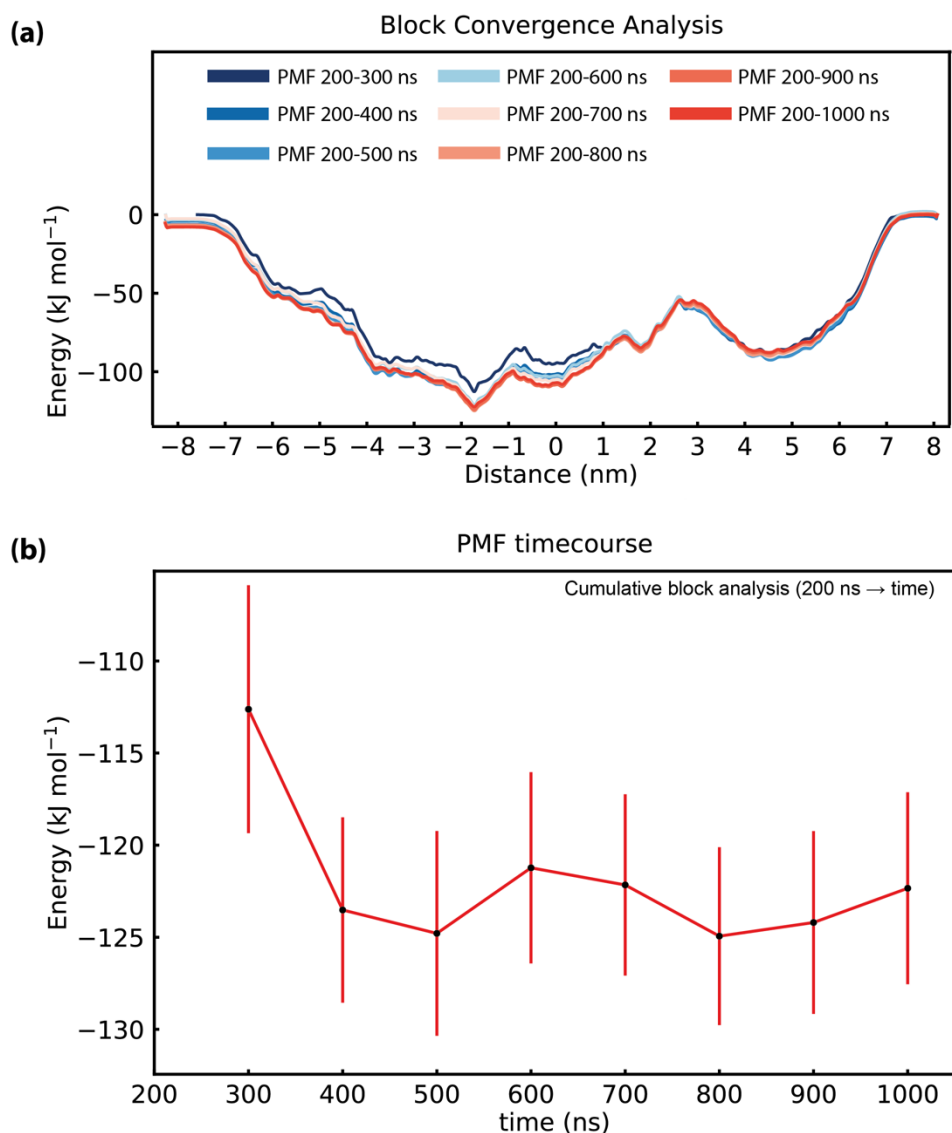


Figure 4.17 Binding energy of lipid A with the LptA₃ portion of the chute. (a) PMF profile of lipid A moving through LptA₃, areas shaded in grey indicates the error and is obtained by bootstrapping the profile 200 times. Cartoon representation of the LptA₃ trimer aligned with the energy profile in the background for reference. (b) Total counts summed across the collective variable of lipid A being pulled from the β -taco into solvent from the raw histograms, where the dataset is normalised to 1. The red line is

a 'rule of thumb' marker for acceptable sampling positioned at 0.2 of total counts. **(c)** The number of overlapping histograms from the umbrella sampling across the collective variable. A count is considered when the overlap is >10% of the total histogram height. The red line corresponds to a 'rule of thumb' for sufficient overlap set to 3. **(d)** The histograms from the WHAM analysis, where every histogram corresponds to a separate sampling position of lipid A from the lateral 'card swiping' through the LptA₃ trimer.



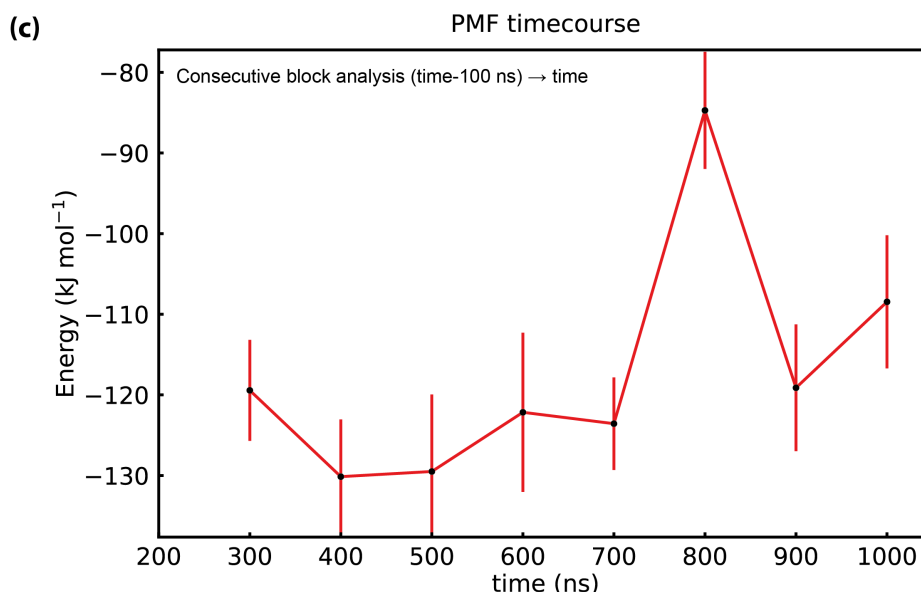


Figure 4.18 LptA₃ PMF convergence analysis. (a) PMF profiles of increasing blocks simulation time by 100 ns for up to one microsecond. Colour scale from blue to red illustrates how the profile evolves through time. The first 200 ns was removed from each window to ensure that the system had fully equilibrated first. (b) The binding values (the lowest energy value) of each PMF profile over the course of the simulation time for increasing blocks of 100 ns. Errors calculated by bootstrapping each PMF profile 10 times. For instance, 200 – 300 ns, 200 – 400 ns, ..., 200 – 1000 ns. (c) The binding values (the lowest energy value) of each PMF profile over the course of the simulation time for consecutive blocks of 100 ns. For instance, 200 – 300 ns, 300 – 400 ns, ..., 900 – 1000 ns.

I obtained this profile (**Figure 4.17**) by sampling across the collective variable 337 times where convergence was reached after 600 ns (**Figure 4.18**). The result of the LptA₃ PMF profile is surprising, despite three of the same LptA units being coordinated together, the landscape is asymmetrical. Curiously, the landscape features an energy sink at -1.8 nm, before encountering an energy barrier at +2.7 nm before another energy sink at +4.7 nm. At the points where lipid A experiences a particularly favourable energy minima (at -1.8 nm and at +4.7 nm) both phosphate headgroups directly coordinates with soluble residues from the loops of the β -strands (Asn24/Ser141 and Thr3/Thr6 respectively) (**Figure 4.19**). At +2.7 nm, only one phosphate group is involved in a favourable interaction with the LptA soluble loops and is therefore higher in energy. Lipid A binds to the same monomer unit in different

binding poses, which is dependent on whether the phosphate headgroups have found favourable interactions with the soluble loops of the bridging units and whether the lipid tails are comfortably inside the hydrophobic groove (and not squeezed out). As before, each end of the chute increases in energy (as seen in the monomer lateral PMF calculations) due to the lipid experiencing more unfavourable solvation interactions.

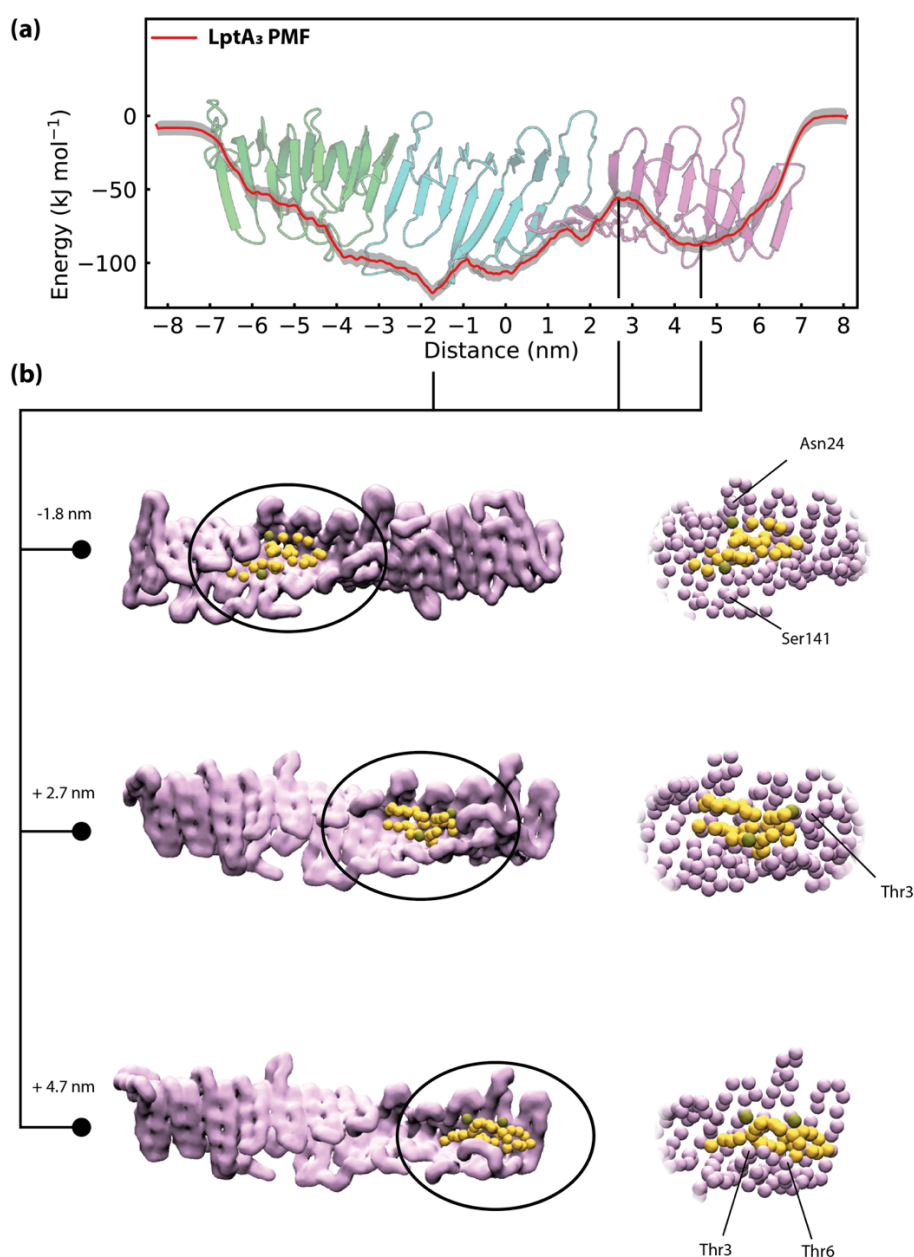


Figure 4.19 Lipid A binding conformation within the LptA₃ trimer. (a) PMF profile of lipid A moving through LptA₃, areas shaded in grey indicates the error and is obtained by bootstrapping the profile 200 times. Cartoon representation of the LptA subunit aligned with the energy profile in the background for reference. (b) Lipid A binding pose at -1.8 nm, +2.7 nm and at +4.7 nm are labelled onto the LptA₃ PMF binding profile, where in each instance, the binding pose has been highlighted. The residues which interact with the phosphate groups are highlighted on the structure.

The binding energy of lipid A through LptA₃ is more favourable than the binding energy with just one LptA unit: $-120 \pm 5 \text{ kJ mol}^{-1}$ versus $-86 \pm 2 \text{ kJ mol}^{-1}$. To account for this, I observe that the hydrophobic groove of the central LptA unit of the trimer fits more tightly with lipid A (approximately 25 kJ mol^{-1} more favourably) due to the structural rigidity provided by the presence of the flanking LptA units. Furthermore, the complex formed by lipid A and one LptA β -taco is more highly solvated than when the lipid is bound in the centre of the LptA₃ trimer.

4.5 Binding energy of lipid A across the periplasmic chute LptA₃CDF

So far, laterally pulling lipid A through sections of the periplasmic chute has allowed me to understand how open the hydrophobic groove is required to be to allow lipid transport (at least 2 nm). Between comparably sized β -taco units, LptC has demonstrated far greater binding energy than LptA ($-131 \pm 2 \text{ kJ mol}^{-1}$ versus $-86 \pm 2 \text{ kJ mol}^{-1}$), where the binding strength between LptA and the lipid is enhanced in the presence of neighbouring β -taco units which provide structural rigidity. The strength of the binding interaction also depends on whether the lipid tails sit comfortably (where the tails align in a row) within the hydrophobic groove and whether the phosphate headgroups are interacting with the soluble loops of the bridging units.

At this stage, it is important to demonstrate how all these factors work together by deriving the thermodynamic landscape across the entire periplasmic chute itself. As with the preceding lateral pulls, lipid A was pulled from the centre of the LptA₃CDF out into solvent in both the forward (towards the OM) and reverse (towards the IM) direction.

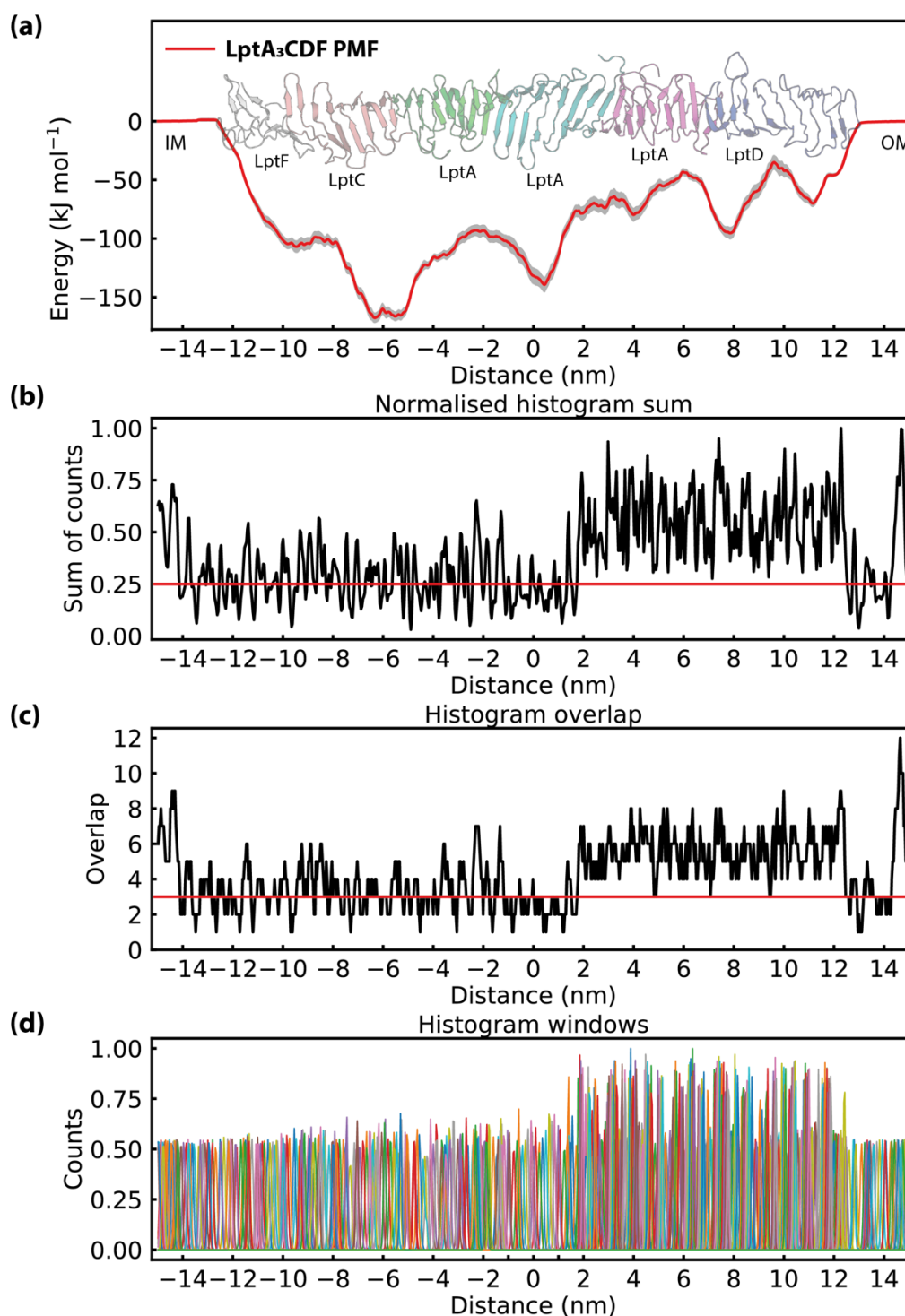


Figure 4.20 Binding energy of lipid A with the LptA₃CDF periplasmic chute. (a) PMF profile of lipid A moving through LptA₃CDF, areas shaded in grey indicates the error and is obtained by bootstrapping the profile 200 times. Cartoon representation of the LptA₃CDF complex it aligned with the energy profile in the background for reference. **(b)** Total counts summed across the collective variable of lipid A being pulled from the β -taco into solvent from the raw histograms, where the dataset is normalised to 1. The red line is a 'rule of thumb' marker for acceptable sampling positioned at 0.2 of total counts. **(c)** The number of overlapping histograms from the umbrella sampling across the collective variable. A count is considered when the overlap is >10% of the total histogram height. The red line corresponds to a 'rule of thumb' for sufficient overlap set to 3. **(d)** The histograms from the WHAM analysis, where every histogram corresponds to a separate sampling position of lipid A.

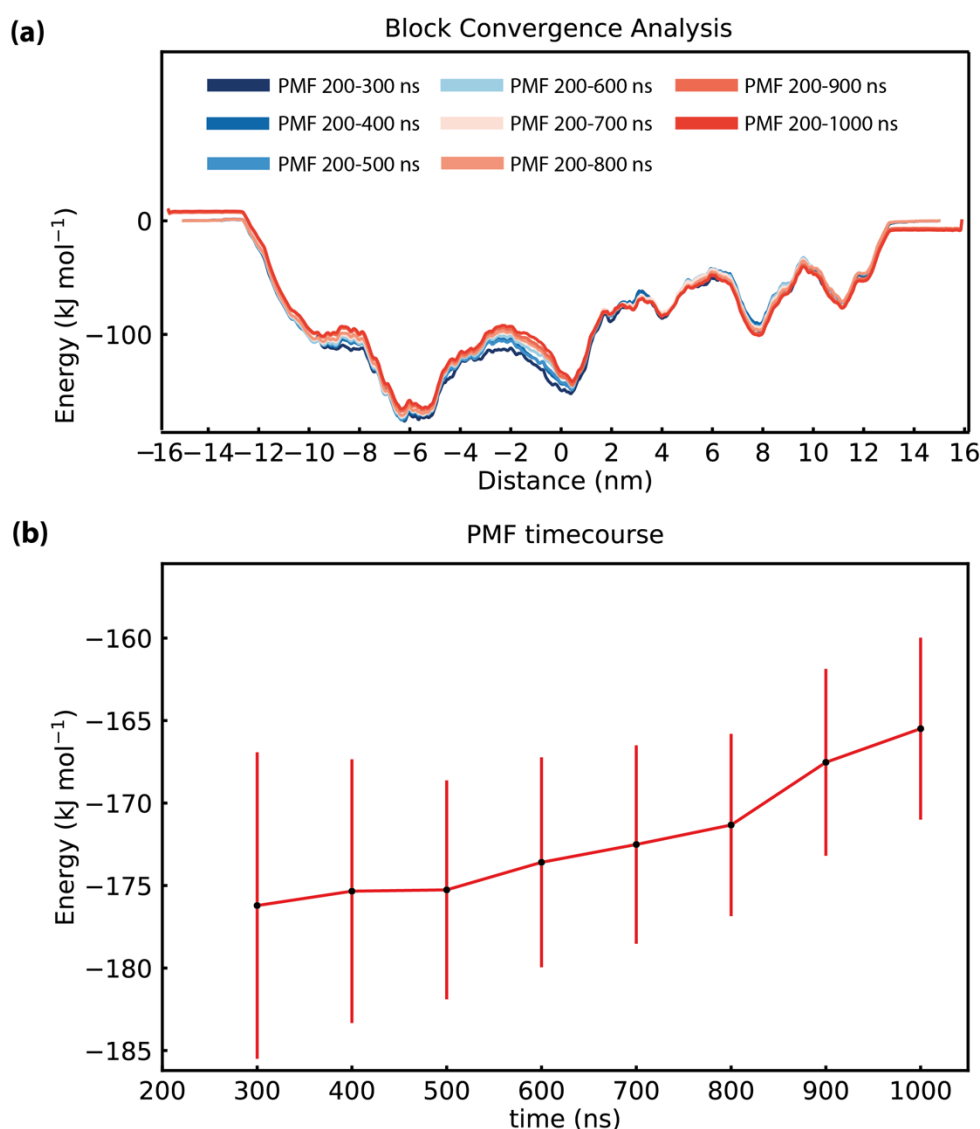


Figure 4.21 LptA₃CDF PMF convergence analysis. (a) PMF profiles of increasing blocks simulation time by 100 ns for up to one microsecond. Colour scale from blue to red illustrates how the profile evolves through time. The first 200 ns was removed from each window to ensure that the system had fully equilibrated first. **(b)** The binding values (the lowest energy value) of each PMF profile over the course of the simulation

time for increasing blocks of 100 ns. Errors calculated by bootstrapping each PMF profile 10 times. For instance, 200 – 300 ns, 200 – 400 ns, ..., 200 – 1000 ns. **(c)** The binding values (the lowest energy value) of each PMF profile over the course of the simulation time for consecutive blocks of 100 ns. For instance, 200 – 300 ns, 300 – 400 ns, ..., 900 – 1000 ns.

Here, LptA₃CDF PMF profile (**Figure 4.20**) was obtained by 573 separate windows to sample along the chute where convergence appears to have been reached after 800 ns of simulation time (**Figure 4.21**). Lipid A experiences a huge energy sink (-165 ± 6 kJ mol⁻¹) inside LptC when moving towards the LptA trimer, which certainly helps to solidify my observations of the more favourable binding energy between lipid A with LptC than with LptA in isolation. Here also, the landscape is not 'V-shaped' where the position at 0 nm is the most favourable (due to being most distal from the solvated end of the chute) but instead shows a very sharp relaxation into the centre of the LptC subunit, which again, further emphasises this point. Notably, the landscape inside the LptA section is different from the previously shown LptA trimer (from -5 nm to +7 nm). It appears, then, that the additions of the subunits LptD and LptC (on either side of LptA₃) has affected the conformation of the LptA trimer and accordingly, the thermodynamics of what lipid A experiences. It is also important to note the simplification that this LptA₃CDF model is constructed in isolation, without the IM, OM or peptidoglycan layer. However, if the peptidoglycan layer was present (and assuming that it was situated the middle of the periplasm, 7-8 nm thick⁴⁸) it would span from -4 nm to +4 nm on the scale in **Figure 4.22**.

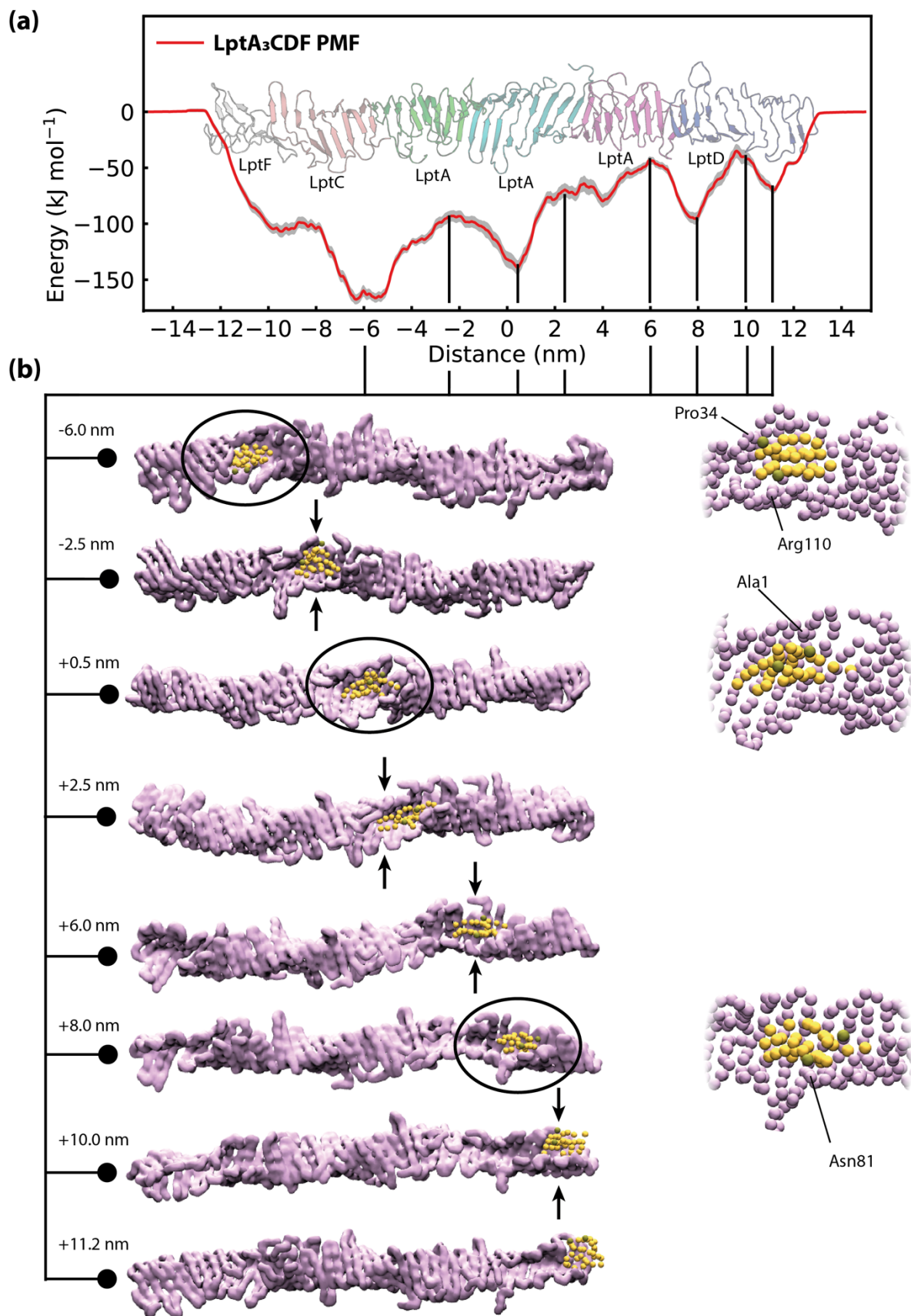


Figure 4.22 Lipid A binding conformation within the LptA₃CDF complex. (a) PMF profile of lipid A moving through LptA₃CDF, areas shaded in grey indicates the error and is obtained by bootstrapping the profile 10 times. Cartoon representation of the LptA subunit aligned with the energy profile in the background for reference. **(b)** Lipid A binding pose at -6.0 nm, -2.5 nm, +0.5 nm, +2.5 nm, +6.0 nm, +8.0 nm, +10.0 nm and at +11.2 nm labelled onto the LptA₃CDF PMF binding profile, where for positions -6.0 nm, 0.5 nm and +8 nm the binding pose has been highlighted. The residues which interact with the phosphate groups are labelled onto the structure.

After lipid A moves from solvent to the chute itself, the lipid instantly relaxes into the hydrophobic groove of LptF and LptC. At -6 nm both phosphate head groups of lipid A interact favourably with the beta-strand loops of the domain (Pro34, Arg110), which emphasise the driving force the lipid experiences to move into the periplasmic bridge for transport (**Figure 4.22**). The energy barrier at -2.5 nm is caused by the lipid tails being squeezed out into solvent due to a slight constriction in the chute, however phosphate groove interactions relax the energy the lipid experiences once more at +0.5 nm (interacting with residue Ala1). The energy sharply rises again at +2.5 nm to +6 nm (to -50 kJ mol⁻¹) due to the disruption of the phosphate group interaction and because of another chute constriction causing the tails of the lipid to rise out of a comfortable interaction within the hydrophobic groove.

Lipid A experiences another energy sink at +8 nm when it moves into the final subunit of the periplasmic bridge, LptD because of phosphate group coordination with residue Asn81. The landscape of the LptD portion of the chute is very similar to the lateral card swipe of LptD in isolation. This again, can be accounted for by a restriction in the chute which forces the lipid tails out to sample more solvent than otherwise would physiologically. However, the energies of the minima are more favourable than in the individual card swipe which is owing to the addition of LptA coordinated with LptD and opening the domain. Adding subunits to the monomers changes the conformation of

the chute allowing for more favourable interactions with lipid A, which is especially acute when comparing the profile for the LptA₃ energy landscape against the PMF calculation for LptA₃CDF.

To understand the nature of these changes, I wanted to see how the landscape of lipid A transport would change by complexing a larger group of LptA units together. To study this, the LptA₄ complex, obtained from the unit cell of the crystal structure (PDB: 2R19)³⁰¹ was opened as before, by growing lipid A inside the hydrophobic groove.

4.6 The binding energy of lipid A through LptA₄

Lipid A was laterally pushed through the LptA tetramer using steered MD to better understand how the landscape may change with another neighbouring subunit.

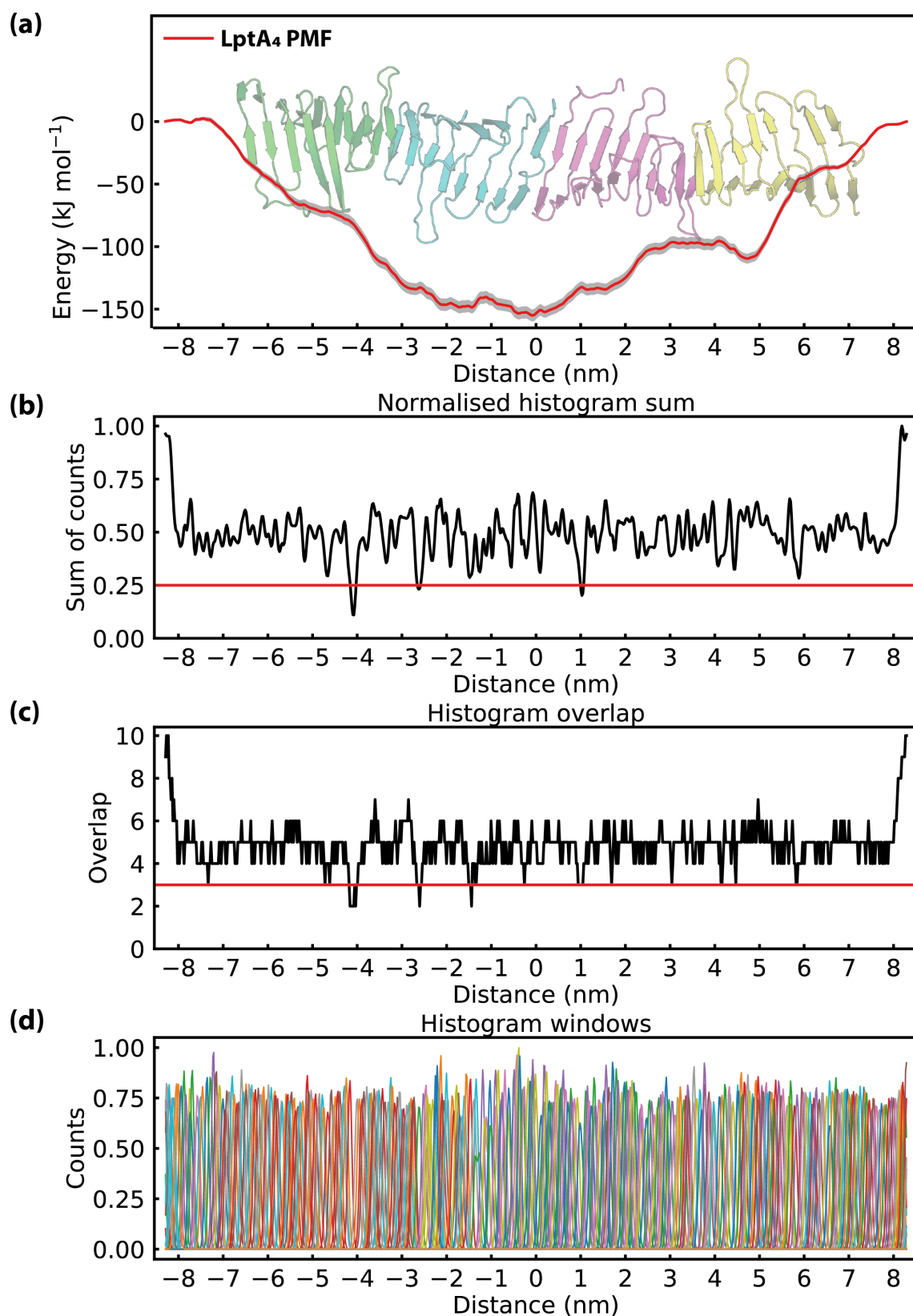
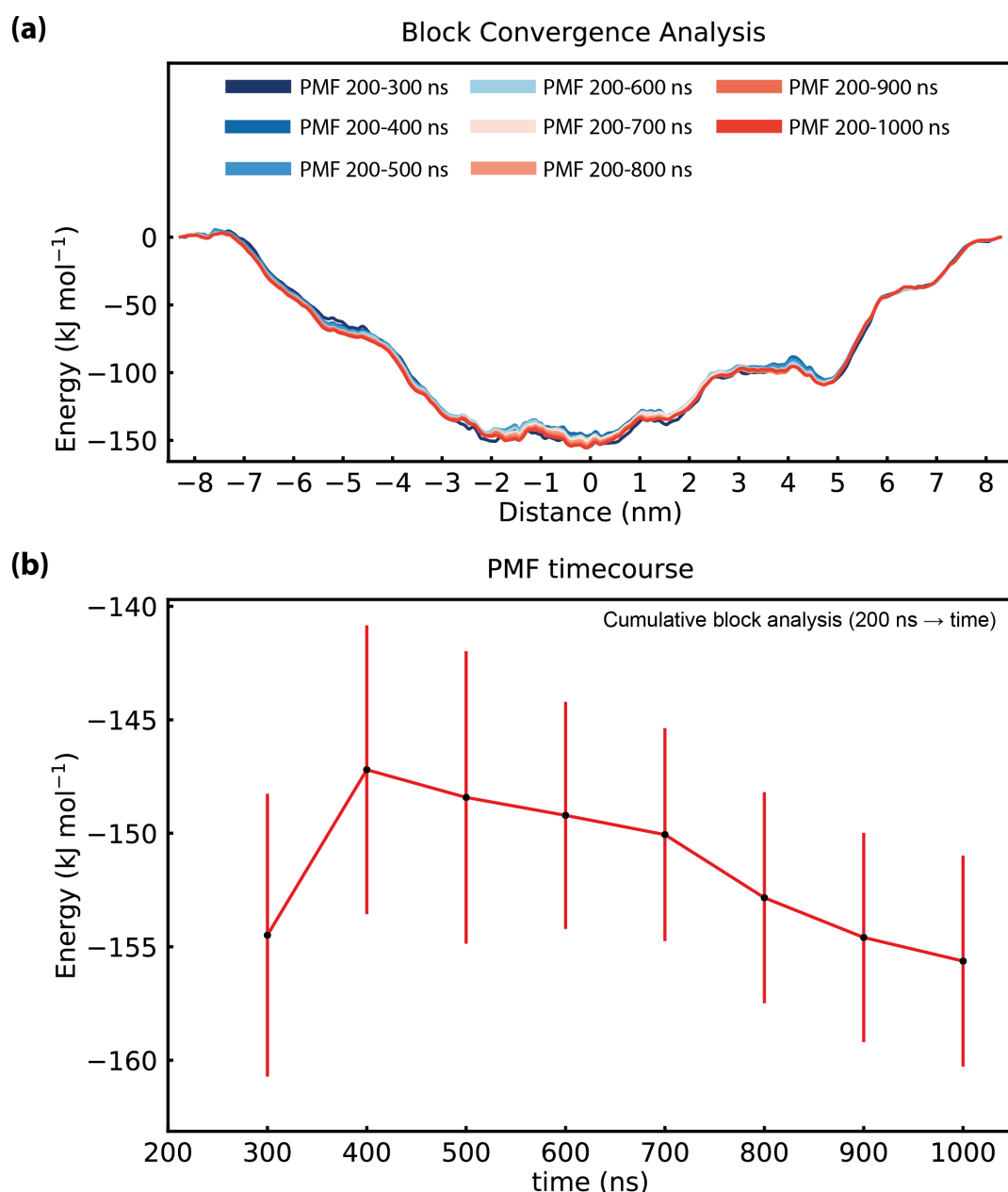


Figure 4.23 binding energy of lipid A with the LptA₄ portion of the chute. (a) PMF profile of lipid A moving through LptA₄, areas shaded in grey indicates the error and is obtained by bootstrapping the profile 200 times. Cartoon representation of the LptA₄ tetramer complex it aligned with the energy profile in the background for reference.

(b) Total counts summed across the collective variable of lipid A being pulled from the β -taco into solvent from the raw histograms, where the dataset is normalised to 1. The red line is a 'rule of thumb' marker for acceptable sampling positioned at 0.2 of total counts. **(c)** The number of overlapping histograms from the umbrella sampling across the collective variable. A count is considered when the overlap is $>10\%$ of the total histogram height. The red line corresponds to a 'rule of thumb' for sufficient overlap set to 3. **(d)** The histograms from the WHAM analysis, where every histogram corresponds to a separate sampling position of lipid A.



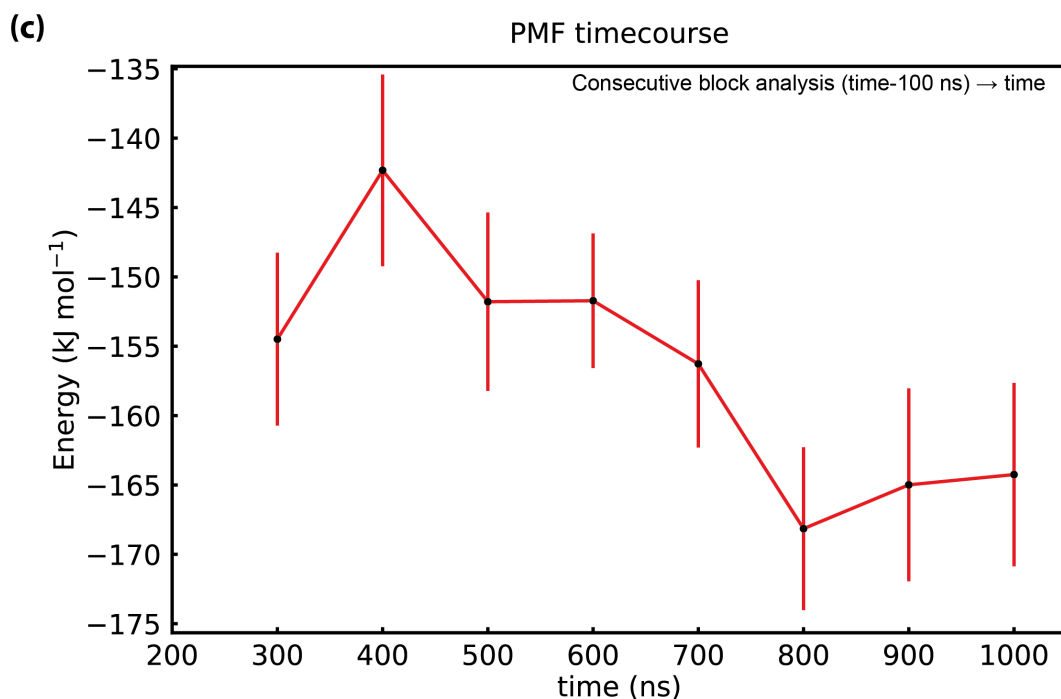


Figure 4.24 LptA₄ PMF convergence analysis. (a) PMF profiles of increasing blocks simulation time by 100 ns for up to one microsecond. Colour scale from blue to red illustrates how the profile evolves through time. The first 200 ns was removed from each window to ensure that the system had fully equilibrated first. (b) The binding values (the lowest energy value) of each PMF profile over the course of the simulation time for increasing blocks of 100 ns. Errors calculated by bootstrapping each PMF profile 10 times. For instance, 200 – 300 ns, 200 – 400 ns, ..., 200 – 1000 ns. (c) The binding values (the lowest energy value) of each PMF profile over the course of the simulation time for consecutive blocks of 100 ns. For instance, 200 – 300 ns, 300 – 400 ns, ..., 900 – 1000 ns.

I obtained the PMF profile (**Figure 4.23**) by using 332 windows across the hydrophobic groove into solvent and assessed convergence had been reached after 800 ns (**Figure 4.24**). The minimum of the LptA₄ (-155 ± 5 kJ mol⁻¹) landscape is greater than that of LptA₃ (-120 ± 5 kJ mol⁻¹) which is again greater than the LptA monomer by itself (-86 ± 2 kJ mol⁻¹). However, the resultant PMF landscape is unexpectedly curved and asymmetric, if lipid A is traversing across the same hydrophobic track four times, then the landscape should be flatter as well as revealing a consistent motif. Again, this is most likely caused by solvation effects where lipid A is more exposed to solvent within one of the LptA monomers, but less exposed inside the LptA₃ trimer and even less

exposed with the LptA₄ tetramer. The LptA₄ thermodynamic landscape is also different (and more symmetrical) to the one I obtained for LptA₃, which is due to the extra conformational rigidity afforded by the additional LptA subunit. However, I cannot comment upon whether my data supports LptA₄, rather than LptA₃ being present in the cell as this may vary according to periplasm size, though the thermodynamics of lipid A transport are more favourable when the other periplasmic units of the chute are present. To obtain a better understanding of lipid A thermodynamics within the chute, it is important to eliminate solvation effect which is artificially biasing the edges of the chute to be higher in energy than it 'should be'. It is also critical to enforce artificial rigidity of subunit neighbours to prevent higher conformational sampling towards either end of the chute under analysis.

4.7 Building the model for a continuous chute of LptA units

Within the LptA crystal structure itself (PDB: 2R19)³⁰¹, the LptA units repeat themselves exactly after every four monomers (the unit cell). To mimic this, I made the boundary condition of the simulation box interface between the end of the LptA tetramer and allowed the rubber band contacts between units to extend over the periodic boundary. This created a continuous chute of LptA units together and allows for a PMF of lipid A to be laterally pushed through the chute itself in the direction which lipid A moves through the system. This is unlike the other PMF calculations where lipid A is made to move counter to what it would be doing physiologically when moving through the periplasm. Doing this also imposes conformational rigidity between the subunits which prevents the outside of the chute from splaying out towards solvent like what is observed in the isolated portions of the chute itself.

4.7.1 Thermodynamic landscape of lipid A through the continuous LptA chute

This time a flat, albeit asymmetrical, PMF landscape was observed (**Figure 4.25**) with a deviation of $\sim 15 \text{ kJ mol}^{-1}$. 289 separate windows were required to sample along the four LptA units, where convergence was assessed to have been reached after 600 ns (**Figure 4.26**).

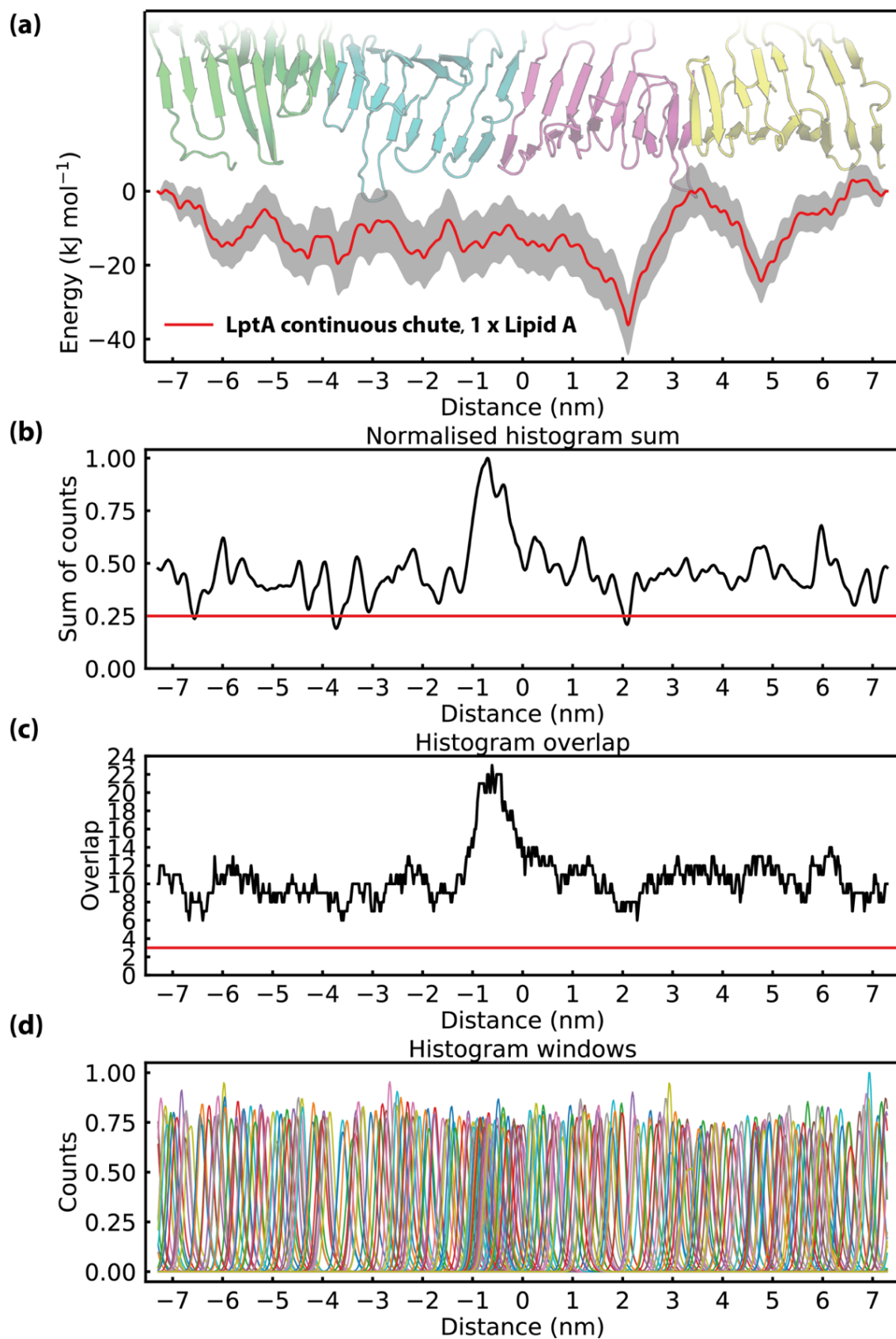
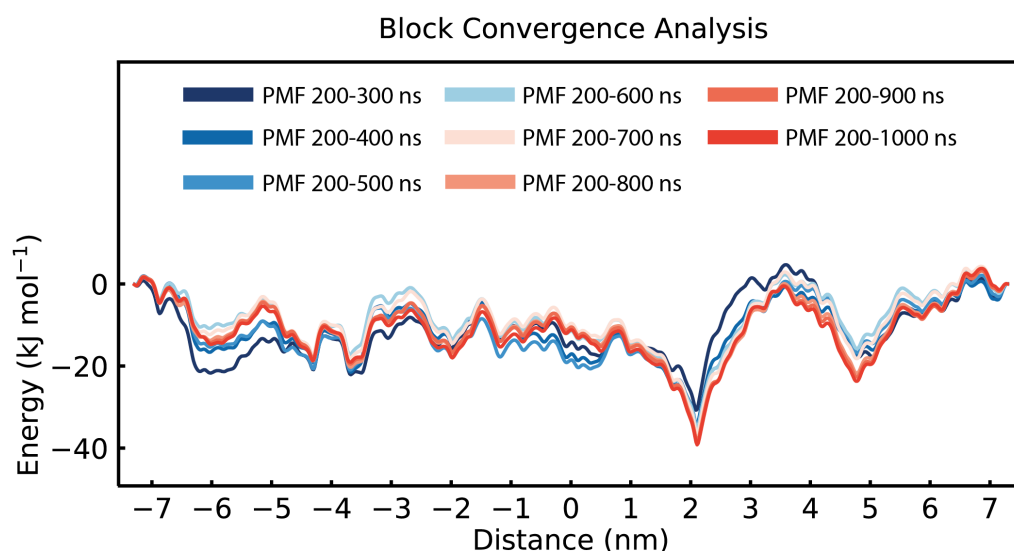


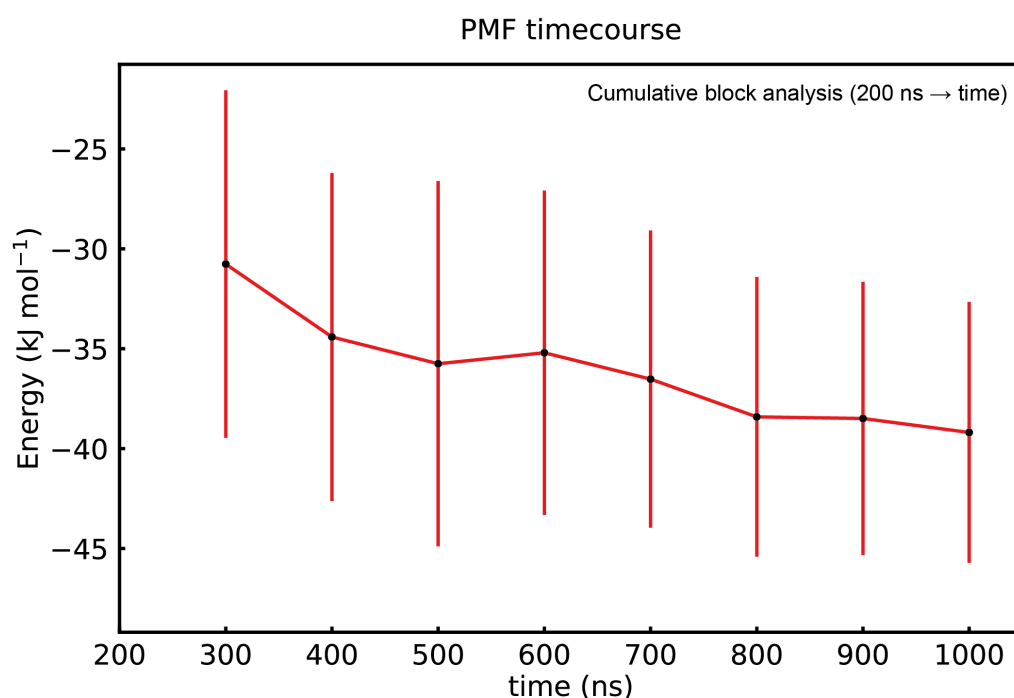
Figure 4.25 Binding energy of lipid A with continuous LptA chute. (a) PMF profile of lipid A moving through LptA₄, areas shaded in grey indicates the error and is

obtained by bootstrapping the profile 200 times. **(b)** Total counts summed across the collective variable of lipid A being pulled from the β -taco into solvent from the raw histograms, where the dataset is normalised to 1. The red line is a 'rule of thumb' marker for acceptable sampling positioned at 0.2 of total counts. **(c)** The number of overlapping histograms from the umbrella sampling across the collective variable. A count is considered when the overlap is $>10\%$ of the total histogram height. The red line corresponds to a 'rule of thumb' for sufficient overlap set to 3. **(d)** The histograms from the WHAM analysis, where every histogram corresponds to a separate sampling position of lipid A.

(a)



(b)



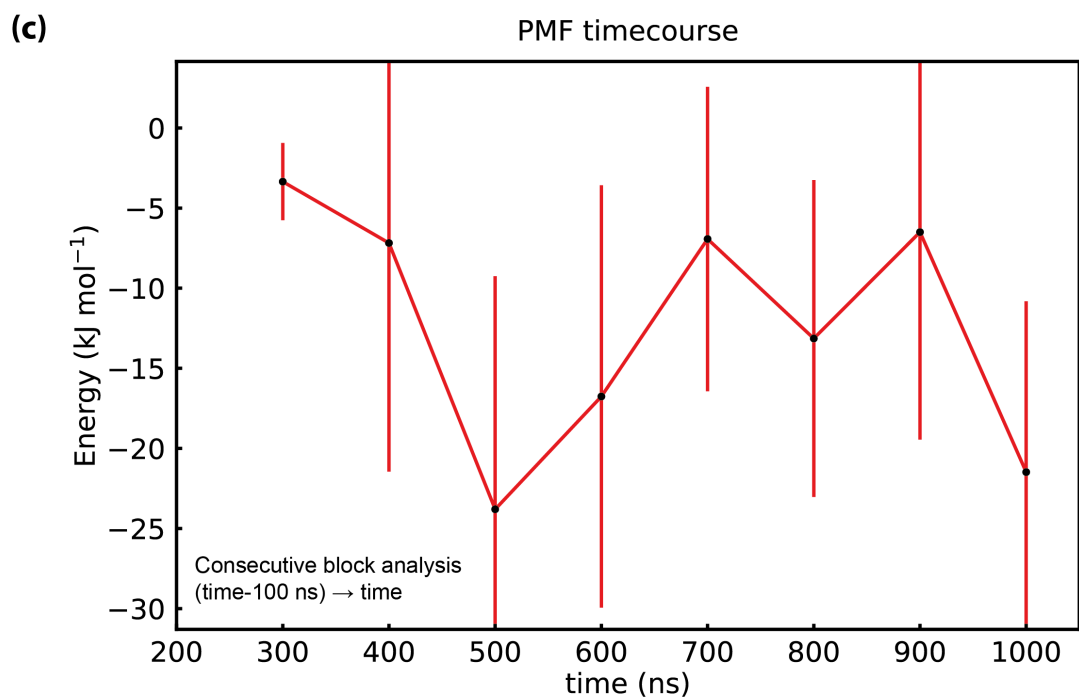


Figure 4.26 Continuous LptA chute PMF convergence analysis. (a) PMF profiles of increasing blocks simulation time by 100 ns for up to one microsecond. Colour scale from blue to red illustrates how the profile evolves through time. The first 200 ns was removed from each window to ensure that the system had fully equilibrated first. (b) The binding values (the lowest energy value) of each PMF profile over the course of the simulation time for increasing blocks of 100 ns. Errors calculated by bootstrapping each PMF profile 10 times. For instance, 200 – 300 ns, 200 – 400 ns, ..., 200 – 1000 ns. (c) The binding values (the lowest energy value) of each PMF profile over the course of the simulation time for consecutive blocks of 100 ns. For instance, 200 – 300 ns, 300 – 400 ns, ..., 900 – 1000 ns.

Between -7 nm and +1 nm the energy landscape is relatively flat where lipid A experiences a minimally fluctuating thermodynamics of binding (~ -10 kJ mol⁻¹) for the first two LptA units. However, the sharp dip (at +2 nm), rise, then dip (at +4.8 nm) in the PMF profile is caused by a specific constriction in the chute, which generates two favourable lipid binding pockets between two restrictions (**Figure 4.27**).

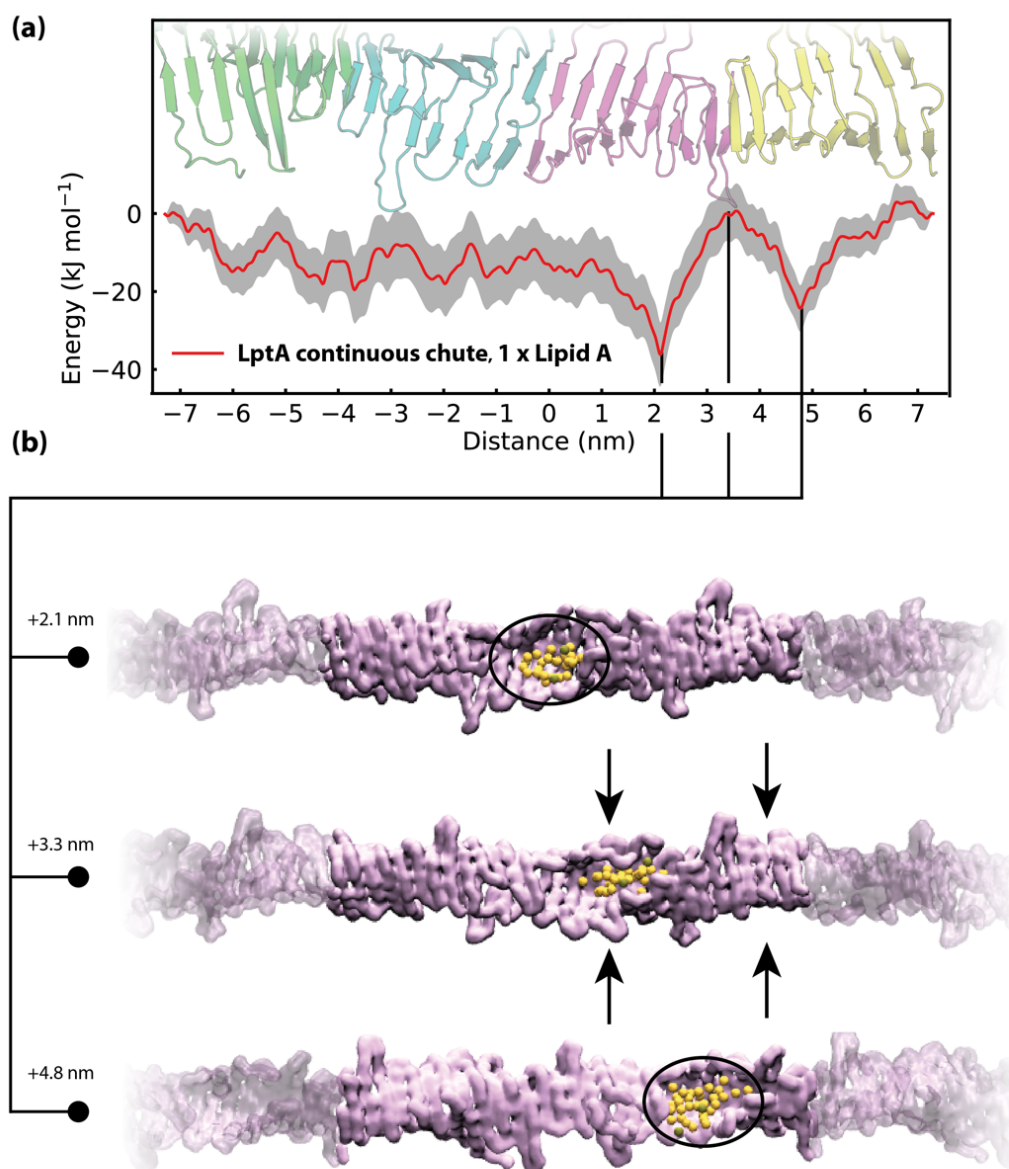


Figure 4.27 Lipid A binding conformation within the LptA continuous chute. (a) PMF profile of lipid A moving through LptA continuous chute areas shaded in grey indicates the error and is obtained by bootstrapping the profile 200 times. Cartoon representation of the LptA₄ tetramer aligned with the energy profile in the background for reference. **(b)** Lipid A binding pose at -6.0 nm, -2.5 nm, +11.2 nm labelled onto the LptA continuous chute PMF binding profile.

Successfully, then, solvation effects were removed from the PMF calculation of binding energy with the periplasmic chute. Furthermore, far greater conformational rigidity across these units granted much more consistent lipid A binding energy when translocated. In this model, however, the chute 'is empty' of other lipids as I have only

calculated the energy of one lipid A molecule through the periplasmic bridging proteins. As mentioned earlier, the 'PEZ model' proposes that the entire chute itself is loaded with lipids as they are transported through the system⁸³. To correct for this, I kept the lipid A molecules within the chute when laterally pulling on one of the lipids using steered MD.

4.7.2 Thermodynamic landscape of lipid A through the continuous LptA chute with lipid neighbours

The PMF was repeated with the continuous LptA chute. However, this time, I moved lipid A through with its lipid neighbours. I linked the lipid A molecules together with position restraints so they would move uniformly when I laterally pulled one of the lipids. Simply, from a steric perspective, two lipid A molecules could 'comfortably' fit in any LptA subunit and so eight lipid A molecules were placed within the repeating LptA₄ tetramer unit. I obtained the landscape (**Figure 4.28**) by sampling along the collective variable with 292 separate windows, where convergence was reached after 800 ns (**Figure 4.29**).

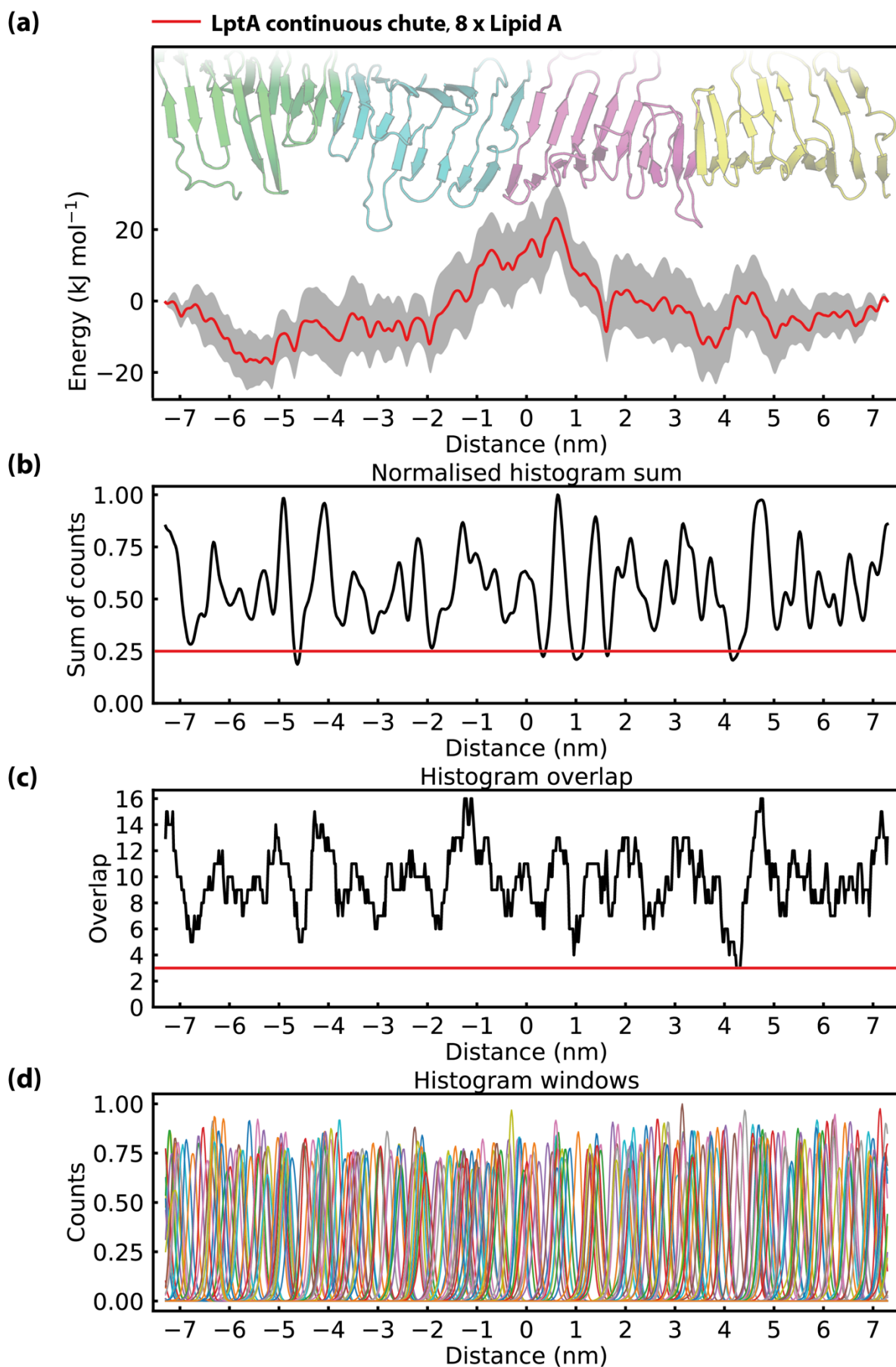
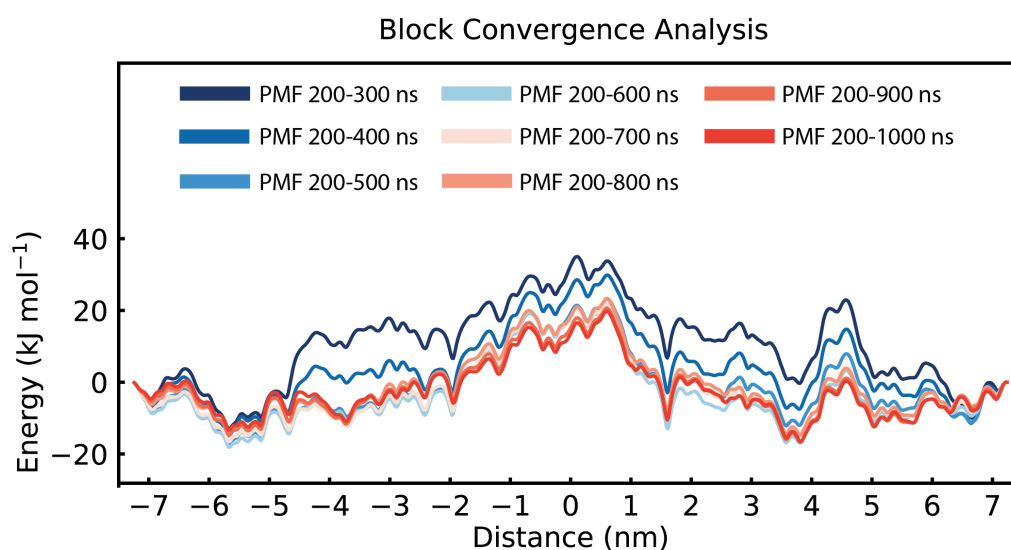


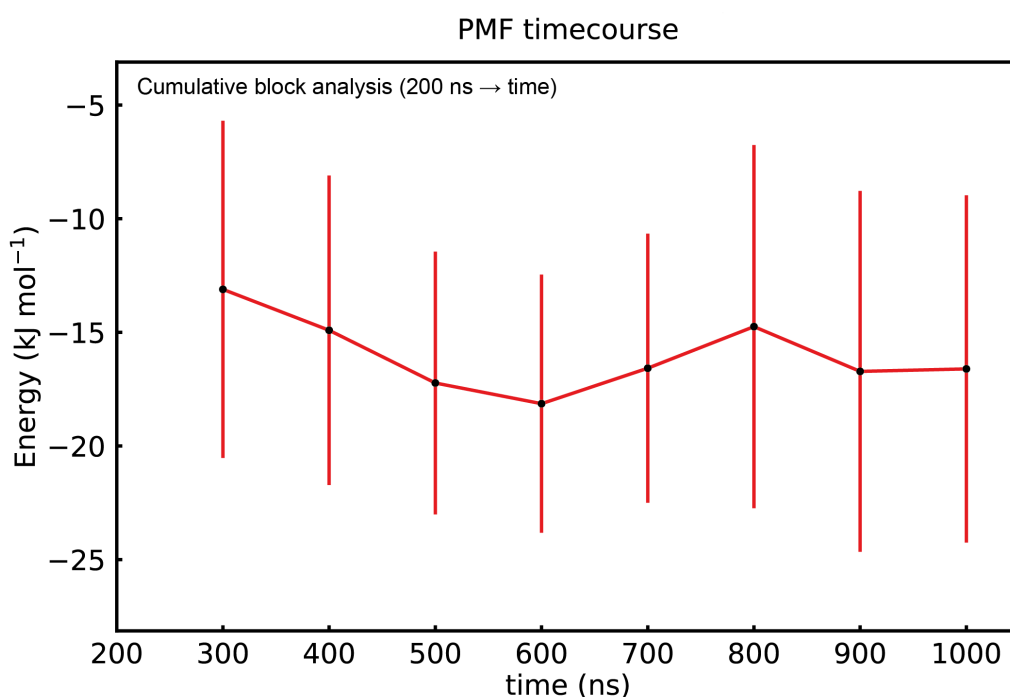
Figure 4.28 binding energy of lipid A with continuous, fully lipidated LptA chute.

(a) PMF profile of lipid A moving through a fully lipidated LptA₄, areas shaded in grey indicates the error and is obtained by bootstrapping the profile 10 times. **(b)** Total counts summed across the collective variable of lipid A being pulled from the β -taco into solvent from the raw histograms, where the dataset is normalised to 1. The red line is a 'rule of thumb' marker for acceptable sampling positioned at 0.2 of total counts. **(c)** The number of overlapping histograms from the umbrella sampling across the collective variable. A count is considered when the overlap is >10% of the total histogram height. The red line corresponds to a 'rule of thumb' for sufficient overlap set to 3. **(d)** The histograms from the WHAM analysis, where every histogram corresponds to a separate sampling position of lipid A.

(a)



(b)



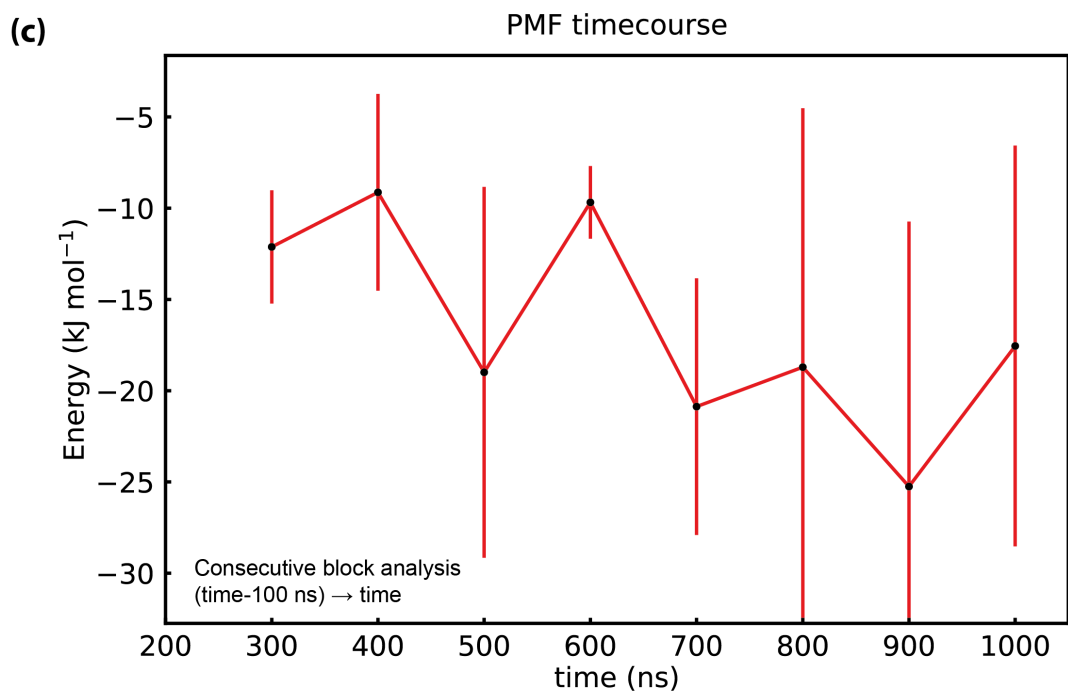


Figure 4.29 Continuous, fully lipidated LptA chute PMF convergence analysis. (a) PMF profiles of increasing blocks simulation time by 100 ns for up to one microsecond. Colour scale from blue to red illustrates how the profile evolves through time. The first 200 ns was removed from each window to ensure that the system had fully equilibrated first. (b) The binding values (the lowest energy value) of each PMF profile over the course of the simulation time for increasing blocks of 100 ns. Errors calculated by bootstrapping each PMF profile 10 times. For instance, 200 – 300 ns, 200 – 400 ns, ..., 200 – 1000 ns. (c) The binding values (the lowest energy value) of each PMF profile over the course of the simulation time for consecutive blocks of 100 ns. For instance, 200 – 300 ns, 300 – 400 ns, ..., 900 – 1000 ns.

Similar to the previous landscape, the energy profile is relatively flat with a deviation of 15 kJ mol⁻¹ (**Figure 4.28**). The energy hump from -2 nm to +2 nm is caused by a non-uniform distribution of lipids within the hydrophobic groove.

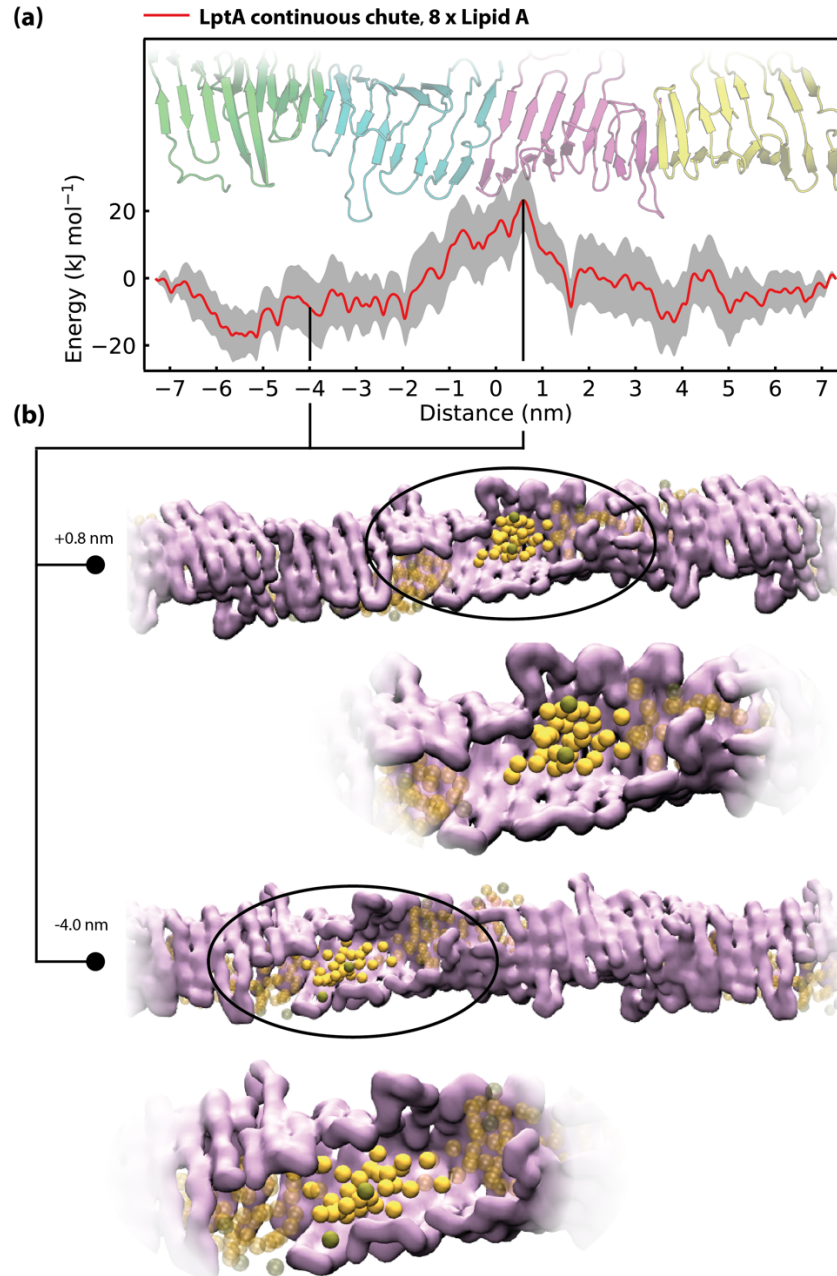


Figure 4.30 Lipid A binding conformation within the fully lipidated LptA continuous chute. (a) PMF profile of lipid A moving through LptA continuous chute areas shaded in grey indicates the error and is obtained by bootstrapping the profile 200 times. Cartoon representation of the LptA₄ tetramer aligned with the energy profile in the background for reference. (b) Lipid A binding pose at -4.0 nm and +0.8 nm labelled onto the fully lipidated LptA continuous chute PMF binding profile.

At +0.8 nm lipid A has separated away from its lipid neighbours and therefore does not benefit from those favourable hydrophobic interactions which it enjoys at -4 nm (for instance) (**Figure 4.30**).

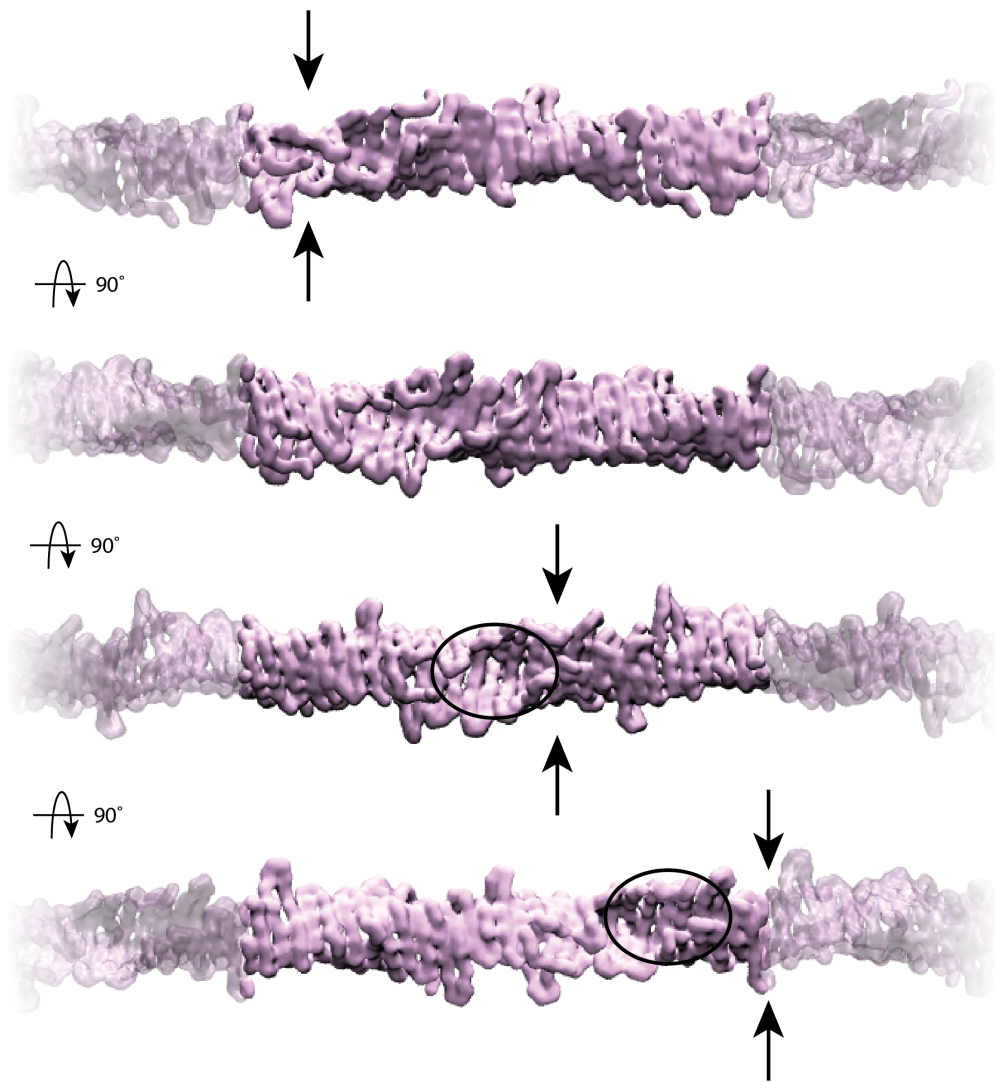


Figure 4.31 LptA continuous chute. Rendered images of the chute taken every 90° rotation through the chute's principal axis. Restrictions along the chute are noted by arrows and binding pockets are highlighted in circles.

The non-uniform distribution of lipids was caused by the restriction in the chute as depicted in **Figure 4.31**. Therefore, it is important for lipid transport that the chute is open evenly so (i) that any singular lipid with all six tails can easily fit within the hydrophobic groove (not squeezed out) and so (ii) lipids do not bundle up and overcrowd the chute, leaving some lipids isolated without adjacent hydrophobic interactions.

4.8 Discussion

For the hydrophobic groove of the β -taco subunits to open, LPS must be accommodated inside otherwise the subunit will immediately close (**Figure 4.2**). I propose that this subunit closure is driven by (i) hydrophobic residues of the fold interacting with themselves and (ii) electrostatic salt bridges on the edges of the β -taco effectively 'zipping' the unit shut. I have shown that the lipid must open the sides of the subunit to at least 2 nm for the subunit to remain open and not close again (**Figure 4.15**), lest salt bridges reform and squeeze the lipid tails out from the groove.

I have found that the binding energy of lipid A with the β -taco units is variable. For now, I shall discount the binding profile of LptF (**Figure 4.5**) when comparing it with LptC/A/D due to its smaller size which exposes lipid A to solvent. Comparatively sized β -taco units LptC and LptA have different binding energies with lipid A, at $-131 \pm 2 \text{ kJ mol}^{-1}$ (**Figure 4.6**) and $-86 \pm 2 \text{ kJ mol}^{-1}$ (**Figure 4.9**) respectively which helps to clarify the confusion between which subunit has a greater affinity to LPS^{90,315,316}. This is further emphasised in the LptA₃CDF complex where lipid A experience a huge energy sink ($-165 \pm 6 \text{ kJ mol}^{-1}$, **Figure 4.20**) when translated into LptC. The binding energy of lipid A with LptD ($87 \pm 5 \text{ kJ mol}^{-1}$, **Figure 4.12**) is very similar to that of LptA ($86 \pm 2 \text{ kJ mol}^{-1}$), however it is very likely that the binding energy of lipid A with LptD would be more favourable than the PMF profile I obtained for the subunit. This is because the LptD β -taco had not been opened up enough to allow facile passage of lipid A through the subunit. Therefore, it would be important to redo this PMF, ensuring at least 2 nm distance between the sides of the groove, to obtain a 'true' value of lipid A binding to

LptD. Regardless, this PMF approach has certainly shed light on the binding values of lipid A with each of the bridging subunits, where it is clear that there is a strong driving force drawing lipid A into LptC, requiring ATP activity to push the lipid into LptA subunit.

I have observed that larger complexes of the bridging units have stronger binding interactions with lipid A compared to singular β -taco units. An obvious example of this is the increasingly favourable binding interactions of lipid A with LptA ($-86 \pm 2 \text{ kJ mol}^{-1}$, **Figure 4.9**), LptA₃ ($-120 \pm 5 \text{ kJ mol}^{-1}$, **Figure 4.17**) and LptA₄ ($-155 \pm 5 \text{ kJ mol}^{-1}$, **Figure 4.24**). Furthermore, the shape of the PMF profile for LptA₃ is different to that of LptA₄ where there is no (expected) repeating motif which implies that additional subunits change the conformational rigidity of the chute which in turn changes what the lipid experiences as it moves through the bridge. Certain tightness with the chute squeezes the lipid tails out of the hydrophobic groove which causes higher energy in certain PMF profiles such as that of LptA₃ (**Figure 4.17**) and LptA₃CDF (**Figure 4.20**). Accordingly, site specific EPR spectroscopy which examined LptA binding with LPS indicates that conformational movements may play a part in how well LPS binds within the subunit itself³¹⁵.

For favourable interactions with the chute the lipid requires (i) all hydrophobic tails to be firmly inside the fold of the β -taco subunits and (ii) for the phosphate head groups to interact with the soluble, electrostatic/polar residues pertaining to the edges and loops of the fold. These observations are apparent across the variable energy profiles of LptA₃ and LptA₃CDF, where the instances of most favourable energy are those with all lipid tails in the hydrophobic groove and where there are phosphate headgroup interactions. Experimental work could be done to validate the binding energy of lipid A

with the periplasmic subunits. Native-MS can be used to calculate the dissociation constant with LptC, LptA and LptD and therefore establish the free energy of binding between these units. Additionally, mutating residues found to interact with the phosphate groups of lipid A (such as Pro34 and Arg110 of LptC or Asn24 and Ser141 of LptA) may reduce the free energy of binding and corroborate the findings of these PMF studies.

Regardless of whether lipid A was sampling a singular subunit in isolation or a complex in isolation, every PMF profile compared the energy of binding the chute with that of solvent. In each of these cases, solvation effects were observed, which was responsible for causing 'V'-shaped looking profiles where the lipid gradually desolvated and re-solvated through the passage of the units in isolation. To circumvent this, I constructed a continuous version of the LptA₄ chute, whereby the edges of one LptA₄ unit interacted with the mirror image of itself through the periodic boundary conditions of the simulation box. Subsequently, I sampled along the continuous chute, where the binding energy of lipid A within LptA was calculated relative to the environment within the hydrophobic groove itself. The energy profile of the LptA 'continuous chute' eliminated the 'V'-shaped topology of the previous PMF landscapes. The profile itself, deviated by approximately 15 kJ mol⁻¹ across the complex, but can be considered 'flat' when factoring in the statistical error obtained by bootstrapping the profile. Larger deviations were, again, due to restrictions within the chute which caused larger hydrophobic pockets for lipid A to nestle inside more favourably compartmented the other sections of the chute. Here, future work should include redoing this PMF calculation by opening up the restricted section of the chute and thereby proving that the energy landscape is affected by 'local' conformational effects.

The currently, proposed functional model of the Lpt system is known as the 'PEZ' model whereby the entire chute is filled with lipids and is pushed up into the OM by the LptB₂FGC ABC transporter. To attempt to replicate this, I filled the LptA₄ continuous chute with eight lipid A molecules, i.e., two lipids per LptA unit. This is in contrast to EPR binding data which suggests that singular units (non-oligomerised) of LptA binds lipid A in a 1:1 ratio³¹⁵ (though oligomerisation of these unit would certainly allow for more space to accommodate another lipid molecule within the subunit). Interestingly, the PMF profile revealed that energy differences across the chute were due to reduced lipid A-lipid A interactions, which was itself caused by a restriction in the chute and 'bunching up' of the lipid molecule inside the groove. This itself implies that lipid A neighbours are required (energetically) to be transported together, which accords with the importance of the periplasmic chute needing to be open 'enough' for lipid transport. Neighbouring lipids would assist opening the chute and keeping the groove open for its counterparts to frictionlessly reach the OM. My observations here, support the PEZ model of transport. To lend credence to the PEZ model, further thermodynamic studies would be required, this would have to include repeating the PMF calculations with different numbers of lipid A neighbouring units to understand how more or less lipid neighbours would affect the landscape. From observing the trajectories of these simulations, it's quite possible that more lipids could easily fit in within the LptA₄ repeating unit. This could help shed light on the correct physiological lipid A: LptA stoichiometry. Furthermore, a PMF which stretches from the bulk lipid membrane in the IM, through the system and out into the OM, would capture a full picture of the thermodynamics of the Lpt system and contextualise lipid A binding within the system compared to the native 'bulk' membrane.

4.9 Methods

4.9.1 Modification of Lpt structures used

Here is a table referencing all the structures taken from the PDB used to model different sections of the Lpt periplasmic chute:

Table 4.1 PDB IDs for Lpt complex structures. Protein structures used directly in simulations or used to generate the homology models of the different subcomplexes of the Lpt system.

Protein	Organism	PDB code	Resolution (Å)
LptB ₂ FG ²⁹⁹	<i>P. aeruginosa</i>	5X5Y	3.46
LptB ₂ FG ³⁰⁰	<i>K. pneumoniae</i>	5L75	3.70
LptB ₂ FG ²⁹³	<i>S. flexneri</i>	6S8H	3.70
LptB ₂ FG ²⁹³	<i>S. flexneri</i>	6S8G	3.50
LptA ³⁰¹	<i>E. coli</i>	2R19	2.16
LptC ³⁰²	<i>E. coli</i>	3MY2	2.20
LptDE ³¹³	<i>S. flexneri</i>	4Q35	2.39
LptDE ³¹⁹	<i>K. pneumoniae</i>	5IV9	4.37

4.9.2 The periplasmic bridging proteins LptA₃CDF

For individual bridging subunits of the periplasmic bridge, modifications to the crystal structure were required to enable easy passage of lipid A across the domains. For this to occur, these units needed (i) to have any transmembrane domain's removed and (ii) to be opened enough to accommodate their substrate within the hydrophobic groove.

For LptF and LptD, the β -taco domains were excised in PyMOL³⁰⁴ at residues 147, 234 and 50, 218 respectively. To open each of the β -taco domains (LptF, LptC, LptA and LptD) the alchembed methodology was inverted which is process where the van der Waal potential is gradually switched on between the Lipid A and the β -taco²³⁶. Typically, alchembed 'expands' a protein within a membrane (often to fix steric clashes in AT simulation setup), so by inverting this the lipid could expand within the domain itself gradually opening up the β -taco in the process.

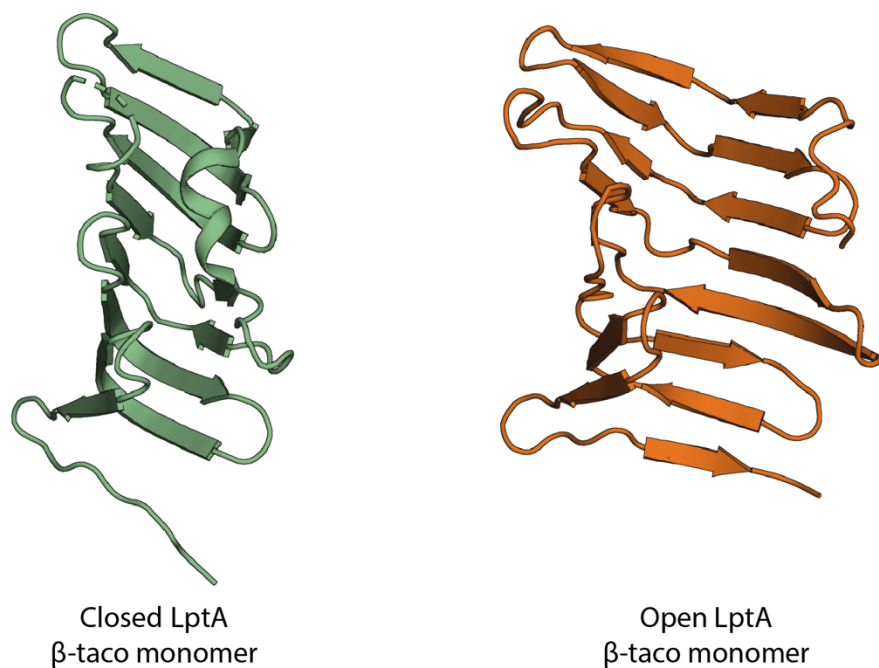


Figure 4.32 Opening LptA with alchembed: A before (left) and after (right) β -taco opening of the LptA monomer

The domains were allowed to open (around 2 nm) with the secondary structure maintained, to fit a lipid A molecule within the hydrophobic groove without any steric clashes (**Figure 4.31**). The initial docking of lipid A to the domains was performed by AutoDock Vina³²⁰. In opening LptD in particular, side chains Arg191, Arg145 and Gln116 were rotated to prevent steric clash with lipid A³²¹. In a similar manner, subcomplex LptA₃, LptA₄ and LptA₃CDF were all aligned together by their constituent subunits in PyMOL and opened uniformly at the same time. I generated CG representations of each of the subunits and subcomplexes by using a Martinize python script²²⁹ for the PMF studies.

4.9.3 Constructing the LptA₄ continuous chute

It was noted that the LptA₄ was a repeating unit within the unit cell of the crystal structure (PDB: 2R19). Therefore, I assembled the structure by taking four monomers

together and opened the hydrophobic groove in the same manner as described above (**Figure 4.31**)³⁰¹. After converting the structure to coarse grain using a Martinize python script,²²⁹ the periodic boundary conditions of the box were placed along the edges of the tetramer where the elastic network was allowed to run over these boundary conditions. This model could then be used to study the interaction of lipid A across the hydrophobic groove of the chute as outlined in this chapter.

4.9.4 Lpt periplasmic subunits CG-MD simulations

Every window in the PMF calculation along the collective variable (CV) for each of the individual subunits and subcomplexes was simulated in the MARTINI 2.2 force field³⁰⁵ which was run on GROMACS 5³²², where the windows each ran for 1 μ s to reach a converged state. The conditions and settings used were identical to the simulations ran for the IM proteins.

However, for the PMFs conducted with the continuous LptA tetramer model, windows were run with NVT conditions to ensure that the periodic boundary conditions remained completely fixed.

4.10 Steered MD pulling experiments

I performed PMF calculations with lipid A and the subunits of the periplasmic chute as outlined previously^{137,142}. To set up these calculations, steered MD is employed on lipid A to move the lipid from inside the hydrophobic groove out into solvent. This trajectory defines the CV for the PMF calculations. Here, the steered MD pulling a distant dependent force acting on the centre-of-mass (COM) of lipid A and the protein. Snapshots along the CV are taken 0.05 nm apart between lipid A centres-of-mass to achieve optimal histogram overlap, where each snapshot itself is a seed for an

independent simulation. For every snapshot (or window) along the CV a harmonic (or umbrella) potential is imposed on the COM of lipid A constraining the lipid to sample that position along the reaction coordinate. A force of $1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ was used for the umbrella potential, which was increased to $1500 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ for particular positions along the CV which lipid A needed additional force to sample sufficiently. I placed (xyz) position restraints of $1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ on the along the backbone beads which make up the base of the hydrophobic groove to allow the sides of the chute to flex accordingly around lipid A.

In order to pull lipid A through the continuous LptA chute along with its neighbours I used plumed UPPER/LOWERWALLS for the steered MD³²³. These restraints were set such that (i) the tails for the lipids within the chute would not move a distance greater than 0.5 nm away from the track of the hydrophobic groove (to ensure no tails left the chute on pulling) and (ii) that lipid A molecules would not move further than a distance of 1.5 nm from its adjacent neighbours.

Each window was run for 1 μs in order for a converged energy landscape, where the first 200 ns of each window was discarded as equilibration. I combined the final 800 ns of every window to form the PMF profile by applying the weighted histogram analysis method (WHAM). In the analysis I used 200 rounds for Bayesian bootstrapping to highlight the error in the energy profiles^{135,275}. For each energy profile the solvent is considered to have a free energy of 0 kJ mol^{-1} .

4.11 Author contributions

FF and JBS made the **Figure 4.1a** and **c** together. All other work including MD studies and analysis was carried out by JBS.

CHAPTER 5 Defining the conformational changes of LptDE with LPS and an inhibitor

Work carried out in chapter has been published in the paper:

Dynamics of an LPS translocon induced by substrate and an antimicrobial peptide⁵¹

Published: 16 November 2020

Francesco Fiorentino, Joshua B. Sauer, Xingyu Qiu, Robin A. Corey, C. Keith Cassidy, Benjamin Mynors-Wallis, Shahid Mehmood, Jani R. Bolla, Phillip J. Stansfeld & Carol V. Robinson

Nature Chemical Biology volume **17**, pages 187–195 (2021)

5.1 Background to the LptDE translocon complex

The outer membrane (OM) of Gram-negative bacteria, is an asymmetric bilayer which features lipopolysaccharide (LPS) molecules on the outer leaflet and phospholipids in the inner leaflet⁷⁷. Here, LPS is instrumental in protecting the cell from toxic hydrophobic compounds^{71,276} due to its amphipathic nature. LPS has a tripartite

structure comprising lipid A, core oligosaccharide and an O-antigen glycan of up to 200 sugars (**Figure 5.1**)⁸³.

Once the full LPS molecule has been formed, the Lpt system facilitates its transport through the periplasm to the OM⁸³. The essential OM heterodimer²⁷⁷, formed by the integral β -barrel LptD and the lipoprotein LptE, is concerned with finally installing LPS from the periplasm into the outer leaflet of the OM (**Figure 5.1**)^{97,99}.

The LptD structure consist of a 26-stranded β -barrel with a lateral LPS exit gate formed between the strands β 1 and β 26 which sits in the OM bilayer, whereas the periplasmic region is characterised by a β -taco domain. This particular protein fold, known as the β -taco fold, is common to both LptA and LptC, of which, LptD forms a periplasmic complex with the former.

Crystal structures of the outer membrane LptDE translocon complex reveal a plug and barrel heterodimer between LptD with lipoprotein LptE, where LptE is situated inside the β -barrel of LptD (**Figure 5.1**)^{21,324}. In order to accommodate LPS for transport to the OM, a number of studies involving crosslinking³²⁵ as well as structural³¹⁹ and genetic work⁹³ reveal that the lipid tails of LPS interact within the hydrophobic groove of the β -taco.

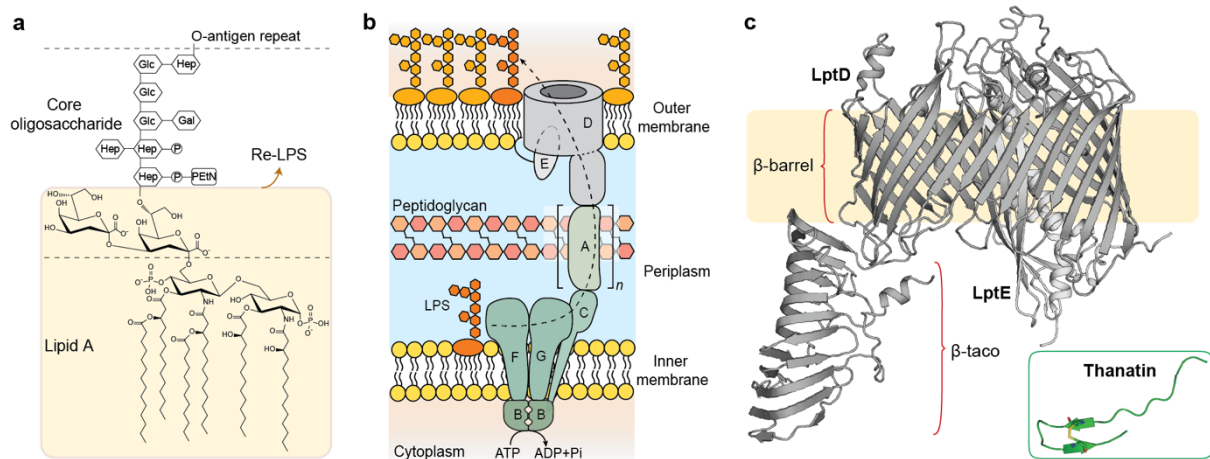


Figure 5.1 Cartoon schematic describing the LPS transport pathway. Figure adapted from Fiorentino *et al*⁵¹, all work here completed by myself and FF. (a) A skeletal representation of LPS where its tripartite structure is described: lipid A, core oligosaccharide and the O-antigen repeating unit. The orange box highlights a derivative of LPS, called ‘Re-LPS’ which consists of the lipid A unit connected to two ketodeoxyoctonic acid (Kdo) moieties. **(b)** cartoon schematic of the Lpt system: LPS is extracted from the inner membrane (IM) by the ABC transporter complex LptB₂FGC, where it is subsequently shuttled through the periplasm via a soluble bridge of LptA units to be finally installed in the OM by LptDE. **(c)** The LptDE X-ray crystal structure (PDB ID: 5IV9)³¹⁹ from *Klebsiella pneumoniae*. The soluble β-taco domain of LptD is known to interact with LptA to form the periplasmic bridge and the β-barrel accommodates LptE inside to form the plug-and-barrel complex. Inset: cartoon structure of the antimicrobial molecule thanatin (PDB ID: 5XO4).

The LptDE heterodimer plays a vital role in the outer membrane, however it is still unclear how precisely the complex handles its substrate as well as the conformational changes induced on LPS binding. Mutagenesis, molecular dynamic studies as well as functional assays suggest that the β-barrel lateral gate is critical for the transport of LPS to the OM³²⁵. These studies also highlight the importance of the β-taco domain forming an intramembrane hole with the inner leaflet of the OM. The proposed mechanism here is that the LPS hydrophobic tails are threaded through the intramembrane hole at the same time as the hydrophilic headgroup is being tracked through the lumen of the β-barrel. To allow for this, the lateral gate is proposed to unzip to allow for lateral transport into the OM (**Figure 5.2**)³²⁵.

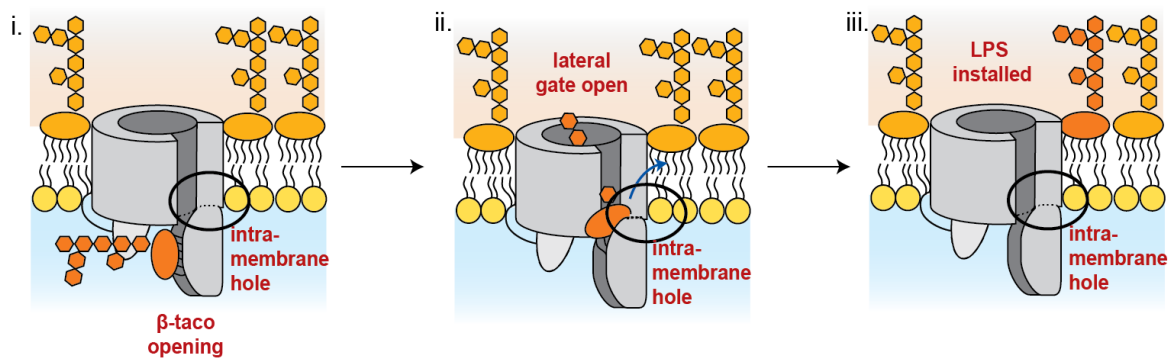


Figure 5.2 Schematic describing proposed mechanism of LPS installation into the OM. Figure adapted from Fiorentino *et al*⁵¹, all work here completed by myself and FF. The open β -taco forms an intramembrane hole in the inner leaflet of the outer membrane to allow the LPS tails to be threaded through as the headgroup moves through the lumen of the β -barrel. The lateral gate is proposed to be open to accommodate the lipid's lateral extraction into the OM.

As alluded to above, LPS transport is a crucial pathway for the viability of Gram-negative bacteria and so understanding the molecular mechanism will be critical for the development of novel antibiotics which can target this pathway. So far, two known antimicrobial peptides have shown to interact with LptD: murepavadin and thanatin. Murepavadin is peptidomimetic and specifically binds LptD derived from *P. aeruginosa*^{326,327}, and has also demonstrated antibacterial properties in a mouse septicemia infection model³²⁸. Thanatin itself is an insect-derived cyclic peptide and has shown to bind LptA and LptD, which consequently disrupts protein-protein complex formation of the periplasmic bridge and therefore inhibits LPS transport³²⁹. The LptA residues which interact with thanatin are known and are described by the NMR solution structure³²⁹, less clear however, is how thanatin interacts with LptD and potentially what conformational changes the inhibitor may incite on the LptDE complex. Thanatin is particularly interesting because of the range of pathogens it exhibits high levels of antibacterial activity in, such as *S. typhimurium*, *E. cloacae* and *K. pneumoniae* where, in each instance, the minimal inhibitory concentrations (MICs) are less than 1.5 μM ³³⁰. Here, thanatin could offer an important lead antibiotic because

of its promiscuity with the pathogens it targets and therefore further emphasises the importance of understanding how it interacts with LptDE.

Here, I integrate molecular dynamic (MD) simulations with native mass spectrometry (MS) hydrogen-deuterium exchange mass spectrometry (HDX-MS) work carried out by collaborators: Francesco Fiorentino (FF), Xingyu Qiu (XQ), Shahid Mehmood (SM), Jani Reddy Bolla (JRB), Carol V. Robinson (CVR) to understand how the conformational dynamics of LptDE is affected by substrate and inhibitor binding⁵¹. The MD simulations performed here study the solvent accessibility and hydrogen bond position of LptDE in different conformations to inform on HDX-MS experimental results.

The OM complex of the Lpt system, LptDE, is known to translocate LPS to its final destination in the OM, however the precise mechanism of how the complex achieves this is not clear. Furthermore, the conformational changes of the transporter on binding to LPS is not known.

5.2 LptDE conformational change with Re-LPS

5.2.1 HDX observations of LptDE substrate binding effects

An experimental tool which can help differentiate conformational changes within a protein system is HDX. Collaborators: FF, XQ, SM, JRB, CRV⁵¹ used HDX to probe the binding of the substrate LPS as well as an inhibitor thanatin with the translocon LptDE, in order to understand conformational changes made upon binding. The observed results obtained by collaborators (FF, XQ, SM, JRB, CVR)⁵¹ from these experiments was surprising and counterintuitive on first inspection (**Figure 5.3**).

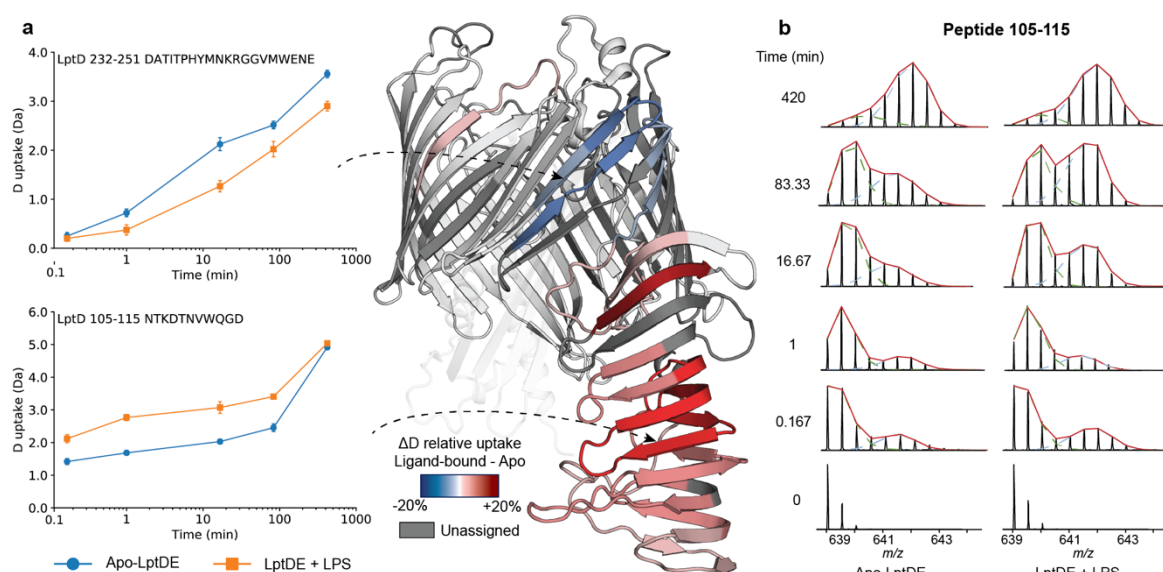


Figure 5.3 Conformational dynamics of LptDE with Re-LPS. Figure modified from Fiorentino *et al*⁵¹, all work here completed by collaborators (FF, XQ, SM, JRB, CVR). **(a)** Left: Peptides illustrating increased deuterium uptake (105–115, β -taco) and decreased deuterium uptake (232–247, β -barrel) and their corresponding labelling time (0.167–420 min) plotted for apo-LptDE (purple circles) and LPS-bound LptDE (orange squares). Here the error bars specify standard deviation. ($n_{\text{biological}} = 2$; $n_{\text{technical}} = 3$). These peptides are mapped upon the LptDE structure according to the difference in deuterium uptake at the 16.67-min labelling time, where only peptides which demonstrate a significant difference in deuteration are shown. Respectively, red peptides emphasise an increase in HDX and blue peptides represent a decrease. **(b)** Mass spectra for peptides 105–115 are illustrated here for the ‘apo’ LptDE and ‘LPS-bound’ states. The best fit for the spectra is two binomial isotopic envelopes, producing low-mass (green) and high-mass (light blue) populations.

Collaborators (FF, XQ, SM, JRB, CVR) found that on binding LptDE (in equilibration buffer: 200 mM NaCl, 25 mM Tris-HCl (pH 7.5), 10% (vol/vol) glycerol and 0.03% (wt/vol) DDM) to LPS (15-fold excess), the β -strands next to the putative lateral gate (located between $\beta 1$ and $\beta 26$) of the β -barrel reveal a HDX decrease, ‘protection’ on strands $\beta 1$ – $\beta 2$ (peptides 195–211, 195–218 and 207–218) and $\beta 3$ – $\beta 4$ (peptides 232–247 and 232–251) (**Figure 5.3**).

It is surprising to observe protection in these regions, as previous modelling data³²⁵ suggests that the β -barrel lateral gate ‘opens up’ to insert LPS into the OM. One would

surmise that an opening of the lateral gate (if that really were to occur) would demonstrate 'deprotection' – where these peptides would be more exposed to solvent. Furthermore, these regions displayed a classical shift in their isotopic distributions, characteristic of EX2 kinetics, which may be indicative of a binding event.

Another observation is peptides pertaining to the β -taco show an overall increase in deuterium uptake on binding to LPS. This is strange because the β -taco is assumed to be the domain which accommodates LPS to facilitate transport, therefore, one would expect peptides in the region 46-204 to be 'protected' to signal the presence of LPS interactions. This region also showcases bimodal isotopic distributions which alludes to EX1 or mixed EX1/EX2 (EXX) kinetics, suggesting concerted conformational motion slower than the HDX reaction. An open β -taco with LPS inside 'should' demonstrate EX2 kinetics, but this is not what is primarily observed (**Figure 5.3**). The domain is clearly influenced by an EX1 kinetic regime which also needs to be explained in order to understand the conformational changes within the complex.

5.2.2 MD approach for interpreting HDX results

As alluded to previously, HDX is a powerful method for observing the positions of ligand binding as well as conformational changes brought about by such binding. Additionally, MD affords the observation of and characterisation of structural detail which may shed light upon these experimental observations. Therefore, here presents a unique opportunity to employ MD *in silico* experiments to establish, in further structural detail, how LPS specifically affects LptDE dynamics to produce the observed HDX-MS results. To do this I developed an *in silico* method for comparing predicted LptDE conformational states with HDX experiment. Simply, this involves comparing

the solvent accessibility differences between two states of the protein (apo and bound) to help understand and interpret observed HDX data. These two ‘before-and-after’ states of the protein were my ‘best guess’ of the conformation of LptDE without LPS present (the apo state) and with LPS present (the bound state). I assessed the solvent accessibility of both states by considering only the interactions between the oxygen of the solvent and nitrogen of the backbone to model likelihood of ‘exchange’. It is important to understand that this method is limited to understanding EX2 kinetics of the system and the direct effects of substrate binding to the complex. (Later within this chapter, I employ hydrogen bond studies to understand the EX1 behaviour of the system).

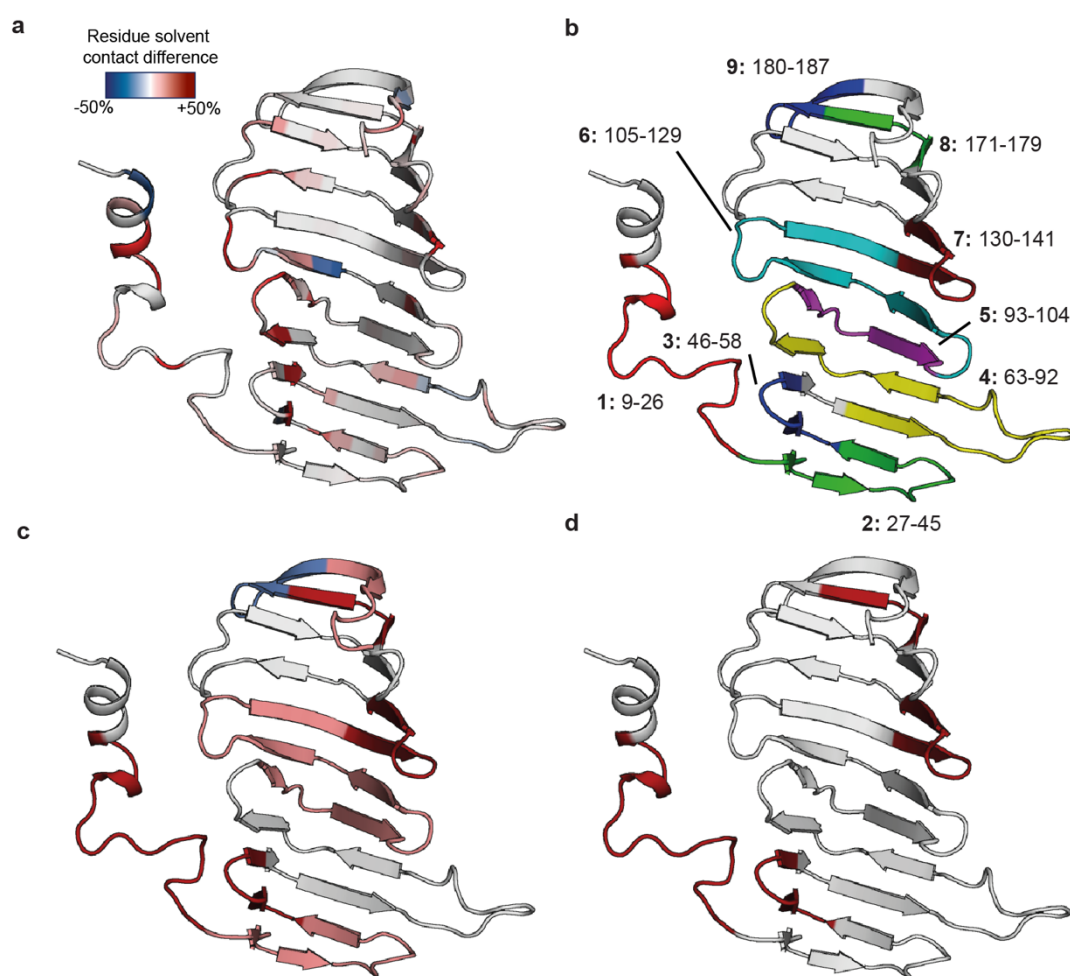


Figure 5.4 Overview of protocol to produce differential solvent maps. Figure adapted from Fiorentino *et al*⁵¹, all work here completed by myself. (a) A residue-

by-residue differential solvent uptake map is generated by subtracting occupancy values for the bound state away from the apo state, including all <95% significant values. **(b)** Experimentally determined peptide coverage of β -taco. Each peptide is coloured and numbered (1-9) for clarity. Averaging the values of the differential solvent contacts across these peptides (1-9) produces **(c)** which is termed the 'peptide differential solvent uptake map' and includes all <95% significant values. The peptides which have found to be at least 95% statistically significant are coloured in red, those that do not are shown in pink or light blue). **(d)** The resultant differential solvent uptake map after filtering out peptides which do not pass the 95% confidence interval test

Briefly, I average the resultant solvent accessibility difference for each residue over the same peptide coverage obtained in experiment by collaborators (FF, XQ, SM, JRB, CVR)⁵¹ (**Figure 5.4**), where a positive difference is used represent 'deprotected' peptides and a negative difference used to model 'protected' peptides accordingly. In a similar approach to the experimental work, I use Woodplots to extract peptides of 95% significance, which I subsequently map to the structure (for further details on the pipeline and statistical analysis, please see **methods**). I refer to the simulated HDX data produced in this pipeline as 'differential solvent contact plots'.

To attempt to replicate or imitate the HDX results obtained by my collaborators (FF, XQ, SM, JRB, CVR)⁵¹, I needed to generate a model for the apo-state LptDE structure and the LPS bound structure. For the apo state, trivially, this is 'just' the crystal structure of *Kp*LptDE (PDB ID: 5IV9)³¹⁹ as this structure was obtained without inclusion of any LPS substrate where the LptD β -barrel and β -taco domain are closed. Henceforth, I refer to this apo state structure as the 'closed state' of the LptDE complex.

Naturally, for the LPS bound state, obtaining a model was less straightforward and I made a number of assumptions in order to do so, which would either corroborate the

experimental data or would contrast the observations and therefore would require revision. To reiterate, deuterium uptake was noted to increase in the peptides pertaining to the β -taco itself and therefore, seems ‘counterintuitive’ to be caused by LPS binding within. Nevertheless, the β -taco is a likely site to specifically incorporate LPS to accommodate translocation into the OM as proposed by current mechanisms of transport³²⁵. Accordingly, then, I used alchemical methods³³¹ to grow Re-LPS inside the β -taco, opening the binding site up in the process.

The current model (also mentioned earlier)³²⁵ proposes that the putative lateral gate (located between β 1 and β 26) of the LptD barrel opens to allow passage of LPS for insertion. To model this, I templated the closed *Kp*LptDE structure onto one which had previously been modelled open (a *Salmonella typhimurium* LptDE structure)³²⁵

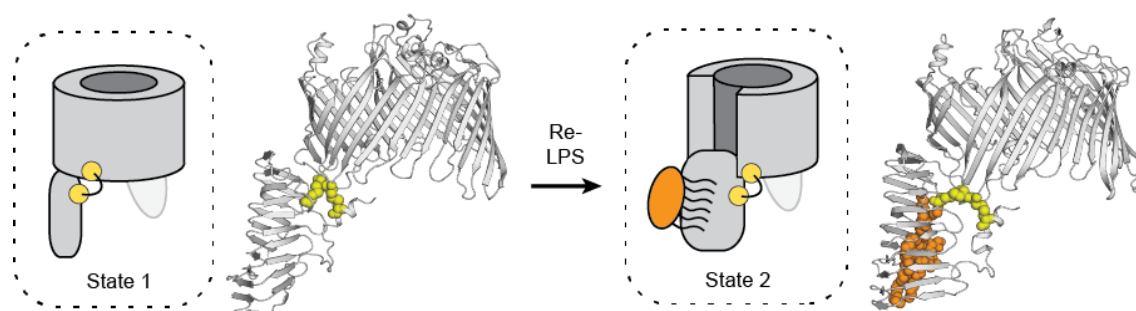


Figure 5.5 Predicted conformational change on binding to Re-LPS. Figure adapted from Fiorentino *et al*⁵¹, all work here completed by myself. Predicted conformational change on binding to Re-LPS. Left: apo ‘closed’ state and Right: lipid bound ‘open’ state. Cysteine bonds between Cys7–Cys696 and Cys149–Cys697 are represented as yellow spheres. Re-LPS coloured orange and shown as sticks.

I found that the movement of the beta strands coordinates in a concerted fashion with the opening of the β -taco by two disulphide bonds between cysteine residues at positions Cys7–Cys696 and (almost entirely conserved) Cys149–Cys697 (see **Figure 5.5**) for the structure of the apo ‘closed’ state and the lipid-bound ‘open’ state). I predict

therefore, that LptDE adopts this conformational change when binding LPS (and derivatives thereof, including Re-LPS) inside the soluble β -taco domain. I term the Re-LPS bound conformation of the protein the ‘open’ state of the LptDE complex.

Together, I ran five repeats of apo ‘closed’ state and the lipid bound ‘open’ state for 300ns, with the latter half (150 – 300ns) assessed to be converged using global protein solvation as a test parameter for convergence i.e. the average solvent contact over all backbone LptDE protein residues (**Figure 5.6a**).

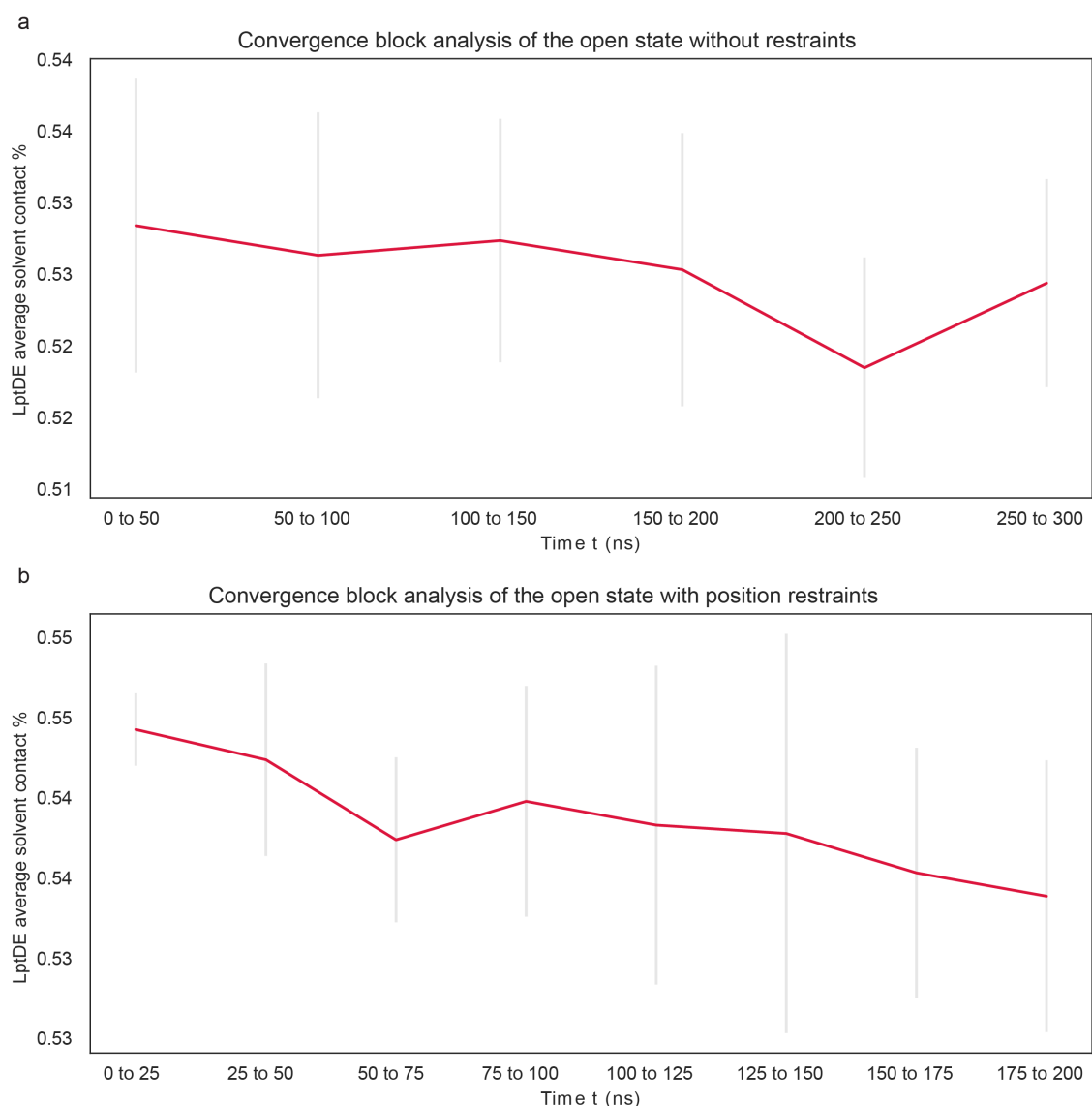


Figure 5.6 Block convergence analysis. Figure adapted from Fiorentino *et al*⁵¹, all work here completed by myself. (a) Convergence block analysis of the open Re-

LPS bound state without PLUMED restraints where convergence has considered to be reached at 150 ns. LptDE average solvent contact is calculated by averaging the total solvent contacts over all the residues of the protein. Here, the error bars represent the standard deviation over the five repeats **(b)** Convergence block analysis of the open state incorporating PLUMED upper and lower wall restraints upon the lateral gate, where a convergence is considered after 50 ns.

However, I found the lateral gate of the open β -barrel was 'unstable' for the duration of the simulation over three of the five repeats by closing back together over the course of the simulation period. However, in one of the replicates it remained open in a stable open state at an average lateral gate distance of ~ 2.25 nm (between four adjacent C α atoms on $\beta 1$ and $\beta 26$). So, the simulations were extended with PLUMED³²³ upper and lower wall restraints to ensure it remained in its stable open state (**Figure 5.6b**). I added these restraints added to an already converged simulation, and therefore only required only a further 50 ns equilibration for convergence (**Figure 5.6b**).

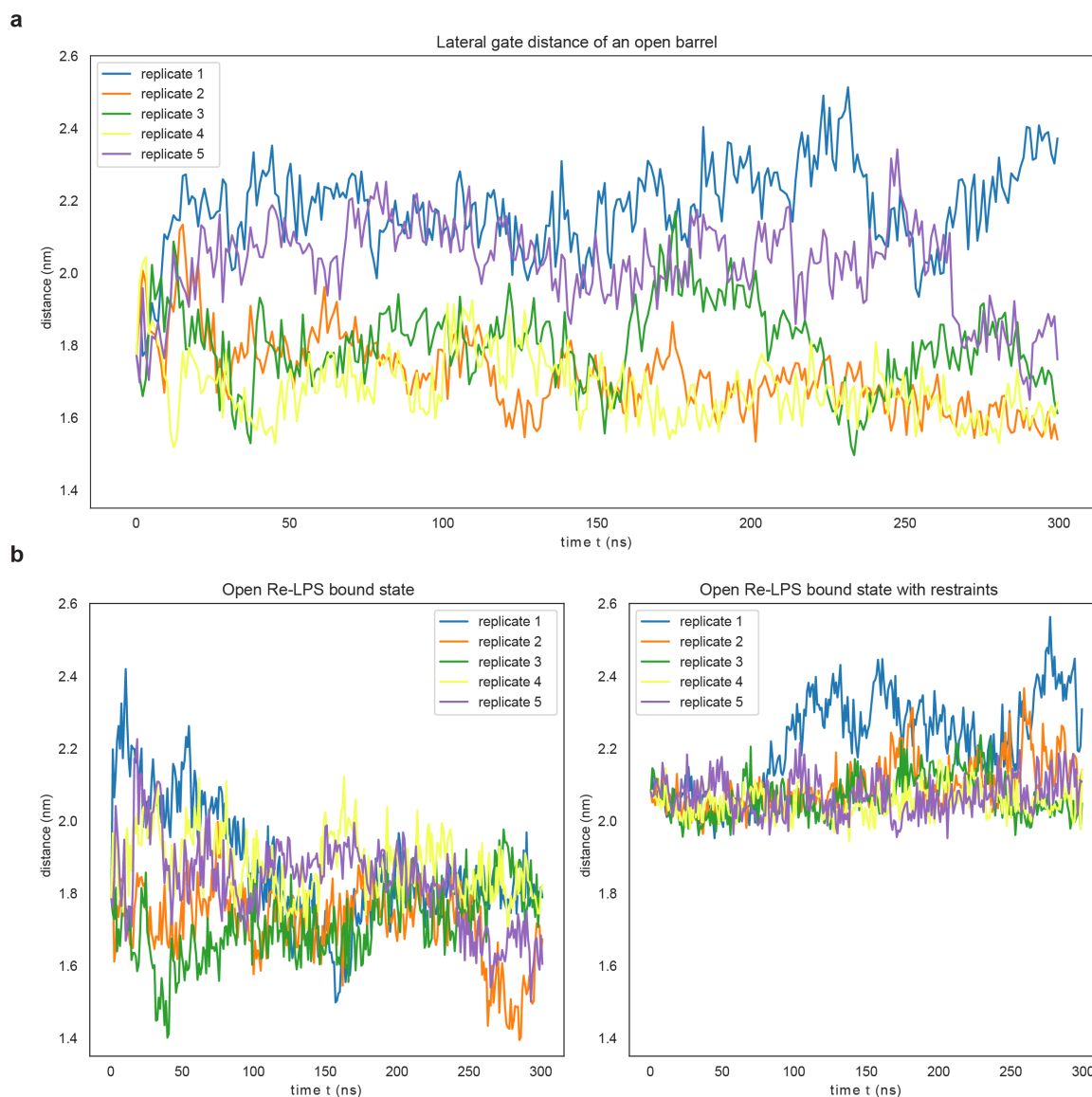


Figure 5.7 β -barrel lateral gate distance analysis. Figure adapted from Fiorentino *et al*⁵¹, all work here completed by myself. **(a)** The distance of the lateral LptDE β -barrel exit gate LptDE was measured over five replicates without Re-LPS bound to the β -taco. Over the simulation period replicate 1 lateral gate distance was consistent over after sampling a distance above 2.25 nm. **(b)** The lateral gate distance of the β -barrel without (left) and with (right) PLUMED restraints³²³.

5.3 Differential contact plot of LptDE bound with Re-LPS

The differential contact plot produced from the difference of the ‘open’ and ‘closed’ state (**Figure 5.8**) has a number of similar features to that of its experimental

counterpart, namely deprotected peptides in the jellyroll (peptides 46–58, 130–141 and 171–179) as well as protected peptide at the exit gate (peptide 232–251).

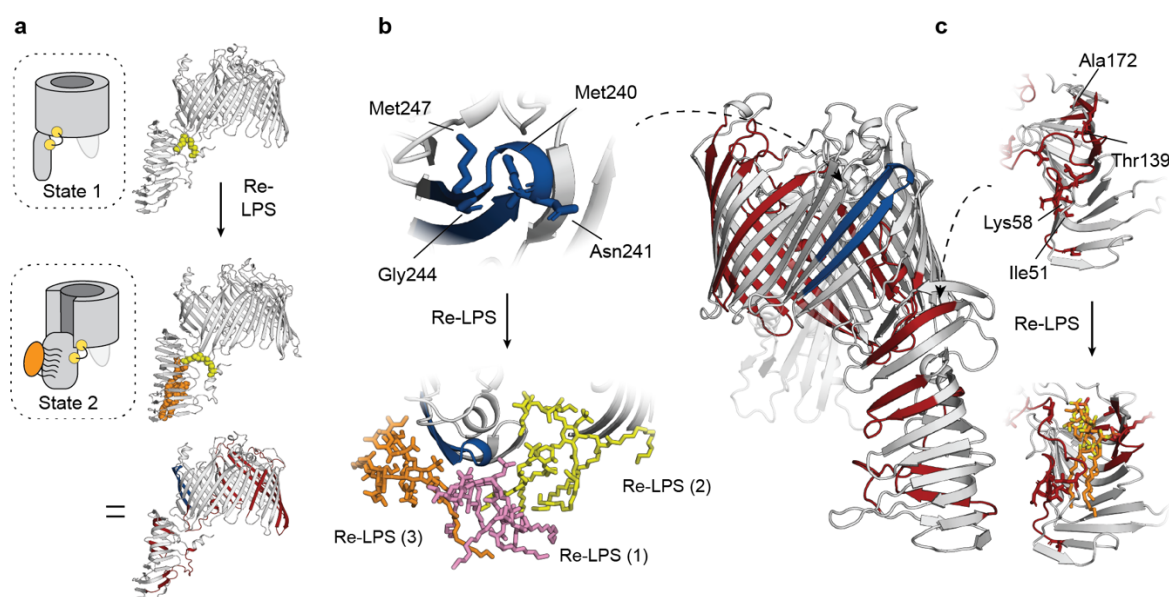


Figure 5.8 Differential solvent plot of LptDE on binding to Re-LPS. Figure adapted from Fiorentino *et al*⁵¹, all work here completed by myself. (a) Cartoon schematic representing the open and closed states of the structure depicted with their respective protein structures with resultant differential contact map. (b) Here the differential contact plot is mapped upon the LptDE structure generated from the difference between the ‘open’ and ‘closed’ state of the complex. Similar to the experimental HDX plots, regions which are more exposed to solvent are coloured red where regions which are shielded and less exposed are coloured blue. The lateral gate of the plot is highlighted here to draw attention to the particularly significant residues which contribute most the peptide being deprotected. Simulation snapshot of the lateral gate interacting with Re-LPS(1-3) (represented in sticks) which contributes to solvent shielding is shown. (c) Likewise, for the β -taco, residues highlighted as sticks are particularly involved in increasing solvent contact for the domain. A representative rendering of Re-LPS bound within the β -taco is also depicted here to illustrate how the lipid opens up the domain.

In a similar way to the peptides, each of the residues can be evaluated for its solvent contact difference, where only those residues of 95% significance are mapped onto the structure (Figure 5.9).

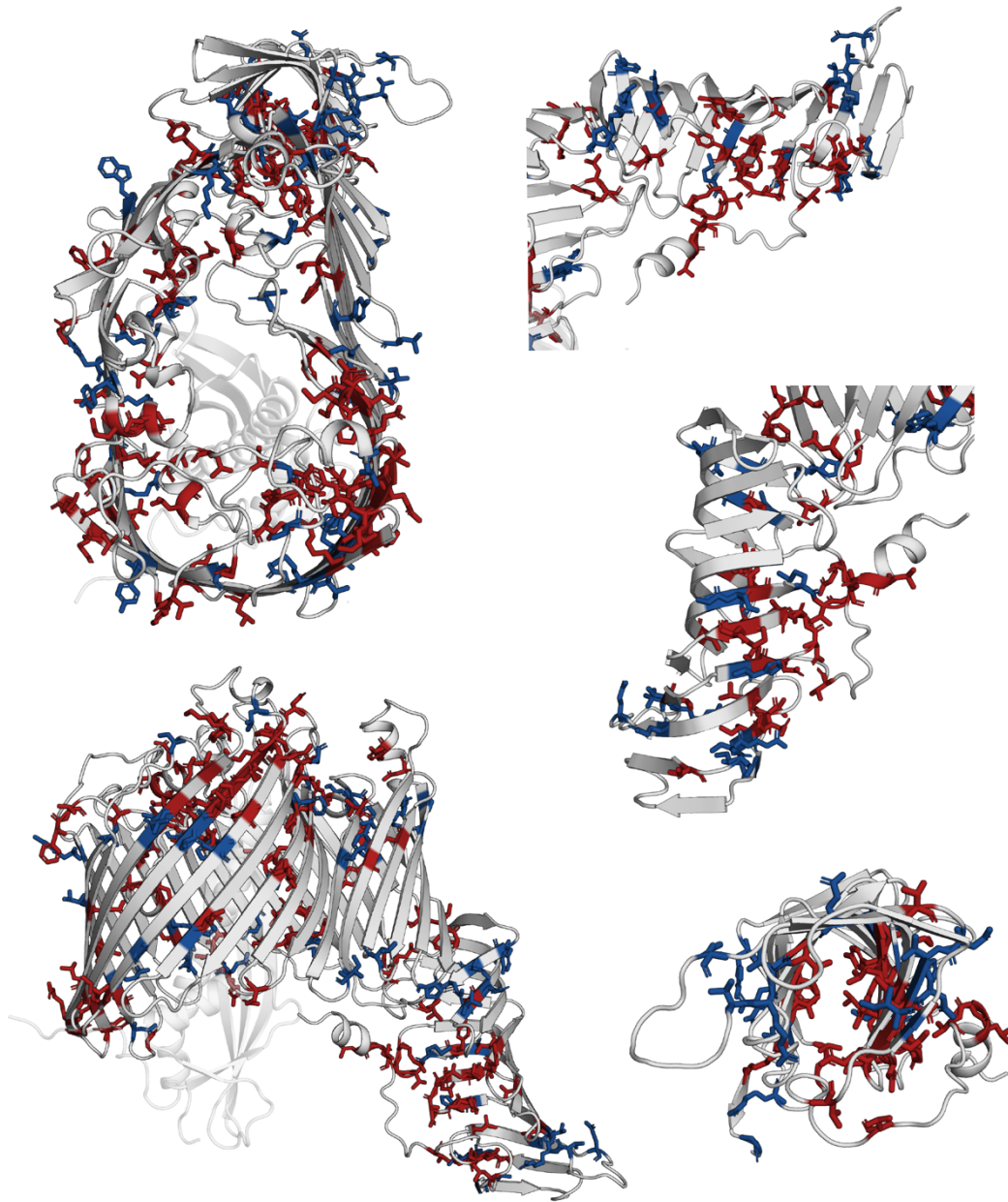


Figure 5.9 Residue-by-residue differential contact map for LptD produced by the difference between ‘open’ and ‘closed’ state. Figure adapted from Fiorentino *et al*⁵¹, all work here completed by myself. Different angles of the residue-by-residue solvent contact map. Residues highlighted in sticks have passed a greater than 95% confidence interval.

One difference of note between the differential solvent plot and the experimental HDX profile is that the extracellular β -barrel loops pointing towards the extracellular milieu

are solvated highly in my simulations. These regions will be probably more solvent exposed in the simulated bilayer than in detergent which the experiments were conducted in. Therefore, this would result in some peptides, when averaged over the residues across the peptide, (e.g. 394-425 and 462-497) to be overly deprotected (**Figure 5.9**).

Furthermore, from the residue-by-residue differential contact map, I highlight those residues, which are responsible for causing particular peptides to become statistically significant, in a stick representation (**Figure 5.8**). This model suggests that on binding to Re-LPS, the β -taco opens up and allows enough solvent access, particularly for inward facing residues on the sides of the β -taco (shown as sticks). This specific binding event causes a significant amount of deprotection in the β -taco by virtue of an open conformation. However, I observe only partial deprotection in our differential contact plot on peptides 46–58, 130–141 and 171–179 when compared to experiment.

The differences in the β -taco differential maps between experiment and MD can be accounted for by restricted sampling of Re-LPS along the β -taco, an inherent limitation of MD simulations. In solution there is a statistical distribution of different binding poses in the periplasmic domain which would be far too computationally costly to sample with MD. Therefore, this does indicate that the Re-LPS bound open state simulated does not fully reproduce the real conformation, however I argue that it is representative and helps to explain the experimental result

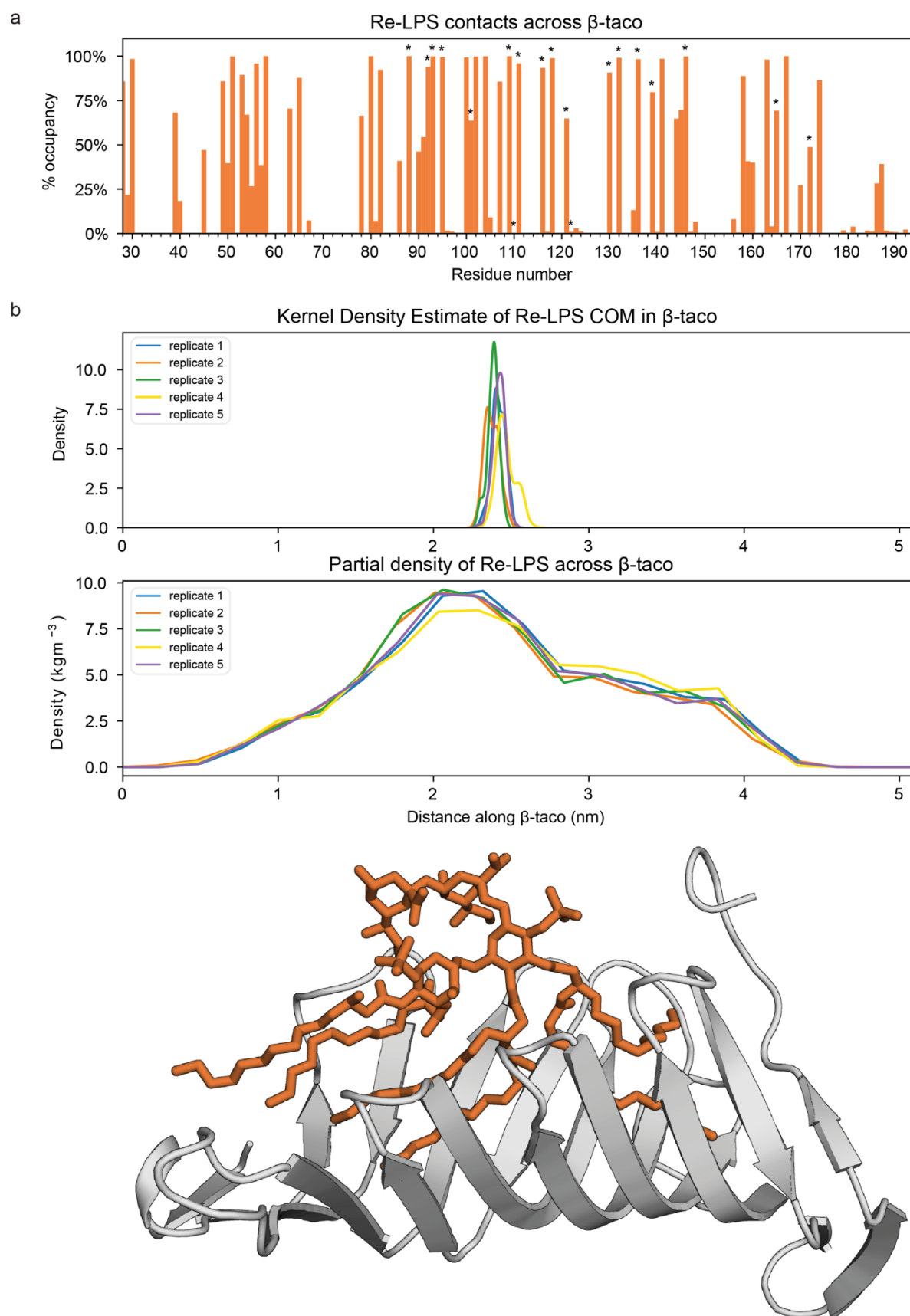


Figure 5.10 Re-LPS contacts and position within the β -taco. Figure adapted from Fiorentino *et al*⁵¹, all work here completed by myself. (a) The average %

occupancy of Re-LPS (heavy atoms) contacts made between the lipid and β -taco within 4 Å. Contacts made by the detergent in the 4Q35 structure are marked on with an asterisk. 90% of these contacts correlate with those made by Re-LPS. This shows good agreement between Re-LPS docked into the β -taco in my model and the position of the detergent from the structural data. **(b)** Top: The centre of mass (COM) position plotted as a kernel density estimate over the five repeats reveal minimal movement of Re-LPS during the simulations. Middle: Similarly, partial density of Re-LPS across the length of the β -taco reveals that the lipid position is consistent and non-diffuse across the five replicates. This accounts for the poor sampling of Re-LPS within the β -taco where the lipid does not sample different poses within the domain and consequently does not reproduce fully the deprotection observed in experiment (**Figure 5.11**). Bottom: cartoon representation of the β -taco for reference

I modelled Re-LPS into the β -taco using both C₈E₄ and LDAO detergents as a guide (without the presence of bound detergent the β -taco will rapidly close, it will not remain partially open – see **Figure 4.2** from **Chapter 4**). I envisaged that the LPS molecule would bind with all six tails in a row to permit the most subtle structural changes to the β -taco, and therefore an idealised conformation of Re-LPS to be then equilibrated in the MD simulations.

As mentioned above, I have quantified the residues of the β -taco in contact with Re-LPS, which we have plotted by percentage occupancy over the course of the simulation (**Figure 5.10a**). We have compared these contacts to those made by the detergent in the 4Q35 structure (marked by an asterisk) and found that 90% of residues which interact with detergent are also in contact with Re-LPS, demonstrating reasonable agreement between the detergent and Re-LPS binding sites. Furthermore, after analysing the centre of mass (COM) position of Re-LPS as well as the partial density of the lipid across the β -taco I can confirm that the diffusion is minimal over the course of Re-LPS bound simulations (**Figure 5.10b**). As suggested previously, this implies that the Re-LPS bound within the β -taco does not have enough simulation time to sample the entire soluble domain in different binding poses, which would

account for the discrepancy in deprotected coverage across this domain when compared to experiment.

However, to better clarify the comparison to the experimental data, I quantified the contacts made between Re-LPS within the β -taco and found that most of these contacts are associated with greater solvent accessibility (highlighted as sticks) when compared to the closed state (**Figure 5.11**).

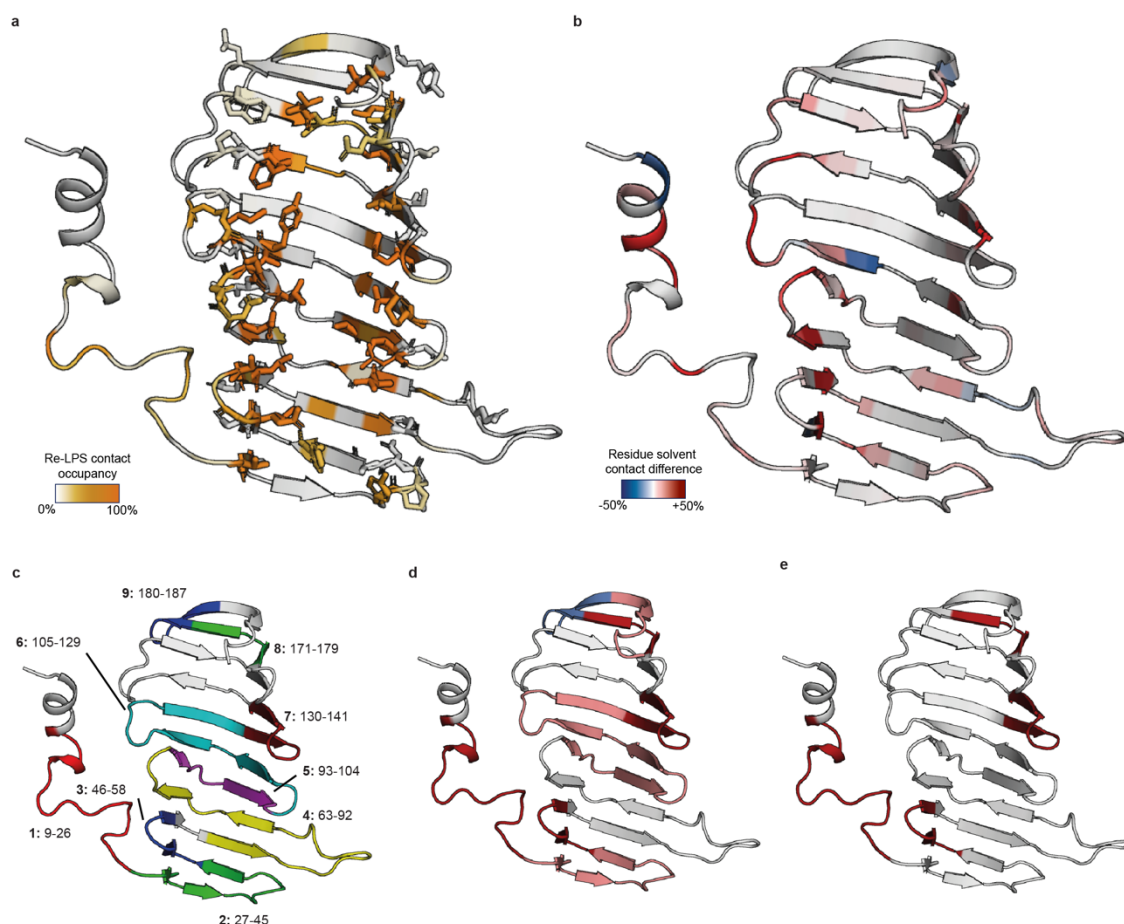


Figure 5.11 Comparison with the Re-LPS contacts on the β -taco with the differential solvent plot. Figure adapted from Fiorentino *et al*⁵¹, all work here completed by myself. (a) Structure of the β -taco indicating the contacts made with Re-LPS. Those residues highlighted in sticks correspond to residues which have also increased in their exposure to solvent and therefore contribute to the deprotection of these peptides. (b) As before, to compare to the contact map in (a) the residue differential contact is plotted here where all residues with less than 95% significance are visualised. (c) The experimental peptide coverage of the β -taco. Each peptide is coloured and number 1-9 to distinguish between areas which have been averaged over to produce (d). The peptides which pass a 95% confidence interval test are finally

plotted on the resultant differential contact map **(e)**. Here, I can observe that Peptide 4 reveals no overall increase in average solvent contact which can be rationalised by comparing with the comparatively low Re-LPS contacts in this area, which has been shown to increase solvent accessibility. Slight increase in solvent contacts in the region mostly likely is 'averaged out' over 29 residues and therefore shows not overall increase across the peptide. Low contacts in this region are associated with poor sampling of Re-LPS within the β -taco which has shown to be rigid over the short timescales of the simulation (see **Figure 5.10**). Also, I observe that peptide 9 reveals a small overall protection which is attributed to transient interactions with the inner leaflet of the OM (**Figure 5.13**).

Therefore, there is a direct correlation between specific residues contacting Re-LPS and solvent access. After averaging these differential residue values over the peptide coverage (**Figure 5.11c**), and without applying statistical significance test on the peptides, I observe that the majority of the β -taco has on average greater solvent exposure (**Figure 5.11d**) (peptides which pass the 95% confidence interval are shown in red, those that do not are shown in pink). Peptide 4 (**Figure 5.11c**) shows no average increase in solvent uptake at all which is due to very few Re-LPS contacts correlated with an increase in solvent access, especially being averaged over comparatively large residue coverage (29 residues). This implies that this area would become more solvent exposed given more Re-LPS sampling in that region, however, this is not possible due to the short timescales of the simulation as well as the rigidity of Re-LPS in the β -taco (see above). Thus, after I apply the 95% confidence interval test on the peptides pertaining to the β -taco, it results in the final coverage observed in (**Figure 5.11e**).

It is also worthwhile to note that peptide 9 shows a slight overall protection owing to infrequent interactions with the inner leaflet of the outer membrane. As explained later (**Figure 5.13**), this is caused by Re-LPS being trapped by the open lateral gate of the β -barrel causing the lipid to laterally penetrate the inner leaflet.

In conclusion, therefore, my model indicates that in order to facilitate binding of Re-LPS the β -taco opens up which in turn increases solvent accessibility of inward facing residues Ile51, Lys58, Thr139 and Ala172). Furthermore, this observation helps to account for the role of EX2 kinetic component of the mixed EXX kinetics observed in the experimental data obtained by the collaborators (FF, XQ, SM, JRB, CRV)⁵¹.

Regarding the lateral gate, another counterintuitive observation suggests that an open barrel causes less exposure to solvent and therefore results in protection of the peptide 232–251. Using the solvent contact difference map on an individual residue resolution, we found that the two significant residues on the outside of the barrel (Asn241, Gly245) particularly interact with Re-LPS (labelled Re-LPS(1–3)). Two of the Lipids labelled Re-LPS(1) and Re-LPS(2) (which itself is caught within the open barrel) protect Asn241, where lipids Re-LPS(1) and Re-LPS(3) protect Gly245 as shown in **(Figure 5.8)**.

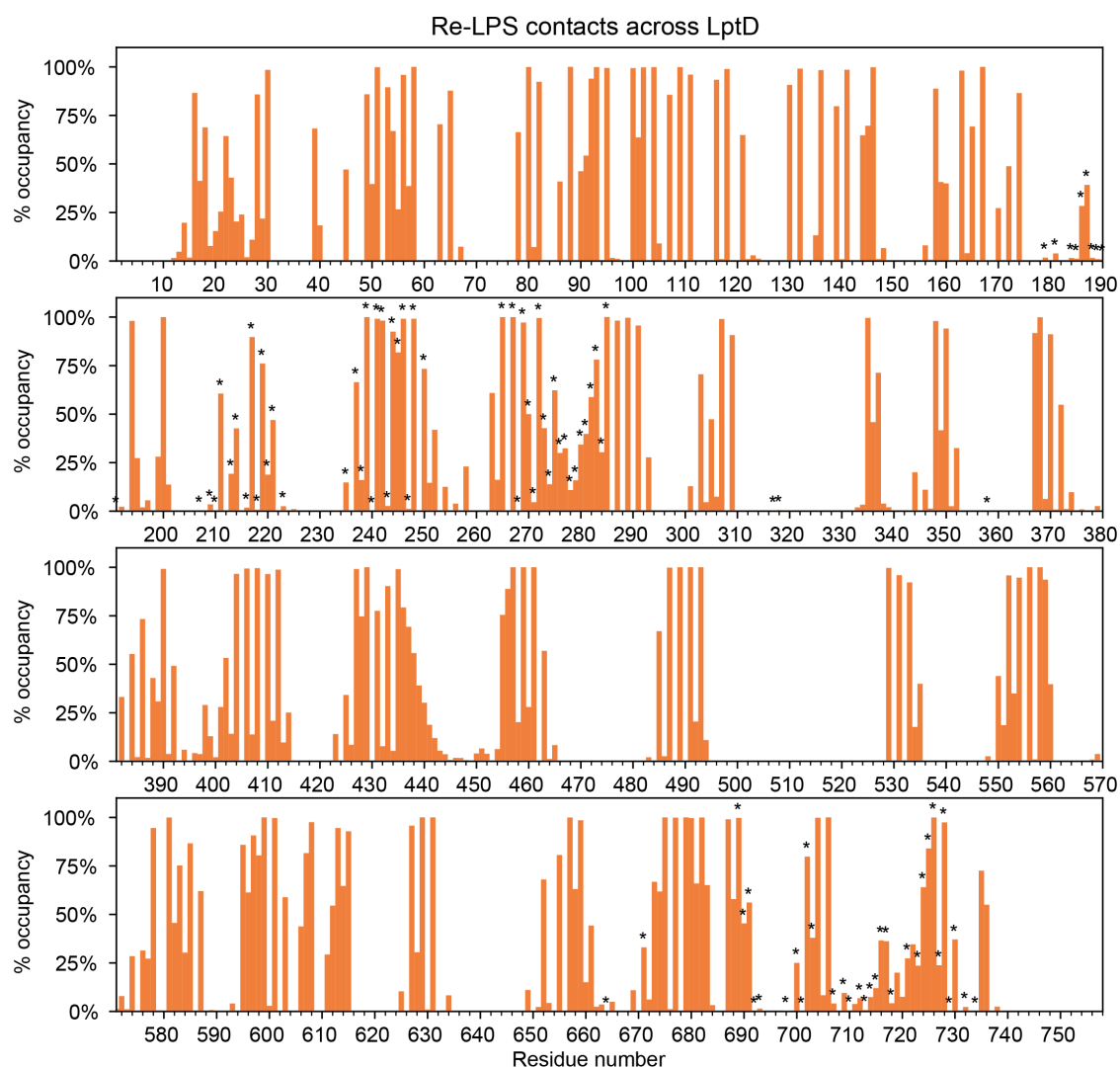


Figure 5.12 Re-LPS contacts made with LptD. Figure adapted from Fiorentino *et al*⁵¹, all work here completed by myself. Average % occupancy of contacts made with Re-LPS and LptD over five repeats. Here, those contacts made with Re-LPS(1-3) referenced in (Figure 8b, 13d) are marked with an asterisk.

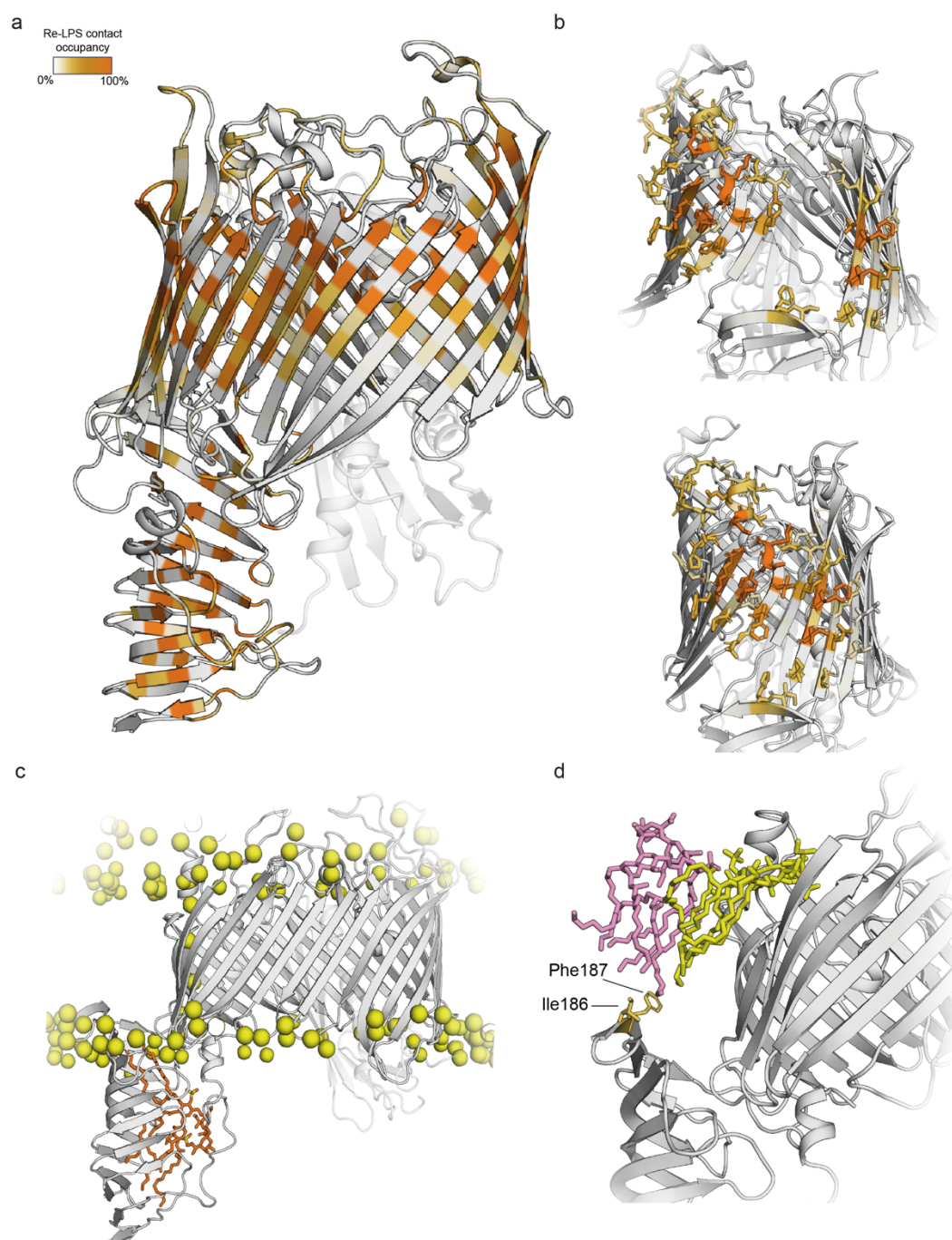


Figure 5.13 Contact map of Re-LPS plotted on LptDE structure. Figure adapted from Fiorentino *et al*⁵¹, all work here completed by myself. (a) Re-LPS % occupancy contacts averaged over five repeats and mapped onto LptDE (b) Residues which contact with three Re-LPS lipids referenced in **Figure 5.8 (Re-LPS(1-3)) are represented as sticks and mapped onto both the closed and open states of the β -barrel for comparison. (c) Side profile view of LptDE where the C-terminal strand of the β -taco interacts with the inner leaflet of the outer membrane (**Figure 5.11**). (d) Re-LPS lipids contact residues 186 and 187 associated with the lateral gate of the β -barrel. The lateral gate appears to pull Re-LPS lipids (coloured yellow and pink) into**

the β -barrel which causes them to sink lower into the inner leaflet and facilitates a transient interaction with the terminal strand of the β -taco.

For completeness and to characterise the contacts made by Re-LPS(1-3) in the context of Re-LPS binding with LptD, I have included the average Re-LPS contacts on LptD by percentage occupancy over the course of the simulation (**Figure 5.12**). I have also highlighted (with asterix) on the contact plot the contacts made by the three Re-LPS lipids reference in **Figure 5.8b**. For clarity, I have mapped these total Re-LPS contacts onto cartoon representations of LptDE (**Figure 5.13a**). Additionally, I have shown (in a stick representation) the residues which interact with the three Re-LPS lipids referenced in **Figure 5.8b**.

As noted earlier, these lipids Re-LPS(1-2) are responsible for transiently interacting with the C-terminal strand of the β -taco. The open gate causes Re-LPS lipids (shown in yellow and pink) to be laterally pulled into the β -barrel sinking them into the inner leaflet of the membrane and allowing the tails to interact with the C-terminal strand of the β -taco. This may suggest how Re-LPS gets laterally extruded into the OM during the translocation process. Furthermore, this observation seems to imply the role of the β -taco facilitating a 'hydrophobic hole' in the inner leaflet of the outer membrane to assist installation of LPS in the OM³²⁵.

Later, I observe that, within the course of the simulation (150-300 ns), Re-LPS(1-3) moves from the lateral gate and out towards bulk membrane. This diffusion away from the putative gate means that the peptide 232-251 is not shielded from solvent as it was before and does not show 'protection' in the differential contact map (**Figure 5.14**).

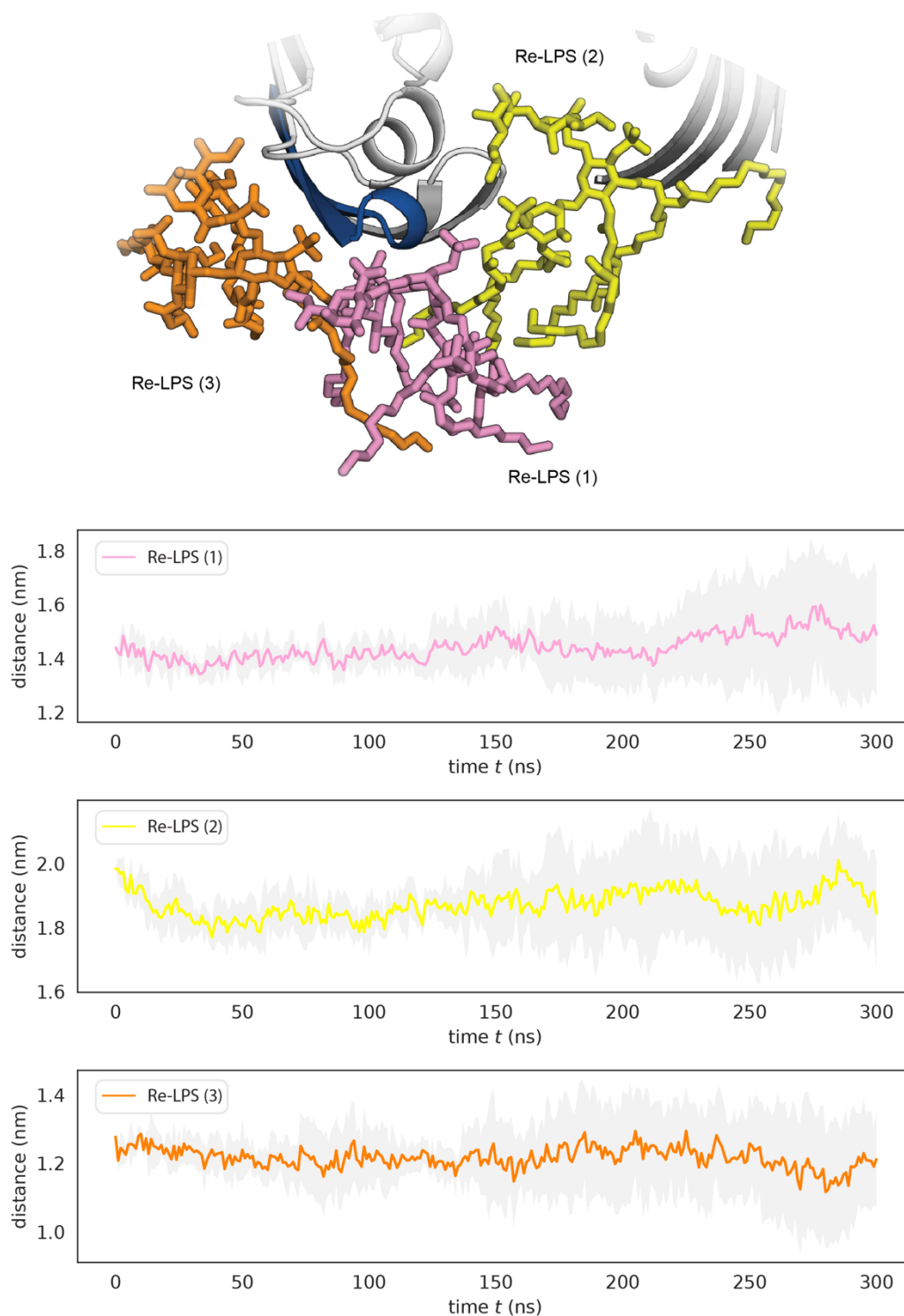


Figure 5.14 Distance of Re-LPS away from lateral gate. Figure adapted from Fiorentino *et al*⁵¹, all work here completed by myself. Top: Rendered image of simulation capturing the interaction of Re-LPS(1-3) with the lateral gate peptide (232-251). Bottom: Distance from the geometric centre of the peptide (232-251) to the geometric centre of the lipids plotted over the five repeats for Re-LPS (1, 2 and 3). The standard deviation is shaded grey between the replicates.

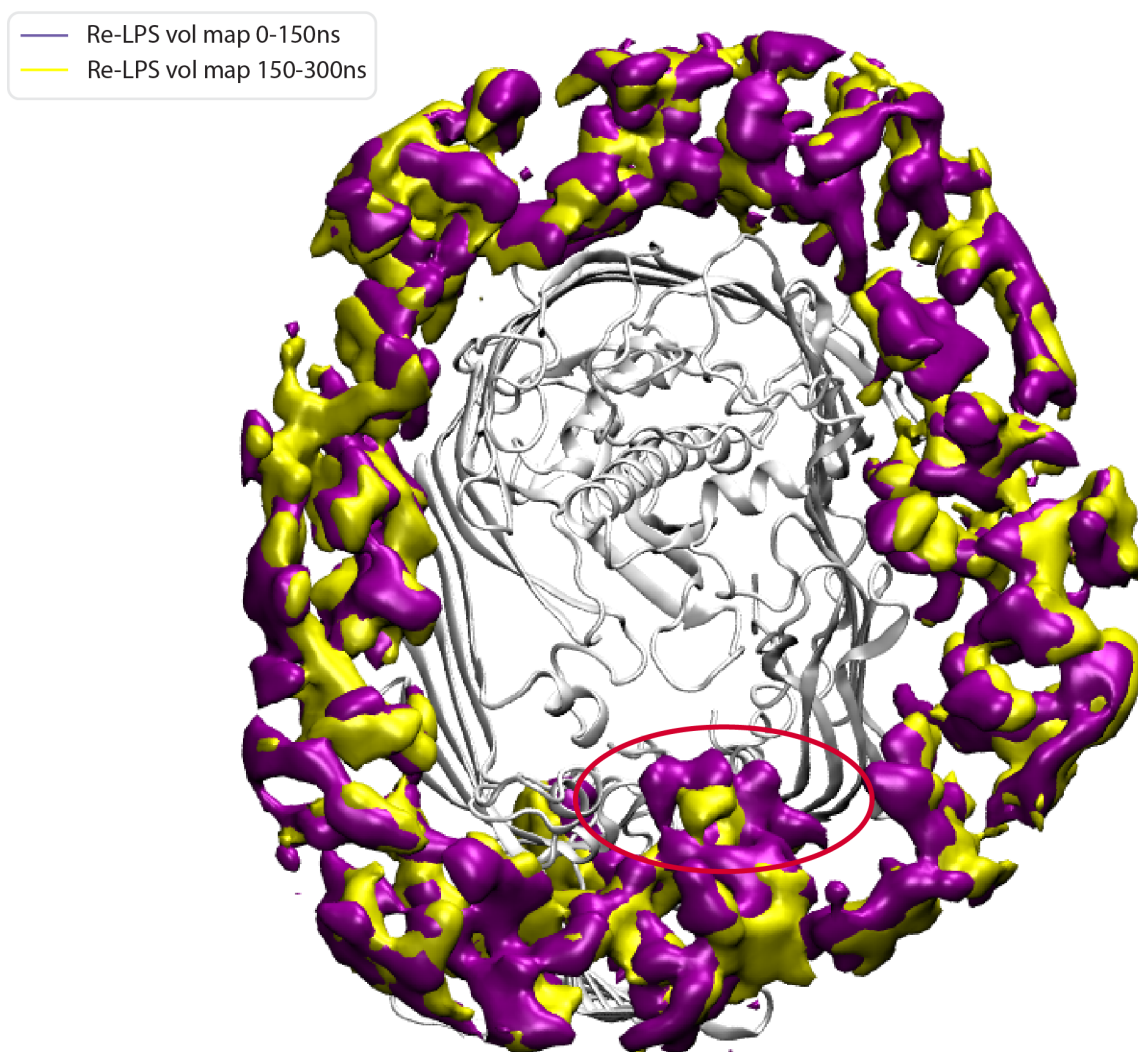


Figure 5.15 Vol-map density of Re-LPS. Density of Re-LPS plotted as a vol-map over the first half of the simulation (0 - 150 ns) coloured in purple compared to the density in the second half of the simulation (150 - 300 ns) coloured in yellow. The circle highlights the density of the Re-LPS lipids engaging with the lateral gate.

Additionally, this can also be appreciated from vol-map density of the simulation (shown in yellow) withdrawing from the exit gate in the latter half of the simulation (150 – 300 ns) compared to the purple density corresponding to the first half of the simulation (0 - 150 ns) (**Figure 5.15**). Moreover, this may suggest how Re-LPS gets laterally extruded into the OM during the translocation process as these lipids do not

just 'get stuck' in this position but, rather, can slowly diffuse away and get installed within the OM.

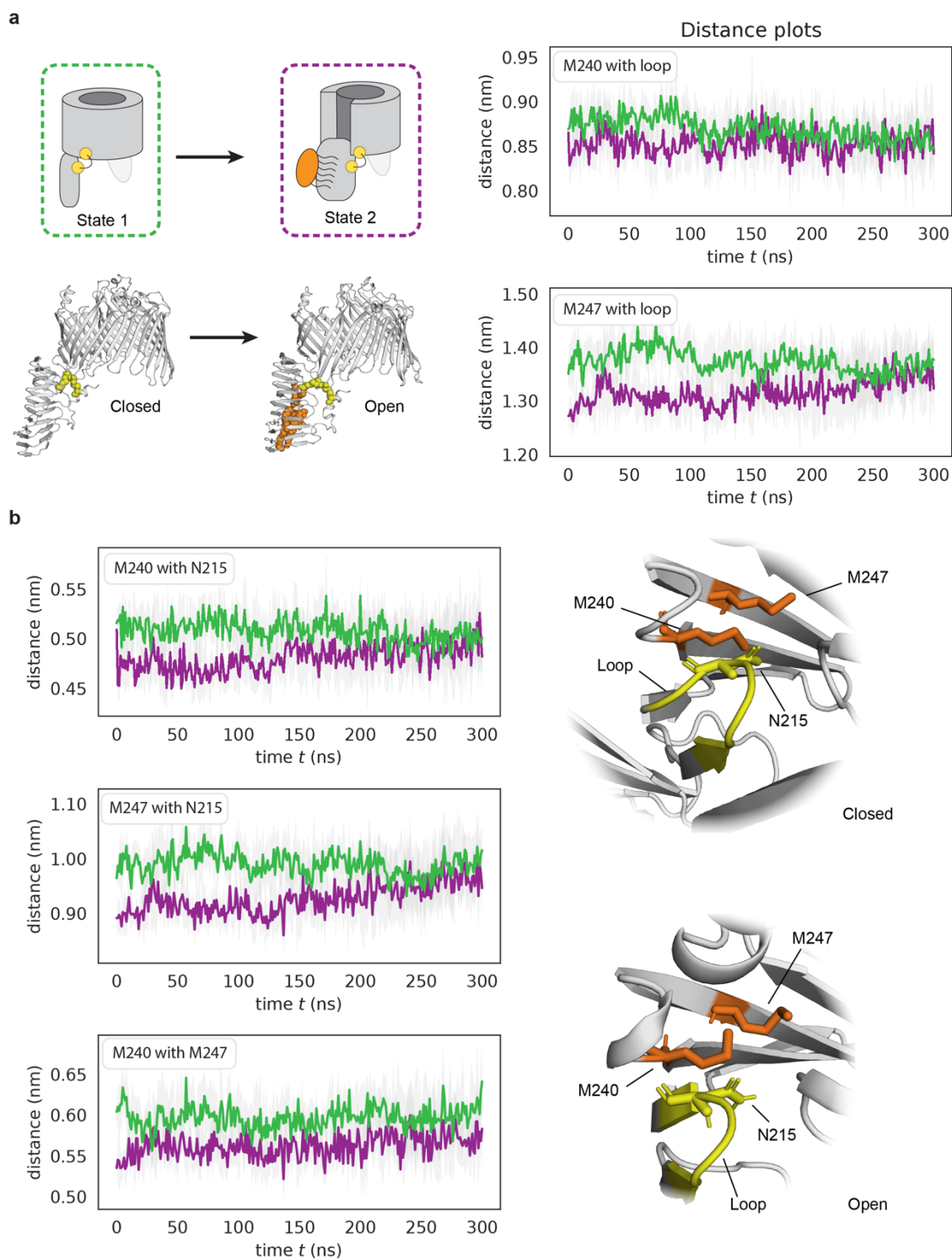


Figure 5.16 Residue-Residue distance plots. Figure adapted from Fiorentino *et al.*⁵¹ all work here completed by myself. (a) Left: schematic representation of the

apo closed state (green) and the Re-LPS bound state (purple). Right: plots measuring the distance between residues Met240 and Met247 from the $\beta 1/\beta 2$ loop of the β -barrel. **(b)** Left: residue distance plots between Met240 and Asn215 (from the $\beta 1/\beta 2$ loop), Met247 and Asn215 as well as between Met240 and Met247 where they are being drawn closer together. Right: structures of the open and closed state conformation of LptD. Here Met240 and Met247 are coloured in orange and Asn215 and the $\beta 1/\beta 2$ loop are coloured yellow.

When the barrel gate opens, I can observe that the loop between beta strands $\beta 1$ and $\beta 2$ (in particular, residue Asn215 of the loop) bending towards Met240 inside the barrel which further increases contact with to Met247 sterically shielding these residues (**Figure 5.16**). Therefore, I can conclude that local Re-LPS interactions with the exit gate are not wholly responsible for solvent shielding but also include contributions from small internal conformational changes. Again, as highlighted with the observations made regarding the β -taco, the differential contact map helps to rationalise the HDX EX2 kinetics observed by peptide 232–251, which can primarily be accounted for by steric effects.

It is also important to note that the differential HDX plot for a lateral gate distance of less than 2 nm did not reveal protection of the peptide 232–251. This implies that the opening of the gate to at least 2.25 nm is required to insert LPS into the OM.

5.4 LptE solvent differential HDX plots

Differential HDX plots were generated comparing the open and closed states of LptDE with the additional alpha-helix modelled in and found no statistically significant peptides on LptE to compare with experiment. In particular, the alpha helix was found to be highly flexible in both the closed and open states of the protein. It is likely that

the timescale of the simulation is not long enough to realise significant solvation differences in such a dynamic region of the protein.

5.5 LptDE and thanatin differential HDX plot

To further understand the conformational behaviour of the LptDE complex, it is important not only to characterise the structural mechanics of substrate binding but also changes induced on inhibitor binding. The cyclical peptide thanatin has shown antimicrobial activity against *Escherichia coli* and *Klebsiella pneumoniae*³³⁰. Furthermore, studies have also demonstrated strong binding of thanatin with β -taco domains LptA and LptD from *E. coli*³³⁰.

Subsequently, collaborators (FF, XQ, JRB, CRV) probed inhibitor binding effect using HDX-MS (**Figure 5.17**). They noted that peptides from the centre of the β -taco towards the C-terminal (near the exit gate of the β -barrel) display a different set of kinetics, EX1/EXX, compared to the N-terminal region (peptides 27–42 and 27–45) which demonstrated EX2 kinetics.

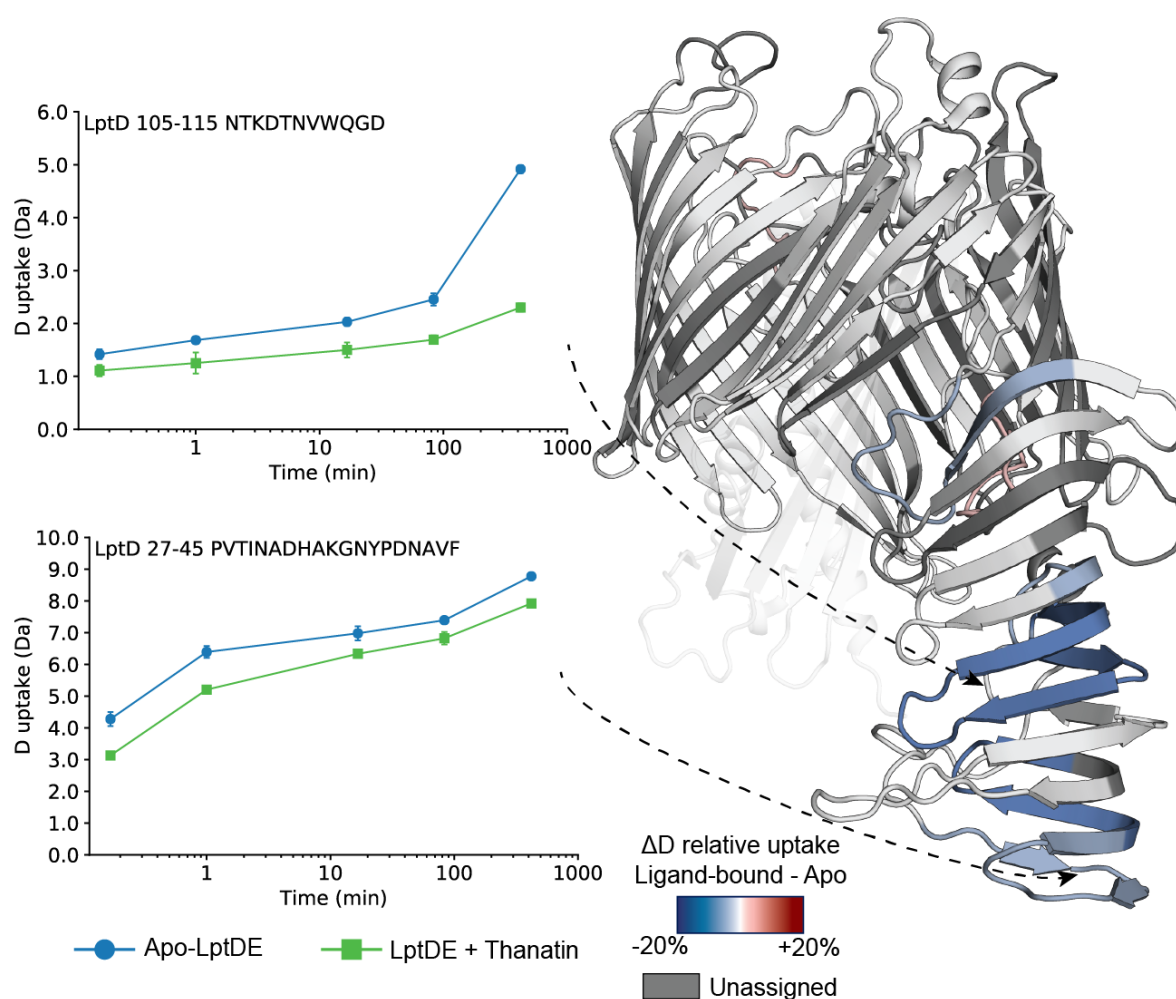


Figure 5.17 HDX plot of LptDE binding to the inhibitor thanatin. Figure modified from Fiorentino *et al*⁵¹, all work here completed by collaborators (FF, XQ, SM, JRB, CVR). As before, representative peptides (27–45, β -taco N terminus; 105–115, β -taco central region) deuteration uptake plotted as a function of labelling time for apo and thanatin-bound LptDE (purple circles and green squares respectively). The error bars represent standard deviation (nbiological = 2; ntechnical = 3). These peptides are mapped upon the LptDE structure according to the difference in deuterium uptake at the 16.67 min labelling time, where only peptides which demonstrate a significant difference in deuteration are shown. Respectively, red peptides emphasise an increase in HDX and blue peptides represent a decrease.

Alignment of thanatin to the β -taco formed the basis of further *in silico* studies to determine the most likely binding site on the β -taco. Due to such marked protection across the β -taco, it is not clear 'where' thanatin binds, whether it is to the N- or C-terminal of the domain.

Therefore, to investigate the likelihood of thanatin binding to both the top and bottom of the β -taco, I docked thanatin to the N- and C-terminal of the β -taco and assessed the stability of the docking to either position respectively. Thanatin was positioned at either end of the LptD β -taco by aligning *Ec*LptA (which has been shown to bind thanatin)³²⁹ with the domain.

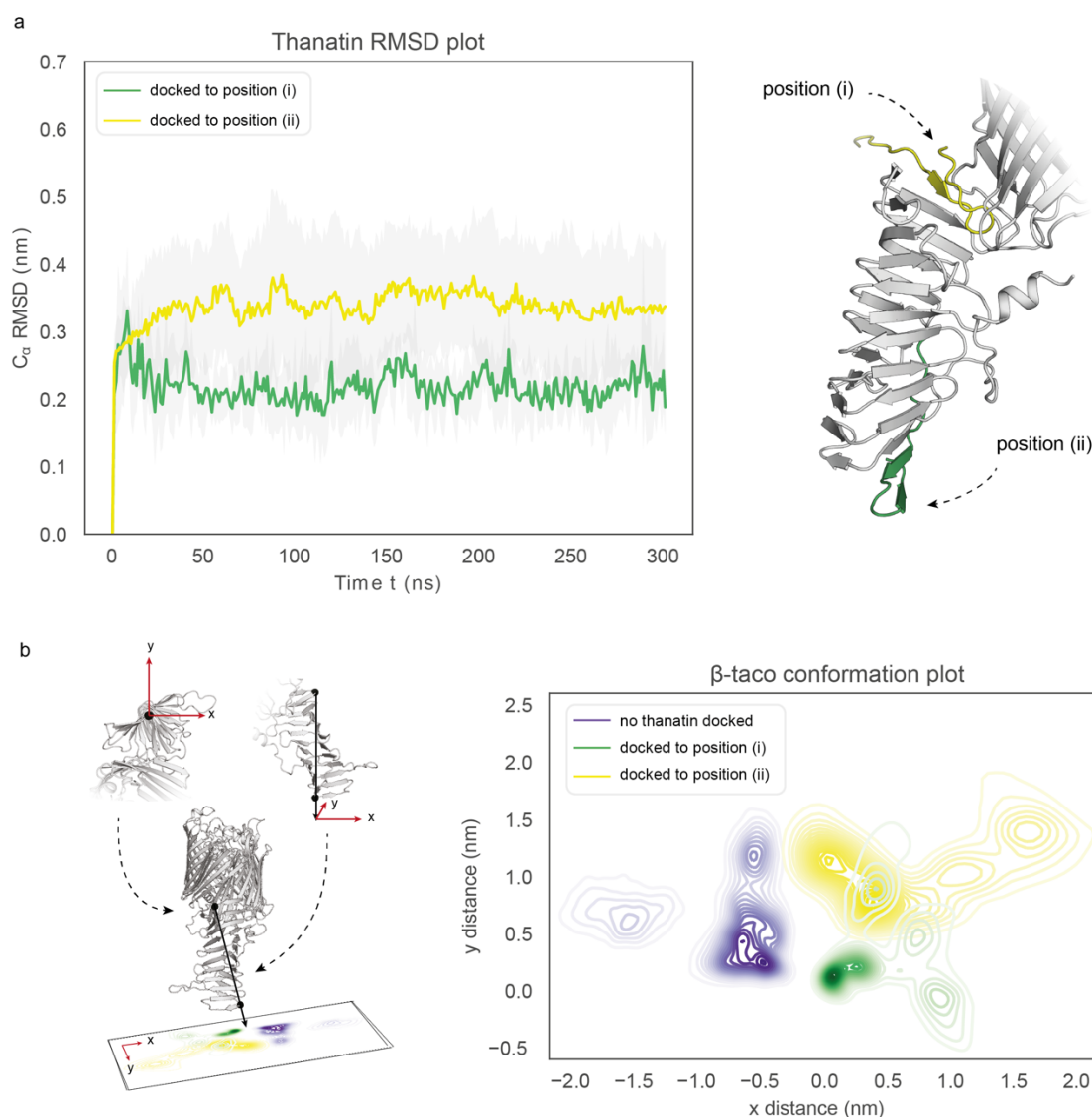


Figure 5.18 Stability of thanatin bound to the closed state of LptDE. Figure adapted from Fiorentino *et al*⁵¹, all work here completed by myself. (a) RMSD plot of thanatin position at N-terminal (i) of the β -jellyroll without Re-LPS bound (closed conformation) compared to being positioned at the C-terminal (ii). Here, standard deviation is coloured grey calculated over the five replicates. **(b)** Left: β -taco vector drawn through the C_{α} of residues Asp33 and Pro190, where x, y displacement is depicted as red arrows. Right: 2D kernel density plot of the x and y displacement of the β -taco vector mapped relative to the origin 0.0 nm, 0.0 nm (which is defined as the

'resting' position of β -taco in the crystal structure). The plot maps the conformational space sampled by β -taco in different states, where the apo closed state is coloured purple and the (i) bottom and (ii) top thanatin docked states are coloured green and yellow respectively.

The root mean square deviation (RMSD) plot (**Figure 5.18a**) shows that thanatin bound to the C-terminus of the β -taco (the membrane-facing end; RMSD = 0.34 ± 0.1 nm) was found to have a higher RMSD and a greater standard deviation compared to when it was docked to the N-terminal end of the β -taco (RMSD = 0.22 ± 0.05 nm).

I also studied the conformation space sampled by the β -taco (**Figure 5.18b**) and found that a large deviation in the position of the β -taco from the crystal structure (distance = 1.13 ± 0.5 nm) is caused when thanatin is docked at the C-terminal of the β -taco. Conversely, when docked to the N-terminal, the β -taco seems to explore less conformational space only deviating (distance = 0.65 ± 0.2 nm) from its original position when thanatin is not bound.

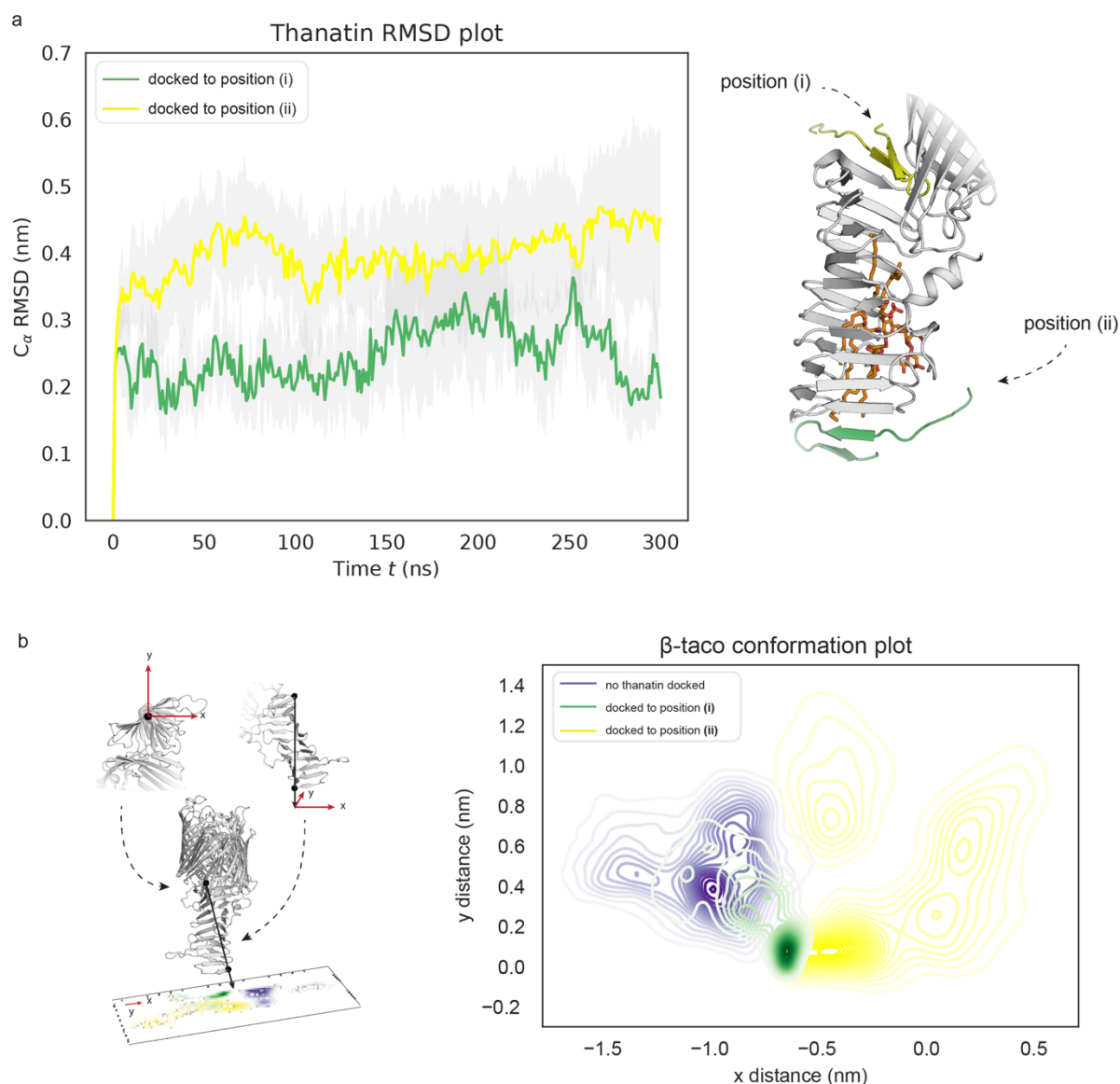


Figure 5.19 Stability of thanatin bound to the open state of LptDE. Figure adapted from Fiorentino *et al*⁵¹, all work here completed by myself. (a) RMSD plot of thanatin position at N-terminal (i) of the β -jellyroll with Re-LPS bound (open conformation) compared to being positioned at the C-terminal (ii). Here, standard deviation is coloured grey calculated over the five replicates. (b) Left: β -taco vector drawn through the C_{α} of residues Asp33 and Pro190, where x, y displacement is depicted as red arrows. Right: 2D kernel density plot of the x and y displacement of the β -taco vector mapped relative to the origin 0.0 nm, 0.0 nm (which is defined as the 'starting' position of β -taco in the open state structure). The plot maps the conformational space sampled by β -taco in different states, where the apo closed state is coloured purple and the (i) bottom and (ii) top thanatin docked states are coloured green and yellow respectively.

Thanatin at the lateral gate position is similarly as unstable when docked to the open state β -taco with Re-LPS is bound inside, following the same trend observed with the RMSD plots as well as the conformational plots (**Figure 5.19**). These observations imply that the β -taco would have to shift considerably and therefore unfavourably to bind thanatin at the C-terminal end. It is likely, then, that thanatin binding takes place at the more accessible N-terminal position on the β -taco.

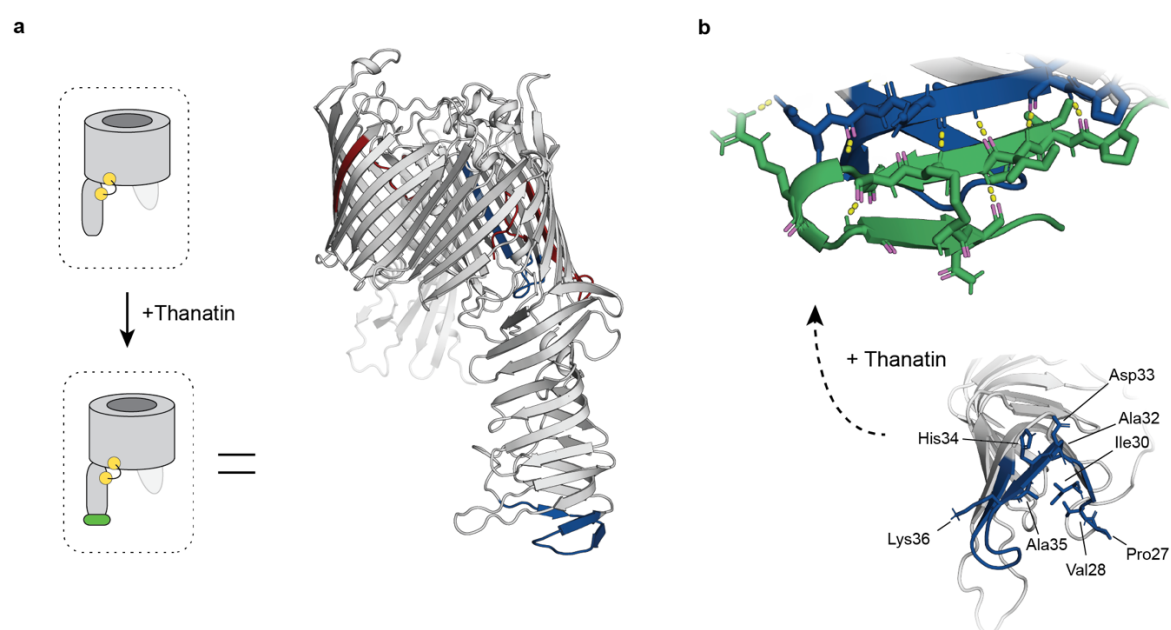


Figure 5.20 Differential contact plot of LptDE binding to the inhibitor thanatin. Figure adapted from Fiorentino *et al*⁵¹, all work here completed by myself. **(a)** Cartoon schematic representing the apo closed state and thanatin bound state of the complex depicted with their respective protein structures and resultant differential contact map. **(b)** Top: the final frame of the simulation with thanatin bound to the N-terminal of the β -taco, where the hydrogen bonds between the inhibitor (coloured green) and the hydrogen bonds are highlighted by yellow dashed lines with the residues drawn as sticks. Bottom: Here the differential contact plot is mapped upon the LptDE structure generated from the difference between the apo closed state and thanatin bound state of the complex. Like the experimental HDX plots, regions which are more exposed to solvent are coloured red where regions which are shielded and less exposed are coloured blue. The N-terminal region of the β -taco of the plot is highlighted here to draw attention to the particularly significant residues which contribute most the peptide being deprotected which are the same as those involved with the hydrogen bonding with thanatin.

Additionally, the differential solvent plot of thanatin bound to the closed state of the protein (**Figure 5.20**) reveals that the residues which cause the peptide to be protected are those that primarily form hydrogen bonds with thanatin (residues Val28, Ile30, Ala32 and Asp33). The MD finding corroborate well with the EX2 experimental kinetics deduced by collaborators (FF, XQ, SM, JRB, CRV) found with the N-terminal peptide of the β -taco⁵¹.

5.6 Conformation dynamics of the complex with both the substrate LPS and the inhibitor thanatin

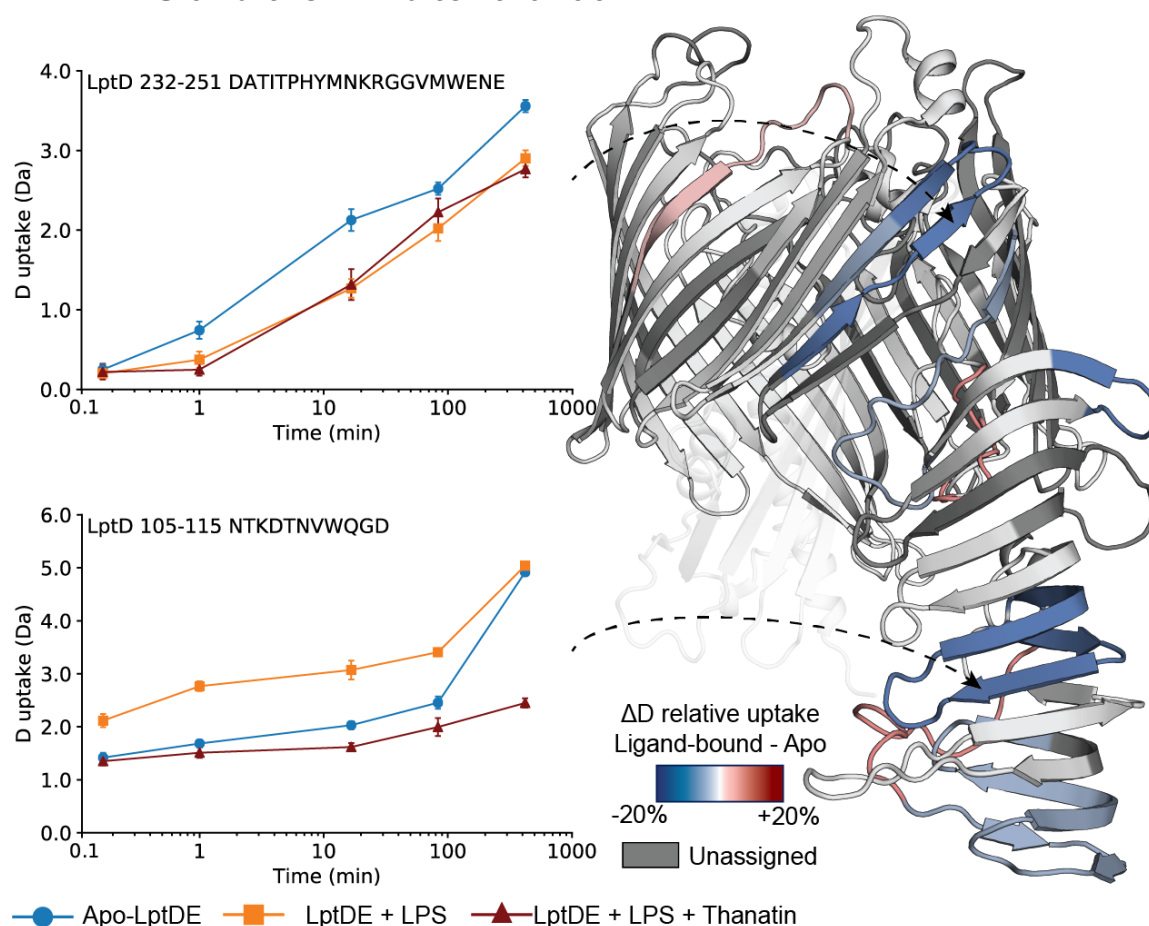


Figure 5.21 HDX plot of LptDE binding to the inhibitor thanatin and Re-LPS. Figure modified from Fiorentino *et al*⁵¹, all work here completed by collaborators (FF, XQ, SM, JRB, CVR). As before, representative peptides (105–115, β -taco and 232–247, β -barrel) deuteration uptake plotted as a function of labelling time for apo and thanatin-bound LptDE (purple circles and green squares respectively). The error bars represent standard deviation (nbiological = 2; ntechnical = 3). These peptides are mapped upon the LptDE structure according to the difference in deuterium uptake at the 16.67-min labelling time, where only peptides which demonstrate a significant

difference in deuteration are shown. Respectively, red peptides emphasise an increase in HDX and blue peptides represent a decrease.

To probe cooperative affects between conformational changes observed by substrate and the inhibitor individually, the combined effect of both LPS and thanatin was observed with LptDE. HDX experiments conducted and analysed by collaborators (FF, XQ, SM, JRB, CRV) generated a profile with similar features to both previous HDX data of Re-LPS binding and thanatin binding (**Figure 5.21**)⁵¹. As with the Re-LPS HDX profile, protection was detected on $\beta 1$ – $\beta 2$ and $\beta 3$ – $\beta 4$ strands near the lateral gate. Inclusion of thanatin also reveals protection of the β -taco as observed before. Collaborators (FF, XQ, SM, JRB, CRV) noted reduced dynamics of this domain, implying that Re-LPS may be bound inside but stabilised by the inclusion of thanatin bound to the N-terminus of the β -taco – which is consistent with the native-MS data, where both are found to bind at the same time.⁵¹

To verify these observations, I generated a differential solvent contact plot between a fully open state of LptDE (bound to Re-LPS) and an open state of LptDE with Re-LPS bound inside and thanatin docked to the N-terminus (**Figure 5.22**). As before (with the closed state), I observe the N-terminal of the β -taco being directly protected. Interestingly a residue-by-residue differential solvent plot also revealed protection across the full length of the soluble domain.

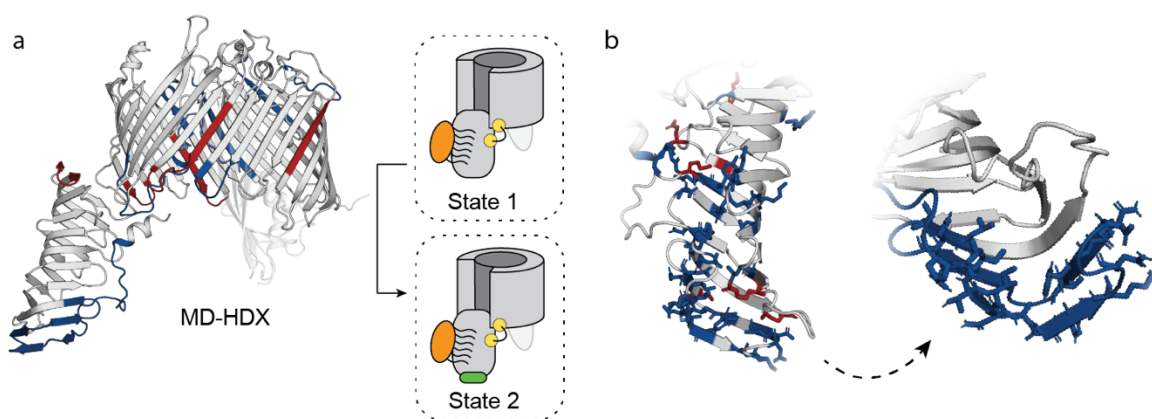


Figure 5.22 Differential contact plot of LptDE binding to the inhibitor thanatin in the open state. Figure adapted from Fiorentino *et al*⁵¹, all work here completed by myself. (a) The differential solvent contact map generated from the difference between the solvent contacts of the open Re-LPS bound state with and without thanatin docked to the bottom of the β -taco. (b) Side-by-side comparison of the residue by residue (left) and the peptide (right) differential solvent map of the β -taco. Residues which contribute to the most solvent deprotection highlighted as sticks.

The differential solvent plot I obtained in **Figure 5.23** shows the solvent contact difference between a completely closed state and a Re-LPS bound open state with thanatin docked to the bottom of the β -taco. In a similar fashion to **Figure 5.8** the protected peptide at the exit gate is restored as opening the barrel causes conformational protection on the peptide as well as encourages Re-LPS lipids to become trapped within the exit gate, agreeing with the experimental HDX plot. Curiously, even though the β -taco is opened by virtue of Re-LPS binding inside, and therefore becoming more exposed to the solvent the inclusion of thanatin causes the β -taco to become protected. Furthermore, this agrees with experiment as the β -taco becomes protected when bound to thanatin despite also binding Re-LPS.

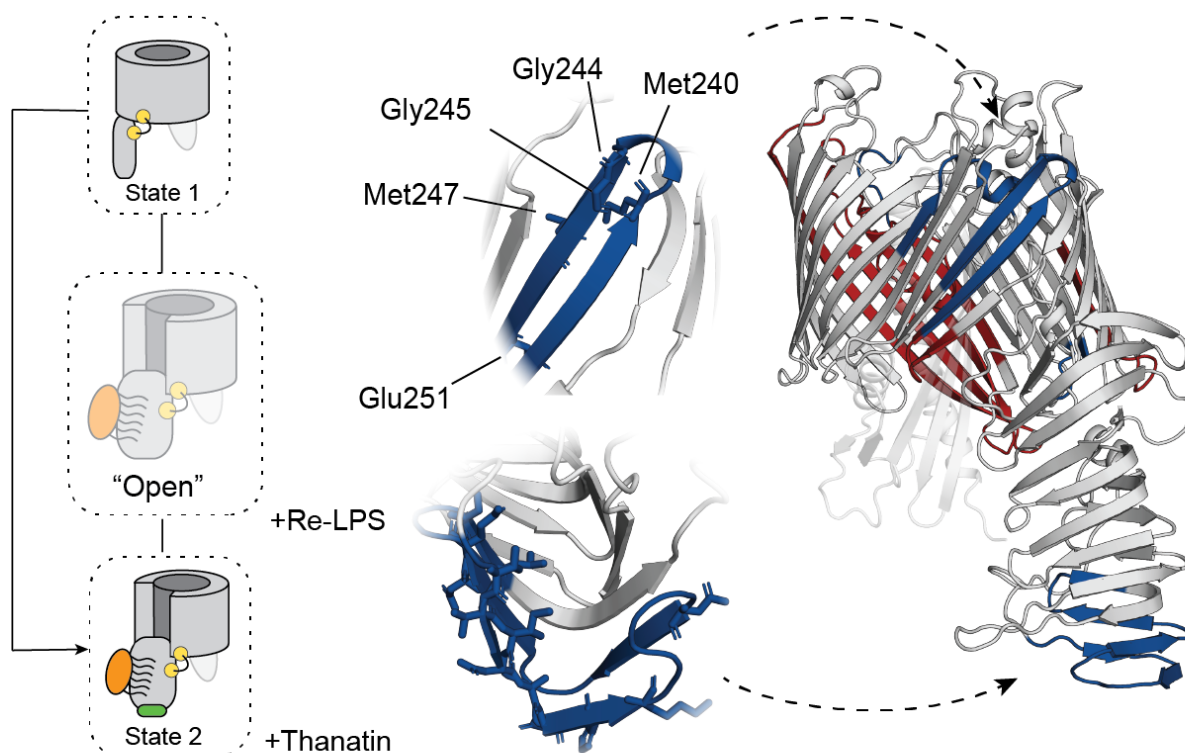


Figure 5.23 Differential contact plot of LptDE between the closed state and LptDE binding to the inhibitor thanatin in the open state. Figure adapted from Fiorentino *et al*⁵¹, all work here completed by myself. Left: cartoon schematic illustrating the conformational change in adopting Re-LPS from the closed state, and then finally the proposed fully bound substrate and inhibitor open state. Right: HDX differential plot between the fully bound substrate and inhibitor open state (state 2) and the closed apo state (state 1). Peptides highlighted on β -taco and β -barrel which show protection where significant residues that contribute to solvent shielding are drawn as sticks.

5.7 Hydrogen bond analysis of the β -taco

Whilst much of the differential solvent plots correlate with experimental HDX profiles, the cannot, by limitation of simulation timescales, comment upon the slow EX1 kinetic component of the β -taco observed by collaborators (FF, XQ, SM, JRB, CRV)⁵¹. HDX data implies that the presence of LPS or Re-LPS in the β -taco increases the rate of conformational change whereas the inclusion of thanatin reduces it. Fundamentally, the ligands included within the experiment either stabilise or destabilise the β -taco in some way which needs to be explained. Of course, EX1 behaviour is generally

associated with slow, concerted conformational shift which could not be detailed in the short timescales of the MD simulations. In order to account for the observations highlighted by the collaborators (FF, XQ, SM, JRB, CRV)⁵¹, I turned to correlate differences in hydrogen bond network with EX1 kinetics as a prerequisite for future conformational changes (which, I am not able to determine over such small timescales). As shown in **Figure 5.3**, the presence of Re-LPS increases the rate of EX1 kinetics (described above as β -strand separation) whereas the presence of thanatin decreases it.

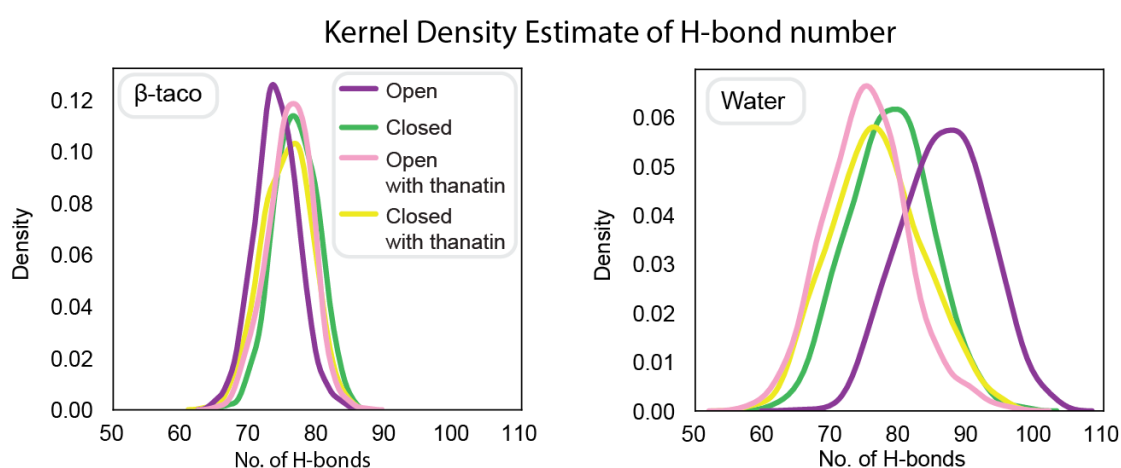


Figure 5.24 KDE of hydrogen bond number. Figure adapted from Fiorentino *et al*⁵¹, all work here completed by myself. Hydrogen-bond number plotted as kernel density estimate (KDE) per frame of simulation for each state of the β -taco (closed, open, closed with thanatin and open with thanatin). Left: the hydrogen-bond network pertaining to the β -taco itself. Right: the lower plot considers the hydrogen bonds which the β -taco makes with the solvent.

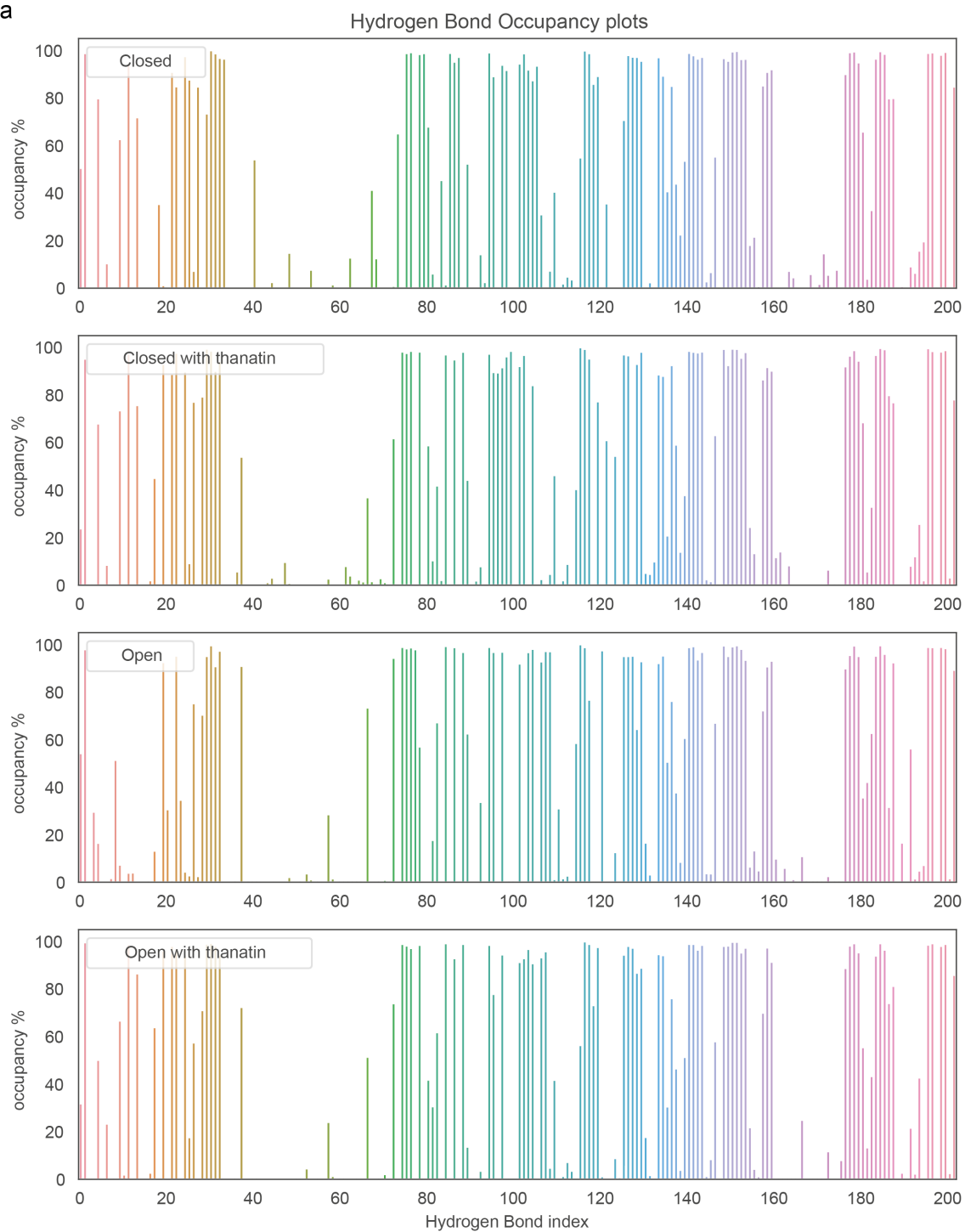
In computing the number of backbone hydrogen bonds in the β -taco over all the simulation data (**Figure 5.24**), I observe a decrease in the number of hydrogen bonds when Re-LPS is docked inside when compared to the number of hydrogen bonds calculated in a closed β -taco domain. However, inclusion of thanatin to the open state of the β -taco ‘restores’ the number of hydrogen bonds on the domain akin to that of the closed state – implying that thanatin’s presence induces a stabilising effect. It is

important to note here that hydrogen bonds between thanatin and the β -taco domain itself was not included in the calculation, where I consider only the backbone hydrogen bonds between residues 28 to 182.

This trend is mirrored by considering the number of hydrogen bonds which are made with solvent, which increases when Re-LPS is accommodated inside the β -taco but decreases when thanatin is present. Therefore, I can explain experimental behaviour by showing that fewer hydrogen bonds exist between strands on the β -taco for the Re-LPS bound LptDE complex compared with the apo state. Conversely, when thanatin binds to either state, more hydrogen bonds are observed, stabilising the domain and reducing the rate of such a conformational change to a level we cannot quantify experimentally. Therefore, I interpret hydrogen bond number and position in helping explain potential conformational changes between β -strands in an EX1 regime.

Granted, it is still unclear just 'how' the hydrogen bond network changes to incur such a kinetic and conformational change. To do this, it is important to ascertain the positions of hydrogen bonds in the β -taco before and after adoption of ligands. I plotted H-bond occupancy across an index of 'all possible' hydrogen bonds (between 4 Å residue separation) in different states of the β -taco: closed, closed with thanatin, open (with Re-LPS) and open (with Re-LPS) and thanatin (**Figure 5.25a**). I generated the difference plots between the states which highlights hydrogen bonds which are 'more present' on the ligand bound state than the apo state (**Figure 5.25b**).

a



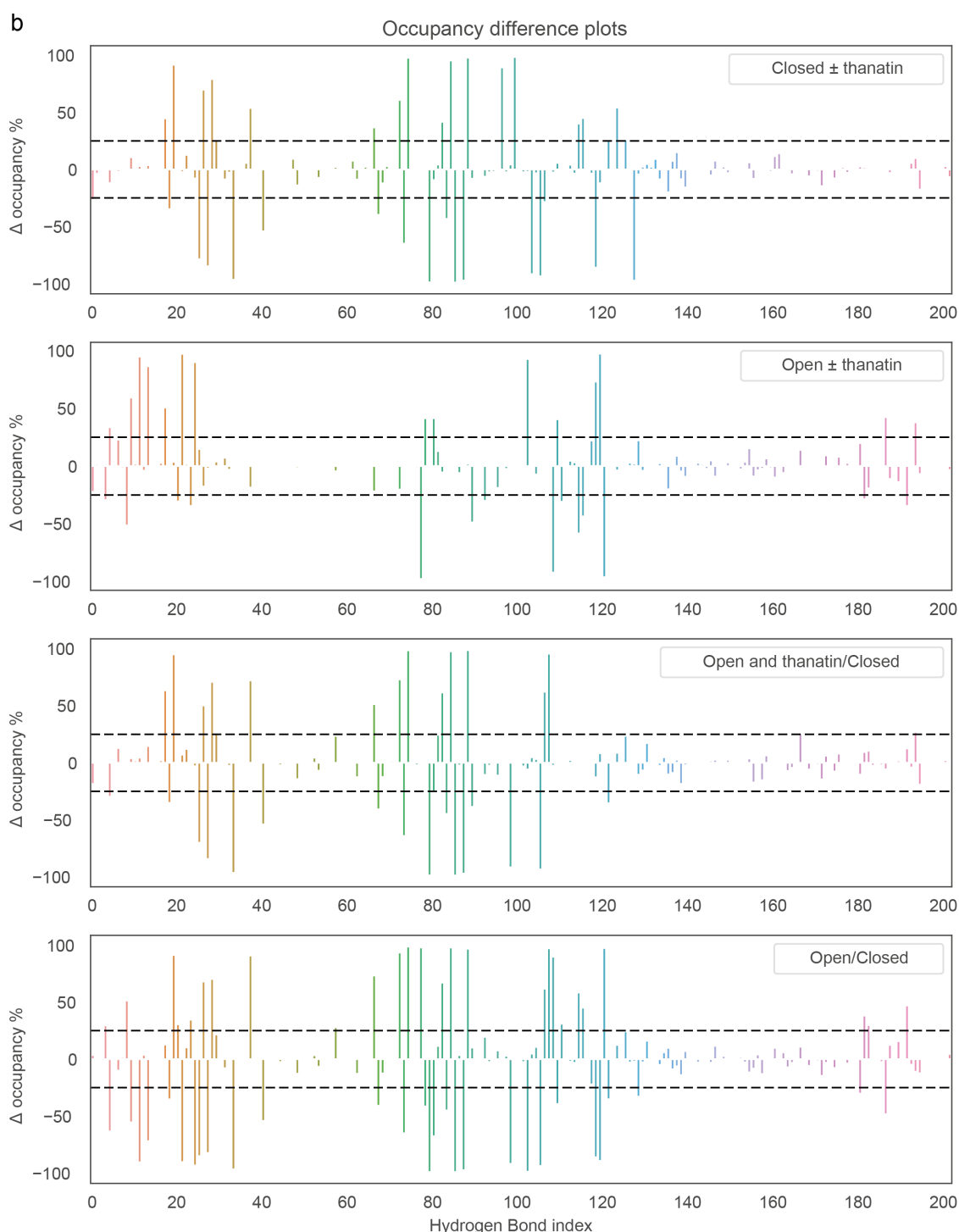


Figure 5.25 Hydrogen bond occupancy. Figure adapted from Fiorentino *et al*⁵¹, all work here completed by myself. (a) Hydrogen bond % occupancy of the β -taco between residues 28 to 182 (i.e. not considering the terminal strand which forms hydrogen bonds with thanatin). The % hydrogen bond occupancy was averaged over five repeats for the four different states of the β -taco (closed, open, closed with thanatin and open with thanatin). (b) % difference occupancy plots for hydrogen bonds between different states of the β -taco. From top to bottom: the plots correspond to the difference between the closed state of the β -taco with and without thanatin bound, the open state of the β -taco with and without thanatin bound, the open state with thanatin

bound versus the closed state and finally the difference between the open and closed state. For each plot the dashed lines represent the arbitrary cut-off at 25% and -25% occupancy difference.

Positive occupancy represents those hydrogen bonds which have present predominantly on ligand binding, negative occupancy indicates the hydrogen bonds located primarily on the 'apo' or the 'before' state (**Figure 5.25b**). I considered only those hydrogen bonds with a 'greater than' or 'less than' 25% occupancy time (as an arbitrary cut-off) and plotted the position of these bonds on the β -taco structure (**Figure 5.26**).

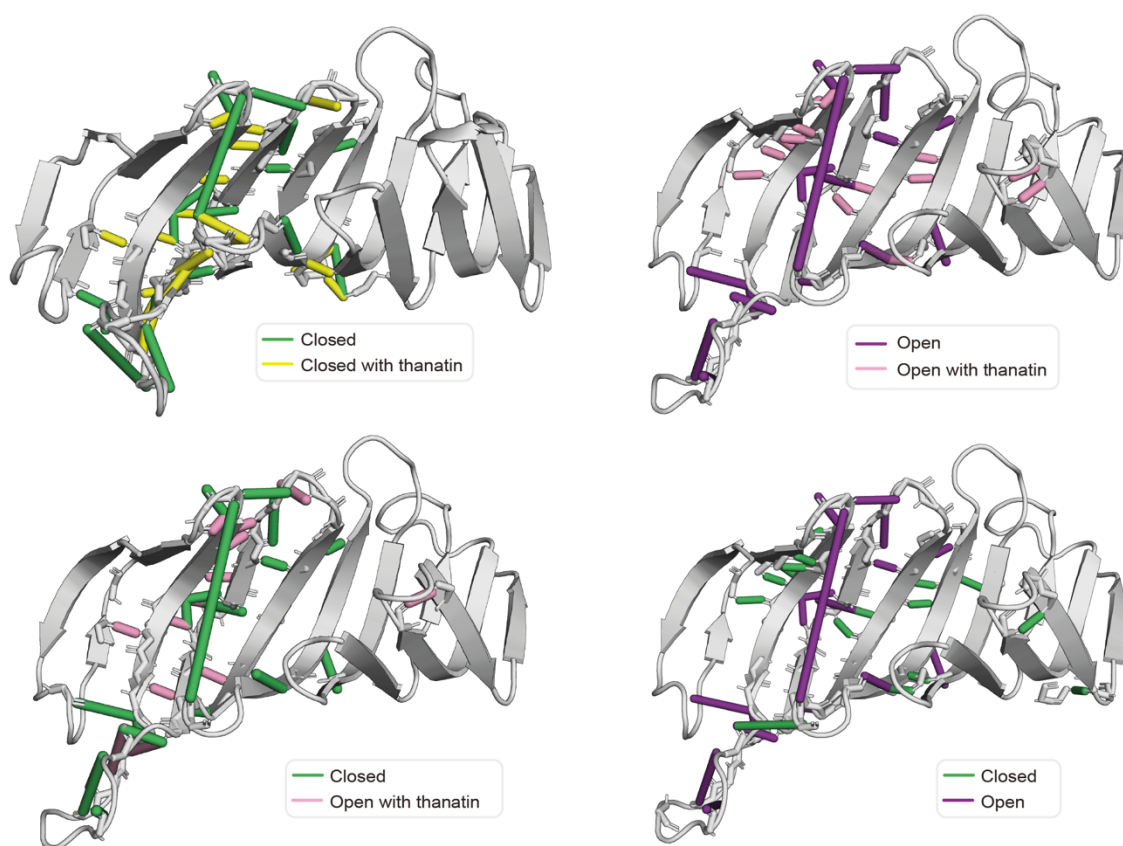


Figure 5.26 Hydrogen bond position. Figure adapted from Fiorentino *et al*⁵¹, all work here completed by myself. Position of the hydrogen bonds which occupy more than 25% in one state of the β -taco compared to the other state as referenced in the difference plots Figure 25b. Top left: difference between the closed state of the β -taco with and without thanatin bound. Top right: difference between the open state of the

β -taco with and without thanatin bound. Bottom left: difference between the open state with thanatin bound and the closed state. Bottom right: difference between the open and closed state.

For those structures with thanatin present, in all cases, a predominancy of new hydrogen bonds were formed between strands of the β -taco. Thanatin appears to strengthen the hydrogen bond network on binding by, not only increasing the number of hydrogen bonds present on the domain, but also by the position of this network across the length of the β -taco between β -strands. This, therefore, means that strands are less likely to separate and become susceptible to exchange with solvent which is why we observe such clear and global protection on this domain whenever thanatin binds to the N-terminal strand of the β -taco. It is also clear that hydrogen bonds between strands are ruptured when the β -taco opens to accommodate the lipid tails of Re-LPS which allows for HDX exchange with the solvent and also why collaborators (FF, XQ, SM, JRB, CRV) observe such deprotection across this domain.⁵¹ It is also interesting that thanatin can completely counteract the destabilisation of Re-LPS in the β -taco domain, where again, its presence anneals the β -strands back together.

In order to account for the kinetics of the HDX experiment we propose two types of concerted motion to describe the behaviour of the β -taco region. The first considers the concerted opening and closing of the β -taco (EX2 kinetics) in order to accommodate Re-LPS binding and the second considers the separation of β -strands on the β -taco, similar to a spring stretching and compressing (EX1 kinetics). Therefore, I can qualitatively account for the observed mixed kinetics EXX demonstrated by the β -taco domain when Re-LPS is bound. The EX2 kinetics derive from the presence of

Re-LPS opening the β -taco and the EX1 kinetics originate from the slow separation of β -strands.

5.8 Discussion

The use of MD simulations provided in this work provided a platform to verify conformational changes observed by HDX-MS by comparing solvent exposure from the simulations with experiment as a proxy for 'deuterium uptake'. MD also granted me better resolution (residue-by-residue) than HDX-MS (peptide resolution), providing a molecular framework to explain why particular regions of the protein are more, or less, solvent exposed after undergoing conformational change.

Here I found that when Re-LPS binds inside the β -taco it yields an open conformation which promotes solvent exposure, particularly with residues that are directed towards the lipid. Being convinced of the reality of the open Re-LPS bound β -taco leads us to question how such a conformation affects the barrel itself. As noted previously, a small network of cysteines coordinates the β -taco with the lateral gate of the β -barrel (**Figure 5.5**). Opening the β -taco and maintaining disulphide bonds between the cysteines on the β -barrel would appear to open the lateral gate in the process. Consequently, an open β -barrel was modelled in tandem, and the simulations showed in correspondence with experiment that this causes significant protection of $\beta 1$ – $\beta 2$ and $\beta 3$ – $\beta 4$ next to the exit gate (**Figure 5.8**). Here I pinpointed particular residues responsible for such deprotection and found that those on the outside of the barrel were shielded from solvent contact by Re-LPS.

Opening the lateral gate in all replicates resulted in a Re-LPS lipid being stuck between beta strands $\beta 1$ and $\beta 26$ encouraging protection of the peptide $\beta 1$ – $\beta 2$ and $\beta 3$ – $\beta 4$. We saw that these lipid interactions were also transient and diffused away during the course of the simulation (**Figure 5.14**). It seems likely then, that the second concentration dependent binding event found in native-MS experiments is at the lateral gate of LptDE⁵¹. Further to this, as the lateral gate opens, the loop between strands $\beta 1$ – $\beta 2$ dips down into the barrel shielding residues inside, suggesting that deprotection is not entirely due to Re-LPS activity and provides further evidence that such an open conformation exists (**Figure 5.16**). These findings were further confirmed by MD simulations, where we observed that the presence of thanatin causes residues throughout the Re-LPS bound β -taco to be significantly protected.

By placing distance restraints on the lateral gate, we can also suggest that this open state captured by HDX and MD simulations occurs when the gate is open by at least 2.25 nm (**Figure 5.7**). This distance may indicate the minimum openness of the gate to facilitate LPS transport and insertion. I also note that the BAM complex β -barrel opens to a similar distance when handling its own OMP substrate³³². Excitingly, a recently published Cryo-EM structure of LptDE from *N. gonorrhoeae* reports a closed and open structure of the complex³³³. The lateral gate distance here shows a separation of ~1 nm between the extracellular ends and ~1.5 nm between the periplasmic ends of $\beta 1$ and $\beta 26$ strands.

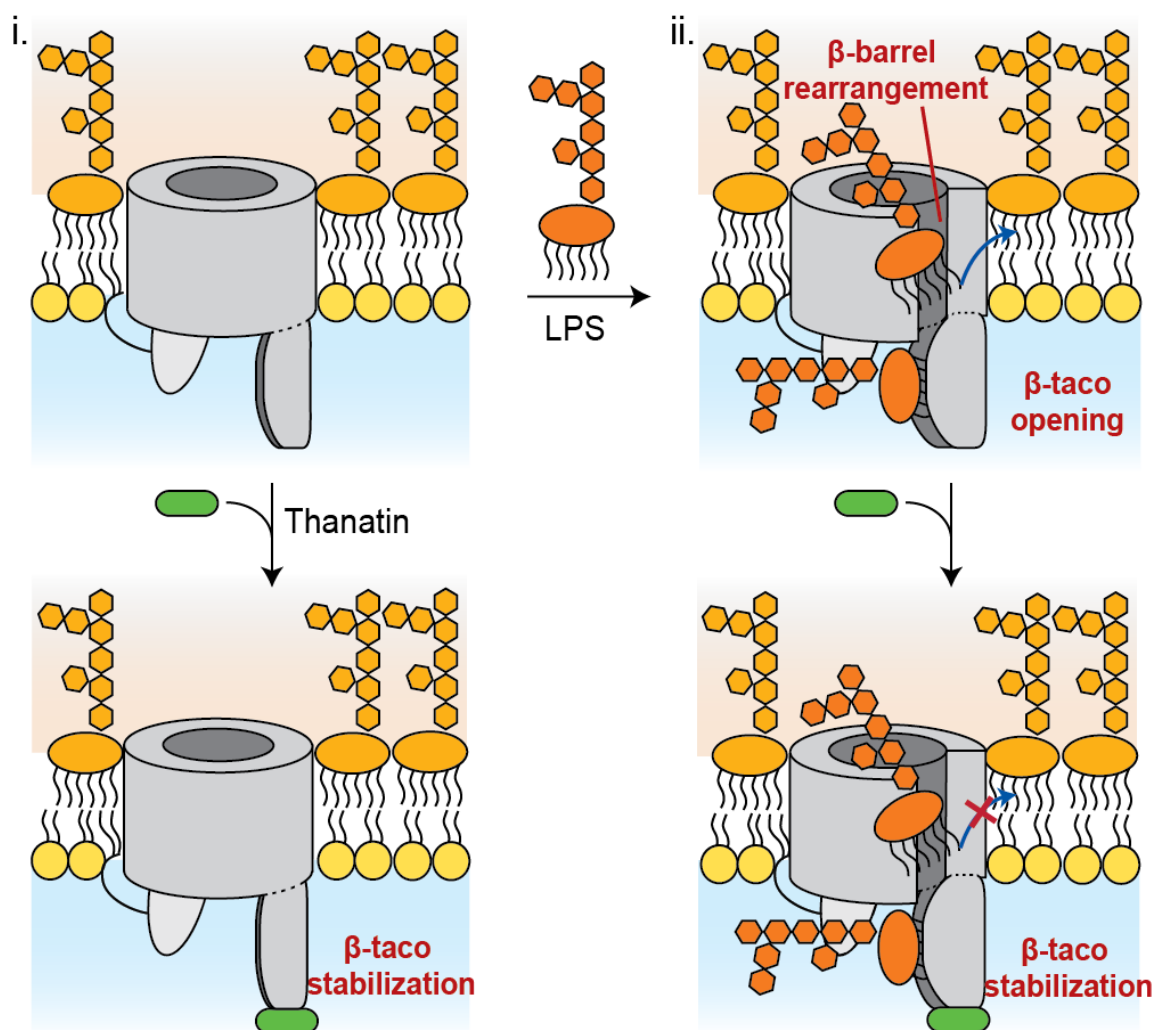


Figure 5.27 Cartoon schematic of LptDE conformations induced by its substrate and inhibitor. Figure adapted from Fiorentino *et al*,⁵¹ all work here completed by myself and FF. (i) binding of thanatin to the apo state of the complex causes β -taco stabilisation. (ii) The interaction between the substrate LPS and LptDE opens the β -taco and therefore pries open the β strands between $\beta 26$ and $\beta 1$ of the lateral gate, coordinate by cysteine bonds. This strand separation accommodates an LPS lipid within the β -barrel. This conformational change creates enough space to accommodate an LPS molecule within the β -barrel for the lipid to be laterally inserted into the OM. Thanatin can also bind to the LPS bound state of the complex stabilising the β -taco and capturing a non-physiological complex of LptDE. Thanatin binding here could conceivably block interactions with LptA and therefore inhibit LPS transport by breaking the formation of the periplasmic bridge.

The existence of a non-physiological complex of LptDE bound to both Re-LPS and thanatin implies that thanatin can bind to LptDE in the open state just as much as the closed state. Therefore, it acts a non-competitive inhibitor to the complex. I propose

that thanatin binds to the N-terminal peptide of β -taco domain and blocks interaction with the periplasmic bridge (**Figure 5.27**). If this occurs, there would not be an instance of LPS binding the LptDE as the lipid would not have been transported to be received by LptDE if the construction of the periplasmic bridge has been disrupted. It is still not yet known whether thanatin can disrupt an already formed complex between β -taco of LptD and LptA.

Furthermore, from hydrogen bond studies and the collaborator's (FF, XQ, SM, JRB, CRV) HDX data it is clear the thanatin has a stabilising influence on the β -taco (**Figure 26**)⁵¹. Not only does thanatin cause an extension of the H-bond network by forming hydrogen bonds with the domain itself but it also increases the number of hydrogen bonds between the β -strands of the domain. This would imply the periplasmic bridge formation of the system would be driven by strong thermodynamic stability caused by hydrogen bond formation between β -taco bridging units as well as between the β -strands of the bridge.

Solvent contacts alongside hydrogen bond studies help to explain observed experimental HDX data obtained by collaborators (FF, XQ, SM, JRB, CRV)⁵¹. However solvent accessibility assessed by oxygen contacts with the protein backbone is admittedly, a fairly crude metric for HDX. Certainly, it is much more qualitative and much simpler than the method used for protection factor calculation by Persson & Halle³³⁴. Initially, I considered including H-bond propensity with solvent based contacts to improve the prediction model and therefore the correlation between predicted and observed HDX. There are a range of different metrics used for comparing MD and HDX^{335,336} but determining the relative contributions of solvent access and hydrogen

bonding can be non-trivial. Indeed, recent work suggests that a full reweighting of the MD statistical ensemble is necessary to faithfully reproduce a HDX profile from MD³³⁷. I reasoned that the risk of introducing artefacts from inadvertently overfitting or misinterpreting our data is greater than the potential gains from a more quantitative comparison. Therefore, as is common for large and dynamic membrane proteins, such as the bacterial Sec translocon³³⁸, we decided instead to apply a more qualitative approach.

5.9 Methods

5.9.1 Modification of Lpt structures used

Here is a table referencing all the structures taken from the PDB used to model LptDE:

Table 5.1 PDB IDs for Lpt complex structures. Protein structures used directly in simulations or used to generate the homology models of the different subcomplexes of the Lpt system.

Protein	Organism	PDB code	Resolution (Å)
LptDE ³¹³	<i>S. flexneri</i>	4Q35	2.39
LptDE ³¹⁹	<i>K. pneumoniae</i>	5IV9	4.37

5.9.2 The OM translocon heterodimer LptDE

To help shed light on experimental HDX work conducted with LptDE discussed in this chapter, I required a structure of LptDE with Re-LPS bound within its β -taco domain. Here, I could build upon the work conducted with the individual subunits of the bridge by growing in Re-LPS within the domain.

5.9.2.1. The open Re-LPS bound state of LptDE

The crystal structure of KpLptDE (PDB ID: 5IV9) is considered a 'closed' state of the protein where the β -taco is not open to allow access to Re-LPS occupation. Interestingly, other structures of LptD such as PDB ID: 4Q35 from *Shigella flexneri* have detergents (tetraethylene glycol mono-octyl ether (C₈E₄) and *N,N*-dimethyldodecylamine N-oxide (LDAO)) bound to this domain.



Figure 5.28 Electron density maps showing detergent molecules within the β -taco. Figure adapted from Qiao *et al.*³²⁴. Electron density of C₈E₄ and LDAO within the hydrophobic groove of the LptD β -taco.

I can therefore dock Re-LPS using these sites as a guide to position the substate within the domain (**Figure 5.28**). Similar to lipid A with the *Ec*LptD subunit, the tails face into the hydrophobic groove whereas the soluble units (Kdo₂ sugars and

phosphate headgroup) face out into solvent. As before, Re-LPS was ‘grown’ within the domain using the inverted alchembed methodology²³⁶.

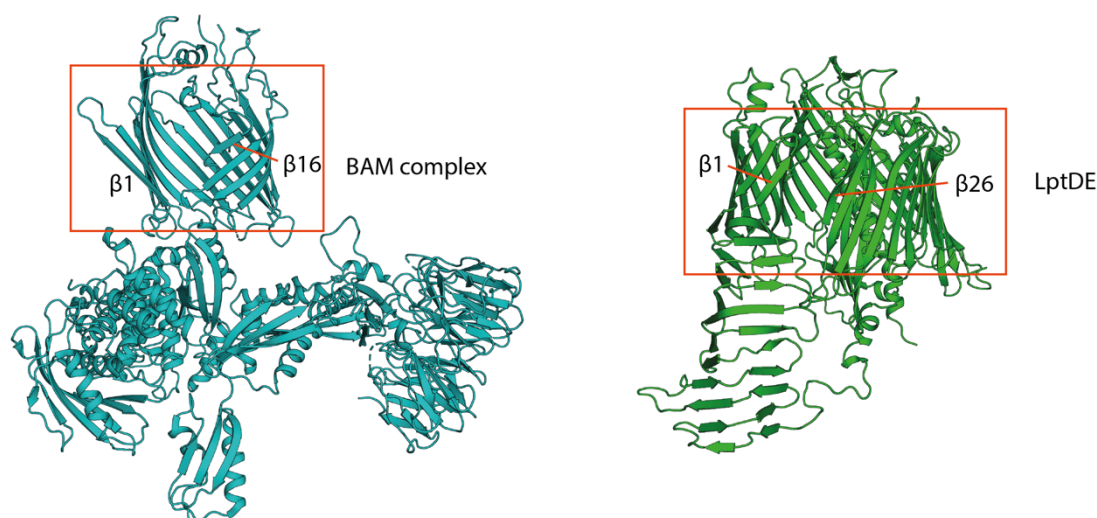


Figure 5.29 Comparison of open β -barrel lateral gates. Left: BAM complex open lateral gate with the OMP substrate removed from the structure for clarity (PDB: 6V05). Right: final open *kplptDE* lateral gate.

For the β -barrel domain, which most likely opens through the lateral gate between $\beta 1$ and $\beta 26$ strands, similar to the observed opening between $\beta 1$ and $\beta 16$ strands found with the BAM complex to allow for OMP installation within the OM (**Figure 5.29**)³³². LptDE structures derived from *Salmonella typhimurium*, from previous MD studies had an open lateral gate due to the application of a negative pressure of -100 bar upon the membrane²¹. This, then, provided a starting point to build an open state structure of the *KpLptD* β -barrel, where various open structures of different distances between the lateral gate strands were obtained. Ultimately, a distance of around 2 nm between $\beta 1$ and $\beta 16$ strands was chosen to reflect the open distance found between the side of β -taco LptD domain to accommodate Re-LPS.

I note that in comparing the open structure of the LptD β -taco domain as well as the β -barrel domain, a pair of cysteine residues on both could be kept proximal if I consider

‘concerted opening’ between the domains. Therefore, two potential disulphide bonds could be maintained as both the β -taco and β -barrel domains open at once. Here, the cysteine pair of residues are between Cys721 (located at the C-terminal end of the barrel) and Cys173 of the β -taco which is a highly conserved cysteine pair. A further pair of cysteines, which could form another disulphide bond are between Cys31 and Cys720. This suggests that these domains are coupled and open together in order to accommodate Re-LPS as the lateral gate unzips to insert LPS into the OM³²⁵.

5.9.2.2. *Thanatin docking to the LptDE β -taco domain*

Native MS and HDX studies, suggest that the cyclic peptidomimetic antibiotic thanatin binds to LptDE⁵¹. However, the experimental data was unclear as to where binding occurs, therefore I required thanatin bound LptDE structures to validate where this may be.

I docked thanatin to LptD (PDB ID: 5IV9) by taking the LptA-thanatin complex and aligning this to the LptD β -taco domain (PDB ID: 6GD5) using PyMOL. The hydrogen bond between thanatin and the β -taco domain were retained in the process, where the coordinates of the antibiotic were then extracted to be incorporated with the *Kp*LptDE crystal structure. In this way thanatin was docked to both the N- and C- terminal of the β -taco so to later verify its stability at both these positions.

5.9.3 *Kp*LptDE AT simulations

Every state of LptDE constructed was initially converted to CG by a Martinize python script²²⁹ so that each representation could be embedded in a bilayer by INSANE³³⁹, so that bilayer would have the chance to be equilibrated and relax around the protein.

In every case the bilayer was composed of (i) pure Re-LPS for the outer leaflet and (ii) a mixture of cardiolipin, POPG and POPE at a 1:2:7 ratio for the inner leaflet. This is to reflect the same percentage composition of lipid species outlined in **chapter 1** for Gram-negative bacteria. To equilibrate the system, I applied elastic bonds between the beads of the protein, between 5 and 10 Å apart, with a force of 1,000 kJ mol⁻¹ nm⁻². Every box was solvated with Martini water with a NaCl concentration of 0.15 M to ensure charge neutrality and the systems were energy minimised by steepest descents methodology to a target force of 1,000 kJ mol⁻¹ nm⁻². Each system was equilibrated with 100 ns of simulations time with a timestep of 20 fs in a Martini 2.2 forcefield²²⁹. After equilibration the structures were converted back to atomistic structures aligning the AT protein structure to the CG coordinates³⁴⁰. Five replicates for 300 ns were run for each state of the LptDE complex (as outlined above with the LptB₂FGC AT simulations) with the velocities reset for each replicate in GROMACS 2019.1²⁵⁶. However, for any open state which required PLUMED³²³ to hold the lateral gate open at a consistent distance, GROMACS 2018.6 was used²⁵⁶. PLUMED is a PLUgin for MoLEcular Dynamics which provides a wide range of tools to set up and analyse simulations²⁵⁶. I applied PLUMED LOWER and UPPERWALLS to the putative lateral gate at a distance of 2.25 nm and 2.75 nm respectively. The open LptDE system which had already been equilibrated for 150 ns, and only required a further 50 ns equilibrations to reach convergence. Restraints were placed between the centre of mass of grouped residues on the β1 and β26 strand respectively: considering the Cα atoms from residues Asn208, Ala209, Lys210 and Tyr211 for β1 and Thr725, Phe726, Gly727 and Ile728 for β26. The force constant for the wall restraint was set to a high value of 10,000 kJ mol⁻¹ to ensure that the ‘open state’ lateral gate would be fixed in

place and not have a chance to sample below 2.25 nm. The default exponent, offset and rescaling factors were used with these restraints.

5.10 Differential solvent uptake analysis

Here I report on an *in silico* method for comparing predicted LptDE conformational states with HDX experiment. This method involves running simulations to produce solvent contact maps on proposed conformations to compare regions of greater or lesser solvation with HDX experiments. I have found this approach a useful tool in rationalising the conformational changes observed in HDX-MS undertaken by LptDE upon binding to Re-LPS as well as the inhibitor thanatin.

I can take these both these two states and perform a 'nominal HDX experiment' to produce the differential solvent contact maps and so 'replicate' the experimental HDX data.

Both states, closed (state 1) and Re-LPS bound open (state 2) conformation are placed in a bilayer (where the inner leaflet is composed of cardiolipin: POPG : POPE in a 1:2:7 ratio and the outer leaflet is composed only of Re-LPS) which is simulated for 300 ns (over five replicates).

I calculated the solvent contacts between the nitrogen of the protein backbone and the oxygen of the water at a cut-off distance of 0.4 nm using a modified MDAnalysis contact script, to highlight those residues most likely to 'exchange' in deuterated solvent as an approximation of experimental HDX^{33,311,341}. Before subtracting the contact maps of the closed conformation (state 1) from the Re-LPS bound open

conformation (state 2), the solvent contacts for prolines are set to zero. The solvent subtraction (or difference) maps were averaged over the five repeats, which produces an average differential contact value for each residue over LptD (**Figure 5.30**).

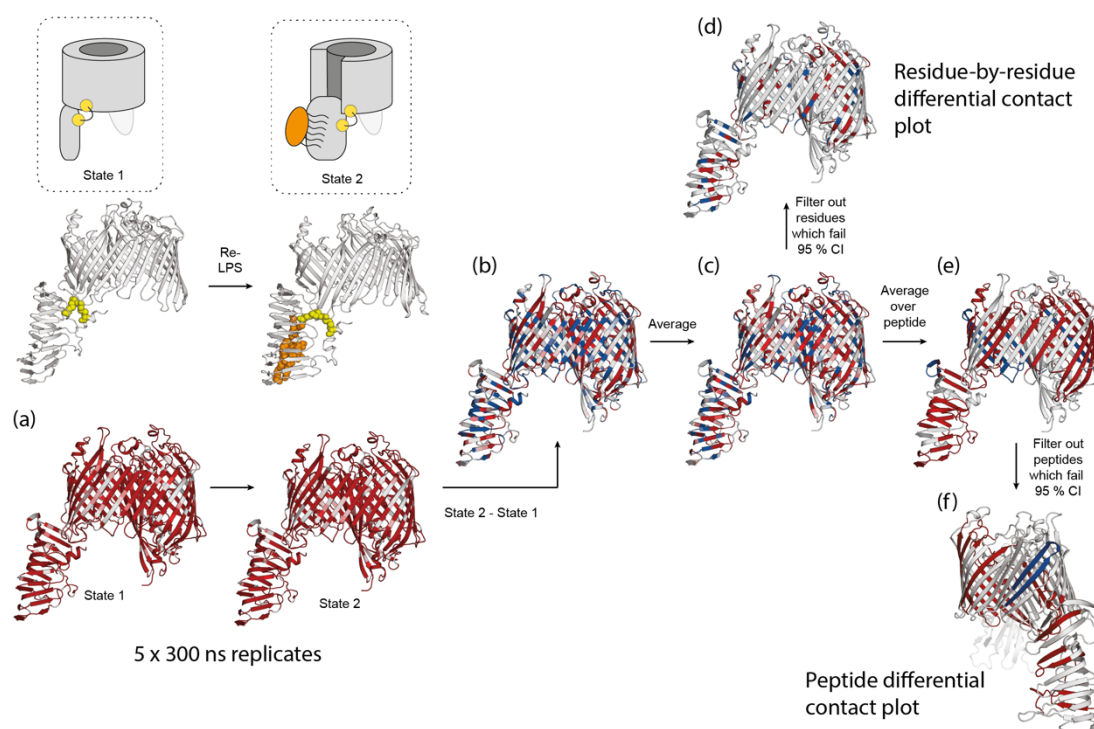


Figure 5.30 Pipeline for producing differential solvent plots. (a) solvent contacts are calculated for both states of the protein (b) a difference map is then generated for each replicate and (c) averaged which can be filtered using a 95% confidence interval to produce (d) a residue-by-residue differential solvent plot. (e) Alternatively, the residues are then averaged over the experimentally determined peptide map and then subsequently filtered by a 95% confidence interval to produce (f) the final differential solvent plot.

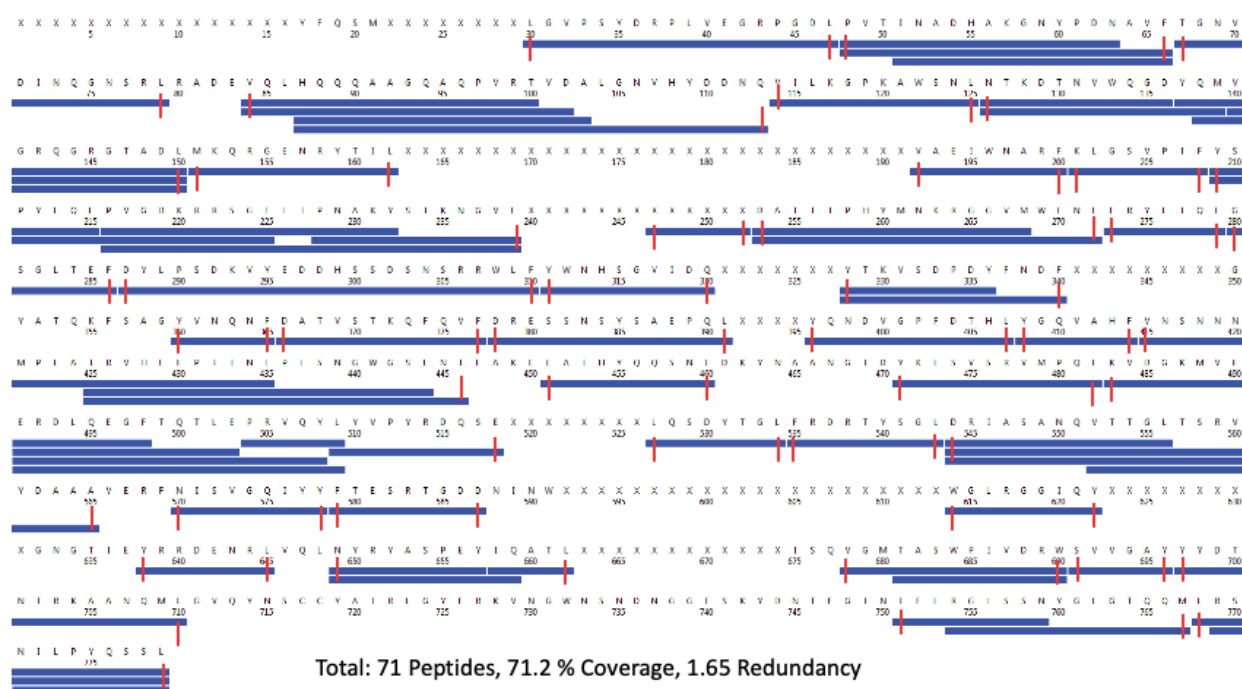


Figure 5.31 Peptide coverage. The experimental coverage map is shown above for the Re-LPS bound state, which is representative over all the HDX coverage maps for different states. The MD coverage map used for producing the differential coverage maps are drawn on top of the experimental peptides. I considered the start and end of a peptide (cleavage sites) when there is no overlap between experimentally determined peptides (marked with red lines).

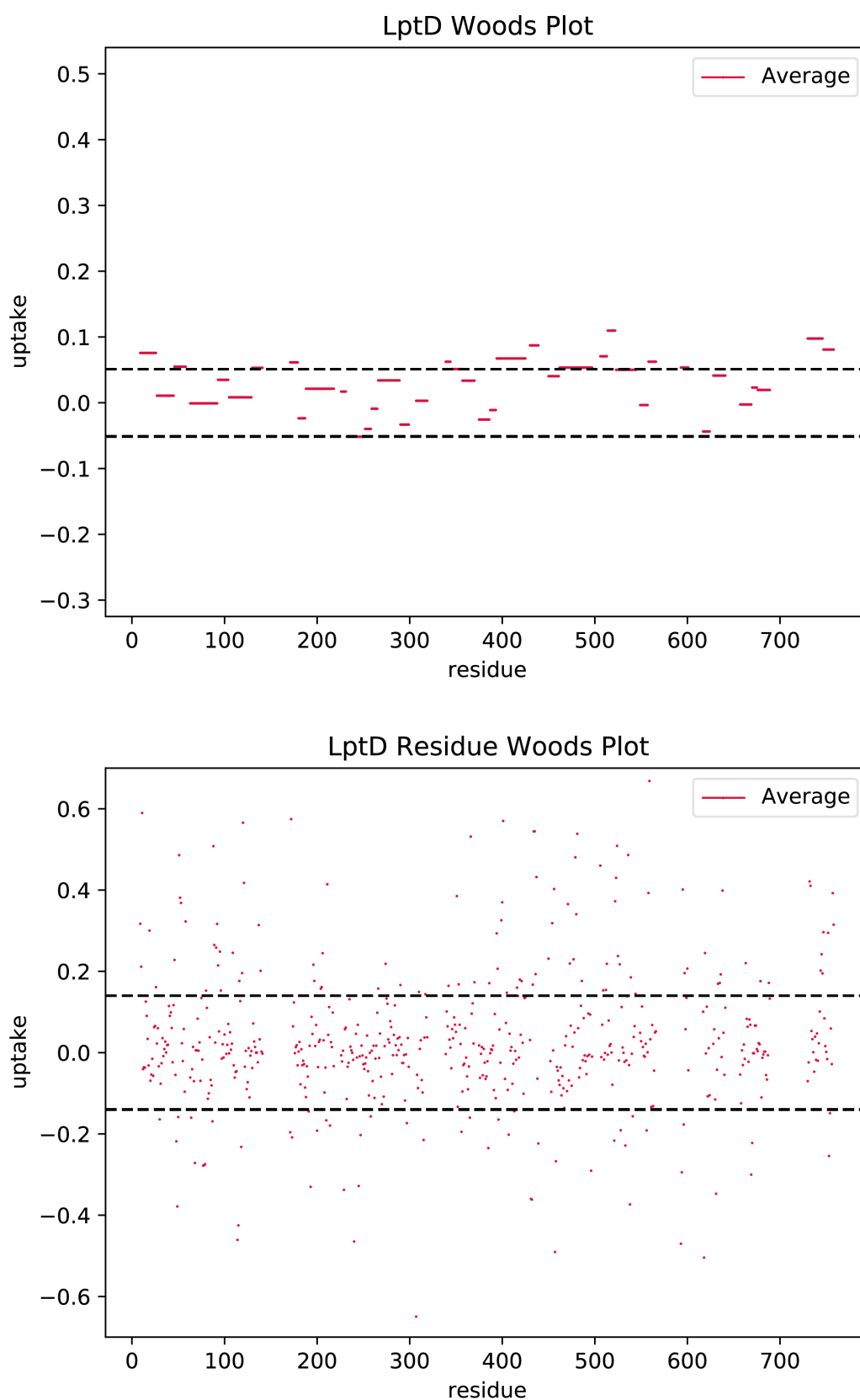


Figure 5.32 Differential solvent Woodsplots. Shown here is an example of a 'Woodsplot' for the differential solvent plot of the Re-LPS bound open state with the

closed state of the protein on an individual peptide (top) and residue (bottom) resolution.

For each of these residue values, the same statistical analysis used by the Deuterios software for experimental HDX is employed here where confidence intervals are calculated using a two tailed test with a 95% significance level³⁴².

For this analysis variance is pooled between state 1 and state 2 of the protein. Since, the number of peptides or residues (where $n_x = n_{pep}$ or n_{res}) are the same for both states, the equation simplifies:

$$\sigma^2 = \frac{\sum_{i=1}^{n_x} \sigma_{state\ 1}^2 + \sum_{i=1}^{n_x} \sigma_{state\ 2}^2}{2n_x} \quad 5.1$$

Where σ is the standard deviation of the differential solvent contact value for either a residue or a peptide for a particular state. Using this pooled variance, the confidence intervals can be calculated:

$$0 \pm \left(\sqrt{\frac{\sigma^2}{N}} \right) \cdot \alpha \quad 5.21$$

Here, $N = 5$ which is the number of replicates of the simulation and $\alpha = 2.776$, the critical value for a two tailed t-test for a 95% confidence interval, with four degrees of freedom ($N-1$) (**Figure 5.32**).

After applying this statistical analysis, only the 95% significant differential values are mapped onto the crystal structure of LptDE as a B-factor. This map is then coloured accordingly in PyMOL where red regions are those which correspond to a positive

difference in solvent contact between state 1 and state 2, whereas blue represents a negative difference. Colours are chosen to reflect 'deprotected' (red) and 'protected' (blue) regions of the protein. As mentioned, these can be mapped on to the structure on an individual residue basis or these values can be averaged over peptides according to the experimentally obtained peptide coverage map. To simplify matters, I considered the start and end of a peptide (cleavage sites) when there is no overlap between experimentally determined peptides (**Figure 5.31**). Areas which are not resolved in experiment are also not considered as part of the differential solvent contact plot. Again, those peptides which are considered 95% significant are calculated in precisely the same way (see above) as the individual residues would be and mapped onto the LptD structure with the same colour scheme.

5.11 Author contributions

FF made **Figures 5.3, 5.17 and 5.21**. FF and JBS made **Figure 5.1, 5.2 and 5.27** together. All HDX work was carried out by FF, JRB, XQ, CVR. All other work including MD studies and analysis was carried out by JBS supported by PJS, RC, KCK.

CHAPTER 6 Summary and future directions

Lipopolysaccharide (LPS) is an endotoxic lipid located in the outer leaflet of the outer membrane (OM) of Gram-negative bacteria⁵⁵. It is responsible for the resistance of pathogenic bacteria to drugs as well as host-immune responses^{58,71,72}. Understanding its interactions with membrane proteins and transporter complex will assist antimicrobial development, by designing small molecules (such as peptidomimetics) which can block the LPS transport pathway⁵¹.

In **chapter 3**, I examined how LPS interacts with apparent phospholipid (PL) binding and LPS binding inner membrane proteins (IMPs). I correlated *in silico* binding studies with native mass spectrometry (native-MS) binding studies to demonstrate comparable binding interactions and behaviours with an LPS-binding protein BtuCD²⁸⁰ and a PL binding protein AmtB²⁸³. Furthermore, I highlighted the similarity in LPS derivatives interacting with BtuCD and LptB₂FGC with the ATP binding domains, implying that LPS has potentially a functional role to play with other IMPs (corresponding with recent experimental assays). This suggests that core-LPS may have a larger role to play in the inner membrane (IM) than otherwise understood. Here, further simulations with other IMPs such as MscL²⁷⁸ and AcrB²⁷⁹ are required to support this hypothesis. Simulations with far larger membrane patches, over much larger timescales could help identify LPS lipid clustering effects and help characterise

the lipidome³⁴³. Experimental mutagenesis and binding studies on the core-LPS saccharide (the mature LPS lipid without the O-antigen) interaction sites on the nucleotide binding domains (NBDs) would verify the residues identified by MD simulations. LptB₂FGC ATP hydrolysis could also be monitored in the presence of lipid A and Re-LPS so to verify whether it reveals the same ATPase trend exhibited by BtuCD²⁸³.

A recently obtained novel Cryo-electron microscopy (Cryo-EM) structure of LptB₂FGC with an unusually skewed periplasmic LptC domain afforded the opportunity to understand how Re-LPS interacts with the subcomplex, compared to an upright LptC conformation of LptB₂FGC. Within the LptC-skewed conformation, Re-LPS and lipid A are housed equidistant between transmembrane (TM) domains LptF and LptG. However, the position of these LPS derivatives markedly shifts towards LptF in the LptC-upright state of the transporter. Mono-phosphate lipid A (MPLA) does not follow such a trend which implies that this shift in position is orchestrated by the phosphate headgroup interactions. Cardiolipin, behaves very similarly to Re-LPS in both its binding interactions and movement within the cavity which supports the competitive kinetics it exhibits in parallel native-MS binding studies. Regardless of the lipid inside the cavity, the LptF and LptG β -taco domains barely fluctuate in the case of the LptC-skewed conformation, compared to the LptC-upright conformations where the soluble domains shift noticeably and cooperatively. This trend, however, is not observed with Re-LPS inside the LptC-upright transporter, implying that the two ketodeoxyoctonic (Kdo₂) sugar groups are used for substrate recognition to stabilise the domain.

To examine the relevancy of the LptC TM helix (TM_c) as well as the competitive effect of cardiolipin further studies could be considered in the form of well-tempered metadynamics (WT-MetaD) calculations¹⁴². Here, separate cardiolipin and Re-LPS thermodynamic binding maps could be obtained both in the presence and absence of the TM_c helix to help identify (i) binding sites at the entrance cavity (ii) the regulatory role of the helix⁸⁶ and (iii) the strength of binding between cardiolipin and lipid A.

The periplasmic bridging section of the Lpt system was examined in **chapter 4**. MD simulations revealed that the β -taco domains would close on removal of the lipid inside the domain. Potential of mean force (PMF) calculations identified that LptC has a greater binding energy than LptA, implying a thermodynamic sink towards the first bridging unit. Large oligomer units of LptA exhibited greater binding energy with lipid A which infers that neighbouring bridging units bring structural rigidity and tighter binding. This also reveals the presence of solvation effects where lipid A experiences less solvent contact (and therefore more favourable binding energies) in larger LptA oligomers. The lipid A PMF calculation along the hydrophobic groove of LptA₃CDF complex confirmed greater binding to LptC than LptA. Here also, I can observe that most favourable lipid A binding energies occurred when all tails were firmly inside the hydrophobic groove and the phosphate headgroups interact with the soluble loops of the β -hairpin strands. To eliminate solvation effects, I modelled a continuous chute of four LptA subunits together where the edges of the tetrameric portion interfaced with its mirror image across the PBC. The resultant PMF profile of lipid A moving across the continuous LptA chute is relatively flat. Furthermore, I modelled lipid A moving through the same system alongside its neighbouring lipid A molecules. Similarly, the PMF energy profile was reasonably flat, however, areas of higher energy were

observed when there were gaps between lipid A molecules due to ‘tightness’ along certain points of the chute.

Further energetic studies should be carried out on this system for further characterisation of LPS transport across the periplasm. For instance, PMF calculations should be repeated with the chute open to a wider extent to observe whether thermodynamic barriers or sinks change in magnitude. It would also be informative to perform a PMF from the hydrophobic cavity of LptB₂FGC along through to LptD in the fully assembled double membrane system. This would (i) eliminate solvation effects and (ii) give a fuller picture of the lipid A transport landscape. Additionally, simulations with more lipid A molecules inside the continuous chute could help predict the number required in the physiological system as proposed by the PEZ model⁸³.

Chapter 5 examines the conformational changes captured by hydrogen deuterium exchange MS (HDX-MS) of LptDE (the translocon complex responsible for installing LPS into the OM) on binding with a LPS derivative, Re-LPS, and an inhibitor, thanatin⁵¹. I developed a simple pipeline to compare the ‘before’ or ‘apo’ state of the protein with the ‘after’ or ‘ligand-bound’ state of the protein by subtracting and averaging the solvation contacts across experimentally determined peptides, to compare with the HDX profiles. Through this *in silico* methodology, I was able to explain areas of protection and deprotection observed experimentally, verifying an ‘open’ Re-LPS bound conformational state of the LptDE complex, supplementing previous modelling studies and validated by a recently published Cryo-EM structure^{21,333}. This conformation exhibited an open β -barrel, where the lateral gate of

the complex (between β -strand 1 and 26) was open by more than 2.25 nm, with an associated and disulphide coordinated open β -taco domain bound to Re-LPS. Areas of the complex which were protected by experiment was rationalised via intramolecular conformational shielding and interaction of Re-LPS lipids at the LptDE exit gate in my MD simulations.

Thanatin bound structures were also modelled and simulated to suggest that the inhibitor binds selectively to the N-terminal of the LptDE β -taco domain. Here also, I showed that thanatin has a stabilising effect on the β -taco domain, extending and increasing the hydrogen bond network between the β -hairpins of the unit, and reveals overall 'protection' on the domain as observed in HDX experiment. This, contrasts with Re-LPS disrupting this hydrogen bond network on opening the β -taco domain (which manifests as deprotection in the HDX profile of the unit). Further studies upon this system should include improved *in silico* HDX modelling techniques, either by developing a further pipeline which combines hydrogen bond and solvation effects or by using recently developed methodologies such as 'HDXer' which re-weighs the MD statistical ensemble³³⁷.

Overall then, I have characterised LPS binding interaction between proteins within the inner membrane, including the ABC transporter subcomplex LptB₂FGC, providing evidence of Re-LPS movement between 'resting' and 'active' conformations of the subcomplex. I have characterised the binding of lipid A across the periplasmic chute and identified an energy sink into the bridging unit LptC, before experiencing a flat energy landscape towards LptDE. Here I have validated an open conformation of the complex which facilitates movement and insertion of Re-LPS into the OM. I have

considered in this thesis how LPS interacts with each section of the Lpt system, from the inner membrane, through the periplasm to the outer membrane. This work is important for the development of antimicrobials which target the LPS transport system, such as thanatin, where I have been able to use MD to rationalise HDX plots and describe its binding action with LptDE.

CHAPTER 7 References

- 1 Alberts, B., Johnson, A., Lewis, A., Raff, M., Roberts, K., & Walter, P. (Garland Science, 2002).
- 2 Eisenberg, D., Marcotte, E. M., Xenarios, I. & Yeates, T. O. Protein function in the post-genomic era. *Nature* **405**, 823-826, doi:10.1038/35015694 (2000).
- 3 Biswas, A., Karmakar, S., Chowdhury, A. & Das, K. P. Interaction of α -crystallin with some small molecules and its effect on its structure and function. *Biochimica et Biophysica Acta* **1860**, 211-221, doi:10.1016/j.bbagen.2015.06.002 (2016).
- 4 Schlessinger, J. Ligand-induced, receptor-mediated dimerization and activation of EGF receptor. *Cell* **110**, 669-672, doi:10.1016/s0092-8674(02)00966-2 (2002).
- 5 Duan, G. & Walther, D. The roles of post-translational modifications in the context of protein interaction networks. *PLOS Computational Biology* **11**, e1004049, doi:10.1371/journal.pcbi.1004049 (2015).
- 6 von Heijne, G. Membrane-protein topology. *Nature Reviews Molecular Cell Biology* **7**, 909-918, doi:10.1038/nrm2063 (2006).
- 7 Lingwood, D. & Simons, K. Lipid rafts as a membrane-organizing principle. *Science* **327**, 46-50, doi:10.1126/science.1174621 (2010).
- 8 Lodish H, B. A., Zipursky SL. in *Molecular Cell Biology* (ed W. H. Freeman) Ch. Section 3.4, Membrane Proteins, (2000).
- 9 Wallin, E. & von Heijne, G. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein science : a publication of the Protein Society* **7**, 1029-1038, doi:10.1002/pro.5560070420 (1998).

- 10 Wickenden, A., Priest, B. & Erdemli, G. Ion channel drug discovery: challenges and future directions. *Future Medicinal Chemistry* **4**, 661-679, doi:10.4155/fmc.12.4 (2012).
- 11 Hong, K. H. & Miller, C. The lipid-protein interface of a Shaker K(+) channel. *Journal of General Physiology* **115**, 51-58, doi:10.1085/jgp.115.1.51 (2000).
- 12 Hughes, T. E. T., Lodowski, D. T., Huynh, K. W., Yazici, A., Del Rosario, J., Kapoor, A., Basak, S., Samanta, A., Han, X., Chakrapani, S., Zhou, Z. H., Filizola, M., Rohacs, T., Han, S. & Moiseenkova-Bell, V. Y. Structural basis of TRPV5 channel inhibition by econazole revealed by cryo-EM. *Nature Structural & Molecular Biology*. **25**, 53-60, doi:10.1038/s41594-017-0009-1 (2018).
- 13 Gao, Y., Cao, E., Julius, D. & Cheng, Y. TRPV1 structures in nanodiscs reveal mechanisms of ligand and lipid action. *Nature* **534**, 347-351, doi:10.1038/nature17964 (2016).
- 14 Powl, A. M., East, J. M. & Lee, A. G. Importance of Direct Interactions with Lipids for the Function of the Mechanosensitive Channel MscL. *Biochemistry* **47**, 12175-12184, doi:10.1021/bi801352a (2008).
- 15 <https://www.nobelprize.org/prizes/lists/all-nobel-prizes-in-chemistry/> (2022).
- 16 Freeman, M. Rhomboid proteases and their biological functions. *Annual Review of Genetics* **42**, 191-210, doi:10.1146/annurev.genet.42.110807.091628 (2008).
- 17 Wolfersberger, M. G. Uniporters, symporters and antiporters. *The Journal of experimental biology* **196**, 5-6 (1994).
- 18 Locher, K. P. Mechanistic diversity in ATP-binding cassette (ABC) transporters. *Nature Structural & Molecular Biology* **23**, 487-493, doi:10.1038/nsmb.3216 (2016).

- 19 Mi, W., Li, Y., Yoon, S. H., Ernst, R. K., Walz, T. & Liao, M. Structural basis of MsbA-mediated lipopolysaccharide transport. *Nature* **549**, 233-237, doi:10.1038/nature23649 (2017).
- 20 Luo, Q., Yang, X., Yu, S., Shi, H., Wang, K., Xiao, L., Zhu, G., Sun, C., Li, T., Li, D., Zhang, X., Zhou, M. & Huang, Y. Structural basis for lipopolysaccharide extraction by ABC transporter LptB2FG. *Nature Structural & Molecular Biology* **24**, 469-474, doi:10.1038/nsmb.3399 (2017).
- 21 Dong, H., Xiang, Q., Gu, Y., Wang, Z., Paterson, N. G., Stansfeld, P. J., He, C., Zhang, Y., Wang, W. & Dong, C. Structural basis for outer membrane lipopolysaccharide insertion. *Nature* **511**, 52-56, doi:10.1038/nature13464 (2014).
- 22 Overington, J. P., Al-Lazikani, B. & Hopkins, A. L. How many drug targets are there? *Nature Reviews Drug Discovery* **5**, 993-996, doi:10.1038/nrd2199 (2006).
- 23 Wymann, M. P. & Schneider, R. Lipid signalling in disease. *Nature Reviews Molecular Cell Biology* **9**, 162-176, doi:10.1038/nrm2335 (2008).
- 24 Miller, S. I. Antibiotic Resistance and Regulation of the Gram-Negative Bacterial Outer Membrane Barrier by Host Innate Immune Molecules. *mBio* **7**, doi:10.1128/mBio.01541-16 (2016).
- 25 Chen, Z., Shi, T., Zhang, L., Zhu, P., Deng, M., Huang, C., Hu, T., Jiang, L. & Li, J. Mammalian drug efflux transporters of the ATP binding cassette (ABC) family in multidrug resistance: A review of the past decade. *Cancer Letters* **370**, 153-164, doi:10.1016/j.canlet.2015.10.010 (2016).
- 26 Amaral, L., Spengler, G., Martins, A. N. A., Armada, A. N. A., Handzlik, J., Kiec-Kononowicz, K. & Molnar, J. Inhibitors of Bacterial Efflux Pumps that also Inhibit Efflux Pumps of Cancer Cells. *Anticancer Research* **32**, 2947 (2012).

- 27 Pielak, R. M. & Chou, J. J. Influenza M2 proton channels. *Biochimica et Biophysica Acta* **1808**, 522-529, doi:10.1016/j.bbamem.2010.04.015 (2011).
- 28 Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T. & Veer, D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **181**, 281-292.e286, doi:10.1016/j.cell.2020.02.058 (2020).
- 29 Pandey, A., Shin, K., Patterson, R. E., Liu, X.-Q. & Rainey, J. K. Current strategies for protein production and purification enabling membrane protein structural biology. *Biochemistry and cell biology = Biochimie et biologie cellulaire* **94**, 507-527, doi:10.1139/bcb-2015-0143 (2016).
- 30 Henderson, R. & Unwin, P. N. T. Three-dimensional model of purple membrane obtained by electron microscopy. *Nature* **257**, 28-32, doi:10.1038/257028a0 (1975).
- 31 Deisenhofer, J., Epp, O., Miki, K., Huber, R. & Michel, H. Structure of the protein subunits in the photosynthetic reaction centre of *Rhodospseudomonas viridis* at 3 Å resolution. *Nature* **318**, 618-624, doi:10.1038/318618a0 (1985).
- 32 Rosenbaum Daniel, M., Cherezov, V., Hanson Michael, A., Rasmussen Søren, G. F., Thian Foon, S., Kobilka Tong, S., Choi, H.-J., Yao, X.-J., Weis William, I., Stevens Raymond, C. & Kobilka Brian, K. GPCR Engineering Yields High-Resolution Structural Insights into β 2-Adrenergic Receptor Function. *Science* **318**, 1266-1273, doi:10.1126/science.1150609 (2007).
- 33 Newport, T. D., Sansom, M. S P. & Stansfeld, P. J. The MemProtMD database: a resource for membrane-embedded protein structures and their lipid interactions. *Nucleic Acids Research* **47**, D390-D397, doi:10.1093/nar/gky1047 (2018).

- 34 Wagner, S., Bader, M. L., Drew, D. & de Gier, J. W. Rationalizing membrane protein overexpression. *Trends in Biotechnology* **24**, 364-371, doi:10.1016/j.tibtech.2006.06.008 (2006).
- 35 Lacapère, J. J., Pebay-Peyroula, E., Neumann, J. M. & Etchebest, C. Determining membrane protein structures: still a challenge! *Trends in Biotechnology* **32**, 259-270, doi:10.1016/j.tibs.2007.04.001 (2007).
- 36 Zorman, S., Botte, M., Jiang, Q., Collinson, I. & Schaffitzel, C. Advances and challenges of membrane-protein complex production. *Current Opinion in Structural Biology* **32**, 123-130, doi:10.1016/j.sbi.2015.03.010 (2015).
- 37 Rawson, S., Davies, S., Lippiat, J. D. & Muench, S. P. The changing landscape of membrane protein structural biology through developments in electron microscopy. *Molecular membrane biology* **33**, 12-22, doi:10.1080/09687688.2016.1221533 (2016).
- 38 Bolla, J. R., Agasid, M. T., Mehmood, S. & Robinson, C. V. Membrane Protein-Lipid Interactions Probed Using Mass Spectrometry. *Annual Review of Biochemistry* **88**, 85-111, doi:10.1146/annurev-biochem-013118-111508 (2019).
- 39 Marsh, D. Protein modulation of lipids, and vice-versa, in membranes. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1778**, 1545-1575, doi:https://doi.org/10.1016/j.bbamem.2008.01.015 (2008).
- 40 Barrera, N. P., Zhou, M. & Robinson, C. V. The role of lipids in defining membrane protein interactions: insights from mass spectrometry. *Trends in Cell Biology* **23**, 1-8, doi:10.1016/j.tcb.2012.08.007 (2013).

- 41 Drew, D., Lerch, M., Kunji, E., Slotboom, D. J. & de Gier, J. W. Optimization of membrane protein overexpression and purification using GFP fusions. *Nature Methods* **3**, 303-313, doi:10.1038/nmeth0406-303 (2006).
- 42 Newby, Z. E., O'Connell, J. D., 3rd, Gruswitz, F., Hays, F. A., Harries, W. E., Harwood, I. M., Ho, J. D., Lee, J. K., Savage, D. F., Miercke, L. J. & Stroud, R. M. A general protocol for the crystallization of membrane proteins for X-ray structural investigation. *Nature Protocols* **4**, 619-637, doi:10.1038/nprot.2009.27 (2009).
- 43 Privé, G. G. Detergents for the stabilization and crystallization of membrane proteins. *Methods* **41**, 388-397, doi:10.1016/j.ymeth.2007.01.007 (2007).
- 44 le Maire, M., Champeil, P. & Møller, J. V. Interaction of membrane proteins and lipids with solubilizing detergents. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1508**, 86-111, doi:https://doi.org/10.1016/S0304-4157(00)00010-1 (2000).
- 45 Bechara, C. & Robinson, C. V. Different Modes of Lipid Binding to Membrane Proteins Probed by Mass Spectrometry. *Journal of the American Chemical Society* **137**, 5240-5247, doi:10.1021/jacs.5b00420 (2015).
- 46 Martens, C., Shekhar, M., Borysik, A. J., Lau, A. M., Reading, E., Tajkhorshid, E., Booth, P. J. & Politis, A. Direct protein-lipid interactions shape the conformational landscape of secondary transporters. *Nature communications* **9**, 4151, doi:10.1038/s41467-018-06704-1 (2018).
- 47 Sherman, D. J., Xie, R., Taylor, R. J., George, A. H., Okuda, S., Foster, P. J., Needleman, D. J. & Kahne, D. Lipopolysaccharide is transported to the cell surface by a membrane-to-membrane protein bridge. *Science* **359**, 798-801, doi:10.1126/science.aar1886 (2018).
- 48 Purcell, A. in *Bacteria* (2016).

- 49 Raetz, C. R. & Whitfield, C. Lipopolysaccharide endotoxins. *Annual Review of Biochemistry*. **71**, 635-700, doi:10.1146/annurev.biochem.71.110601.135414 (2002).
- 50 Sohlenkamp, C. & Geiger, O. Bacterial membrane lipids: diversity in structures and pathways. *FEMS Microbiology Reviews* **40**, 133-159, doi:10.1093/femsre/fuv008 (2016).
- 51 Fiorentino, F., Sauer, J. B., Qiu, X., Corey, R. A., Cassidy, C. K., Mynors-Wallis, B., Mehmood, S., Bolla, J. R., Stansfeld, P. J. & Robinson, C. V. Dynamics of an LPS translocon induced by substrate and an antimicrobial peptide. *Nature Chemical Biology* **17**, 187-195, doi:10.1038/s41589-020-00694-2 (2021).
- 52 Blount, Z. D. The unexhausted potential of *E. coli*. *eLife* **4**, e05826, doi:10.7554/eLife.05826 (2015).
- 53 Raetz, C. R. & Dowhan, W. Biosynthesis and function of phospholipids in *Escherichia coli*. *The Journal of biological chemistry* **265**, 1235-1238 (1990).
- 54 Renner, L. D. & Weibel, D. B. Cardiolipin microdomains localize to negatively curved regions of *Escherichia coli* membranes. *Proceedings of the National Academy of Sciences* **108**, 6264, doi:10.1073/pnas.1015757108 (2011).
- 55 Nikaido, H. in *Escherichia Coli and Salmonella: Cellular and Molecular Biology* Ch. Outer membrane, 29-47 (1996).
- 56 Malinverni, J. C. & Silhavy, T. J. An ABC transport system that maintains lipid asymmetry in the gram-negative outer membrane. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 8009-8014, doi:10.1073/pnas.0903229106 (2009).
- 57 Benn, G., Mikheyeva, I. V., Inns, P. G., Forster, J. C., Ojkic, N., Bortolini, C., Ryadnov, M. G., Kleanthous, C., Silhavy, T. J. & Hoogenboom, B. W. Phase separation in the

- outer membrane of Escherichia coli. *Proceedings of the National Academy of Sciences* **118**, e2112237118, doi:10.1073/pnas.2112237118 (2021).
- 58 Heinrichs, D. E., Yethon, J. A. & Whitfield, C. Molecular basis for structural diversity in the core regions of the lipopolysaccharides of *Escherichia coli* and *Salmonella enterica*. *Molecular Microbiology* **30**, 221-232, doi:10.1046/j.1365-2958.1998.01063.x (1998).
- 59 Whitfield, C., Kaniuk, N. & Frirdich, E. Molecular insights into the assembly and diversity of the outer core oligosaccharide in lipopolysaccharides from *Escherichia coli* and *Salmonella*. *Journal of Endotoxin Research* **9**, 244-249, doi:10.1179/096805103225001440 (2003).
- 60 Klein, G. & Raina, S. Regulated Control of the Assembly and Diversity of LPS by Noncoding sRNAs. *BioMed Research International* **2015**, 153561, doi:10.1155/2015/153561 (2015).
- 61 Kalynych, S., Morona, R. & Cygler, M. Progress in understanding the assembly process of bacterial O-antigen. *FEMS Microbiology Reviews* **38**, 1048-1065, doi:10.1111/1574-6976.12070 (2014).
- 62 Hong, Y., Liu, M. A. & Reeves, P. R. Progress in Our Understanding of Wzx Flippase for Translocation of Bacterial Membrane Lipid-Linked Oligosaccharide. *Journal of Bacteriology*. **200**, e00154-00117, doi:10.1128/JB.00154-17 (2017).
- 63 Wang, L., Wang, Q. & Reeves, P. R. The variation of O antigens in gram-negative bacteria. *Subcellular Biochemistry* **53**, 123-152, doi:10.1007/978-90-481-9078-2_6 (2010).

- 64 Liu, B., Knirel, Y. A., Feng, L., Perepelov, A. V., Senchenkova, S. N., Reeves, P. R. & Wang, L. Structural diversity in Salmonella O antigens and its genetic basis. *FEMS Microbiology Reviews* **38**, 56-89, doi:10.1111/1574-6976.12034 (2014).
- 65 Iguchi, A. A complete view of the Escherichia coli O-antigen biosynthesis gene cluster and the development of molecular-based O-serogrouping methods. *Nihon Saikingaku Zasshi* **71**, 209-215, doi:10.3412/jsb.71.209 (2016).
- 66 Raetz, C. R. H., Reynolds, C. M., Trent, M. S. & Bishop, R. E. Lipid A modification systems in gram-negative bacteria. *Annual Review of Biochemistry* **76**, 295-329, doi:10.1146/annurev.biochem.76.010307.145803 (2007).
- 67 Maldonado, R. F., Sá-Correia, I. & Valvano, M. A. Lipopolysaccharide modification in Gram-negative bacteria during chronic infection. *FEMS Microbiology Reviews* **40**, 480-493, doi:10.1093/femsre/fuw007 (2016).
- 68 Scott, A. J., Oyler, B. L., Goodlett, D. R. & Ernst, R. K. Lipid A structural modifications in extreme conditions and identification of unique modifying enzymes to define the Toll-like receptor 4 structure-activity relationship. *Biochimica et biophysica acta. Molecular and cell biology of lipids* **1862**, 1439-1450, doi:10.1016/j.bbalip.2017.01.004 (2017).
- 69 Bertani, B. & Ruiz, N. Function and Biogenesis of Lipopolysaccharides. *EcoSal Plus* **8**, 10.1128/ecosalplus.ESP-0001-2018, doi:10.1128/ecosalplus.ESP-0001-2018 (2018).
- 70 Stevenson, G., Neal, B., Liu, D., Hobbs, M., Packer, N. H., Batley, M., Redmond, J. W., Lindquist, L. & Reeves, P. Structure of the O antigen of Escherichia coli K-12 and the sequence of its rfb gene cluster. *Journal of Bacteriology* **176**, 4144-4156, doi:10.1128/jb.176.13.4144-4156.1994 (1994).

- 71 Nikaido, H. Molecular basis of bacterial outer membrane permeability revisited. *Microbiology and Molecular Biology Reviews*. **67**, 593-656 (2003).
- 72 Carpenter, T. S., Parkin, J. & Khalid, S. The Free Energy of Small Solute Permeation through the Escherichia coli Outer Membrane Has a Distinctly Asymmetric Profile. *The Journal of Physical Chemistry Letters* **7**, 3446-3451, doi:10.1021/acs.jpcllett.6b01399 (2016).
- 73 Labischinski, H., Barnickel, G., Bradaczek, H., Naumann, D., Rietschel, E. T. & Giesbrecht, P. High state of order of isolated bacterial lipopolysaccharide and its possible contribution to the permeation barrier property of the outer membrane. *Journal of Bacteriology* **162**, 9-20, doi:10.1128/jb.162.1.9-20.1985 (1985).
- 74 Sassi, N., Paul, C., Martin, A., Bettaieb, A. & Jeannin, J. F. Lipid A-induced responses in vivo. *Advances in Experimental Medicine and Biology* **667**, 69-80, doi:10.1007/978-1-4419-1603-7_7 (2010).
- 75 Montminy, S. W., Khan, N., McGrath, S., Walkowicz, M. J., Sharp, F., Conlon, J. E., Fukase, K., Kusumoto, S., Sweet, C., Miyake, K., Akira, S., Cotter, R. J., Goguen, J. D. & Lien, E. Virulence factors of Yersinia pestis are overcome by a strong lipopolysaccharide response. *Nature Immunology* **7**, 1066-1073, doi:10.1038/ni1386 (2006).
- 76 Saldías, M. S., Ortega, X. & Valvano, M. A. Burkholderia cenocepacia O antigen lipopolysaccharide prevents phagocytosis by macrophages and adhesion to epithelial cells. *Journal of Medical Microbiology* **58**, 1542-1548, doi:10.1099/jmm.0.013235-0 (2009).

- 77 Whitfield, C. & Trent, M. S. Biosynthesis and export of bacterial lipopolysaccharides. *Annual Review of Biochemistry*. **83**, 99-128, doi:10.1146/annurev-biochem-060713-035600 (2014).
- 78 Coggins, B. E., Li, X., McClerren, A. L., Hindsgaul, O., Raetz, C. R. & Zhou, P. Structure of the LpxC deacetylase with a bound substrate-analog inhibitor. *Nature Structural & Molecular Biology* **10**, 645-651 (2003).
- 79 Erwin, A. L. Antibacterial drug discovery targeting the lipopolysaccharide biosynthetic enzyme LpxC. *Cold Spring Harbor perspectives in medicine* **6**, a025304 (2016).
- 80 Cipolla, L., Polissi, A., Airoidi, C., Gabrielli, L., Merlo, S. & Nicotra, F. New targets for antibacterial design: Kdo biosynthesis and LPS machinery transport to the cell surface. *Current medicinal chemistry* **18**, 830-852 (2011).
- 81 Greenfield, L. K. & Whitfield, C. Synthesis of lipopolysaccharide O-antigens by ABC transporter-dependent pathways. *Carbohydrate research* **356**, 12-24 (2012).
- 82 Silhavy, T. J., Kahne, D. & Walker, S. The bacterial cell envelope. *Cold Spring Harbor perspectives in biology* **2**, a000414-a000414, doi:10.1101/cshperspect.a000414 (2010).
- 83 Okuda, S., Sherman, D. J., Silhavy, T. J., Ruiz, N. & Kahne, D. Lipopolysaccharide transport and assembly at the outer membrane: the PEZ model. *Nature Reviews Microbiology*. **14**, 337-345, doi:10.1038/nrmicro.2016.25 (2016).
- 84 Sherman, D. J., Lazarus, M. B., Murphy, L., Liu, C., Walker, S., Ruiz, N. & Kahne, D. Decoupling catalytic activity from biological function of the ATPase that powers lipopolysaccharide transport. *Proceedings of the National Academy of Sciences of the United States* **111**, 4982-4987, doi:10.1073/pnas.1323516111 (2014).

- 85 Okuda, S., Freinkman, E. & Kahne, D. Cytoplasmic ATP hydrolysis powers transport of lipopolysaccharide across the periplasm in *E. coli*. *Science* **338**, 1214-1217, doi:10.1126/science.1228984 (2012).
- 86 Li, Y., Orlando, B. J. & Liao, M. Structural basis of lipopolysaccharide extraction by the LptB2FGC complex. *Nature* **567**, 486-490, doi:10.1038/s41586-019-1025-6 (2019).
- 87 Owens, T. W., Taylor, R. J., Pahil, K. S., Bertani, B. R., Ruiz, N., Kruse, A. C. & Kahne, D. Structural basis of unidirectional export of lipopolysaccharide to the cell surface. *Nature* **567**, 550-553, doi:10.1038/s41586-019-1039-0 (2019).
- 88 Thomas, C. & Tampé, R. Multifaceted structures and mechanisms of ABC transport systems in health and disease. *Current Opinion in Structural Biology* **51**, 116-128, doi:10.1016/j.sbi.2018.03.016 (2018).
- 89 Josts, I., Stubenrauch, C. J., Vadlamani, G., Mosbahi, K., Walker, D., Lithgow, T. & Grinter, R. The Structure of a Conserved Domain of TamB Reveals a Hydrophobic β Taco Fold. *Structure (London, England : 1993)* **25**, 1898-1906.e1895, doi:10.1016/j.str.2017.10.002 (2017).
- 90 Tran, A. X., Dong, C. & Whitfield, C. Structure and functional analysis of LptC, a conserved membrane protein involved in the lipopolysaccharide export pathway in *Escherichia coli*. *The Journal of biological chemistry* **285**, 33529-33539, doi:10.1074/jbc.M110.144709 (2010).
- 91 Suits, M. D., Sperandio, P., Deho, G., Polissi, A. & Jia, Z. Novel structure of the conserved gram-negative lipopolysaccharide transport protein A and mutagenesis analysis. *Journal of Molecular Biology.* **380**, 476-488, doi:10.1016/j.jmb.2008.04.045 (2008).

- 92 Freinkman, E., Okuda, S., Ruiz, N. & Kahne, D. Regulated assembly of the transenvelope protein complex required for lipopolysaccharide export. *Biochemistry* **51**, 4800-4806, doi:10.1021/bi300592c (2012).
- 93 Li, X., Gu, Y., Dong, H., Wang, W. & Dong, C. Trapped lipopolysaccharide and LptD intermediates reveal lipopolysaccharide translocation steps across the Escherichia coli outer membrane. *Scientific Reports* **5**, 11883, doi:10.1038/srep11883 (2015).
- 94 Merten, J. A., Schultz, K. M. & Klug, C. S. Concentration-dependent oligomerization and oligomeric arrangement of LptA. *Protein Science* **21**, 211-218, doi:10.1002/pro.2004 (2012).
- 95 Laguri, C., Sperandio, P., Pounot, K., Ayala, I., Silipo, A., Bougault, C. M., Molinaro, A., Polissi, A. & Simorre, J. P. Interaction of lipopolysaccharides at intermolecular sites of the periplasmic Lpt transport assembly. *Scientific Reports* **7**, 9715, doi:10.1038/s41598-017-10136-0 (2017).
- 96 Cohen, E. J., Ferreira, J. L., Ladinsky, M. S., Beeby, M. & Hughes, K. T. Nanoscale-length control of the flagellar driveshaft requires hitting the tethered outer membrane. *Science (New York, N.Y.)* **356**, 197-200, doi:10.1126/science.aam6512 (2017).
- 97 Freinkman, E., Chng, S. S. & Kahne, D. The complex that inserts lipopolysaccharide into the bacterial outer membrane forms a two-protein plug-and-barrel. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 2486-2491, doi:10.1073/pnas.1015617108 (2011).
- 98 Malojčić, G., Andres, D., Grabowicz, M., George, A. H., Ruiz, N., Silhavy, T. J. & Kahne, D. LptE binds to and alters the physical state of LPS to catalyze its assembly at the

- cell surface. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 9467-9472, doi:10.1073/pnas.1402746111 (2014).
- 99 Chng, S. S., Ruiz, N., Chimalakonda, G., Silhavy, T. J. & Kahne, D. Characterization of the two-protein complex in *Escherichia coli* responsible for lipopolysaccharide assembly at the outer membrane. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 5363-5368, doi:10.1073/pnas.0912872107 (2010).
- 100 Chimalakonda, G., Ruiz, N., Chng, S.-S., Garner, R. A., Kahne, D. & Silhavy, T. J. Lipoprotein LptE is required for the assembly of LptD by the beta-barrel assembly machine in the outer membrane of *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 2492-2497, doi:10.1073/pnas.1019089108 (2011).
- 101 Gupta, K., Donlan, J. A. C., Hopper, J. T. S., Uzdavinyas, P., Landreh, M., Struwe, W. B., Drew, D., Baldwin, A. J., Stansfeld, P. J. & Robinson, C. V. The role of interfacial lipids in stabilizing membrane protein oligomers. *Nature* **541**, 421-424, doi:10.1038/nature20820 (2017).
- 102 Parker, J. L., Corey, R. A., Stansfeld, P. J. & Newstead, S. Structural basis for substrate specificity and regulation of nucleotide sugar transporters in the lipid bilayer. *Nature communications* **10**, 4657, doi:10.1038/s41467-019-12673-w (2019).
- 103 Leyton, D. L., Belousoff, M. J. & Lithgow, T. The β -Barrel Assembly Machinery Complex. *Methods in Molecular Biology* **1329**, 1-16, doi:10.1007/978-1-4939-2871-2_1 (2015).
- 104 Wen, P. C., Mahinthichaichan, P., Trebesch, N., Jiang, T., Zhao, Z., Shinn, E., Wang, Y., Shekhar, M., Kapoor, K., Chan, C. K. & Tajkhorshid, E. Microscopic view of lipids and

- their diverse biological functions. *Current Opinion in Structural Biology* **51**, 177-186, doi:10.1016/j.sbi.2018.07.003 (2018).
- 105 Corradi, V., Sejdiu, B. I., Mesa-Galloso, H., Abdizadeh, H., Noskov, S. Y., Marrink, S. J. & Tieleman, D. P. Emerging Diversity in Lipid-Protein Interactions. *Chemical Reviews* **119**, 5775-5848, doi:10.1021/acs.chemrev.8b00451 (2019).
- 106 Bushell, S. R., Pike, A. C. W., Falzone, M. E., Rorsman, N. J. G., Ta, C. M., Corey, R. A., Newport, T. D., Christianson, J. C., Scofano, L. F., Shintre, C. A., Tessitore, A., Chu, A., Wang, Q., Shrestha, L., Mukhopadhyay, S. M. M., Love, J. D., Burgess-Brown, N. A., Sitsapesan, R., Stansfeld, P. J., Huiskonen, J. T., Tammara, P., Accardi, A. & Carpenter, E. P. The structural basis of lipid scrambling and inactivation in the endoplasmic reticulum scramblase TMEM16K. *Nature communications* **10**, 3956, doi:10.1038/s41467-019-11753-1 (2019).
- 107 Caffalette, C. A., Corey, R. A., Sansom, M. S. P., Stansfeld, P. J. & Zimmer, J. A lipid gating mechanism for the channel-forming O antigen ABC transporter. *Nature communications* **10**, 824, doi:10.1038/s41467-019-08646-8 (2019).
- 108 Pebay-Peyroula, E., Dahout-Gonzalez, C., Kahn, R., Trézéguet, V., Lauquin, G. J. M. & Brandolin, G. Structure of mitochondrial ADP/ATP carrier in complex with carboxyatractyloside. *Nature* **426**, 39-44, doi:10.1038/nature02056 (2003).
- 109 Hansen, S. B., Tao, X. & MacKinnon, R. Structural basis of PIP₂ activation of the classical inward rectifier K⁺ channel Kir2.2. *Nature* **477**, 495-498, doi:10.1038/nature10370 (2011).
- 110 Norimatsu, Y., Hasegawa, K., Shimizu, N. & Toyoshima, C. Protein-phospholipid interplay revealed with crystals of a calcium pump. *Nature* **545**, 193-198, doi:10.1038/nature22357 (2017).

- 111 Hagn, F., Nasr, M. L. & Wagner, G. Assembly of phospholipid nanodiscs of controlled size for structural studies of membrane proteins by NMR. *Nature Protocols* **13**, 79-98, doi:10.1038/nprot.2017.094 (2018).
- 112 Bada Juarez, J. F., Harper, A. J., Judge, P. J., Tonge, S. R. & Watts, A. From polymer chemistry to structural biology: The development of SMA and related amphipathic polymers for membrane protein extraction and solubilisation. *Chemistry and Physics of Lipids* **221**, 167-175, doi:10.1016/j.chemphyslip.2019.03.008 (2019).
- 113 Corey, R. A., Stansfeld, P. J. & Sansom, M. S. P. The energetics of protein–lipid interactions as viewed by molecular simulations. *Biochemical Society Transactions* **48**, 25-37, doi:10.1042/bst20190149 (2019).
- 114 Teo, A. C. K., Lee, S. C., Pollock, N. L., Stroud, Z., Hall, S., Thakker, A., Pitt, A. R., Dafforn, T. R., Spickett, C. M. & Roper, D. I. Analysis of SMALP co-extracted phospholipids shows distinct membrane environments for three classes of bacterial membrane protein. *Scientific Reports* **9**, 1813, doi:10.1038/s41598-018-37962-0 (2019).
- 115 Lee, A. G. A Database of Predicted Binding Sites for Cholesterol on Membrane Proteins, Deep in the Membrane. *Biophysical Journal* **115**, 522-532, doi:10.1016/j.bpj.2018.06.022 (2018).
- 116 Hedger, G. & Sansom, M. S. P. Lipid interaction sites on channels, transporters and receptors: Recent insights from molecular dynamics simulations. *Biochimica et Biophysica Acta* **1858**, 2390-2400, doi:10.1016/j.bbamem.2016.02.037 (2016).
- 117 Ingólfsson, H. I., Arnarez, C., Periole, X. & Marrink, S. J. Computational ‘microscopy’ of cellular membranes. *Journal of cell science* **129**, 257-268 (2016).

- 118 Chavent, M., Duncan, A. L. & Sansom, M. S. Molecular dynamics simulations of membrane proteins and their interactions: from nanoscale to mesoscale. *Current Opinion in Structural Biology* **40**, 8-16, doi:10.1016/j.sbi.2016.06.007 (2016).
- 119 Marrink, S. J. & Tieleman, D. P. Perspective on the Martini model. *Chemical Society Reviews* **42**, 6801-6822, doi:10.1039/c3cs60093a (2013).
- 120 Shinoda, W., DeVane, R. & Klein, M. L. Computer simulation studies of self-assembling macromolecules. *Current Opinion in Structural Biology* **22**, 175-186, doi:https://doi.org/10.1016/j.sbi.2012.01.011 (2012).
- 121 Pöyry, S. & Vattulainen, I. Role of charged lipids in membrane structures - Insight given by simulations. *Biochimica et Biophysica Acta* **1858**, 2322-2333, doi:10.1016/j.bbamem.2016.03.016 (2016).
- 122 Aponte-Santamaría, C., Briones, R., Schenk, A. D., Walz, T. & de Groot, B. L. Molecular driving forces defining lipid positions around aquaporin-0. *Proceedings of the National Academy of Sciences* **109**, 9887-9892 (2012).
- 123 Stansfeld, P. J., Jefferys, E. E. & Sansom, M. S. Multiscale simulations reveal conserved patterns of lipid interactions with aquaporins. *Structure* **21**, 810-819, doi:10.1016/j.str.2013.03.005 (2013).
- 124 Arnarez, C., Mazat, J. P., Elezgaray, J., Marrink, S. J. & Periole, X. Evidence for cardiolipin binding sites on the membrane-exposed surface of the cytochrome bc1. *Journal of the American Chemical Society* **135**, 3112-3120, doi:10.1021/ja310577u (2013).
- 125 Arnarez, C., Marrink, S. J. & Periole, X. Identification of cardiolipin binding sites on cytochrome c oxidase at the entrance of proton channels. *Scientific Reports* **3**, 1263, doi:10.1038/srep01263 (2013).

- 126 Duncan, A. L., Robinson, A. J. & Walker, J. E. Cardiolipin binds selectively but transiently to conserved lysine residues in the rotor of metazoan ATP synthases. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 8687-8692, doi:10.1073/pnas.1608396113 (2016).
- 127 Corey, R. A., Pyle, E., Allen, W. J., Watkins, D. W., Casiraghi, M., Miroux, B., Arechaga, I., Politis, A. & Collinson, I. Specific cardiolipin-SecY interactions are required for proton-motive force stimulation of protein secretion. *Proceedings of the National Academy of Sciences of the United States of America* **115**, 7967-7972, doi:10.1073/pnas.1721536115 (2018).
- 128 Schmidt, M. R., Stansfeld, P. J., Tucker, S. J. & Sansom, M. S. Simulation-based prediction of phosphatidylinositol 4,5-bisphosphate binding to an ion channel. *Biochemistry* **52**, 279-281, doi:10.1021/bi301350s (2013).
- 129 Kasimova, M. A., Zaydman, M. A., Cui, J. & Tarek, M. PIP₂-dependent coupling is prominent in Kv7.1 due to weakened interactions between S4-S5 and S6. *Scientific Reports* **5**, 7474, doi:10.1038/srep07474 (2015).
- 130 Yen, H. Y., Hoi, K. K., Liko, I., Hedger, G., Horrell, M. R., Song, W., Wu, D., Heine, P., Warne, T., Lee, Y., Carpenter, B., Pluckthun, A., Tate, C. G., Sansom, M. S. P. & Robinson, C. V. PtdIns(4,5)P₂ stabilizes active states of GPCRs and enhances selectivity of G-protein coupling. *Nature* **559**, 423-427, doi:10.1038/s41586-018-0325-6 (2018).
- 131 Wang, Q., Corey, R. A., Hedger, G., Aryal, P., Grieben, M., Nasrallah, C., Baronina, A., Pike, A. C., Shi, J. & Carpenter, E. P. Lipid interactions of a ciliary membrane TRP channel: Simulation and structural studies of polycystin-2. *Structure* **28**, 169-184. e165 (2020).

- 132 Javanainen, M., Hammaren, H., Monticelli, L., Jeon, J.-H., Miettinen, M. S., Martinez-Seara, H., Metzler, R. & Vattulainen, I. Anomalous and normal diffusion of proteins and lipids in crowded lipid membranes. *Faraday Discussions* **161**, 397-417 (2013).
- 133 Hedger, G., Rouse, S. L., Domański, J., Chavent, M., Koldsø, H. & Sansom, M. S. Lipid-Loving ANTs: Molecular Simulations of Cardiolipin Interactions and the Organization of the Adenine Nucleotide Translocase in Model Mitochondrial Membranes. *Biochemistry* **55**, 6238-6249, doi:10.1021/acs.biochem.6b00751 (2016).
- 134 Song, W., Yen, H. Y., Robinson, C. V. & Sansom, M. S. P. State-dependent Lipid Interactions with the A2a Receptor Revealed by MD Simulations Using In Vivo-Mimetic Membranes. *Structure* **27**, 392-403.e393, doi:10.1016/j.str.2018.10.024 (2019).
- 135 Souaille, M. & Roux, B. t. Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Computer Physics Communications* **135**, 40-57, doi:https://doi.org/10.1016/S0010-4655(00)00215-0 (2001).
- 136 Li, P. C., Miyashita, N., Im, W., Ishido, S. & Sugita, Y. Multidimensional umbrella sampling and replica-exchange molecular dynamics simulations for structure prediction of transmembrane helix dimers. *Journal of computational chemistry* **35**, 300-308, doi:10.1002/jcc.23494 (2014).
- 137 Domański, J., Hedger, G., Best, R. B., Stansfeld, P. J. & Sansom, M. S. P. Convergence and Sampling in Determining Free Energy Landscapes for Membrane Protein Association. *The Journal of Physical Chemistry B* **121**, 3364-3375, doi:10.1021/acs.jpcb.6b08445 (2017).

- 138 Shirts, M. R. & Mobley, D. L. in *Biomolecular Simulations: Methods and Protocols* (eds Luca Monticelli & Emppu Salonen) 271-311 (Humana Press, 2013).
- 139 Aldeghi, M., Heifetz, A., Bodkin, M. J., Knapp, S. & Biggin, P. C. Accurate calculation of the absolute free energy of binding for drug molecules. *Chemical Science* **7**, 207-218, doi:10.1039/c5sc02678d (2016).
- 140 Klimovich, P. V., Shirts, M. R. & Mobley, D. L. Guidelines for the analysis of free energy calculations. *Journal of Computer-Aided Molecular Design* **29**, 397-411, doi:10.1007/s10822-015-9840-9 (2015).
- 141 Barducci, A., Bussi, G. & Parrinello, M. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Physical Review Letters* **100**, 020603, doi:10.1103/PhysRevLett.100.020603 (2008).
- 142 Corey, R. A., Vickery, O. N., Sansom, M. S. P. & Stansfeld, P. J. Insights into Membrane Protein–Lipid Interactions from Free Energy Calculations. *Journal of Chemical Theory and Computation* **15**, 5727-5736, doi:10.1021/acs.jctc.9b00548 (2019).
- 143 Martens, C., Shekhar, M., Lau, A. M., Tajkhorshid, E. & Politis, A. Integrating hydrogen-deuterium exchange mass spectrometry with molecular dynamics simulations to probe lipid-modulated conformational changes in membrane proteins. *Nature Protocols* **14**, 3183-3204, doi:10.1038/s41596-019-0219-6 (2019).
- 144 Jerabek-Willemsen, M., Wienken, C. J., Braun, D., Baaske, P. & Duhr, S. Molecular interaction studies using microscale thermophoresis. *ASSAY and Drug Development Technologies* **9**, 342-353, doi:10.1089/adt.2011.0380 (2011).

- 145 Olaru, A., Bala, C., Jaffrezic-Renault, N. & Aboul-Enein, H. Y. Surface plasmon resonance (SPR) biosensors in pharmaceutical analysis. *Critical Reviews in Analytical Chemistry* **45**, 97-105, doi:10.1080/10408347.2014.881250 (2015).
- 146 van Berkel, W. J., van den Heuvel, R. H., Versluis, C. & Heck, A. J. Detection of intact megaDalton protein assemblies of vanillyl-alcohol oxidase by mass spectrometry. *Protein science : a publication of the Protein Society* **9**, 435-439, doi:10.1110/ps.9.3.435 (2000).
- 147 Tito, M. A., Tars, K., Vlegard, K., Hajdu, J. & Robinson, C. V. Electrospray Time-of-Flight Mass Spectrometry of the Intact MS2 Virus Capsid. *Journal of the American Chemical Society* **122**, 3550-3551, doi:10.1021/ja993740k (2000).
- 148 Rostom, A. A., Fucini, P., Benjamin, D. R., Juenemann, R., Nierhaus, K. H., Hartl, F. U., Dobson, C. M. & Robinson, C. V. Detection and selective dissociation of intact ribosomes in a mass spectrometer. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 5185-5190, doi:10.1073/pnas.97.10.5185 (2000).
- 149 Barrera Nelson, P., Di Bartolo, N., Booth Paula, J. & Robinson Carol, V. Micelles Protect Membrane Complexes from Solution to Vacuum. *Science* **321**, 243-246, doi:10.1126/science.1159292 (2008).
- 150 Rose, R. J., Damoc, E., Denisov, E., Makarov, A. & Heck, A. J. High-sensitivity Orbitrap mass analysis of intact macromolecular assemblies. *Nature Methods* **9**, 1084-1086, doi:10.1038/nmeth.2208 (2012).
- 151 Gault, J., Donlan, J. A., Liko, I., Hopper, J. T., Gupta, K., Housden, N. G., Struwe, W. B., Marty, M. T., Mize, T., Bechara, C., Zhu, Y., Wu, B., Kleanthous, C., Belov, M., Damoc, E., Makarov, A. & Robinson, C. V. High-resolution mass spectrometry of small

- molecules bound to membrane proteins. *Nature Methods* **13**, 333-336, doi:10.1038/nmeth.3771 (2016).
- 152 Aguilar-Gurreri, C., Larabi, A., Vinayachandran, V., Patel, N. A., Yen, K., Reja, R., Ebong, I.-O., Schoehn, G., Robinson, C. V., Pugh, B. F. & Panne, D. Structural evidence for Nap1-dependent H2A-H2B deposition and nucleosome assembly. *The EMBO journal* **35**, 1465-1482, doi:10.15252/embj.201694105 (2016).
 - 153 Jackson, V. A., Mehmood, S., Chavent, M., Roversi, P., Carrasquero, M., del Toro, D., Seyit-Bremer, G., Ranaivoson, F. M., Comoletti, D., Sansom, M. S. P., Robinson, C. V., Klein, R. & Seiradake, E. Super-complexes of adhesion GPCRs and neural guidance receptors. *Nature communications* **7**, 11184, doi:10.1038/ncomms11184 (2016).
 - 154 Abbas, Y. M., Wu, D., Bueler, S. A., Robinson, C. V. & Rubinstein, J. L. Structure of V-ATPase from the mammalian brain. *Science (New York, N.Y.)* **367**, 1240-1246, doi:10.1126/science.aaz2924 (2020).
 - 155 Liko, I., Degiacomi, M. T., Mohammed, S., Yoshikawa, S., Schmidt, C. & Robinson, C. V. Dimer interface of bovine cytochrome oxidase is influenced by local posttranslational modifications and lipid binding. *Proceedings of the National Academy of Sciences* **113**, 8230, doi:10.1073/pnas.1600354113 (2016).
 - 156 Su, C.-C., Klenotic, P. A., Bolla, J. R., Purdy, G. E., Robinson, C. V. & Yu, E. W. MmpL3 is a lipid transporter that binds trehalose monomycolate and phosphatidylethanolamine. *Proceedings of the National Academy of Sciences* **116**, 11241, doi:10.1073/pnas.1901346116 (2019).
 - 157 Bolla, J. R., Sauer, J. B., Wu, D., Mehmood, S., Allison, T. M. & Robinson, C. V. Direct observation of the influence of cardiolipin and antibiotics on lipid II binding to MurJ. *Nature Chemistry* **10**, 363-371, doi:10.1038/nchem.2919 (2018).

- 158 Gupta, K., Li, J., Liko, I., Gault, J., Bechara, C., Wu, D., Hopper, J. T. S., Giles, K., Benesch, J. L. P. & Robinson, C. V. Identifying key membrane protein lipid interactions using mass spectrometry. *Nature Protocols* **13**, 1106-1120, doi:10.1038/nprot.2018.014 (2018).
- 159 Liko, I., Allison, T. M., Hopper, J. T. S. & Robinson, C. V. Mass spectrometry guided structural biology. *Current Opinion in Structural Biology* **40**, 136-144, doi:https://doi.org/10.1016/j.sbi.2016.09.008 (2016).
- 160 Lübbert, C. & Peukert, W. Characterization of Electrospray Drop Size Distributions by Mobility-Classified Mass Spectrometry: Implications for Ion Clustering in Solution and Ion Formation Pathways. *Analytical Chemistry* **93**, 12862-12871, doi:10.1021/acs.analchem.1c00727 (2021).
- 161 Banerjee, S. & Mazumdar, S. Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte. *International Journal of Analytical Chemistry* **2012**, 282574, doi:10.1155/2012/282574 (2012).
- 162 Mehmood, S., Allison, T. M. & Robinson, C. V. Mass Spectrometry of Protein Complexes: From Origins to Applications. *Annual Review of Physical Chemistry* **66**, 453-474, doi:10.1146/annurev-physchem-040214-121732 (2015).
- 163 Juraschek, R., Dülcks, T. & Karas, M. Nanoelectrospray—More than just a minimized-flow electrospray ionization source. *Journal of the American Society for Mass Spectrometry* **10**, 300-308, doi:10.1016/S1044-0305(98)00157-3 (1999).
- 164 Wilm, M. & Mann, M. Analytical properties of the nanoelectrospray ion source. *Analytical Chemistry* **68**, 1-8, doi:10.1021/ac9509519 (1996).

- 165 Kebarle, P. & Verkerk, U. H. Electrospray: from ions in solution to ions in the gas phase, what we know now. *Mass Spectrometry Reviews* **28**, 898-917, doi:10.1002/mas.20247 (2009).
- 166 Fernandez de la Mora, J. Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism. *Analytica Chimica Acta* **406**, 93-104, doi:https://doi.org/10.1016/S0003-2670(99)00601-7 (2000).
- 167 Konermann, L., Ahadi, E., Rodriguez, A. D. & Vahidi, S. Unraveling the Mechanism of Electrospray Ionization. *Analytical Chemistry* **85**, 2-9, doi:10.1021/ac302789c (2013).
- 168 Benesch, J. L. P. Collisional activation of protein complexes: Picking up the pieces. *Journal of the American Society for Mass Spectrometry* **20**, 341-348, doi:10.1016/j.jasms.2008.11.014 (2009).
- 169 Borysik, A. J., Hewitt, D. J. & Robinson, C. V. Detergent release prolongs the lifetime of native-like membrane protein conformations in the gas-phase. *Journal of the American Chemical Society* **135**, 6078-6083, doi:10.1021/ja401736v (2013).
- 170 Ruotolo, B. T., Hyung, S. J., Robinson, P. M., Giles, K., Bateman, R. H. & Robinson, C. V. Ion mobility-mass spectrometry reveals long-lived, unfolded intermediates in the dissociation of protein complexes. *Angewandte Chemie International Edition* **46**, 8001-8004, doi:10.1002/anie.200702161 (2007).
- 171 Loo, J. A. Studying noncovalent protein complexes by electrospray ionization mass spectrometry. *Mass Spectrometry Reviews* **16**, 1-23, doi:10.1002/1098-2787 (1997).
- 172 Hoffmann, E. d. S. V. *Mass spectrometry : principles and applications*. (J. Wiley, 2007).
- 173 Zubarev, R. A. & Makarov, A. Orbitrap Mass Spectrometry. *Analytical Chemistry* **85**, 5288-5296, doi:10.1021/ac4001223 (2013).

- 174 Makarov, A. Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. *Analytical Chemistry* **72**, 1156-1162, doi:10.1021/ac991131p (2000).
- 175 Wales, T. E. & Engen, J. R. Hydrogen exchange mass spectrometry for the analysis of protein dynamics. *Mass Spectrometry Reviews* **25**, 158-170, doi:10.1002/mas.20064 (2006).
- 176 Zhang, Z. & Smith, D. L. Determination of amide hydrogen exchange by mass spectrometry: a new tool for protein structure elucidation. *Protein science : a publication of the Protein Society* **2**, 522-531, doi:10.1002/pro.5560020404 (1993).
- 177 Niimura, N., Chatake, T., Kurihara, K. & Maeda, M. Hydrogen and hydration in proteins. *Cell Biochemistry and Biophysics* **40**, 351-369, doi:10.1385/CBB:40:3:351 (2004).
- 178 Guttman, M. & Lee, K. K. Isotope Labeling of Biomolecules: Structural Analysis of Viruses by HDX-MS. *Methods in enzymology* **566**, 405-426, doi:10.1016/bs.mie.2015.05.021 (2016).
- 179 Hvidt, A. & Nielsen, S. O. Hydrogen exchange in proteins. *Advances in Protein Chemistry* **21**, 287-386, doi:10.1016/s0065-3233(08)60129-1 (1966).
- 180 Konermann, L., Tong, X. & Pan, Y. Protein structure and dynamics studied by mass spectrometry: H/D exchange, hydroxyl radical labeling, and related approaches. *Journal of Mass Spectrometry* **43**, 1021-1036, doi:10.1002/jms.1435 (2008).
- 181 Konermann, L., Pan, J. & Liu, Y. H. Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chemical Society Reviews* **40**, 1224-1234, doi:10.1039/c0cs00113a (2011).

- 182 Weis, D. D., Wales, T. E., Engen, J. R., Hotchko, M. & Ten Eyck, L. F. Identification and characterization of EX1 kinetics in H/D exchange mass spectrometry by peak width analysis. *Journal of the American Society for Mass Spectrometry* **17**, 1498-1509, doi:10.1016/j.jasms.2006.05.014 (2006).
- 183 Merkle, P. S., Gotfryd, K., Cuendet, M. A., Leth-Espensen, K. Z., Gether, U., Loland, C. J. & Rand, K. D. Substrate-modulated unwinding of transmembrane helices in the NSS transporter LeuT. *Science Advances*. **4**, eaar6179, doi:10.1126/sciadv.aar6179 (2018).
- 184 Zhou, J., Yang, L., DeColli, A., Freel Meyers, C., Nemeria, N. S. & Jordan, F. Conformational dynamics of 1-deoxy-d-xylulose 5-phosphate synthase on ligand binding revealed by H/D exchange MS. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 9355-9360, doi:10.1073/pnas.1619981114 (2017).
- 185 Nielsen, A. K., Moller, I. R., Wang, Y., Rasmussen, S. G. F., Lindorff-Larsen, K., Rand, K. D. & Loland, C. J. Substrate-induced conformational dynamics of the dopamine transporter. *Nature Communications* **10**, 2714, doi:10.1038/s41467-019-10449-w (2019).
- 186 Xiao, H., Hoerner, J. K., Eyles, S. J., Dobo, A., Voigtman, E., Mel'cuk, A. I. & Kaltashov, I. A. Mapping protein energy landscapes with amide hydrogen exchange and mass spectrometry: I. A generalized model for a two-state protein and comparison with experiment. *Protein Science*. **14**, 543-557, doi:10.1110/ps.041001705 (2005).
- 187 Yan, X., Zhang, H., Watson, J., Schimerlik, M. I. & Deinzer, M. L. Hydrogen/deuterium exchange and mass spectrometric analysis of a protein containing multiple disulfide bonds: Solution structure of recombinant macrophage colony stimulating factor-beta

- (rhM-CSFbeta). *Protein science : a publication of the Protein Society* **11**, 2113-2124, doi:10.1110/ps.0204402 (2002).
- 188 Zhang, H.-M., McLoughlin, S. M., Frausto, S. D., Tang, H., Emmett, M. R. & Marshall, A. G. Simultaneous reduction and digestion of proteins with disulfide bonds for hydrogen/deuterium exchange monitored by mass spectrometry. *Analytical Chemistry* **82**, 1450-1454, doi:10.1021/ac902550n (2010).
- 189 Gallagher, E. S. & Hudgens, J. W. in *Methods in enzymology* Vol. 566 (ed Zvi Kelman) 357-404 (Academic Press, 2016).
- 190 Engen, J. R. & Wales, T. E. Analytical Aspects of Hydrogen Exchange Mass Spectrometry. *Annual Review of Analytical Chemistry (Palo Alto Calif)* **8**, 127-148, doi:10.1146/annurev-anchem-062011-143113 (2015).
- 191 Senn, H. M. & Thiel, W. QM/MM Methods for Biomolecular Systems. *Angewandte Chemie International Edition* **48**, 1198-1229, doi:https://doi.org/10.1002/anie.200802019 (2009).
- 192 Hug, S. Classical molecular dynamics in a nutshell. *Methods in Molecular Biology* **924**, 127-152, doi:10.1007/978-1-62703-017-5_6 (2013).
- 193 Gundelach, L., Fox, T., Tautermann, C. S. & Skylaris, C.-K. Protein–ligand free energies of binding from full-protein DFT calculations: convergence and choice of exchange–correlation functional. *Physical Chemistry Chemical Physics* **23**, 9381-9393, doi:10.1039/D1CP00206F (2021).
- 194 Monticelli, L. & Tieleman, D. P. Force fields for classical molecular dynamics. *Methods in Molecular Biology* **924**, 197-213, doi:10.1007/978-1-62703-017-5_8 (2013).

- 195 Born, M. & Oppenheimer, R. Zur Quantentheorie der Molekeln. *Annalen der Physik* **389**, 457-484, doi:<https://doi.org/10.1002/andp.19273892002> (1927).
- 196 Lifson, S. & Warshel, A. Consistent Force Field for Calculations of Conformations, Vibrational Spectra, and Enthalpies of Cycloalkane and n-Alkane Molecules. *The Journal of Chemical Physics* **49**, 5116-5129, doi:10.1063/1.1670007 (1968).
- 197 Best, R. B., Zhu, X., Shim, J., Lopes, P. E. M., Mittal, J., Feig, M. & Mackerell, A. D., Jr. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ , ψ and side-chain $\chi(1)$ and $\chi(2)$ dihedral angles. *Journal of Chemical Theory and Computation* **8**, 3257-3273, doi:10.1021/ct300400x (2012).
- 198 Jorgensen, W. L. & Tirado-Rives, J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society* **110**, 1657-1666, doi:10.1021/ja00214a001 (1988).
- 199 Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., Guvench, O., Lopes, P., Vorobyov, I. & Mackerell, A. D., Jr. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of computational chemistry* **31**, 671-690, doi:10.1002/jcc.21367 (2010).
- 200 Leach, A. R. & Leach, A. R. *Molecular modelling: principles and applications*. (Pearson education, 2001).
- 201 Wang, Y. & McCammon, J. A. in *Computational Modeling of Biological Systems: From Molecules to Pathways* (ed Nikolay V. Dokholyan) 3-30 (Springer US, 2012).

- 202 Tieleman, P., Vattulainen, I., Tajkhorshid, E., Essex, J. W., Ulmschneider, M. B., Schulten, K., Robertson, J. L., Roux, B. & Khalid, S. *Molecular Simulations and Biomembranes: From Biophysics to Function*. (Royal Society of Chemistry, 2010).
- 203 Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An $N\log(N)$ method for Ewald sums in large systems. *Journal of Chemical Physics* **98**, 10089, doi:10.1063/1.464397 (1993).
- 204 Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H. & Pedersen, L. G. A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **103**, 8577-8593, doi:10.1063/1.470117 (1995).
- 205 Ewald, P. P. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Annalen der Physik* **369**, 253-287, doi:<https://doi.org/10.1002/andp.19213690304> (1921).
- 206 Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of computational chemistry* **4**, 187-217, doi:<https://doi.org/10.1002/jcc.540040211> (1983).
- 207 Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., de Groot, B. L., Grubmüller, H. & MacKerell, A. D., Jr. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods* **14**, 71-73, doi:10.1038/nmeth.4067 (2017).
- 208 Scott, W. R. P., Hünenberger, P. H., Tironi, I. G., Mark, A. E., Billeter, S. R., Fennen, J., Torda, A. E., Huber, T., Krüger, P. & van Gunsteren, W. F. The GROMOS Biomolecular Simulation Program Package. *The Journal of Physical Chemistry A* **103**, 3596-3607, doi:10.1021/jp984217f (1999).

- 209 Reif, M. M., Hünenberger, P. H. & Oostenbrink, C. New Interaction Parameters for Charged Amino Acid Side Chains in the GROMOS Force Field. *Journal of Chemical Theory and Computation* **8**, 3705-3723, doi:10.1021/ct300156h (2012).
- 210 Case, D. A., Cheatham, T. E., 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K. M., Jr., Onufriev, A., Simmerling, C., Wang, B. & Woods, R. J. The Amber biomolecular simulation programs. *Journal of computational chemistry* **26**, 1668-1688, doi:10.1002/jcc.20290 (2005).
- 211 Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E. & Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation* **11**, 3696-3713, doi:10.1021/acs.jctc.5b00255 (2015).
- 212 Jorgensen, W. L., Maxwell, D. S. & Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *Journal of the American Chemical Society* **118**, 11225-11236, doi:10.1021/ja9621760 (1996).
- 213 Pastor, R. W. & Mackerell, A. D., Jr. Development of the CHARMM Force Field for Lipids. *The Journal of Physical Chemistry Letters* **2**, 1526-1532, doi:10.1021/jz200167q (2011).
- 214 Berger, O., Edholm, O. & Jähnig, F. Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature. *Biophys J* **72**, 2002-2013, doi:10.1016/s0006-3495(97)78845-3 (1997).
- 215 Klauda, J. B., Venable, R. M., Freites, J. A., O'Connor, J. W., Tobias, D. J., Mondragon-Ramirez, C., Vorobyov, I., MacKerell, A. D. & Pastor, R. W. Update of the CHARMM

- All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *The Journal of Physical Chemistry B* **114**, 7830-7843, doi:10.1021/jp101759q (2010).
- 216 Vanommeslaeghe, K. & MacKerell, A. D. CHARMM additive and polarizable force fields for biophysics and computer-aided drug design. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1850**, 861-871, doi:https://doi.org/10.1016/j.bbagen.2014.08.004 (2015).
- 217 Venable, R. M., Brown, F. L. H. & Pastor, R. W. Mechanical properties of lipid bilayers from molecular dynamics simulation. *Chemistry and Physics of Lipids* **192**, 60-74, doi:https://doi.org/10.1016/j.chemphyslip.2015.07.014 (2015).
- 218 Yeh, I. C. & Hummer, G. Peptide loop-closure kinetics from microsecond molecular dynamics simulations in explicit solvent. *Journal of the American Chemical Society* **124**, 6563-6568, doi:10.1021/ja025789n (2002).
- 219 Paton, R. S. & Goodman, J. M. Hydrogen Bonding and π -Stacking: How Reliable are Force Fields? A Critical Evaluation of Force Field Descriptions of Nonbonded Interactions. *Journal of Chemical Information and Modeling* **49**, 944-955, doi:10.1021/ci900009f (2009).
- 220 Schmid, N., Eichenberger, A. P., Choutko, A., Riniker, S., Winger, M., Mark, A. E. & van Gunsteren, W. F. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *European Biophysics Journal* **40**, 843, doi:10.1007/s00249-011-0700-9 (2011).
- 221 Chiu, S.-W., Pandit, S. A., Scott, H. L. & Jakobsson, E. An Improved United Atom Force Field for Simulation of Mixed Lipid Bilayers. *The Journal of Physical Chemistry B* **113**, 2748-2763, doi:10.1021/jp807056c (2009).

- 222 Dama, J. F., Sinitskiy, A. V., McCullagh, M., Weare, J., Roux, B., Dinner, A. R. & Voth, G. A. The Theory of Ultra-Coarse-Graining. 1. General Principles. *Journal of Chemical Theory and Computation* **9**, 2466-2480, doi:10.1021/ct4000444 (2013).
- 223 Noid, W. G. Perspective: Coarse-grained models for biomolecular systems. *The Journal of Chemical Physics* **139**, 090901, doi:10.1063/1.4818908 (2013).
- 224 Mullinax, J. W. & Noid, W. G. Generalized Yvon-Born-Green Theory for Molecular Systems. *Physical Review Letters* **103**, 198104, doi:10.1103/PhysRevLett.103.198104 (2009).
- 225 Ingólfsson, H. I., Lopez, C. A., Uusitalo, J. J., de Jong, D. H., Gopal, S. M., Periole, X. & Marrink, S. J. The power of coarse graining in biomolecular simulations. *WIREs Computational Molecular Science* **4**, 225-248, doi:10.1002/wcms.1169 (2014).
- 226 Bennun, S. V., Hoopes, M. I., Xing, C. & Faller, R. Coarse-grained modeling of lipids. *Chemistry and Physics of Lipids* **159**, 59-66, doi:https://doi.org/10.1016/j.chemphyslip.2009.03.003 (2009).
- 227 Lyubartsev, A. P. & Rabinovich, A. L. Recent development in computer simulations of lipid bilayers. *Soft Matter* **7**, 25-39, doi:10.1039/C0SM00457J (2011).
- 228 Cascella, M. & Vanni, S. in *Chemical Modelling: Volume 12* Vol. 12 1-52 (The Royal Society of Chemistry, 2016).
- 229 Monticelli, L., Kandasamy, S. K., Periole, X., Larson, R. G., Tieleman, D. P. & Marrink, S.-J. The MARTINI Coarse-Grained Force Field: Extension to Proteins. *Journal of Chemical Theory and Computation* **4**, 819-834, doi:10.1021/ct700324x (2008).
- 230 de Jong, D. H., Lopez, C. A. & Marrink, S. J. Molecular view on protein sorting into liquid-ordered membrane domains mediated by gangliosides and lipid anchors. *Faraday Discussions* **161**, 347-363, doi:10.1039/C2FD20086D (2013).

- 231 Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P. & de Vries, A. H. The MARTINI force field: coarse grained model for biomolecular simulations. *The Journal of Physical Chemistry B* **111**, 7812-7824, doi:10.1021/jp071097f (2007).
- 232 López, C. A., de Vries, A. H. & Marrink, S. J. Computational microscopy of cyclodextrin mediated cholesterol extraction from lipid model membranes. *Scientific Reports* **3**, 2071, doi:10.1038/srep02071 (2013).
- 233 López, C. A., Rzepiela, A. J., de Vries, A. H., Dijkhuizen, L., Hünenberger, P. H. & Marrink, S. J. Martini Coarse-Grained Force Field: Extension to Carbohydrates. *Journal of Chemical Theory and Computation* **5**, 3195-3210, doi:10.1021/ct900313w (2009).
- 234 Uusitalo, J. J., Ingólfsson, H. I., Marrink, S. J. & Faustino, I. Martini Coarse-Grained Force Field: Extension to RNA. *Biophysical Journal* **113**, 246-256, doi:10.1016/j.bpj.2017.05.043 (2017).
- 235 Song, W., Corey, R. A., Ansell, T. B., Cassidy, C. K., Horrell, M. R., Duncan, A. L., Stansfeld, P. J. & Sansom, M. S. P. PyLipID: A Python package for analysis of protein-lipid interactions from MD simulations. *bioRxiv*, 2021.2007.2014.452312, doi:10.1101/2021.07.14.452312 (2021).
- 236 Jefferys, E., Sands, Z. A., Shi, J., Sansom, M. S. & Fowler, P. W. Alchembed: A Computational Method for Incorporating Multiple Proteins into Complex Lipid Geometries. *Journal of Chemical Theory and Computation* **11**, 2743-2754, doi:10.1021/ct501111d (2015).
- 237 Vickery, O. N. & Stansfeld, P. J. CG2AT2: An Enhanced Fragment-based approach for Serial Multi-scale Molecular Dynamics simulations. *bioRxiv*, 2021.2003.2025.437005, doi:10.1101/2021.03.25.437005 (2021).

- 238 Stansfeld, P. J., Hopkinson, R., Ashcroft, F. M. & Sansom, M. S. P. PIP2-Binding Site in Kir Channels: Definition by Multiscale Biomolecular Simulations. *Biochemistry* **48**, 10926-10933, doi:10.1021/bi9013193 (2009).
- 239 Reddy, T. & Sansom, M. S. The Role of the Membrane in the Structure and Biophysical Robustness of the Dengue Virion Envelope. *Structure* **24**, 375-382, doi:10.1016/j.str.2015.12.011 (2016).
- 240 Shinoda, W., DeVane, R. & Klein, M. L. Zwitterionic Lipid Assemblies: Molecular Dynamics Studies of Monolayers, Bilayers, and Vesicles Using a New Coarse Grain Force Field. *The Journal of Physical Chemistry B* **114**, 6836-6849, doi:10.1021/jp9107206 (2010).
- 241 Seo, S. & Shinoda, W. SPICA Force Field for Lipid Membranes: Domain Formation Induced by Cholesterol. *Journal of Chemical Theory and Computation* **15**, 762-774, doi:10.1021/acs.jctc.8b00987 (2019).
- 242 Orsi, M. & Essex, J. W. The ELBA Force Field for Coarse-Grain Modeling of Lipid Membranes. *PLOS ONE* **6**, e28637, doi:10.1371/journal.pone.0028637 (2011).
- 243 Darré, L., Machado, M. R., Brandner, A. F., González, H. C., Ferreira, S. & Pantano, S. SIRAH: A Structurally Unbiased Coarse-Grained Force Field for Proteins with Aqueous Solvation and Long-Range Electrostatics. *Journal of Chemical Theory and Computation* **11**, 723-739, doi:10.1021/ct5007746 (2015).
- 244 Brandner, A., Schüller, A., Melo, F. & Pantano, S. Exploring DNA dynamics within oligonucleosomes with coarse-grained simulations: SIRAH force field extension for protein-DNA complexes. *Biochemical and Biophysical Research Communications* **498**, 319-326, doi:https://doi.org/10.1016/j.bbrc.2017.09.086 (2018).

- 245 Machado, M. R., Barrera, E. E., Klein, F., Sónora, M., Silva, S. & Pantano, S. The SIRAH 2.0 Force Field: Altius, Fortius, Citius. *Journal of Chemical Theory and Computation* **15**, 2719-2733, doi:10.1021/acs.jctc.9b00006 (2019).
- 246 Kawamoto, S., Liu, H., Seo, S., Miyazaki, Y., Dixit, M., DeVane, R., MacDermaid, C., Fiorin, G., Klein, M. L. & Shinoda, W. The SPICA Coarse-Grained Force Field for Proteins and Peptides. *bioRxiv*, 2021.2012.2001.470839, doi:10.1101/2021.12.01.470839 (2021).
- 247 Marrink, S. J., de Vries, A. H. & Mark, A. E. Coarse Grained Model for Semiquantitative Lipid Simulations. *The Journal of Physical Chemistry B* **108**, 750-760, doi:10.1021/jp036508g (2004).
- 248 Petrov, D. & Zagrovic, B. Are Current Atomistic Force Fields Accurate Enough to Study Proteins in Crowded Environments? *PLOS Computational Biology* **10**, e1003638, doi:10.1371/journal.pcbi.1003638 (2014).
- 249 Best, R. B., Zheng, W. & Mittal, J. Balanced Protein–Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association. *Journal of Chemical Theory and Computation* **10**, 5113-5124, doi:10.1021/ct500569b (2014).
- 250 Henriques, J., Cragnell, C. & Skepö, M. Molecular Dynamics Simulations of Intrinsically Disordered Proteins: Force Field Evaluation and Comparison with Experiment. *Journal of Chemical Theory and Computation* **11**, 3420-3431, doi:10.1021/ct501178z (2015).
- 251 Miller, M. S., Lay, W. K. & Elcock, A. H. Osmotic Pressure Simulations of Amino Acids and Peptides Highlight Potential Routes to Protein Force Field Parameterization. *The Journal of Physical Chemistry B* **120**, 8217-8229, doi:10.1021/acs.jpcb.6b01902 (2016).

- 252 Schmalhorst, P. S., Deluweit, F., Scherrers, R., Heisenberg, C.-P. & Sikora, M. Overcoming the Limitations of the MARTINI Force Field in Simulations of Polysaccharides. *Journal of Chemical Theory and Computation* **13**, 5039-5053, doi:10.1021/acs.jctc.7b00374 (2017).
- 253 Javanainen, M., Martinez-Seara, H. & Vattulainen, I. Nanoscale Membrane Domain Formation Driven by Cholesterol. *Scientific Reports* **7**, 1143, doi:10.1038/s41598-017-01247-9 (2017).
- 254 Stark, A. C., Andrews, C. T. & Elcock, A. H. Toward Optimized Potential Functions for Protein–Protein Interactions in Aqueous Solutions: Osmotic Second Virial Coefficient Calculations Using the MARTINI Coarse-Grained Force Field. *Journal of Chemical Theory and Computation* **9**, 4176-4185, doi:10.1021/ct400008p (2013).
- 255 Berendsen, H. J. C., van der Spoel, D. & van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications* **91**, 43-56, doi:https://doi.org/10.1016/0010-4655(95)00042-E (1995).
- 256 Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B. & Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2**, 19-25, doi:https://doi.org/10.1016/j.softx.2015.06.001 (2015).
- 257 Hockney, R. W. POTENTIAL CALCULATION AND SOME APPLICATIONS. (1970).
- 258 Verlet, L. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Physical Review* **159**, 98-103, doi:10.1103/PhysRev.159.98 (1967).

- 259 Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* **52**, 7182-7190, doi:10.1063/1.328693 (1981).
- 260 Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics* **126**, 014101, doi:10.1063/1.2408420 (2007).
- 261 Berendsen, H., van Postma, J. P. M., van Gunsteren, W., DiNola, A. D. & Haak, J. R. *Molecular-Dynamics with Coupling to An External Bath*. Vol. 81 (1984).
- 262 Lynch, C. I., Rao, S. & Sansom, M. S. P. Water in Nanopores and Biological Channels: A Molecular Simulation Perspective. *Chemical Reviews* **120**, 10298-10335, doi:10.1021/acs.chemrev.9b00830 (2020).
- 263 Ouyang, J. F. & Bettens, R. P. Modelling Water: A Lifetime Enigma. *Chimia (Aarau)* **69**, 104-111, doi:10.2533/chimia.2015.104 (2015).
- 264 Onufriev, A. V. & Izadi, S. Water models for biomolecular simulations. *WIREs Computational Molecular Science* **8**, e1347, doi:https://doi.org/10.1002/wcms.1347 (2018).
- 265 Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F. & Hermans, J. in *Intermolecular Forces: Proceedings of the Fourteenth Jerusalem Symposium on Quantum Chemistry and Biochemistry Held in Jerusalem, Israel, April 13–16, 1981* (ed Bernard Pullman) 331-342 (Springer Netherlands, 1981).
- 266 Jorgensen, W. L. Quantum and statistical mechanical studies of liquids. 10. Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water. *Journal of the American Chemical Society* **103**, 335-340, doi:10.1021/ja00392a016 (1981).

- 267 Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **79**, 926-935, doi:10.1063/1.445869 (1983).
- 268 Jorgensen, W. L. & Madura, J. D. Temperature and size dependence for Monte Carlo simulations of TIP4P water. *Molecular Physics* **56**, 1381-1392, doi:10.1080/00268978500103111 (1985).
- 269 Izadi, S., Anandakrishnan, R. & Onufriev, A. V. Building Water Models: A Different Approach. *The Journal of Physical Chemistry Letters* **5**, 3863-3871, doi:10.1021/jz501780a (2014).
- 270 Mahoney, M. W. & Jorgensen, W. L. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *The Journal of Chemical Physics* **112**, 8910-8922, doi:10.1063/1.481505 (2000).
- 271 Yesylevskyy, S. O., Schäfer, L. V., Sengupta, D. & Marrink, S. J. Polarizable Water Model for the Coarse-Grained MARTINI Force Field. *PLOS Computational Biology* **6**, e1000810, doi:10.1371/journal.pcbi.1000810 (2010).
- 272 Kirkwood, J. G. Statistical Mechanics of Fluid Mixtures. *The Journal of Chemical Physics* **3**, 300-313, doi:10.1063/1.1749657 (1935).
- 273 Roux, B. The calculation of the potential of mean force using computer simulations. *Computer Physics Communications* **91**, 275-282, doi:https://doi.org/10.1016/0010-4655(95)00053-I (1995).
- 274 Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of computational chemistry* **13**, 1011-1021, doi:https://doi.org/10.1002/jcc.540130812 (1992).

- 275 Hub, J. S., de Groot, B. L. & van der Spoel, D. g_wham—A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *Journal of Chemical Theory and Computation* **6**, 3713-3720, doi:10.1021/ct100494z (2010).
- 276 Ruiz, N., Kahne, D. & Silhavy, T. J. Transport of lipopolysaccharide across the cell envelope: the long road of discovery. *Nature Reviews Microbiology* **7**, 677-683, doi:10.1038/nrmicro2184 (2009).
- 277 Sperandio, P., Lau, F. K., Carpentieri, A., De Castro, C., Molinaro, A., Deho, G., Silhavy, T. J. & Polissi, A. Functional analysis of the protein machinery required for transport of lipopolysaccharide to the outer membrane of Escherichia coli. *Journal of Bacteriology* **190**, 4460-4469, doi:10.1128/jb.00270-08 (2008).
- 278 Reading, E., Walton, Troy A., Liko, I., Marty, Michael T., Laganowsky, A., Rees, Douglas C. & Robinson, Carol V. The Effect of Detergent, Temperature, and Lipid on the Oligomeric State of MscL Constructs: Insights from Mass Spectrometry. *Chemistry & Biology* **22**, 593-603, doi:https://doi.org/10.1016/j.chembiol.2015.04.016 (2015).
- 279 Reading, E., Ahdash, Z., Fais, C., Ricci, V., Wang-Kan, X., Grimsey, E., Stone, J., Malloci, G., Lau, A. M., Findlay, H., Konijnenberg, A., Booth, P. J., Ruggerone, P., Vargiu, A. V., Piddock, L. J. V. & Politis, A. Perturbed structural dynamics underlie inhibition and altered efflux of the multidrug resistance pump AcrB. *Nature communications* **11**, 5565, doi:10.1038/s41467-020-19397-2 (2020).
- 280 Fiorentino, F., Bolla, J. R., Mehmood, S. & Robinson, C. V. The Different Effects of Substrates and Nucleotides on the Complex Formation of ABC Transporters. *Structure* **27**, 651-659, doi:10.1016/j.str.2019.01.010 (2019).

- 281 Urner, L. H., Liko, I., Yen, H.-Y., Hoi, K.-K., Bolla, J. R., Gault, J., Almeida, F. G., Schweder, M.-P., Shutin, D., Ehrmann, S., Haag, R., Robinson, C. V. & Pagel, K. Modular detergents tailor the purification and structural analysis of membrane proteins including G-protein coupled receptors. *Nature communications* **11**, 564, doi:10.1038/s41467-020-14424-8 (2020).
- 282 Thota, B. N. S., Urner, L. H. & Haag, R. Supramolecular Architectures of Dendritic Amphiphiles in Water. *Chemical Reviews* **116**, 2079-2102, doi:10.1021/acs.chemrev.5b00417 (2016).
- 283 Urner, L. H., Fiorentino, F., Shutin, D., Sauer, J. B., Agasid, M., Tarick, J. E.-B., Bolla, J. R., Stansfeld, P. J. & Robinson, C. V. Discovering the lipidome surrounding membrane proteins with tunable detergents and native mass spectrometry. *Under Review* (2021).
- 284 Laganowsky, A., Reading, E., Hopper, J. T. & Robinson, C. V. Mass spectrometry of intact membrane protein complexes. *Nature Protocols* **8**, 639-651, doi:10.1038/nprot.2013.024 (2013).
- 285 Korkhov, V. M., Mireku, S. A., Veprintsev, D. B. & Locher, K. P. Structure of AMP-PNP-bound BtuCD and mechanism of ATP-powered vitamin B12 transport by BtuCD-F. *Nature Structural & Molecular Biology* **21**, 1097-1099, doi:10.1038/nsmb.2918 (2014).
- 286 Mamat, U., Wilke, K., Bramhill, D., Schromm, A. B., Lindner, B., Kohl, T. A., Corchero, J. L., Villaverde, A., Schaffer, L., Head, S. R., Souvignier, C., Meredith, T. C. & Woodard, R. W. Detoxifying *Escherichia coli* for endotoxin-free production of recombinant proteins. *Microbial cell factories* **14**, 57-57, doi:10.1186/s12934-015-0241-5 (2015).

- 287 Shearer, J., Jefferies, D. & Khalid, S. Outer Membrane Proteins OmpA, FhuA, OmpF, EstA, BtuB, and OmpX Have Unique Lipopolysaccharide Fingerprints. *Journal of Chemical Theory and Computation* **15**, 2608-2619 (2019).
- 288 Tran, A. X., Dong, C. & Whitfield, C. Structure and functional analysis of LptC, a conserved membrane protein involved in the lipopolysaccharide export pathway in *Escherichia coli*. *Journal of Biological Chemistry* **285**, 33529-33539, doi:10.1074/jbc.M110.144709 (2010).
- 289 Narita, S. & Tokuda, H. Biochemical characterization of an ABC transporter LptBFGC complex required for the outer membrane sorting of lipopolysaccharides. *FEBS Lett.* **583**, 2160-2164, doi:10.1016/j.febslet.2009.05.051 (2009).
- 290 Tran, A. X., Trent, M. S. & Whitfield, C. The LptA protein of *Escherichia coli* is a periplasmic lipid A-binding protein involved in the lipopolysaccharide export pathway. *Journal of Biological Chemistry* **283**, 20342-20349, doi:10.1074/jbc.M802503200 (2008).
- 291 Bollati, M., Villa, R., Gourlay, L. J., Benedet, M., Deho, G., Polissi, A., Barbiroli, A., Martorana, A. M., Sperandio, P., Bolognesi, M. & Nardini, M. Crystal structure of LptH, the periplasmic component of the lipopolysaccharide transport machinery from *Pseudomonas aeruginosa*. *FEBS Journal* **282**, 1980-1997, doi:10.1111/febs.13254 (2015).
- 292 Thomas, C. & Tampe, R. Structural and Mechanistic Principles of ABC Transporters. *Annual Review of Biochemistry* **89**, 605-636, doi:10.1146/annurev-biochem-011520-105201 (2020).
- 293 Tang, X., Chang, S., Luo, Q., Zhang, Z., Qiao, W., Xu, C., Zhang, C., Niu, Y., Yang, W., Wang, T., Zhang, Z., Zhu, X., Wei, X., Dong, C., Zhang, X. & Dong, H. Cryo-EM

- structures of lipopolysaccharide transporter LptB2FGC in lipopolysaccharide or AMP-PNP-bound states reveal its transport mechanism. *Nature communications* **10**, 4175, doi:10.1038/s41467-019-11977-1 (2019).
- 294 Sadikot, R. T., Blackwell, T. S., Christman, J. W. & Prince, A. S. Pathogen-host interactions in *Pseudomonas aeruginosa* pneumonia. *American Journal of Respiratory and Critical Care Medicine* **171**, 1209-1223, doi:10.1164/rccm.200408-1044SO (2005).
- 295 DK, Augustin., Y, Song., MS, Baek., Y, Sawa., G, Singh., B, Taylor., A, Rubio-Millis., JL, Flanagan., JP, Weiner-Kronish. & SV, Lynch. Presence or Absence of Lipopolysaccharide O Antigens Affects Type III Secretion by *Pseudomonas Aeruginosa*. *Journal of Bacteriology* **189**, 2203-2209, doi:10.1128/JB.01839-06 (2007).
- 296 Webb, B. & Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Current protocols in bioinformatics* **54**, 5.6.1-5.6.37, doi:10.1002/cpbi.3 (2016).
- 297 Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Gallo Cassarino, T., Bertoni, M., Bordoli, L. & Schwede, T. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research* **42**, W252-258, doi:10.1093/nar/gku340 (2014).
- 298 Rassam, P., Copeland, N. A., Birkholz, O., Tóth, C., Chavent, M., Duncan, A. L., Cross, S. J., Housden, N. G., Kaminska, R., Seger, U., Quinn, D. M., Garrod, T. J., Sansom, M. S. P., Piehler, J., Baumann, C. G. & Kleanthous, C. Supramolecular assemblies underpin turnover of outer membrane proteins in bacteria. *Nature* **523**, 333-336, doi:10.1038/nature14461 (2015).

- 299 Luo, Q., Yang, X., Yu, S., Shi, H., Wang, K., Xiao, L., Zhu, G., Sun, C., Li, T., Li, D., Zhang, X., Zhou, M. & Huang, Y. Structural basis for lipopolysaccharide extraction by ABC transporter LptB2FG. *Nature Structural & Molecular Biology* **24**, 469-474, doi:10.1038/nsmb.3399 (2017).
- 300 Dong, H., Zhang, Z., Tang, X., Paterson, N. G. & Dong, C. Structural and functional insights into the lipopolysaccharide ABC transporter LptB2FG. *Nature Communications* **8**, 222, doi:10.1038/s41467-017-00273-5 (2017).
- 301 Suits, M. D., Sperandeo, P., Deho, G., Polissi, A. & Jia, Z. Novel structure of the conserved gram-negative lipopolysaccharide transport protein A and mutagenesis analysis. *Journal of Molecular Biology* **380**, 476-488, doi:10.1016/j.jmb.2008.04.045 (2008).
- 302 Tran, A. X., Dong, C. & Whitfield, C. Structure and functional analysis of LptC, a conserved membrane protein involved in the lipopolysaccharide export pathway in *Escherichia coli*. *Journal of Molecular Biology* **285**, 33529-33539, doi:10.1074/jbc.M110.144709 (2010).
- 303 Shen, M.-y. & Sali, A. Statistical potential for assessment and prediction of protein structures. *Protein Science* **15**, 2507-2524, doi:https://doi.org/10.1110/ps.062416606 (2006).
- 304 The PyMOL Molecular Graphics System v. 1.4 (Schrödinger LLC, 2010).
- 305 de Jong, D. H., Singh, G., Bennett, W. F., Arnarez, C., Wassenaar, T. A., Schafer, L. V., Periole, X., Tieleman, D. P. & Marrink, S. J. Improved Parameters for the Martini Coarse-Grained Protein Force Field. *Journal of Chemical Theory and Computation* **9**, 687-697, doi:10.1021/ct300646g (2013).

- 306 Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics* **126**, 014101, doi:10.1063/1.2408420 (2007).
- 307 UniProt Consortium, T. UniProt: the universal protein knowledgebase. *Nucleic Acids Research* **46**, 2699, doi:10.1093/nar/gky092 (2018).
- 308 Khademi, S., O'Connell, J., Remis, J., Robles-Colmenares, Y., Miercke Larry, J. W. & Stroud Robert, M. Mechanism of Ammonia Transport by Amt/MEP/Rh: Structure of AmtB at 1.35 Å. *Science* **305**, 1587-1594, doi:10.1126/science.1101952 (2004).
- 309 Joseph, B., Korkhov, V. M., Yulikov, M., Jeschke, G. & Bordignon, E. Conformational cycle of the vitamin B12 ABC importer in liposomes detected by double electron-electron resonance (DEER). *The Journal of biological chemistry* **289**, 3176-3185, doi:10.1074/jbc.M113.512178 (2014).
- 310 de Jong, D. H., Singh, G., Bennett, W. F. D., Arnarez, C., Wassenaar, T. A., Schäfer, L. V., Periole, X., Tieleman, D. P. & Marrink, S. J. Improved Parameters for the Martini Coarse-Grained Protein Force Field. *Journal of Chemical Theory and Computation* **9**, 687-697, doi:10.1021/ct300646g (2013).
- 311 Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAAnalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of computational chemistry* **32**, 2319-2327, doi:10.1002/jcc.21787 (2011).
- 312 Sperandeo, P., Cescutti, R., Villa, R., Di Benedetto, C., Candia, D., Dehò, G. & Polissi, A. Characterization of lptA and lptB, two essential genes implicated in lipopolysaccharide transport to the outer membrane of Escherichia coli. *Journal of Bacteriology* **189**, 244-253, doi:10.1128/jb.01126-06 (2007).

- 313 Qiao, S., Luo, Q., Zhao, Y., Zhang, X. C. & Huang, Y. Structural basis for lipopolysaccharide insertion in the bacterial outer membrane. *Nature* **511**, 108-111, doi:10.1038/nature13484 (2014).
- 314 Sestito, S. E., Sperandeo, P., Santambrogio, C., Ciaramelli, C., Calabrese, V., Rovati, G. E., Zambelloni, L., Grandori, R., Polissi, A. & Peri, F. Functional characterization of E. coli LptC: interaction with LPS and a synthetic ligand. *Chembiochem* **15**, 734-742, doi:10.1002/cbic.201300805 (2014).
- 315 Schultz, K. M., Lundquist, T. J. & Klug, C. S. Lipopolysaccharide binding to the periplasmic protein LptA. *Protein Sci* **26**, 1517-1523, doi:10.1002/pro.3177 (2017).
- 316 Schultz, K. M. & Klug, C. S. Characterization of and lipopolysaccharide binding to the E. coli LptC protein dimer. *Protein Science* **27**, 381-389, doi:10.1002/pro.3322 (2018).
- 317 Santambrogio, C., Sperandeo, P., Villa, R., Sobott, F., Polissi, A. & Grandori, R. LptA assembles into rod-like oligomers involving disorder-to-order transitions. *Journal of the American Society for Mass Spectrometry* **24**, 1593-1602, doi:10.1007/s13361-013-0687-9 (2013).
- 318 Asmar, A. T. & Collet, J.-F. Lpp, the Braun lipoprotein, turns 50—major achievements and remaining issues. *FEMS Microbiology Letters* **365**, doi:10.1093/femsle/fny199 (2018).
- 319 Botos, I., Majdalani, N., Mayclin, S. J., McCarthy, J. G., Lundquist, K., Wojtowicz, D., Barnard, T. J., Gumbart, J. C. & Buchanan, S. K. Structural and Functional Characterization of the LPS Transporter LptDE from Gram-Negative Pathogens. *Structure* **24**, 965-976, doi:10.1016/j.str.2016.03.026 (2016).

- 320 Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry* **31**, 455-461, doi:10.1002/jcc.21334 (2010).
- 321 Mynors-Wallis, B. Dissertation. (2018).
- 322 Spoel, D. V. D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E. & Berendsen, H. J. C. GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry* **26**, 1701-1718, doi:10.1002/jcc.20291 (2005).
- 323 Bonomi, M., Branduardi, D., Bussi, G., Camilloni, C., Provasi, D., Raiteri, P., Donadio, D., Marinelli, F., Pietrucci, F., Broglia, R. A. & Parrinello, M. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Computer Physics Communications* **180**, 1961-1972, doi:<https://doi.org/10.1016/j.cpc.2009.05.011> (2009).
- 324 Qiao, S., Luo, Q., Zhao, Y., Zhang, X. C. & Huang, Y. Structural basis for lipopolysaccharide insertion in the bacterial outer membrane. *Nature* **511**, 108-111, doi:10.1038/nature13484 (2014).
- 325 Gu, Y., Stansfeld, P. J., Zeng, Y., Dong, H., Wang, W. & Dong, C. Lipopolysaccharide is inserted into the outer membrane through an intramembrane hole, a lumen gate, and the lateral opening of LptD. *Structure* **23**, 496-504, doi:10.1016/j.str.2015.01.001 (2015).
- 326 Srinivas, N., Jetter, P., Ueberbacher, B. J., Werneburg, M., Zerbe, K., Steinmann, J., Meijden, B. V. d., Bernardini, F., Lederer, A., Dias, R. L. A., Misson, P. E., Henze, H., Zumbunn, J., Gombert, F. O., Obrecht, D., Hunziker, P., Schauer, S., Ziegler, U., Käch, A., Eberl, L., Riedel, K., DeMarco, S. J. & Robinson, J. A. Peptidomimetic Antibiotics

- Target Outer-Membrane Biogenesis in *Pseudomonas aeruginosa*. *Science* **327**, 1010-1013, doi:10.1126/science.1182749 (2010).
- 327 Andolina, G., Bencze, L. C., Zerbe, K., Muller, M., Steinmann, J., Kocherla, H., Mondal, M., Sobek, J., Moehle, K., Malojcic, G., Wollscheid, B. & Robinson, J. A. A Peptidomimetic Antibiotic Interacts with the Periplasmic Domain of LptD from *Pseudomonas aeruginosa*. *ACS Chemical Biology* **13**, 666-675, doi:10.1021/acscchembio.7b00822 (2018).
- 328 Werneburg, M., Zerbe, K., Juhas, M., Bigler, L., Stalder, U., Kaech, A., Ziegler, U., Obrecht, D., Eberl, L. & Robinson, J. A. Inhibition of lipopolysaccharide transport to the outer membrane in *Pseudomonas aeruginosa* by peptidomimetic antibiotics. *Chembiochem* **13**, 1767-1775, doi:10.1002/cbic.201200276 (2012).
- 329 Vetterli, S. U., Zerbe, K., Muller, M., Urfer, M., Mondal, M., Wang, S. Y., Moehle, K., Zerbe, O., Vitale, A., Pessi, G., Eberl, L., Wollscheid, B. & Robinson, J. A. Thanatin targets the intermembrane protein complex required for lipopolysaccharide transport in *Escherichia coli*. *Science Advances* **4**, eaau2634, doi:10.1126/sciadv.aau2634 (2018).
- 330 Fehlbaum, P., Bulet, P., Chernysh, S., Briand, J. P., Roussel, J. P., Letellier, L., Hetru, C. & Hoffmann, J. A. Structure-activity analysis of thanatin, a 21-residue inducible insect defense peptide with sequence homology to frog skin antimicrobial peptides. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 1221-1225, doi:10.1073/pnas.93.3.1221 (1996).
- 331 Jefferys, E., Sands, Z. A., Shi, J., Sansom, M. S. P. & Fowler, P. W. Alchembed: A Computational Method for Incorporating Multiple Proteins into Complex Lipid

- Geometries. *Journal of Chemical Theory and Computation* **11**, 2743-2754, doi:10.1021/ct501111d (2015).
- 332 Tomasek, D., Rawson, S., Lee, J., Wzorek, J. S., Harrison, S. C., Li, Z. & Kahne, D. Structure of a nascent membrane protein as it folds on the BAM complex. *Nature*, doi:10.1038/s41586-020-2370-1 (2020).
- 333 Botte, M., Ni, D., Schenck, S., Zimmermann, I., Chami, M., Bocquet, N., Egloff, P., Bucher, D., Trabuco, M., Cheng, R. K. Y., Brunner, J. D., Seeger, M. A., Stahlberg, H. & Hennig, M. Insight into Lipopolysaccharide Translocation by Cryo-EM structures of a LptDE Transporter in Complex with Pro-Macrobodies. *bioRxiv*, 2021.2003.2023.436624, doi:10.1101/2021.03.23.436624 (2021).
- 334 Persson, F. & Halle, B. How amide hydrogens exchange in native proteins. *Proceedings of the National Academy of Sciences* **112**, 10383, doi:10.1073/pnas.1506079112 (2015).
- 335 Park, I.-H., Venable, J. D., Steckler, C., Cellitti, S. E., Lesley, S. A., Spraggon, G. & Brock, A. Estimation of Hydrogen-Exchange Protection Factors from MD Simulation Based on Amide Hydrogen Bonding Analysis. *Journal of Chemical Information and Modeling* **55**, 1914-1925, doi:10.1021/acs.jcim.5b00185 (2015).
- 336 Markwick, P. R. L., Peacock, R. B. & Komives, E. A. Accurate Prediction of Amide Exchange in the Fast Limit Reveals Thrombin Allostery. *Biophysical Journal* **116**, 49-56, doi:10.1016/j.bpj.2018.11.023 (2019).
- 337 Bradshaw, R. T., Marinelli, F., Faraldo-Gómez, J. D. & Forrest, L. R. Interpretation of HDX Data by Maximum-Entropy Reweighting of Simulated Structural Ensembles. *Biophysical Journal* **118**, 1649-1664, doi:10.1016/j.bpj.2020.02.005 (2020).

- 338 Ahdash, Z., Pyle, E., Allen, W. J., Corey, R. A., Collinson, I. & Politis, A. HDX-MS reveals nucleotide-dependent, anti-correlated opening and closure of SecA and SecY channels of the bacterial translocon. *eLife* **8**, e47402, doi:10.7554/eLife.47402 (2019).
- 339 Wassenaar, T. A., Ingólfsson, H. I., Böckmann, R. A., Tieleman, D. P. & Marrink, S. J. Computational Lipidomics with insane: A Versatile Tool for Generating Custom Membranes for Molecular Simulations. *Journal of Chemical Theory and Computation* **11**, 2144-2155, doi:10.1021/acs.jctc.5b00209 (2015).
- 340 Stansfeld, P. J. & Sansom, M. S. P. From Coarse Grained to Atomistic: A Serial Multiscale Approach to Membrane Protein Simulations. *Journal of Chemical Theory and Computation* **7**, 1157-1166, doi:10.1021/ct100569y (2011).
- 341 R. J. Gowers, M. L., J. Barnoud, T. J. E. Reddy, M. N. Melo, S. L. Seyler, D. L. Dotson, J. Domanski, S. Buchoux, I. M. Kenney, and O. Beckstein. MDAnalysis: A Python package for the rapid analysis of molecular dynamics simulations. *Proceedings of the 15th Python in Science Conference*, 98-105, doi:10.25080/Majora-629e541a-00e (2016).
- 342 Lau, A. M. C., Ahdash, Z., Martens, C. & Politis, A. Deuterios: software for rapid analysis and visualization of data from differential hydrogen deuterium exchange-mass spectrometry. *Bioinformatics* **35**, 3171-3173, doi:10.1093/bioinformatics/btz022 (2019).
- 343 Duncan, A. L., Reddy, T., Koldsø, H., Hélie, J., Fowler, P. W., Chavent, M. & Sansom, M. S. P. Protein crowding and lipid complexity influence the nanoscale dynamic organization of ion channels in cell membranes. *Scientific Reports* **7**, 16647, doi:10.1038/s41598-017-16865-6 (2017).

