Structural and Functional Studies of a Phospholipid Transporter Involved in the Maintenance of the Outer Membrane Asymmetry in Gram-negative Bacteria

A Thesis

Submitted in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

by

Angshu Dutta
(166106020)

Under supervision of

Prof. Shankar Prasad Kanaujia

Department of Biosciences and Bioengineering
Indian Institute of Technology Guwahati
Guwahati-781039, Assam, India

August 2023

Dedicated to all the people who stood by me

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

DEPARTMENT OF BIOSCIENCES AND BIOENGINEERING

DECLARATION

I do hereby declare that the research findings of this thesis entitled "**Structural and Functional Studies of a Phospholipid Transporter Involved in the Maintenance of the Outer Membrane Asymmetry in Gram-negative Bacteria**" are the result of research work carried out by me at the Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, Guwahati, India, under the supervision of **Prof. Shankar Prasad Kanaujia**.

I also declare that the contents of this thesis have not been the basis for an award of any other degree, diploma, fellowship or any other similar title of any University or Institution.

In keeping with the general practice of reporting scientific observations, due acknowledgments have been made wherever the research findings of other researchers have been cited in this thesis.

Angshu Dutta

Date: August 21, 2023 **Angshu Dutta**

Place: IIT Guwahati

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

DEPARTMENT OF BIOSCIENCES AND BIOENGINEERING

CERTIFICATE

This is to certify that the work described in the thesis entitled "**Structural and Functional Studies of a Phospholipid Transporter Involved in the Maintenance of the Outer Membrane Asymmetry in Gram-negative Bacteria**" is an authentic record of the results obtained from the research work carried out by **Angshu Dutta** at the Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, Assam, India, under my supervision and this work in part or as a whole has not been submitted elsewhere for the award of any other degree.

Date: August 21, 2023

Place: IIT Guwahati

Prof. Shankar Prasad Kanaujia

(Thesis Supervisor)

Acknowledgements

Acknowledgements

The doctoral research presented in this thesis involves the significant contributions of numerous people who have touched my life in numerous ways. It is my utmost honor and pleasure to put forward my gratitude to them.

Foremost, I would like to give my warmest thanks to my Ph.D. supervisor, **Prof. Shankar Prasad Kanaujia**, for his constant guidance and support throughout my Ph.D. tenure. This work and the thesis would not have seen the light of day without his persistent support, tremendous guidance and words of encouragement. I shall remain forever in debt to him for giving me the opportunity to work in the lab and providing the necessary resources to conduct scientific research. His zeal for acquiring knowledge and his ability to critically analyze always inspired me to think rationally. His perseverance and efforts have completely transformed me into an aspiring researcher.

I would like to put forward my sincere gratitude to the members of my doctoral committee - **Dr. A. M. Limaye**, **Prof. M. Kumar and Dr. K. P. Bhabak** for their valuable suggestions and healthy discussions over the past years. I would also like to thank **Prof. V. K. Dubey** for his valuable time in evaluating my comprehensive examination. I thank the Heads of the Department of Biosciences and Bioengineering (past and present)- **Prof. V. V. Dasu**, **Prof. K. Pakshirajan**, **Prof. L. Rangan** and **Prof. R. Chaturvedi** for being cooperative and providing all the necessary departmental facilities. I shall always be grateful to the infrastructural support and academic environment of the Department of Biosciences and Bioengineering at IIT Guwahati.

I am grateful to the Central Instruments Facility (CIF), IIT Guwahati, for providing access to sophisticated instruments such as the X-ray diffractometer and mass spectrometer, which was critical for the completion of this work. I thank the Heads of CIF (past and present) for always being so supportive and understanding. I also express my gratitude to the technical team (especially **Dr. A. Malakar** and **Mr. S. K. Deb**) and the non-technical team (security guards and caretakers) for their constant cooperation.

I sincerely acknowledge the financial aid received from the Ministry of Human Resource and Development (MHRD), Government of India (GOI), for providing me a fellowship. I am grateful to the Science and Engineering Research Board (SERB),

Department of Science and Technology, Government of India, for providing a research grant to our laboratory. I am also thankful for various other funding bodies that have supported our laboratory.

I would take this opportunity to thank the past and present members of the Structural and Computational Biology Laboratory (SCBL) for being supportive and creating a scientific environment. I thank my seniors, **Dr. Prerana Gogoi**, **Dr. Monika Chandravanshi** and **Dr. Suraj Kumar Mandal**, for helping me train during my formative days in the lab. I also thank my fellow juniors, **Sayan Saha**, **Pratik Das Gupta**, **Kalyan Ghosh**, **Ritu Tripathi**, **Arpana Gupta**, **Navneeth Sriram**, **Shimpy Nigam** and **Jeancolar Thoudam**, for their immense cooperation and timely help. The thesis would be incomplete without the valuable help of **Smit Patel**, who set up the simulation experiments. I would also like to thank him for his helping hand during the process of thesis writing. I would like to thank other members of the lab **Prerana Mordina**, **Shreya LB**, **Sakshi Chauhan**, **Ankit Kumar Sinha**, **Kanchan Chauhan**, **Dr. Ambuj Srivastava**, **Reshama Samanta**, **Smriti Tapan Dolai**, **Arunabh Sharma** and **Dr. Ankita Gupta** for their support and understanding.

I am fortunate to have well-wishers who never failed to help and inspire me – **Dr. Himanshu Sharma**, **Dr. Rocktotpal Konwarh**, **Dr. Vivek Prakash**, **Dr. Mahesh Nagargoje**, **Dr. Chandi Patra**, **Dr. Ganesh Nehru**, **Dr. Joseph Christakiran Moses**, **Dr. Snigdha Saikia**, **Adhiraj Nath**, **Jonjyoti Kalita**, **Suvankar Ghosh**, **Anurag Priyadarshi**, **Aman Bhardwaj**, **Laipubam Gayatri Sharma**, **Chinmaya Panda**, **Mahesh Das**, **Md. Saddam Hussain**, **Feba Francis**, **Arijit Nath, Dr. Abhishek Suman, Dr Bhuvan Dixit, Vineet Anand, Ankita Saikia, Geeta Ba, Mahi & family and Sharan**. I would also like to express my gratitude to **Vaibhav Pratap Singh** and **Sumit Baraskar** for always being supportive without being judgmental. I shall be forever grateful to **Vardan**, **Akshay**, **Ashwin**, **Haridas** and friends for being so helpful and encouraging.

I thank **Dr. Jagruti Chasmawala**, **Rasesh Chasmawala** and **Jignyasu Chasmawala** for their unconditional love, care and affection.

I shall always be grateful to **Dr. Subrata Sinha** (Dibrugarh University) for constantly encouraging me to purse my passion in the field of scientific research. **Prof. K. Hanumae Gowd** (Central University of Karnataka) has played a critical role in

developing my analytical abilities to conduct scientific investigations during my post-graduation days. I express my gratitude to him for helping me hone my skill sets.

I am grateful to a number of institutes and academies- Tezpur University, Institute of Advanced Study in Science and Technology, CSIR – North East Institute of Science and Technology (NEIST), Central University of Karnataka, Indian Academy of Sciences (Summer Research Fellowship Program) for giving me the opportunity to conduct small-term research during my graduation and post-graduation. These trainings have helped me develop strong scientific reasoning that significantly aided my doctoral study. I am sincerely grateful to Scholarship for Higher Education (SHE) scheme under Innovation in Science Pursuit for Inspired Research (INSPIRE) programme that supported my graduation and post-graduation studies as well as encouraged me to pursue research as a career.

There are people who transcend the idea of flesh, blood and friendship. From being a naïve boy to becoming an adult with a conscience, my journey as a human being, as well as a researcher, would be incomplete if I do not take the name of the person who has been a witness- **Sanjiv Dutta**. Welcoming, warm and inspiring, Sanjiv's impact on my life is incomparable. I shall forever remain grateful to him for his support and encouragement, which always lifted me up whenever I hit the lowest of the lows. My batchmate **Yashavanth PR** who never failed to give his helping hand even before I asked for is a blessing from the heavens. His words of encouragement and his constant belief in me always provided the force to proceed forward during my doctoral journey. I am constantly amazed by the strength and courage of **Sukumar Purohit** to stand against the world for the people he loves and cares. A dedicated researcher but, most importantly, a good human being, Sukumar's love and support for me remain incomparable in my life. I am glad to have met these people as they now have become an extension of my being. I shall always be grateful to **Dr. Barsha Dutta (Kakoty)** and **Dr. Barnali Sharma**, who have always inspired me to think rationally and work hard. Their contributions to my life during my formative years are incomparable. It was through their endeavors I was able to see how physics and mathematics can be the keys to unanswered questions. A realization that has immensely helped me conduct experiments during my Ph.D. My wonderful teachers **Kanishka Dutta** and **Sicily Dutta** have played a significant role in improving my writing and narrating skills which have helped immensely during my Ph.D. I express my gratitude to **Moushumi Madam**, **Sangeeta Madam**, **D'Souza Madam** and **Nidhi Madam**, for they have been

instrumental in molding my love for literature and language, which in turn has helped me in my doctoral studies. I would also like to thank other teachers, **Madhumita Madam**, **Manju Narain**, **Madhumita Barman** and **Nitul Ahmed**, who laid down the foundation stones for biology and chemistry in my life. My gratitude goes to my teachers at Kanoi college (Dept. of Zoology, Botany, Chemistry, Statistics) and Dibrugarh University (Dept. of Life Sciences), who always believed in me and constantly guided me.

I would like to put forward my heartfelt gratitude to all the stalwarts of macromolecular crystallography and structural biology whose works have always inspired me. I would like to thank **Late Prof. M. Vijayan,** whose contributions and legacy never cease to encourage me to explore the beautiful world of protein crystallography. My evolution into a researcher would not have been possible if I had not been touched by the luminous works of genius storytellers and luminaries such as **JRR Tolkien**, **André Aciman**, **Bill Konigsberg**, **Dr. Shashi Tharoor**, **Dr. Jahnavi Barua**, **Dustin Lance Black**, **Prof. Sujata Sharma** and many more. Their literary works never failed to create a safe haven for me to dwell in and introspect. Through their contributions, they have taught me the importance of reading, enjoying one's own company and being self-sufficient. For that, I am forever grateful. I am thankful to the numerous bibliophiles who have always admired my works which I feel are at most mediocre

I would like to express my heartfelt gratitude to my genesis – my beloved **Ma** and **Deta** for their selfless and never-ending love towards me. Their numerous sacrifices and impeccable support enabled me to pursue PhD. They have left no stone unturned to shield me from hardships and always provided me with more than I deserved. It is my privilege to be their son. Equally significant is my elder brother, **Maina**, whose constant care and support inspired me to move ahead in life. My life would be incomplete without the prayers and good wishes of **Mira Mahi** and **Renu Mahi**. Forever grateful.

Last but not least, I would like to thank the Almighty God for showering me with his blessings and strength to achieve the goal.

Angshu Dutta

Angshu Dutta

Table of contents

Table of contents

Section	Content	Page No.
	Abstract	xvii
	Abbreviations and notations	xxv
	List of figures	xxxi
	List of tables	xxxvii
	Chapter 1- Introduction	1
1	Introduction	3
1.1	Bacterial cell envelope	3
1.2	Molecular assembly of the OM	5
1.3	Transport systems involved in the biogenesis of the OM	8
1.3.1	Lpt transporter	8
1.3.2	Lol transporter	10
1.3.3	Bam complex	11
1.3.4.	The process of PL transport to the OM and maintenance of PL is poorly understood	12
1.3.4.1	PL biosynthesis pathway	12
1.3.4.2	PL sorting	13
1.3.4.3	OM enzymes involved in the maintenance of PL asymmetry	13
1.3.4.4	Transporters involved in ferrying PLs between the membranes	15
1.4	The Mla system does not possess a typical ABC transporter construct	19
1.4.1	ABC importers	21
1.4.2	ABC exporters	22
1.4.3	Transmembrane domain	23
1.4.4	Nucleotide-binding domain	24
1.4.5	Substrate (solute)-binding protein (SBP)	26
1.4.5.1	Classification of SBPs	27
1.5	The Mla system shows similarity to the Mce system	32
1.6	Mechanism of substrate translocation by ABC importers	34
1.6.1	Ligand capture	34
1.6.2	Internalization of ligand	35

1.6.3	Mechanism of substrate translocation by ABC exporters	36
1.7	Dilemma of directionality	37
1.7.1	Retrograde transport	37
1.7.2	Anterograde transport	39
1.7.3	Bidirectional/ambiguous transport	40
1.7.4	Auxiliary function	41
1.8	Significance of the study	41
1.9	Objectives	42
	Chapter 2- Materials and methods	43
	The general outline of Chapter 2	45
2.1	Materials	45
2.1.1	Reagents	45
2.1.2	Phospholipid	46
2.2	Methods	46
2.2.1	*In silico* approach	46
2.2.1.1	Sequence-based analysis	46
2.2.1.1.1	Retrieval of sequences	46
2.2.1.1.2	Analysis of domain architecture and amino acid composition	46
2.2.1.1.3	Sequence alignment	47
2.2.1.1.4	Phylogenetic tree	47
2.2.1.1.5	Interactome analysis	47
2.2.1.2	Structure-based analysis	48
2.2.1.2.1	Data retrieval and structural comparison	48
2.2.1.2.2	Analysis of binding pocket and charge distribution	48
2.2.1.2.3	Protein-protein docking	48
2.2.2	Experimental approach	49
2.2.2.1	Construction of recombinants protein overexpression	49
2.2.2.2	Preparation of competent cells for gene cloning and protein overexpression	49
2.2.2.3	Gene amplification	50
2.2.2.4.	Double digestion of the gene of interest and vector	50
2.2.2.5	Ligation of the gene of interest to vector and clone confirmation	51
2.2.2.6	Protein overexpression	51

2.2.2.7	Solubility of proteins	51
2.2.2.8	Affinity chromatography for protein purification	52
2.2.2.9	Protein characterization	52
2.2.2.9.1	Circular dichroism (CD) spectroscopy	52
2.2.2.9.2	Thin layer chromatography (TLC)	53
2.2.2.9.3	Mass spectrometric analysis	53
2.2.2.10	Protein crystallization	54
2.2.2.10.1	Crystallization of the selected proteins	54
2.2.2.10.2	Data collection	55
2.2.2.10.3	Data processing	56
2.2.2.10.4	Structure determination	56
2.2.2.10.5	Model building and structure refinement	56
2.2.2.10.6	Cross-validation	57
2.2.2.10.7	Structure validation	57
2.2.2.10.8	PROCHECK and MolProbity	58
2.2.2.10.9	Structure analysis	58
2.2.2.11	Molecular docking and dynamics simulation	59
2.2.2.11.1	Molecular docking	59
2.2.2.11.2	Molecular dynamics (MD) simulation	60
	Chapter 3- *In silico* characterization of Mla proteins from *Escherichia coli*	63
	Abstract	65
	Graphical abstract	67
3.1	Introduction	67
3.2	Material and methods	69
3.2.1	Data collection	69
3.2.2	Sequence analysis	70
3.2.3	Structure analysis	71
3.2.3.1	Protein-ligand docking	71
3.2.3.2	Protein-protein docking	72
3.3	Results	72
3.3.1	The MlaD domains are organized in different domain architecture types	72
3.3.2	The MlaD domains share a similar amino acid profile across different architectures	74

3.3.3	The N-terminal region of the MlaD domain is conserved	77
3.3.4	The *Ec*MlA_MlaD1 domain is evolutionarily closer to the MlaD domains of *M. tuberculosis* and *A. thaliana* than those of *E. coli*	79
3.3.5	The MlaD domain adopts a conserved structure with varying loop regions	81
3.3.6	The MlaD domain forms a hydrophobic pore	82
3.3.7	The MlaD domains are not involved in determining the oligomeric state of the proteins	86
3.3.8	Nuclear transport factor 2 (NTF2) dependent Mla system might work in coordination with the Lpt system	88
3.3.9	*Ec*MlaE and its orthologues have evolved separately from other TMDs	90
3.3.10	*Ec*MlaE and its orthologue possess an N-terminal interfacial helix	91
3.3.11	*Ec*MlaE and its orthologue show variation in coupling helix	92
3.3.12	*Ec*MlaF and its orthologues have evolved separately from other NBDs	92
3.3.13	*Ec*MlaF and its orthologues show variation in signature motifs	94
3.3.14	The association of *Ec*MlaB gives rise to four additional interfaces	94
3.3.15	*Ec*MlaB possesses a DSSG motif	95
3.3.16	*Ec*MlaA possesses a doughnut-shaped structure	96
3.3.17	A major portion of *Ec*MlaA remains embedded in the OM and gives rise to a dynamic central channel	99
3.3.18	The central channel of *Ec*MlaA can bind PL, while the CTH region might determine the orientation of the protein	100
3.4	Discussion	101
3.4.1	The MlaD domain possesses a diverse domain architecture but a conserved amino acid profile	102
3.4.2	The MlaD domain possesses a conserved N-terminal region with a well-preserved motif	102
3.4.3	*Ec*MlA_MlaD1 is not evolutionarily closer to the Mla domains from *E. coli*	103

3.4.4	MlaD domain has a conserved structure with variable loops	104
3.4.5	MlaD domain imparts flexibility to the protein	104
3.4.6	MlaD domain-containing systems have diverse interactomes	105
3.4.7	*Ec*MlaE has a unique evolutionary lineage	106
3.4.8	*Ec*MlaF-*Ec*MlaB gives rise to a unique functional association and macromolecular arrangement	107
3.4.9	The structural features of *Ec*MlaA might aid in the ATP-independent transfer of PLs	107
3.5	Conclusion	107
	Chapter 4- Structural and mechanistic characterization of MlaC protein from *Escherichia coli*	111
	Abstract	113
	Graphical abstract	114
4.1	Introduction	114
4.2	Materials and methods	117
4.2.1	Cloning, overexpression, and purification	117
4.2.2	Circular dichroism (CD) experiments	119
4.2.3	Lipid extraction and thin-layer chromatography	119
4.2.4	Crystallization, data collection, and structure determination	120
4.2.5	Mass spectrometry analysis	121
4.2.6	Molecular docking	121
4.2.7	Sequence and structure analyses	122
4.3	Results	124
4.3.1	The protein is folded properly and co-purified with phospholipid	124
4.3.2	The overall structure of *Ec*MlaC	125
4.3.3	*Ec*MlaC possesses a unique segmented domain arrangement	128
4.3.4	*Ec*MlaC shows variation in ligand interaction	129
4.3.5	Ligand binding and orientation dictate the volumetric expansion of the binding pocket of MlaC	130
4.3.6	The role of the D1R1 subdomain is ambiguous	132
4.3.7	The D2R1 subdomain modulates the binding-pocket volume	134

4.3.8	The D1R2 subdomain exhibits a change in β-sheet curvature upon ligand binding	136
4.3.9	The D2R2 subdomain plays a significant role in accommodating two phospholipid molecules	139
4.3.10	The phospholipid head groups show variations in the polar interactions	142
4.3.11	MlaC demonstrates polyspecificity towards phospholipids	143
4.3.12	MlaC has a unique ligand-binding mechanism involving ligand re-orientation	144
4.3.13	MlaC constitutes a novel class of non-canonical SBPs	147
4.4	Discussion	148
4.4.1	A unique binding pocket that is extremely plastic	149
4.4.2	Two subdomains mainly execute the opening and closing of the binding pocket	149
4.4.3	A unique mechanism of ligand binding not observed among canonical SBPs	150
4.5	Conclusion	152
	Chapter 5- Structural and mechanistic characterization of MlaD protein from *Escherichia coli*	155
	Abstract	157
	Graphical abstract	158
5.1	Introduction	158
5.2	Materials and methods	160
5.2.1	Cloning, overexpression and protein purification	160
5.2.2	Circular dichroism (CD) experiments	162
5.2.3	Crystallization, data collection and structure determination	162
5.2.4	Sequence and structure analyses	165
5.3	Results	166
5.3.1	The overall structure of *Ec*MlaD	166
5.3.2	MlaD domains have structurally conserved cores	170
5.3.3	MlaD domains have local structural variations	172
5.3.4	MlaD domains give rise to hydrophobic central channels of varying diameters	174
5.3.5	*Ec*MlaD_MlaD1^{P41212} does not form domain-swapped dimers via the β5 strand	175

5.3.6	The *Ec*MlaD_MlaD1^{P41212} does not determine the oligomeric state of the protein	177
5.3.7	*Ec*MlaD undergoes an “asymmetric protomer movement” as well as channel expansion during ligand transfer	178
5.3.8	The “asymmetric protomer movement” mechanism might not be demonstrated by all the MlaD domains-containing proteins	183
5.3.9	The MlaD domain remains associated with diverse translocases	185
5.3.10	The MlaD protein shares structural similarity with EF/AMT-type beta(6)-barrel fold and constitutes a novel SBP cluster	187
5.4	Discussion	190
5.4.1	*Ec*MlaD oligomerizes to give rise to a dynamic central channel suitable for phospholipid transport	191
5.4.2	The MlaD domains have conserved cores but possess local structural differences	191
5.4.3	Phospholipid transport involves two complementary mechanisms	192
5.4.4	*Ec*MlaD constitutes a novel cluster of SBP	194
5.5	Conclusion	195
	Chapter 6- Molecular docking and molecular dynamics simulation studies of *Ec*MlaC protein	197
	Abstract	199
	Graphical abstract	200
6.1	Introduction	200
6.2	Materials and methods	202
6.2.1	Analysis of structural data	202
6.2.2	Molecular docking studies	203
6.2.3	Generation of the molecular systems	204
6.2.4	Molecular dynamics simulations	204
6.2.5	Analysis of simulation data	205
6.3	Results	206
6.3.1	MlaC orthologues can exist in the open unliganded state	206
6.3.2	*Ec*MlaC orthologues can bind different ligands in varying conformations	210

6.3.3	Ligand binding stabilizes the backbone of MlaC	211
6.3.4	The bound ligands demonstrate dynamic movements in MlaC	213
6.3.5	Opening and closing of the phospholipid-binding site is modulated by D2R1 and D1R2 subdomains aided by angular motions	214
6.3.6	The subdomain D2R1 is the most plastic among the other regions	215
6.3.7	Unliganded MlaC systems exhibit more motions	216
6.3.8	Unliganded MlaC systems have larger conformational spaces	218
6.4	Discussion	221
6.4.1	MlaC can exist in different transient states	221
6.4.2	MlaC is multi-specific in nature and can bind to different ligands	222
6.4.3	MlaC is an extremely dynamic conformational landscape	223
6.4.4	MlaC exhibits a series of dynamic movements via the subdomains resulting in a unique ligand-binding mechanism	223
6.5	Conclusion	225
	Summary	227
	Appendix A- Supplementary data to Chapter 3	233
	Appendix B- Supplementary data to Chapter 4	285
	Appendix C- Supplementary data to Chapter 5	321
	Appendix D- Supplementary data to Chapter 6	375
	Bibliography	401
	List of publications, conferences and workshops	417

Abstract

Gram-negative bacteria are generally more resilient than Gram-positive bacteria owing to the presence of a relatively complex cell envelope. The cell envelope of Gram-positive bacteria consists of an inner membrane (IM), a small periplasmic space and a thick peptidoglycan layer. On the other hand, the cell envelope of Gram-negative bacteria comprises an IM, a large periplasmic space that holds a thin peptidoglycan layer and outer membrane (OM). This OM is a salient feature of Gram-negative bacteria which serves as a formidable permeability barrier interrupting the passage of detrimental compounds such as antibiotics, detergents, toxins, etc. The robust barrier function of the OM is attributed to its unique macromolecular composition and arrangement. The inner leaflet of the membrane consists of phospholipids (PLs), while the outer leaflet possesses lipopolysaccharides (LPSs). Such an organization makes the OM asymmetric in nature, which acts as a protective shield against the free entry of antimicrobials. Interestingly, the asymmetric organization of the OM is entropically unfavorable although it is critical for the barrier function. As a consequence, the PLs have a tendency to flip back to the outer leaflet from the inner leaflet in order to attain an entropically favorable state leading to the disruption of the asymmetry and the barrier function. In order to remove these misplaced PLs and restore the OM asymmetry, bacteria utilize various two distinct mechanisms involving OM β-barrel enzymes, viz. PldA and PagP during stress. The enzyme PldA removes the fatty acid chains, whereas PagP transfers the acyl chain to the lipid A moiety of LPS from the unwanted PLs, thereby restoring the asymmetry of the OM. Recently, a third mechanism has been reported, which involves a highly conserved ATP-binding cassette (ABC) transporter, viz. maintenance of the lipid asymmetry (Mla) system. The Mla system comprises six proteins - MlaA (OM lipoprotein), MlaB (cytoplasmic protein), MlaC (periplasmic substrate (solute)-binding protein, SBP), MlaD (IM-associated periplasmic SBP), MlaE (Transmembrane domain, TMD) and MlaF (Nucleotide-binding domain, NBD). Interestingly, the outer membrane protein, OmpC, has been reported to form a complex with MlaA and is critical for the functioning of the Mla system. Together, these proteins give rise to three inter-membrane sub-complexes – OmpC-MlaA (OM), MlaC (periplasm) and MlaFEDB (IM). As per initial reports, the Mla system supposedly follows a unique retrograde movement wherein the misplaced PLs are flipped from the outer leaflet to the inner leaflet of the OM by the OmpC-MlaA complex. The PLs are handed over to MlaC, which ferries them to the MlaFEDB complex for internalization, thereby restoring the ideal PL population in the OM and its asymmetry. Interestingly, studies have also reported the anterograde movement of PL

wherein the Mla system transfers PLs from the IM to the OM. As a consequence, the subject of directionally of PL transport by the Mla system remains a debatable topic. In addition to this, the Mla system does not possess a typical ABC transporter complex. A canonical ABC transporter (exporter and importer) complex consists of two copies, each of NBD and TMD. In the case of the ABC importer, an additional component, SBP, is present that captures ligands and hands them to the TMD for internalization. The SBPs possess a conserved structure with an N-terminal and C-terminal domain (NTD and CTD) linked via a flexible linker region. Surprisingly, in the Mla system, two SBPs – MlaC and MlaD are present. In addition to this, it contains MlaB, which remains associated with MlaF (NBD) and MlaA, which remains associated with the OM. Furthermore, MlaB and MlaD possess the highly conserved Sulphate Transporter and AntiSigma factor antagonist (STAS) and Mammalian cell entry (Mce) domains, respectively. Such a macromolecular organization has not been observed in the case of typical ABC transporters. Owing to these unique features, the components of the Mla system remain poorly characterized.

Hence, this study proposes to characterize the Mla proteins from *Escherichia coli* so as to acquire insights into the machinations of the Mla system. For this purpose, we have utilized *in silico* as well as experimental approaches for the structural and mechanistic characterization of the Mla system. In the IM complex, the MlaD protein (*Ec*MlaD) is considered to be one of the unique components of the Mla system owing to the presence of the Mce domain. Hence, for the *in silico* characterization, *Ec*MlaD is taken as the point of reference. The study reveals that the MlaD domain would possess diverse architectures but has conserved profiles comprising an abundance of glycine and hydrophobic residues and a lack of cysteine residues. Furthermore, the N-terminal region of the domain is well-conserved and possesses a well-preserved glycine residue that constitutes a motif (MlaD domain motif). Surprisingly, the MlaD domain of *Ec*MlaD is evolutionarily similar to the MlaD domains from *Mycobacterium tuberculosis* rather than from *E. coli*. Furthermore, interactome analysis suggests that *Ec*MlaC would serve as a mediator between *Ec*MlaD and *Ec*LptC (transport of LPS), indicating a synergistic relationship between the Mla and Lpt systems.

The *in silico* characterization of MlaE from *E. coli* (*Ec*MlaE) reveals that the protein, along with its orthologs, has evolved separately from other TMDs. This is further supported by the presence of an N-terminal interfacial helix (IF1) and the uncommon EQ loop instead of the EAA loop, as observed in other TMDs. Similar to *Ec*MlaE, the

MlaF from *E. coli* (*Ec*MlaB) demonstrates separate clustering, thus, indicating a distinct evolution from other NBDs. Also, *Ec*MlaF possesses the LSGGM motif instead of LSGGQ and remains functionally associated with MlaB (*Ec*MlaB). Owing to this, the *Ec*MlaF-*Ec*MlaB gives rise to additional interfaces that are critical for the functioning of the Mla system. The study also suggests the presence of DSSG and Walker A-like motif in *Ec*MlaB, indicating that the protein might be able to bind ATP and undergo phosphorylation. The study also shows that MlaA from *E. coli* (*Ec*MlaA) would possess a doughnut-shaped structure and remain embedded in the OM. Such a localization would be aided by stable interaction with OmpC/F. Furthermore, the central channel of *Ec*MlaA is convoluted but extremely flexible, which helps in the passage of PLs. The analysis also suggests that the C-terminal helix might play a critical role in determining the orientation of *Ec*MlaA.

After *in silico* characterization, the proteins *Ec*MlaC and *Ec*MlaD were further studied using experimental approaches. This study reports the crystal structure of *Ec*MlaC in a unique quasi-open with an endogenously bound PL at a resolution of 2.5 Å in the space group *H*3. The tails of the PL remain deeply buried in the hydrophobic core of *Ec*MlaC while the head remains solvent exposed. Interestingly, the bound PL is present in an unusual conformation not before observed in any other available structures of MlaC orthologs. The structural analysis reveals that *Ec*MlaC does not possess an NTD-CTD arrangement. Instead, it has two major domains, viz. nuclear transport factor 2-like (NTF2-like, D1) and phospholipid-binding protein (PBP, D2). Each domain can be further divided into two subdomains arranged in a discontinuous fashion (D1R1-D2R1-D1R2-D2R2). The study also suggests that, unlike typical SBPs that can bind to only one ligand, some MlaC orthologs can bind two PLs simultaneously. Surprisingly, the binding-pocket volume of MlaC increases proportionally with the number of bound PLs, leading to the opening of the binding pocket. This is contrary to the binding mechanism observed in the case of typical SBPs, wherein they make a transition from an open (unliganded) to a closed (liganded) state upon ligand binding. Also, during double ligand binding, the binding pocket of MlaC splits into two sub-sites. Analysis of different liganded states reveals the subdomains contribute significantly to ligand binding, especially D1R2 and D2R1, that modulate the opening/closing of the protein. Based on extensive structural studies, we propose a novel mechanism of ligand binding, viz. “segmented domain movement (SDM)’’, which involves a series of inherent structural changes in order to bind PL. Furthermore, our study establishes that *Ec*MlaC and its

orthologs have atypical features and ancestries which are different from typical SBPs. Hence, we have considered *Ec*MlaC and its orthologs to be non-canonical SBPs. This led us to introduce a new scheme of SBP classification, which consists of two classes – canonical SBPs (possessing NTD-CTD arrangement) and non-canonical SBPs (not possessing NTD-CTD arrangement).

This study also reports the crystal structure of the periplasmic domain of protein *Ec*MlaD protein in three different space groups $P4_12_12$, $I222$ and $P2_122_1$, at a resolution range of 2.3-3.2 Å. Structural analysis reveals that *Ec*MlaD protomers consist of seven β-strands adopting a β-barrel fold (MlaD domain) and a C-terminal α-helical domain (HD). The MlaD domain has a conserved core but possesses local structural variations. The *Ec*MlaD protomer forms a homo-hexameric ring with a central channel and possesses six-fold symmetry. This central channel serves as the main translocation pathway for the passage of PLs during the process of ligand transport. Interestingly, during the process of PL export, three protomers of *Ec*MlaD hexamer exhibit an upward movement. This is accompanied by a change in the geometry of the central channel as well as the outward movement of Pore LooPs (PLPs) that line the channel. These conformational changes provide the necessary force for orienting the PLs in the channel meant to be exported. Eventually, *Ec*MlaC would dock and, following the SDM mechanism, would extract the PLs for the eventual transfer to the OM. By taking into consideration these observations, we have proposed a novel mechanism of PL transfer, viz. the “asymmetric protomer movement (APM)’’ mechanism. This has, in turn, helped us put forward a working model for the ATP-independent export process of PL from *Ec*MlaD to *Ec*MlaC. Additionally, the study also highlights that *Ec*MlaD is structurally homologous to EF/AMT-type beta(6)-barrel and has a unique ancestry, unlike canonical SBPs. This has led us to further update the SBP classification scheme.

Additionally, in order to estimate the affinities for various ligands and identify the protein motions at the global and local scales, molecular docking and molecular dynamics (MD) simulations were performed on *Ec*MlaC and its orthologues. The docking experiments performed using 104 ligands clearly suggest that *Ec*MlaC and its orthologues can bind different ligands in varying conformations. This indicates that *Ec*MlaC orthologues are multi-specific in nature. Also, the ligands within the binding cavity can change orientation depending on the availability of space. On the other hand, the simulation studies performed for 1000 ns on different systems points that

*Ec*MlaC orthologues are extremely dynamic and the bound PL(s) can change confirmations. The findings suggest that the unliganded *Ec*MlaC orthologues exhibit higher motions as compared to their unliganded states. This is because ligand binding reduces the plasticity of the protein. Also, we have been able to identify a twisting motion of α7 and α8 that is crucial for protein dynamics. These observations validate our previously reported SDM mechanism. Based on this, we have to put forward a working model of PL transport by MlaC that highlights the molecular details.

In summary, the findings of this study firmly establish that not all SBPs would possess the typical NTD-CTD arrangement. As a consequence, they would not necessarily follow the most commonly associated mechanism of ligand binding of SBPs, viz. “Venus Fly-trap’’ and its derivatives (‘’asymmetric domain movement’’, ‘’one domain movement’’ and “subdomain movement’’). The in-depth structural characterization of *Ec*MlaC and *Ec*MlaD led to the identification of two new mechanisms (SDM and APM) of ligand binding and transport for SBPs. This has helped us present a detailed molecular picture of the process of PL export. The analysis also highlights the unique features of the Mla proteins, which altogether suggests the versatility of the Mla system as compared to other ABC transporters.

Abbreviations and notations

Abbreviations

AAA	ATPase-associated with diverse cellular activities
ABC	ATP-binding cassette
ANM	Anisotropic network model
APM	Asymmetric protomer movement
APO	Average percentage of occurrence
β-ME	β-mercaptoethanol
CD	Circular dichroism
CL	Cardiolipin
CTD	C-terminal domain
CTH	C-terminal helix
CV	Column-volume
DCCM	Dynamic cross-correlation matrix
IM	Inner membrane
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LB	Luria-Bertani
LPS	Lipopolysaccharide
LTP	Lipid transfer protein
Mam	Mce-associated membrane
Mce	Mammalian cell entry
MD	Molecular dynamics
Mla	Maintenance of lipid asymmetry
MSA	Multiple sequence alignment
NBD	Nucleotide-binding domain
NCBI	National Center for Biotechnology Information
NS	Not significant
NTD	N-terminal domain
NTF2	Nuclear transport factor 2
OM	Outer membrane
Omam	Orphaned Mce-associated membrane
OMP	Outer membrane porins
PBP	Phospholipid-binding protein
PC	Principal component
PCA	Principal component analysis

PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylene glycol
PEF	Phosphatidylethanolamine
PL	Phospholipid
PLP	Pore loop
PMSF	Phenylmethylsulfonyl fluoride
PO	Percentage of occurrence
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
SBP	Substrate (or solute)-binding protein
SDM	Segmented domain movement
STAS	Sulphate Transporter and AntiSigma factor antagonist
TLC	Thin-layer chromatography
TMD	Transmembrane domain

Notations

~	Approximately
%	Percent
Å	Angstrom
a.u.	Arbitrary unit
bp	Base pair
°C	Degree Celsius
g	Relative Centrifugal Force (RCF)
kb	kilobase
kcal	Kilocalories
K_d	Dissociation constant
Da	Dalton
kDa	kilodalton
L	Litre
M	Molar
mdeg	Change in ellipticity
mM	Millimolar
ml	Millilitre
s	Second
T	Temperature
Tm	Melting temperature
λ	Wavelength
µg	Microgram
µl	Microlitre
µM	Micromolar

List of figures

List of figures

Figure No.	Figure title	Page No.
1.1	Overview of the bacterial cell envelope	4
1.2	Organization of the bacterial OM	6
1.3	Overview of the transporters involved in the transport of OM components	9
1.4	Kennedy pathway for biosynthesis of PLs in *E. coli*	13
1.5	Enzymes involved in the maintenance of OM asymmetry	15
1.6	Overview of the transporters involved in the transport of PLs	19
1.7	General architecture of ABC transporters	20
1.8	Overview of ABC importers and exporters	22
1.9	Architecture of NBD	25
1.10	Overview of the SBP	27
1.11	Structure-based classification of SBPs	29
1.12	Organization of *mce* operon and Mce system in *M. tuberculosis*	33
1.13	Overview of the different ligand binding mechanisms of SBP	34
1.14	Schematic representation of substrate translocation by ABC importer	35
1.15	Schematic representation of substrate translocation by ABC exporter	37
1.16	Proposed mechanism of PL transport by Mla system	39
2.1	A general work plan of the methodology used to accomplish the objectives	45
2.2	Schematic representation for molecular cloning	49
2.3	Schematic representation for the phase diagram of protein crystallization	55
3.1	Schematic representation of the MlaD domain architectures.	73
3.2	The distribution of amino acids in the MlaD domains of the proteins from *E. coli* and *M. tuberculosis*	76
3.3	Conserved features of different MlaD domains	78

3.4	Phylogenetic trees displaying the evolutionary relationship among the MlaD domains and their associated TMDs of *E. coli*, *M. tuberculosis* and *A. thaliana*	80
3.5	Overall structural comparison of *Ec*MlA_MlaD1 with various MlaD domains.	82
3.6	Overview of the hexameric assemblies of the MlaD domains and their respective pores in *E. coli*	84
3.7	An overview of the organization of the MlaD domains in the proteins MlaD, PqiB and YebT from *E. coli*	86
3.8	The residues present at the dimeric interface of the protein MlaD	87
3.9	Genetic organization of the genes encoding proteins containing the MlaD domain and the functionally associated components in *E. coli*, *M. tuberculosis* and *A. thaliana*.	89
3.10	Sequence and structure analyses of *Ec*MlaE	91
3.11	Sequence and structure analysis of *Ec*MlaF	93
3.12	Sequence analysis of STAS domain-containing proteins	96
3.13	Structure analysis of *Ec*MlaA	98
3.14	Analysis of protein-ligand and protein-protein docking	100
4.1	Cloning, overexpression, solubility and purification of *Ec*MlaC	118
4.2	Crystallization of *Ec*MlaC	121
4.3	Biophysical characterization of *Ec*MlaC	125
4.4	The overall structure of *Ec*MlaC	127
4.5	Comparison of the binding-pocket volumes	131
4.6	Structural details of the ligand interaction with the D1R1 subdomain.	133
4.7	Structural details of the ligand interaction with the D2R1 subdomain	135
4.8	Structural details of the ligand interaction with the D2R1 subdomain	137
4.9	Structural details of the ligand interaction with the D2R2 subdomain	140
4.10	Phospholipid head group interactions with *Ec*MlaC_H3_holo1 orthologs.	142

4.11	Molecular docking of *Ec*MlaC_H3_holo1 with phospholipids	144
4.12	Overview of phospholipid transport by Mla system	145
4.13	Evolutionary analysis of MlaC	148
4.14	Schematic representation of the major models proposed for ligand-binding mechanisms	152
5.1	Cloning, overexpression, solubility and purification of *Ec*MlaD	161
5.2	Crystallization of *Ec*MlaD	164
5.3	The overall structure of *Ec*MlaD	168
5.4	Structural comparison of the MlaD domains from *E. coli*	171
5.5	Comparison of topologies of the MlaD domains from *E. coli*	174
5.6	Comparison of the diameters of the MlaD-domain hexamers and their central channels from *E. coli*	175
5.7	Overview of the structural features of *Mt*Mce4A	176
5.8	The residues present at the dimeric interface of *Ec*MlaD	178
5.9	Mechanism of PL transport	180
5.10	Evolutionary relationship of the MlaD domain-containing proteins	184
5.11	Structural analysis of *Ec*MlaE	186
5.12	Ancestry analysis of the MlaD domain	188
5.13	Overview of the two major types of ABC transporter mechanisms based on SBP types	193
6.1	Overview of the workflow of molecular dynamics simulation studies	206
6.2	Structural analysis of *Ec*MlaC and its orthologues	208
6.3	Molecular docking analyses	211
6.4	RMSD analysis of MlaC systems	212
6.5	RMSD analysis of PLs bound to MlaC systems	214
6.6	Analysis of the change in distance between residues around the binding pocket	215
6.7	RMSF analysis of MlaC systems	216
6.8	Correlation maps of *Ec*MlaC systems	218
6.9	Free energy landscape (FEL) analysis and porcupine plots of *Ec*MlaC systems	219
6.10	Proposed working model for PL transport by MlaC	225

List of tables

List of tables

Table No.	Table title	Page No.
1.1	Classification of ABC transporter based on TMD folds in prokaryotes	23
1.2	Classification of NBDs	26
1.3	An overview of the SBP classification	30

Chapter 1

Introduction

भारतीय प्रौद्योगिकी संस्थान गुवाहाटी
Indian Institute of Technology Guwahati

1. Introduction

1.1. Bacterial cell envelope

The bacterial cell envelope is an intricate multi-layered organelle that serves to protect the organisms from the unpredictable and hostile environment while allowing the selective passage of nutrients (Silhavy et al., 2010). In addition to this, it gives bacteria their shape, provides the means by which they generate usable forms of energy for growth and division, enhances pathogenesis, aids in the horizontal transfer of plasmids and other mobile elements as well as establishes the interface with their surroundings (Kleanthous and Armitage, 2015). Thus, the envelope imparts specialized functions that are crucial for the survival, sustenance and physiology of bacteria (Viljoen et al., 2020). The bacterial cell envelope can be distinguished into two major types based on their ability to retain Gram stain, Gram negative and Gram positive. The former is not capable of retaining the stain, while the latter is able to do so. This difference in property is due to structural and compositional variations between the two envelopes. The envelope of Gram-negative bacteria is an assimilation of three layers which consist of an inner membrane (IM), a large periplasmic space that accommodates a thin peptidoglycan layer and an outer membrane (OM) (Figure 1.1A). On the other hand, the envelope of Gram-positive bacteria is relatively simpler in composition as it consists of an IM membrane, a reduced periplasm but a considerably thicker peptidoglycan layer (Figure 1.1B) (Henderson et al., 2016). Hence, OM is one of the defining features of Gram-negative bacteria.

The OM serves as a robust permeability barrier and unlike the IM, it has a unique macromolecular composition and arrangement (May and Grabowicz, 2018). This is because the inner leaflet of the OM consists of PL, while the outer leaflet possesses lipopolysaccharide (LPS) (Kamio and Nikaido, 1976; Nikaido, 2003; Henderson et al., 2016). Such an organization makes the OM asymmetric in nature, which acts as a protective shield against the uninterrupted passage of detrimental compounds such as antibiotics, detergents, toxins, etc. The LPS molecules densely pack the outer leaflet with acyl chains making it extremely hydrophobic. Also, the saccharide and the lipid moieties carry a negative charge that facilitates intermolecular bridging by binding divalent cations and imparts a sealing effect to the OM (Nikaido, 2003). Such a barrier

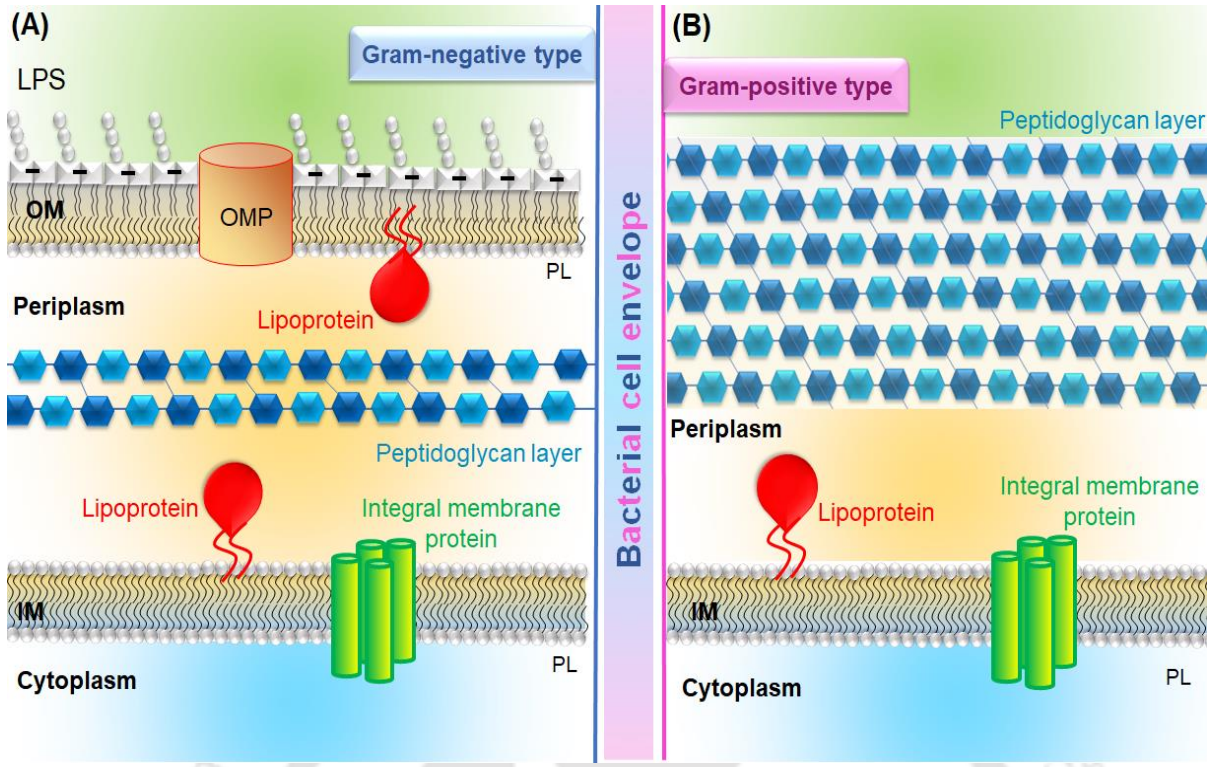


Figure 1.1. Overview of the bacterial cell envelope. *(A) Gram-negative type. It consists of an outer membrane (OM), a large periplasmic space and an inner membrane (IM). The periplasmic space contains a thin peptidoglycan layer. (B) Gram-positive type. It consists of a thick peptidoglycan layer, a reduced periplasmic space and an IM. The OM is asymmetric in nature as it contains lipopolysaccharide (LPS) and phospholipid (PL) in the outer and inner leaflets, respectively. Both IM and OM contain lipoproteins. However, outer membrane proteins (OMPs) possessing β-barrel architecture is confined to the OM and integral membrane proteins with transmembrane helices are present in the IM.*

function originating from the strict asymmetry has been attributed to the enhanced resilience of Gram-negative bacteria towards available anti-bacterial interventions. In addition to PL and LPS molecules, the OM also possesses lipoproteins and β-barrel proteins. The lipoproteins remain associated with the inner leaflet of the OM through a lipid moiety, while the protein component protrudes into the periplasm and does not span the OM. On the other hand, β-barrel proteins are integral, transmembrane components that contain sheets wrapped into cylinders. They are referred to as outer membrane proteins (OMPs) and some of these proteins are involved in the transport of small molecules via passive diffusion (Silhavy et al., 2010).

In contrast to the OM, the IM possesses a symmetric PL bilayer organization and is devoid of LPS molecules. It performs the functions of selective nutrient transport, protein translocation, lipid biosynthesis and oxidative phosphorylation. Furthermore, unlike OM, the integral membrane proteins of the IM have transmembrane (TM) helices that span the membrane (Duong et al., 1997). The IM also accommodates various membrane-associated proteins that remain embedded via helix while the core component protrudes into the periplasmic space, which separates the IM and OM while serving as a multitasking aqueous compartment. It provides a reducing ambience, thereby facilitating protein oxidation and folding (Miller and Salama, 2018). It contains various structural elements such as solute (substrate)-binding proteins (SBPs) for ferrying nutrients and hydrolases that serve a lysosomal function. Additionally, the periplasm also accommodates a thin peptidoglycan layer which is composed of covalently cross-linked linear glycan strands linked via short peptides. Each glycan strand is a polymer of disaccharides containing an N-acetylmuramic acid (MurNAc) and N-acetylglucosamine (GlcNAc) residue that is linked by β-1,4 glycosidic bonds (Garde et al., 2021). Among its various functions, this layer provides cell shape and integrity by withstanding the turgor, serves as a molecular sieve allowing the passage of selective molecules and acts as a scaffold for anchoring cell envelope components (Vázquez-Laslop et al., 2001; Vollmer et al., 2008; Garde et al., 2021). Thus, the three layers of the Gram-negative cell envelope contribute to selective permeability. As most antibiotics are targeted at intracellular processes, the penetration of all the layers for effective outcomes is crucial. Due to this, the components of the cell envelope have garnered significant interest for their potential as novel antibiotic targets. Among them, the OM has been particularly recognized as a promising candidate as it is responsible for bacterial sensitivity and susceptibility to antibiotics owing to its sophisticated macromolecular assembly (Delcour, 2009).

1.2. Molecular assembly of the OM

The OM of Gram-negative bacteria is constituted by four major components, PL, LPS, OMP and lipoprotein (Figure 1.2A). However, the differential presence of PL and LPS molecules in the outer and inner leaflets, respectively, makes the OM asymmetric (Nikaido and Vaara, 1985; Nikaido, 2003). The arrangement of the PL molecules is such that the polar head groups face the periplasm facilitating interaction with the aqueous environment while the hydrophobic tails are protected within the core of the OM. In *E. coli*, the OM is constituted by three major PLs, viz. phosphatidylethanolamine

(PE), phosphatidylglycerol (PG) and cardiolipin (CL) (Raetz and Dowhan, 1990; Sohlenkamp and Geiger, 2016; Hilton et al., 2021). On the other hand, the LPS molecule is relatively larger than the PL and is exclusive to the OM of Gram-negative bacteria. It is a large glycolipid molecule that consists of three distinct domains, lipid A, the core oligosaccharide and the O-antigen (Martorana et al., 2014; Bertani and Ruiz, 2018). The first domain, lipid A is an acylated β-1‘-6-linked glucosamine disaccharide that constitutes the outer leaflet of the OM. The glucosamine residues are phosphorylated at the 1 and 4’ positions and acylated at the 2, 3, 2’ and 3’ positions. The distal glucosamine also contains two extra secondary acyl chains, making mature lipid A hexa-acylated. Thus, compared to a typical PL, lipid A contains four additional acyl chains. This makes the asymmetric OM more hydrophobic and less fluid compared to a typical PL bilayer. The second domain, the core oligosaccharide, is branched containing sugars units along with 3-deoxy-D-manno-oct-2-ulosonic acid (Kdo) residue linked to glucosamines of lipid A. The core region is heterogeneous because of the presence of additional substituents. The third domain, the O-antigen, is an extended immunogenic oligosaccharide consisting of 1-40 repeating units (Delcour, 2009; Bertani and Ruiz, 2018). The differential distribution of PL and LPS in the OM is crucial for the barrier function of the OM (Figure 1. 2B-1.2C).

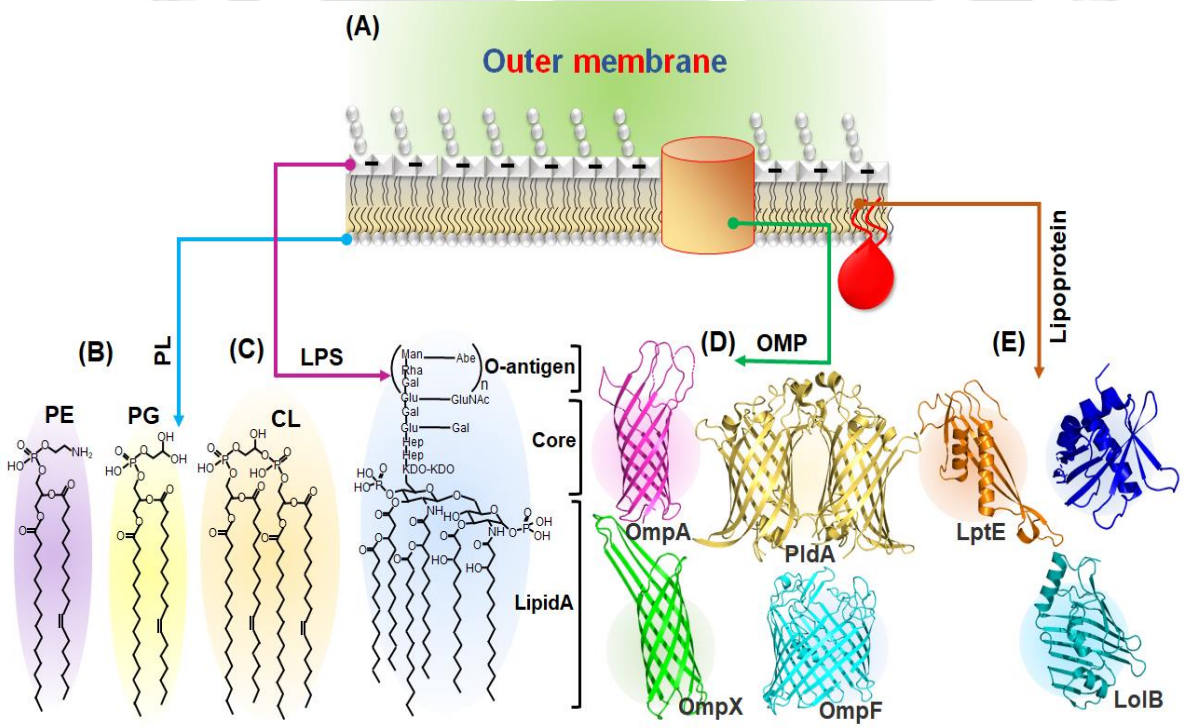


Figure 1.2. Organization of the bacterial OM. *(A) Architecture of the bacterial OM. (B) Structures of different PLs present in the OM: phosphatidylethanolamine (PE), phosphatidylglycerol (PG) and cardiolipin (CL). (C) Structure of LPS molecule. (D) Structures of OMPs: OmpA (Magenta; PDB id: 1BXW), PldA (Yelloworange; PDB id: 1QD6), OmpX (Green; PDB id: 1QJ8) and OmpF (Cyan; PDB id: 2OMF). (E) Structures of lipoproteins: LptE (Orange; PDB id: 4NHR), BamC (Blue; PDB id: 3SNS) and LolB (Teal; PDB id: 1IWM).*

In addition to this, two different types of proteins, viz. OMPs and lipoproteins are present in the OM. The first type, i.e. OMPs, are integral membrane proteins that adopt a signature β-barrel architecture bearing short and long loops on the periplasmic and extracellular sides, respectively (Figure 1.2D). Such an architecture imparts stability to the OMP, helping them withstand hostile and unpredictable environments (Rollauer et al., 2015). They belong to six different protein families, (i) Small β-barrel membrane anchors (OmpA, establishing linkage between OM and peptidoglycan), (ii) Small β-barrel membrane anchors (OmpX, neutralizing host defense mechanism), (iii) Membrane-integral enzymes (PldA, hydrolysis of PLs), (iv) General (non-specific) porins (OmpF, diffusion pore for ions and small molecules), (v) Substrate-specific porins (LamB, Maltose and maltodextrin uptake) and (vi) TonB-dependent receptors (FhuA, uptake of iron-siderophore complexes) (Koebnik et al., 2000). In addition to this, two other OMPs, LptD (insertion of LPS in the outer leaflet) and BAM complex (insertion of β-barrel proteins into the OM), are critical for bacterial physiology (Rollauer et al., 2015).

The second type, i.e. lipoproteins, are OM-associated proteins that remain attached to the membrane by means of N-terminal lipid attachments (Figure 1.2E). As per previous reports, in *E. coli,* around 90 different lipoproteins are synthesized, most of which are localized in the OM (Tokuda and Matsuyama, 2004; Silhavy et al., 2010; Grabowicz, 2019). They are reported to play significant roles such as the maintenance of cell shape, biogenesis of the OM, transport of molecules, signal transduction and cell motility (Miyadai et al., 2004; Narita and Tokuda, 2017). Among the OM lipoproteins, LptE (assembly of LPS), BamC (assembly of OMPs), LolB (assembly of lipoproteins) and Lpp (stabilization of the cell envelope via covalent bonding between the C-terminal lysine residue and peptidoglycan) aid in safeguarding the integrity of bacterial cell

envelope (Narita and Tokuda, 2017). All these four components of the OM are synthesized in the cytoplasm and have respective transporter systems that are critical for OM biogenesis.

1.3. Transport systems involved in the biogenesis of the OM

1.3.1. Lpt transporter

Synthesis of LPS originates in the cytoplasmic leaflet of the IM. Lipid A and core oligosaccharide are synthesized separately in the cytoplasm and are eventually ligated together to form rough LPS (Henderson et al., 2016). Further maturation of rough LPS occurs on the periplasmic leaflet of the IM. The rough LPS is flipped across the IM with the aid of ATP-Binding Cassette (ABC) transporter MsbA in an ATP-dependent manner. It is proposed that the nucleotide-binding domains (NBDs) come together upon binding to ATP, following lipid entry into the cavity, which could force the lipid to move to the periplasmic side of the IM (Figure 1.3A). O-antigen is then ligated to the lipid A-Core oligosaccharide complex at the periplasmic face of the IM to form mature LPS (Bertani and Ruiz, 2018; Giacometti et al., 2022).

The mature LPS molecule is transported from the periplasmic face of the IM to the OM by the trans-envelope Lpt system. The system can be divided into three sub-complexes, IM ABC transporter complex LptBFGC, periplasmic spanning LptA and OM complex LptDE. The IM complex consists of cytoplasmic ATPase (LptB), two TMDs (LptGF) and an IM-anchored bridge subunit (LptC). The periplasmic component (LptA) serves as a bridge subunit between the IM and OM. The OM complex consists of a β-barrel protein (LptD) and a lipoprotein protruding into the periplasm (LptE) (Figure 1.3A). The IM complex collects the LPS molecules laterally from the periplasmic face of IM and transfers them to the LptC and eventually to LptA. The energy for the lateral extraction of LPS is derived from ATP hydrolysis. Eventually, the OM complex LptD/E transports LPS across the OM and inserts it appropriately into the outer leaflet of the OM (Chng et al., 2010; Dong et al., 2017; Giacometti et al., 2022).

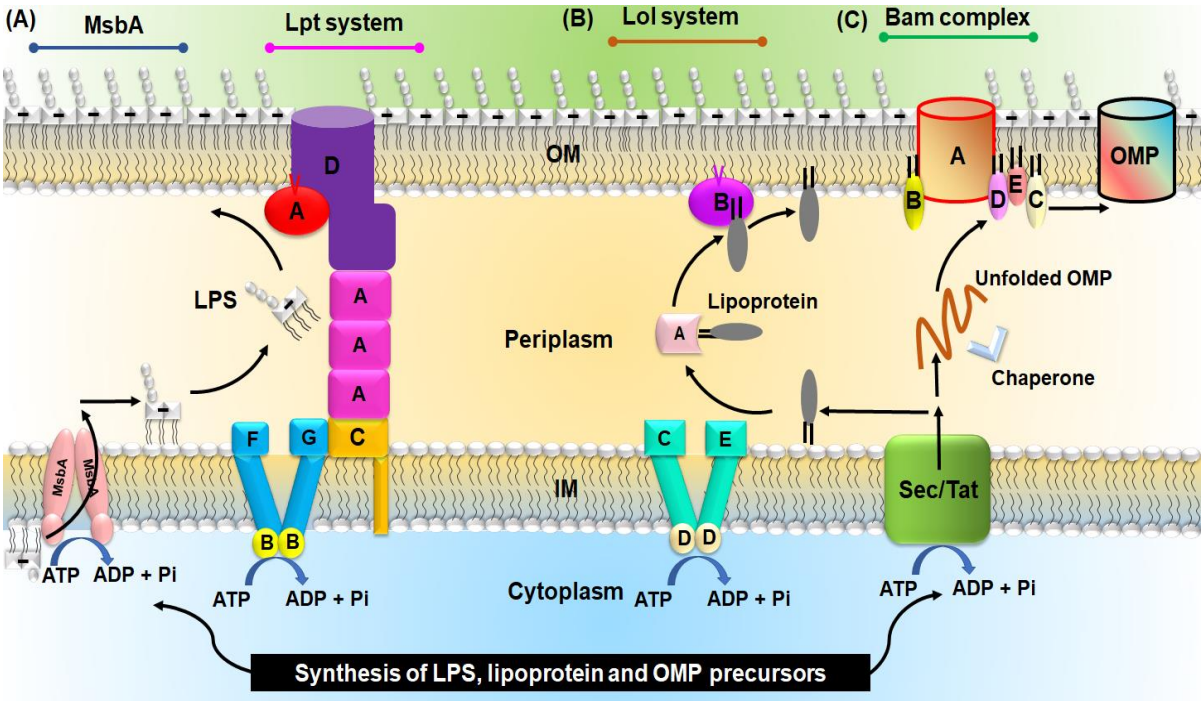


Figure 1.3. Overview of the transporters involved in the transport of OM components. *(A) MsbA and Lpt system (Transport of LPS). Synthesis of rough LPS (LipidA-core) takes place in the cytoplasmic side and is flipped to the periplasmic side by MsbA that hydrolysis ATP molecule. The O-antigen component is ligated to rough LPS, forming the mature LPS. The mature LPS molecule is transported to the OM by the Lpt system with the help of energy obtained from ATP hydrolysis. The system consists of seven proteins and gives rise to three sub-complexes, LptBCFG (IM), LptA (periplasmic) and LptAD (OM). (B) Lol system (Transport of lipoprotein). The lipoproteins are synthesized with N-terminal signal peptide and are translocated to periplasm either by Sec (in the unfolded state using ATP as the source of energy) or Tat pathway (in the folded state without ATP hydrolysis). After translocation, the lipoproteins undergo maturation while remaining anchored to the IM. Once mature, they are transported to the OM by the Lol system with the help of energy obtained from ATP hydrolysis. The Lol system consists of five proteins giving rise to three sub-complexes, LolCDE (IM), LolA (periplasmic) and LolB (OM). (C) Bam complex (sorting of OMPs). The OMPs are synthesized in the cytoplasm and are translocated to the periplasm via the Sec pathway in the unfolded state. To avoid aggregation, the unfolded OMPs are aided by chaperones and are delivered to the Bam complex for sorting. The central component of the complex is the OMP BamA, which has a periplasmic domain. The other four components are BamB-E, which are all*

lipoproteins. The unfolded OMP, on reaching the BAM complex, supposedly binds BamA and brings about the folding of the OMP with the aid of other components. Once the folding is complete, the OMP is incorporated into the OM.

1.3.2. Lol transporter

Lipoproteins have both lipids as well as protein components and are synthesized with signal peptides. At first, the bacterial OM lipoproteins are synthesized as precursors molecules with N-signal peptides. The signal peptides contain conserved motifs, viz. lipobox, that serve as a molecular determinant of lipid modification. The motif contains a preserved cysteine residue that acts as the site of lipidation. The translocation to the periplasm takes place via Sec (secretory) pathway or Tat (twin-arginine translocation) pathway based on the presence of signal peptide (El Rayes et al., 2021). The Tat signal peptides are characteristically longer, have low hydrophobicity and contain an invariable twin-arginine motif (S/T-RR-FLK) in their N-terminal region. Also, Sec pathway translocates proteins in their unfolded states in an ATP-dependent manner, while it is the opposite in the case of the Tat pathway (Adhikari et al., 2017). After the transportation to periplasmic space is complete, the lipoproteins remain anchored to the IM via their signal peptide. They undergo several steps of maturation, including lipidation for anchoring to the OM. After that, the lipoprotein molecule is handed over to the Lol pathway (El Rayes et al., 2021).

The Lol system is a highly conserved transporter that consists of three sub-complexes, LolCDE (IM), LolA (periplasmic) and LolB (OM-associated). The IM complex consists of cytoplasmic ATPase (LolD) and two TMDs (LolCE). The periplasmic component (LolA) serves as a chaperone for guiding the lipoproteins to the OM. The OM consists of lipoprotein (LolB) responsible for the insertion of lipoproteins (Figure 1.3B). LolCDE binds to an OM-targeted lipoprotein, which increases the affinity of LolD for ATP. The ATP binding to LolD results in the weakening of the hydrophobic interaction between the lipoprotein and LolCDE. Finally, ATP hydrolysis causes the transfer of lipoproteins to LolA and opens the hydrophobic cavity of LolA (Ito et al., 2006; Taniguchi and Tokuda, 2008). Lipoproteins are transferred from LolA to LolB mouth-to-mouth manner and are then incorporated into the outer membrane (Okuda and Tokuda, 2011). Lipoprotein sorting to the IM or OM, in general, is based on the presence/absence of

an aspartate residue at the +2 position with respect to cysteine of lipobox. The presence of the aspartate residue at the +2 position causes localization of lipoprotein to the IM, whereas substitution of this position leads to OM localization. The method by which this IM retention signal from the aspartate residue works is that the ATP-dependent IM translocator, LolCDE fails to interact with lipoproteins, containing the aspartate residue (El Rayes et al., 2021).

1.3.3. Bam complex

The OMPs are synthesized in the cytoplasm and the Sec pathway eventually transports them to the periplasm. Following the Sec-dependent transport, the proteins to be incorporated in the OM head towards the BAM complex with the help of periplasmic chaperones. Proteins that have the tendency to form β-sheets are prone to aggregation. Hence, in order to check such a development, the unfolded proteins transverse the periplasmic space with the aid of chaperones and have their signal peptide cleaved by signal peptidase. The periplasmic chaperones, SurA, Skp, and DegP, have been shown to transport the bulk mass of OMPs to the OM (Okuda and Tokuda, 2011; Narita and Tokuda, 2017).

The unfolded OMPs, upon reaching the OM interact with the Bam complex. The Bam complex consists of one β-barrel protein (BamA) and four OM-associated lipoproteins (BamBCDE) (Figure 1.3C). Two of the components of the complex, BamA and BamD, are essential for protein sorting. BamCE is crucial for the stability of the complex. The central component of the complex, BamA, is conserved in Gram-negative bacteria. The soluble domain of all of the proteins in this superfamily contains one or more polypeptide transport-associated (POTRA) domains. In bacteria, these POTRA domains are found in the periplasm and receive the translocated OMPs in the unfolded states (Kim et al., 2012). By binding segments of the unfolded OMP, the POTRA domains of BamA initiate the formation of folding and insertion. The process is further aided by the remaining components of the Bam complex leading to the eventual insertion into the OM (Konovalova et al., 2017). Thus, to sum up, the molecular details of the transport of LPS, lipoproteins and OMP are well studied.

1.3.4. The process of PL transport to the OM and maintenance of PL is poorly understood

Unlike the other three OM components, the subject of PL transport and their eventual insertion in the OM is poorly understood. This is because the subject incorporates multiple aspects, some of which still remain unexplained. These include PL synthesis, sorting, transport machinery and the molecular mechanism associated with the maintenance of this asymmetry.

1.3.4.1. PL biosynthesis pathway

The Kennedy pathway of PL biosynthesis in *E. coli* is considered an ideal model for all Gram-negative bacteria. The enzymes involved are either intrinsic membrane proteins or membrane-associated. The precursor molecule, phosphatidic acid, is *de novo* synthesized on the cytoplasmic surface of the IM. It is converted to the intermediate molecule cytidine diphosphate diacylglycerol (CDP-DAG) by the enzyme CDP-DAG synthase (CdsA) that utilizes cytidine triphosphate (CTP) and releases pyrophosphate. The pathway then bifurcates based on the requirement of PL that the cell is needed to synthesize. CDP-DAG functions as a donor of phosphatidyl moieties for the biosynthesis of phosphatidylserine (PS) and phosphatidylglycerol-3-phosphate (PGP) with the release of cytidine monophosphate (CMP). The PS molecule formed by phosphatidylserine synthase (PssA) is decarboxylated by phosphatidylserine decarboxylase (Psd), producing the abundant PL, i.e. phosphatidylethanolamine. In the other route, PGP is quickly dephosphorylated by inner membrane phosphatases (PgpA-C) to form phosphatidylglycerol (PG). Finally, CL is synthesized by the condensation of PG/PE molecules by cardiolipin synthase (Cls). Both ClsA and B enzymes form CL by the condensation of two PG molecules. However, ClsC catalyzed reaction utilizes PE as a phosphatidyl donor to PG to synthesize CL (Figure 1.4). This hints at an intermingling of the PL synthesis pathway (Tan et al., 2012; Henderson et al., 2016; Sohlenkamp and Geiger, 2016).

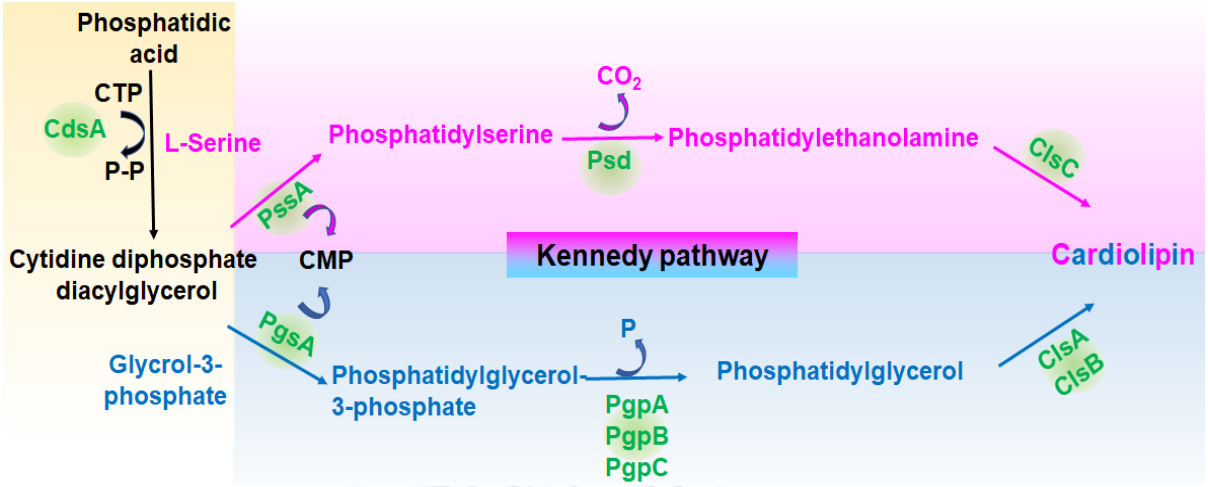


Figure 1.4. Kennedy pathway for biosynthesis of PLs in E. coli. *Phosphatidic acid molecule serves as the precursor for all other PLs. The enzyme CdsA converts this to CDP-diacylglycerol. The enzyme PssA converts CDP-diacylglycerol to phosphatidylserine, which is subsequently converted to phosphatidylethanolamine in the presence of enzyme Psd. On the other hand, the enzyme PgsA converts CDP-diacylglycerol to phosphatidylglycerol-3-phosphate, which is eventually converted to phosphatidylglycerol in the presence of enzymes PgpA-C. Cardiolipin is synthesized by the condensation of two molecules of phosphatidylglycerol or one molecule each of phosphatidylethanolamine and phosphatidylglycerol in the presence of enzymes ClsA-C.*

1.3.4.2. PL sorting

The OM-located lipoproteins and β-barrel proteins contain specific N-terminal signal sequences that direct their transport. In contrast, a strict compartmentalization of the LPS molecules in the outer leaflet of the OM is observed. However, unlike these three OM components, no sorting factor(s) specific to PLs transport and any distinguishing feature of PLs located in the OM has been reported. Also, there is no change in the PL characteristics between the inner leaflet of the OM and the IM, where PLs remain scattered randomly (Henderson et al., 2016).

1.3.4.3. OM enzymes involved in the maintenance of PL asymmetry

The maintenance of OM asymmetry is dependent on the regulation of the ideal PL population in the OM. However, the asymmetric organization of the OM is entropically

unfavourable, though it is critical for the barrier function. Hence, the PLs have a tendency to flip back to the outer leaflet from the inner leaflet in order to attain an entropically favourable state (Chong et al., 2015). But this would result in the compromise of the barrier function. In order to safeguard the permeability barrier, the misplaced PLs are required to flip back or be removed so as to restore the OM integrity (Figure 1.5A). Two distinct mechanisms involving OM β-barrel enzymes, viz. PldA and PagP have been reported to maintain the ideal PL population in the OM during stress conditions (Figure 1.5B-1.5C). The first mechanism involves phospholipase PldA, which remains in the inactive monomeric state when the PL asymmetry is intact. Perturbations in the OM leading to the accumulation of PLs in the outer leaflet induce the dimerization of the PldA to attain the catalytically active state. The active PldA removes the *sn*-1 and *sn*-2 fatty acid side chains from the misplaced PLs and restores the PL asymmetry in the OM. The second mechanism involves the palmitoyltransferase PagP which remains in a dormant monomeric state in unstressed cells. PagP is induced due to the limitation of divalent cations that cause the disruption of the LPS barrier and accumulation of unwanted PLs in the OM. The enzyme cleaves a palmitate moiety from the *sn*-1 position of a PL molecule and transfers it to the lipid A moiety of LPS (Bishop, 2008). This results in the formation of hepta-acylated LPS molecules of increased hydrophobicity, resulting in improved quality of the OM. However, the two mechanisms are destructive by nature and do not explain how PLs from IM reach the OM and come into play during stress.

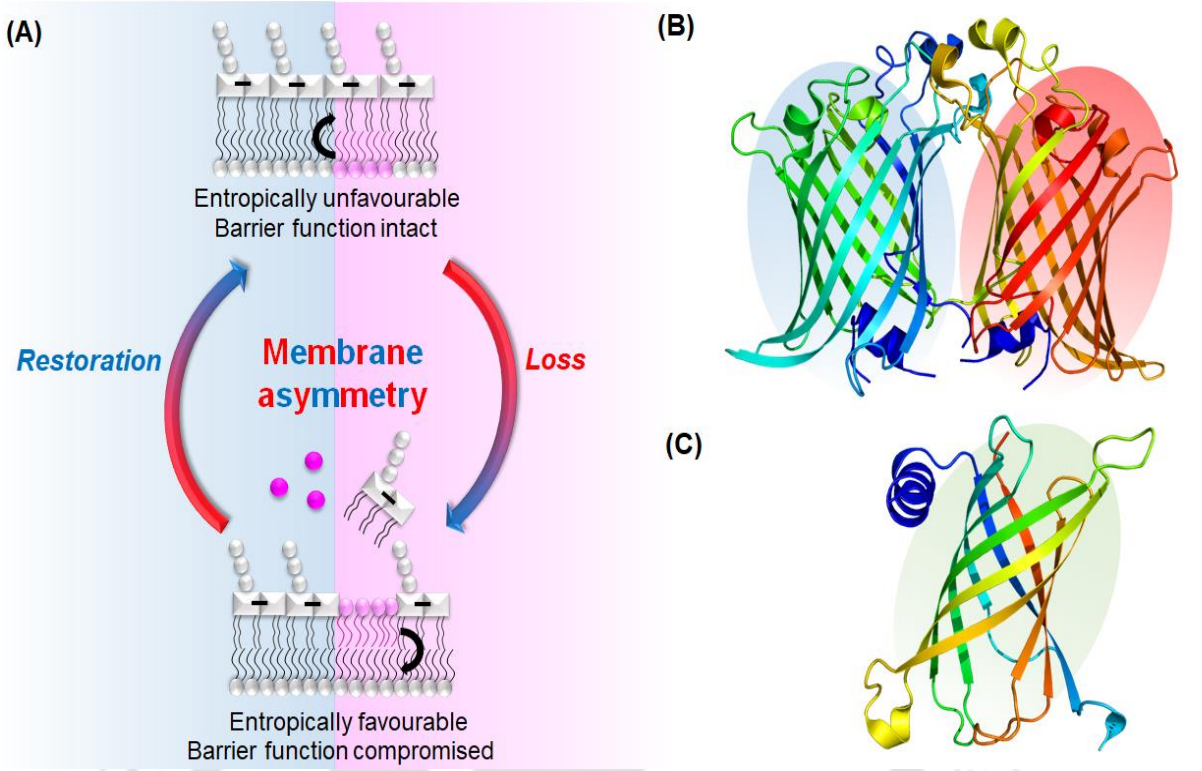


Figure 1.5. Enzymes involved in the maintenance of OM asymmetry. *(A) Loss and restoration of membrane asymmetry. (Top) The asymmetric nature of OM has an intact barrier function but is entropically unfavourable. The PLs have the tendency to flip and compromise this asymmetry. (Bottom) Symmetric OM has PL in the outer leaflet instead of LPS. This is entropically favourable, but the barrier function is compromised. This leads to the uninterrupted passage of drugs into the bacterial cells. To restore the barrier function, misplaced PLs need to be removed or flipped back. (B-C) Crystal structures of PldA and PagP from E. coli (PDB ids: 1QD6 and 1THQ, respectively). Both PldA and PagP are OM β-barrel enzymes involved in the removal of misplaced PLs.*

1.3.4.4. Transporters involved in ferrying PLs between the membranes

Transport of PLs presents a unique challenge in Gram-negative bacteria. Following synthesis in the cytoplasmic side of the IM, the PL molecules are flipped across the IM by yet-to-be-identified transporters. Furthermore, for incorporation in the OM, PLs must be removed from the IM and transported through the aqueous periplasmic space to the inner leaflet of the OM, where they are primarily found. Two mechanistic models have been proposed to explain such a transport process. The first model highlights the

utilization of soluble lipid-protein complexes and continuous physical bridges made of protein between the membranes. The second model proposes the utilization of vesicle budding and physical membrane bridges (Shrivastava and Chng, 2019). The discovery of the maintenance of the lipid asymmetry (Mla) system has provided a new perspective on PL transport and OM asymmetry (Malinverni and Silhavy, 2009). The Mla system represents a non-vesicular transport of PL transport which follows the soluble lipid-protein complex model and is considered to be the third mechanism of maintenance of OM asymmetry. In addition to the Mla transporter, various other systems have been reported to be involved in PL transport, which is discussed below.

The Mla system

The Mla system is a highly conserved inter-membrane transport machinery that consists of six proteins, MlaA (OM lipoprotein), MlaB (cytoplasmic protein), MlaC (periplasmic SBP), MlaD (IM-associated periplasmic protein), MlaE (TMD) and MlaF (NBD). The proteins MlaB and MlaF remain closely associated in the cytoplasmic side (Malinverni and Silhavy, 2009). Eventually, it was reported that MlaA specifically interacts with OmpC. These interactions are critical for the localization of MlaA in the OM, even in the absence of a lipid anchor. It was demonstrated that OmpC is an additional component of the Mla system (Chong et al., 2015). Together, all these proteins give rise to three distinct components, the IM ABC transporter complex (MlaFEDB), periplasmic mediator (MlaC) and OM complex (OmpC-MlaA) (Figure 1.6A). Furthermore, the Mla system contains two unique proteins, MlaB and MlaD, that contain a single copy of the highly conserved Sulphate Transporter and AntiSigma factor antagonist (STAS) and Mammalian cell entry (Mce) domains, respectively (Malinverni and Silhavy, 2009). These proteins are not present in the typical ABC transporters and hence, are considered auxiliary components of the system (Thong et al., 2016). Overexpression and purification studies of the entire *mla* operon reveal that MlaFEDB forms a stable complex wherein MlaB and MlaD are critical for the stability and ATPase activity of the complex (Thong et al., 2016; Ekiert et al., 2017).

Unlike PldA and PagP, the Mla system does not chemically modify the unwanted PLs. As per the proposed mechanism, the surface-exposed PL would be removed by OmpC-Mla complex and handed over to MlaC, which would ferry the PL to MlaFEDB for internalization at the expense of ATP hydrolysis (Malinverni and Silhavy, 2009; Chong et al., 2015). Thus, the Mla system elegantly restores the OM asymmetry by

relocating the PL via a retrograde movement. All the components of the Mla system are critical for the proper functioning of the system. Removal of any of these components results in OM defects and increased permeability as well as susceptibility (Malinverni and Silhavy, 2009; Ekiert et al., 2017). Although the retrograde transport model of the Mla system has been extensively supported, recent reports have also highlighted the anterograde transport of PL. The studies suggest that the Mla system is involved in the transport of PL from the IM to OM (Hughes et al., 2019; Kamischke et al., 2019). As a result of this, the subject of the directionality of transport of the Mla system remains unresolved.

The different components of the Mla system have been structurally characterized. The crystal structure of MlaC from *E. coli* (*Ec*MlaC) reveals the presence of an endogenously bound PL molecule. The tails of the PL molecule remain deeply buried in the hydrophobic pocket of *Ec*MlaC while the head remains solvent exposed. Also, some MlaC orthologs can bind two PLs (Ekiert et al., 2017; Yero et al., 2021). On the other hand, MlaD from *E. coli* (*Ec*MlaD) undergoes oligomerization to give rise to a homo-hexameric ring with a central hydrophobic channel. The PLs would pass through this channel during the process of transport (Ekiert et al., 2017). MlaA is reported to have a doughnut-shaped structure with a central amphipathic channel for the passage of PLs (Abellón-Ruiz et al., 2017). The cryo-EM structures of the MlaFEDB complex from Gram-negative bacteria such as *E. coli*, *Pseudomonas aeruginosa* and *Acinetobacter baumannii* reveal the different conformations of the complex in nucleotide-free and –bound states (Kamischke et al., 2019; Coudray et al., 2020; Zhou et al., 2021). However, in spite of these studies, the detailed molecular picture associated with the process of PL transport is still not available.

The Pqi and Yeb systems

Similar to the Mla system, both Pqi and Yeb systems are inter-membrane transporters. The Pqi system consists of three components, PqiA (TMD), PqiB (IM-associated periplasmic protein) and PqiC (OM lipoprotein). On the other hand, the Yeb system consists of two components, YebS (TMD) and YebT (IM-associated periplasmic protein) (Nakayama and Zhang-Akiyama, 2017). PqiB and YebT systems contain three and seven copies of Mce domains, respectively (Isom et al., 2020). Similar to MlaD, PqiB and YebT systems form a homo-hexameric ring and give rise to a central channel for the passage of PLs (Figure 1. 6B-1.6C). Thus, both systems represent the physical

bridge model of lipid transport. Unlike the Mla system, both Pqi and Yeb systems lack NBD and hence, it has been speculated that the process of transport for both these systems would not depend upon ATP hydrolysis (Ekiert et al., 2017; Isom et al., 2020). The deletion of *pqiABC* and *yebST* operons leads to an increase in OM susceptibility and permeability (Nakayama and Zhang-Akiyama, 2017). The detailed mechanism of action of PL transport, especially the directionality of transport, is not available. However, structural analysis suggests that the flexible central channel would aid in the movement of PLs (Isom et al., 2020; Liu et al., 2020).

PbgA

The protein PbgA is involved in the transport of CL from the IM to the OM. It consists of an N-terminal membrane-spanning domain and a C-terminal globular domain (Figure 1.6D). Deletion of the globular domain causes OM defects and increases OM permeability (Dong et al., 2016). PbgA/YejM is required for the enrichment of CL in the OM in *Salmonella Typhimurium*. (Dalebroux et al., 2015). The crystal structure of the globular domain of PbgA is structurally homologous to the arylsulfatase protein family. It possesses a hydrophobic pocket that accommodates and transports CL. The opening of the pocket is mainly modulated by loop movement that serves as a lid (Dong et al., 2016).

Tol-Pal

The Tol-Pal complex is a highly conserved trans-envelope system in Gram-negative bacteria (Shrivastava and Chng, 2019). Although the physiological role of the system has remained elusive; recently, it was established that the Tol-Pal complex is involved in PL transport and OM lipid homeostasis in *E. coli* (Shrivastava et al., 2017). Cells lacking this complex exhibit OM defects and increased susceptibility (Shrivastava et al., 2017). It consists of five proteins, TolQ (TMD), TolR-TolA (IM-associated protein with the major domains residing in the periplasm), TolB (periplasmic component) and Pal (OM lipoprotein that binds peptidoglycan). These proteins give rise to two sub-complexes, TolQRA (IM) and TolB-Pal (OM). In a proton motive force (pmf)-dependent fashion, TolQR transfers energy to govern conformational changes in TolA, enabling it to reach across the periplasm and make contact with Pal. Moreover, TolA engages in interaction with periplasmic TolB, whose role in the complex is unclear (Shrivastava et al., 2017).

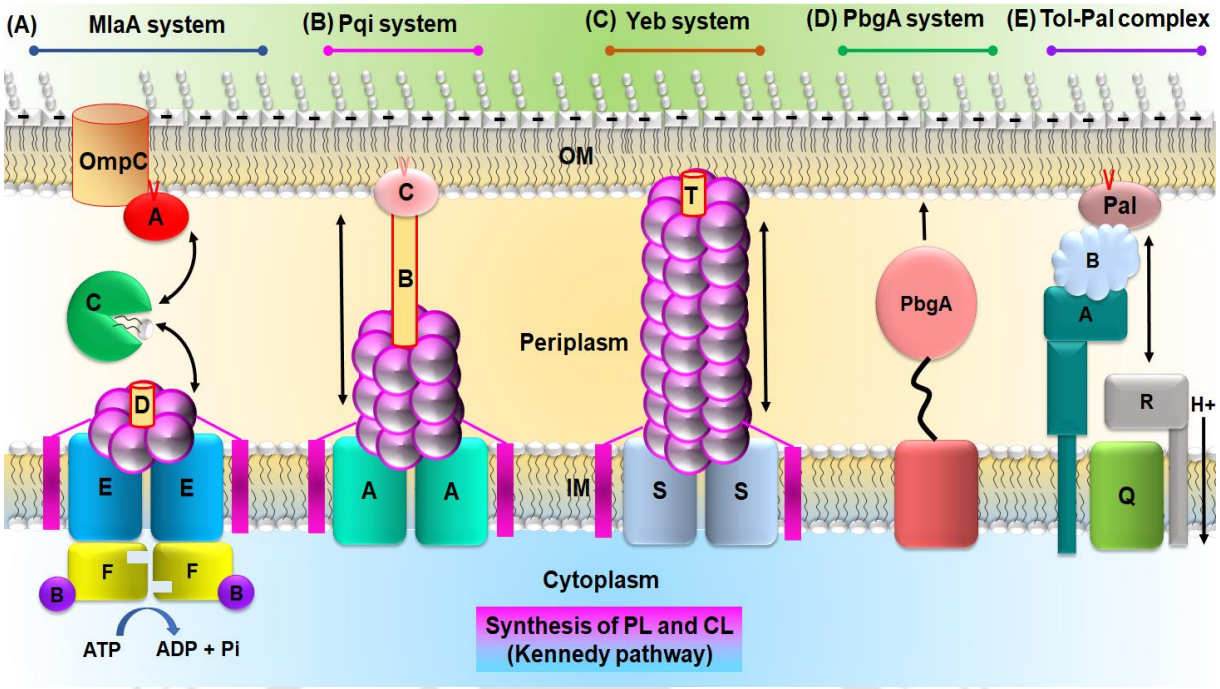


Figure 1.6. Overview of the transporters involved in the transport of PLs. *(A) Mla system. The system consists of seven proteins that form three sub-complexes, MlaFEDB (IM), MlaC (periplasmic) and OmpC-MlaA (OM). (B) Pqi system. It consists of PqiA (IM), PqiB (IM-associated periplasmic) and PqiC (OM-associated). (C) Yeb system. It consists of YebS (IM) and YebT (IM-associated periplasmic). (D) PbgA. It consists of a single protein PbgA with an N-terminal transmembrane region (IM) and a C-terminal globular region (periplasmic). (E) Tol-Pal complex. It consists of five proteins that form three sub-complexes, TolAQR (IM), TolB (periplasmic) and Pal (OM-associated). The directionality of ligand transport is depicted by black arrows.*

1.4. Mla system does not possess a typical ABC transporter construct

The ABC transporter is a highly conserved superfamily of integral membrane proteins that couple the transport of a diverse array of substrates (sugars, amino acids, peptides, vitamins, etc.) across the membrane to the hydrolysis of ATP molecules (Rees et al., 2009; ter Beek, 2014). They are ubiquitous and are present in all domains of life. In addition to transport, the members of this superfamily are involved in various cellular processes such as quorum sensing, signal transduction, bacterial pathogenesis, virulence and the ability to survive in a different environment, drug or antibiotic resistance, etc. (Lewis et al., 2012). Based on the directionality of transport, ABC transporters can be divided into two types,importers and exporters. ABC

importers transport substrates from the extra-cytoplasmic region to the cytoplasmic region, whereas ABC exporters do the same in the reverse direction (Locher, 2009). ABC exporters are present in all domains of life, while ABC importers are confined to prokaryotes, though reports have highlighted their presence in some plants (Terasaka et al., 2005). Regardless of the directionality of transport or type of substrate, all ABC transporters share a common structural architecture (Figure 1.7). They are composed of two copies of integral membrane proteins, viz. transmembrane domains (TMDs) and two copies of soluble proteins associated with the TMDs on the cytoplasmic side, viz. nucleotide-binding domains (NBDs). The TMDs constitute the transport channel for the passage of substrates, while NBDs bind and hydrolyze ATP providing the necessary energy for the transport. In addition to these functional units, ABC importers, in general, are dependent on an additional subunit, referred to as substrate(solute)-binding proteins or domains (SBPs), responsible for substrate capturing and delivering it to the TMDs (Oldham et al., 2008; Moussatova et al., 2008). Each of the three components has been extensively studied and categorized under different classification schemes.

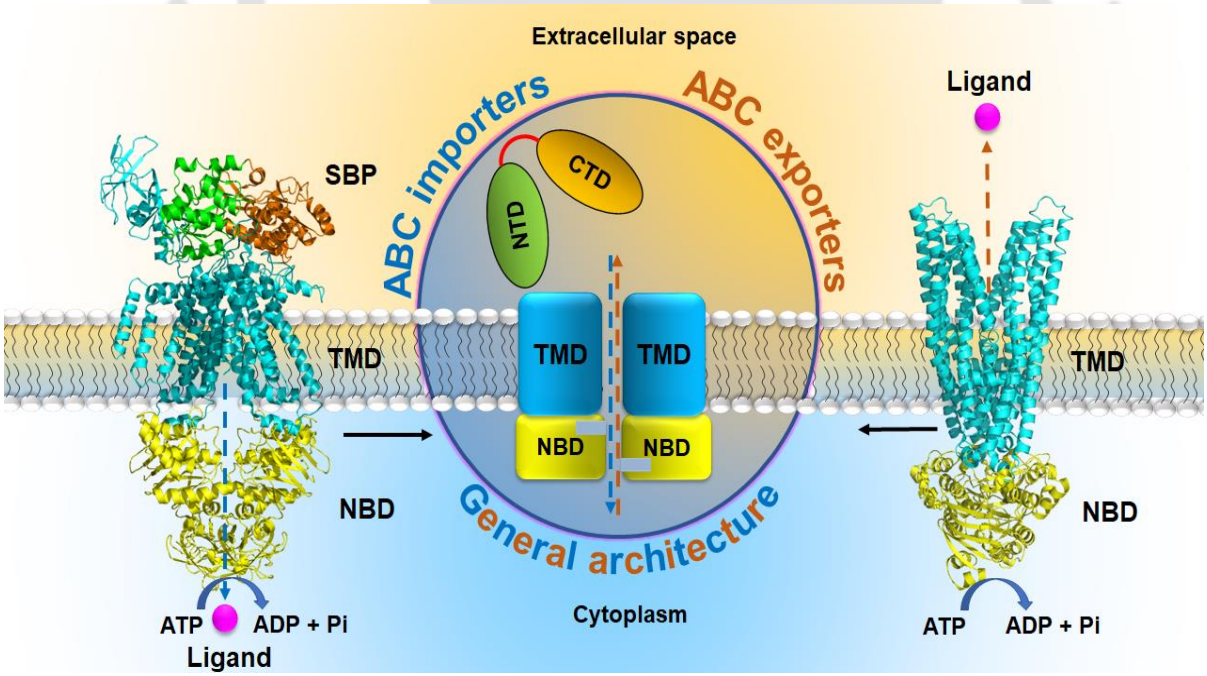


Figure 1.7. General architecture of ABC transporters. *(Left, right) The overall structure of maltose importer from E. coli (PDB id: 4KHZ) and Sav1866 exporter from Staphylococcus aureus (PDB id: 2HYD). (Centre) Schematic representation of ABC transporters. SBP, TMD, NBD and ligands are coloured in green-orange, cyan-teal,*

yellow and magenta, respectively. The directionality of ligand transport is represented by dotted arrows. CTD, C-terminal domain; NBD, nucleotide-binding domain; NTD, N-terminal domain; SBP, substrate-binding protein; TMD, transmembrane domain.

1.4.1. ABC importers

Although all the ABC importers are involved in the internalization of substrates, they exhibit significant diversity among themselves. Based on structural and biochemical evidence, ABC importers can be divided into three types, I, II and III (Energy-coupling factor; ECF) (Rice et al., 2014). Among the three, Type I and II are dependent on SBPs for substrate uptake. However, there are distinct structural and mechanistic features of Type I and II importers. The Type I importer fold has fewer TM helices and is organized around a minimal set of five TM helices, though it can increase to six or eight helices. On the other hand, Type II importer fold has a larger number (10-12) of TM helices that are packed together in an intricate topology (Figure 1.8A-1.8B) (Locher et al., 2002; Rees et al., 2009; Lewinson and Livnat-Levanon, 2017). Also, Type I importers have low basal ATPase rates and, upon stimulation by SBPs, exhibit high ATPase rates. On the other hand, Type II importers, in general, have opposite characteristics (Rice et al., 2014; Lewinson and Livnat-Levanon, 2017). Type III importers are structurally more diverse and possess a transmembrane 'S'-component responsible for substrate binding (Figure 1.8C) (Rice et al., 2014). Even though the Mla system has been reported to import PLs from the OM to IM, it has never been categorized under any of the three types.

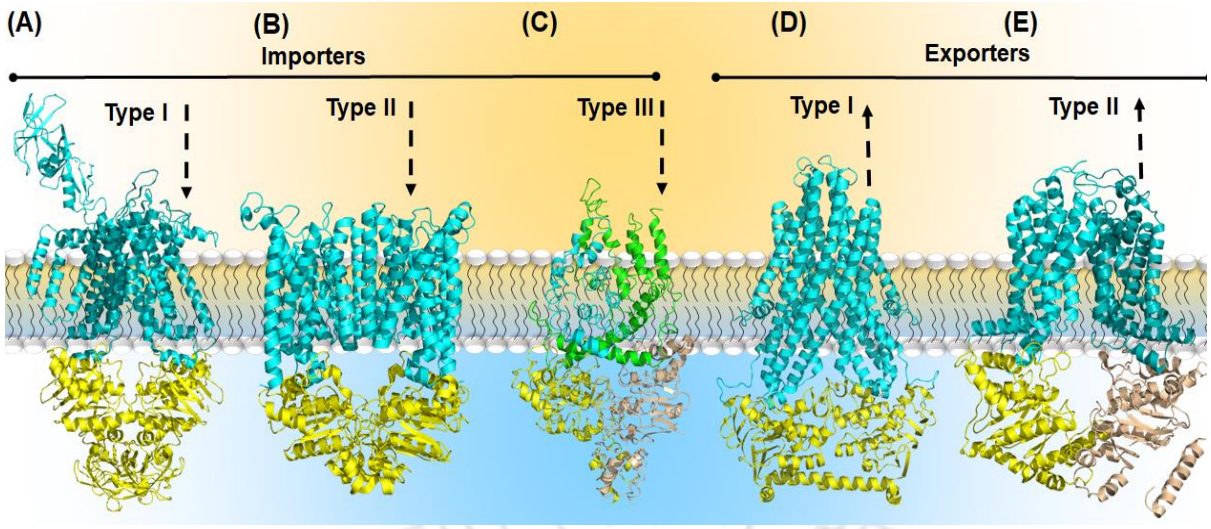


Figure 1.8. Overview of ABC importers and exporters. *(A-B) ABC importers belonging to Type I (Maltose importer MalFGK from E. coli; PDB id: 4KHZ), II (Vitamin B_{12} importer BtuCD from E. coli; PDB id: 1L7V) and III (Cobalamin importer ECF-CbrT from Lactobacillus delbrueckii; PDB id: 6FNP), respectively. (D-E) ABC exporters belonging to Type I (Metal-glutathione exporter Atm1 from Novosphingobium aromaticivorans; PDB id: 4MRV) and II (Sterol exporter ABCG5/ABCG8 from Homo sapiens; PDB id: 5DO7), respectively. The directionality of ligand transport is represented by dotted arrows.*

1.4.2. ABC exporters

Based on structural folds, ABC exporters are divided into two types, I and II. In Type I, the TMD communicates with the cytosolic NBD through a large intracellular domain (ICD) composed of a pair of long hairpin-like loops. Interestingly, this communication can be of two types. The first type involves domain swapping, wherein the coupling helix of the TMD of one monomer comes in contact with the NBD of the other monomer. On the other hand, in the second type, the coupling helix of the TMD of one monomer comes in contact with the NBD of the same monomer. In type II, there is no domain swapping, though the NBDs in close contact with the TMDs. These TMD and NBD are maintained via the coupling helix (Figure 1.8D-1.8E) (Lewinson and Livnat-Levanon, 2017; Banerjee et al., 2021). Interestingly, in spite of being considered an importer, the Mla system possesses the exporter fold (Coudray et al., 2020). However, it has never been incorporated into any one of the two ABC exporter types.

1.4.3. Transmembrane domain (TMD)

TMDs are well-conserved at the tertiary structure level, but they lack conservation at the primary structure level. This lack of conservation is expected owing to the diverse nature of substrates, which interact with specific residues during the process of transport. Interestingly, the translocation pathway constituted by the TMDs exhibits distinct folds and topologies. Also, the structural data of TMDs have increased manyfold in recent years. Hence, a new ABC transporter classification scheme has been proposed in which distinct types, I–VII transporters, have been identified (Table 1.1). The scheme is entirely based on the structural homology of the TMD fold and incorporates quantitative analyses performed by using TM-scores obtained by the structural alignment of TMDs (Thomas et al., 2020). Types I-III are unanimously composed of importers. Also, the structurally diverse ECF transporters are categorized as Type III. Interestingly, Type IV shows significant functional diversity in spite of possessing a common structural fold. Type V is mainly exporters, though transporters with importer functions are also included. Types VI and VII are represented by LptBFG and MacB, respectively. The TMDs of both systems are comparable to Type V. However, MacB possesses an amphipathic N-terminal 'elbow helix'. Also, both the TMDs differ in the number of TM and coupling helices. Hence, they have been allotted two different types. Interestingly, MlaE (5 TM helices) is structurally similar to LptFG (6 TM helices) and MacB (4 TM helices). But owing to the difference in the number of helices, MlaE has not been incorporated either in Type VI or in Type VII and still remains unplaced (Coudray et al., 2020; Thomas et al., 2020).

Table 1.1. Classification of ABC transporter based on TMD folds in prokaryotes.

TMD fold	TM helix organization	Proteins	Function	PDB id
Type I	(5-6) + (5-6/8)	$MalFGK_2$-(E)	Maltose import	2R6G
Type II	10 + 10	$BtuC_2D_2$(-F)	Cobalamin import	1L7V
Type III	4-8 (TMD) + 6-7 (S component)	EcfTAA'-FolT	Folate import	4HUO

Type IV	6 + 6	Sav1866	Multidrug export	2HYD
Type V	6 + 6	Wzm-WztN	O-antigen export	6OIH
Type VI	6 + 6	$LptB_2FG(C)$	LPS extraction	5X5Y
Type VII	4 + 4	MacB	Export of macrolides, polypeptides, virulence factors	5GKO

1.4.4. Nucleotide-binding domain (NBD)

The NBD is the highly conserved module of ABC transporters that binds and hydrolyze ATP molecules. It is present in the cytoplasmic side in a dimeric state (homo or heterodimer) (Figure 1.9A). Owing to their conserved nature, the NBD is considered the hallmark of the ABC transporter superfamily. Overall, an NBD can be divided into two structural domains, a helical domain and a RecA-like domain (Figure 1.9B) (Dassa, 2011). An NBD contains conserved motifs, Walker A motif, Walker B motif, signature motif, D-loop, H-loop and Q-loop (Figure 1.9C) (Moussatova et al., 2008; Dassa, 2011). ATP binding results in the formation of a tight nucleotide sandwich dimer, whereas ATP hydrolysis produces the dissociation of the dimer resulting in a nucleotide-free state (Figure 1.9D). Thus, the NBDs, in general, transition between nucleotide-bound and -free states. The conformational changes in NBD transfer mechanical energy to TMDs to drive transport (Moussatova et al., 2008).

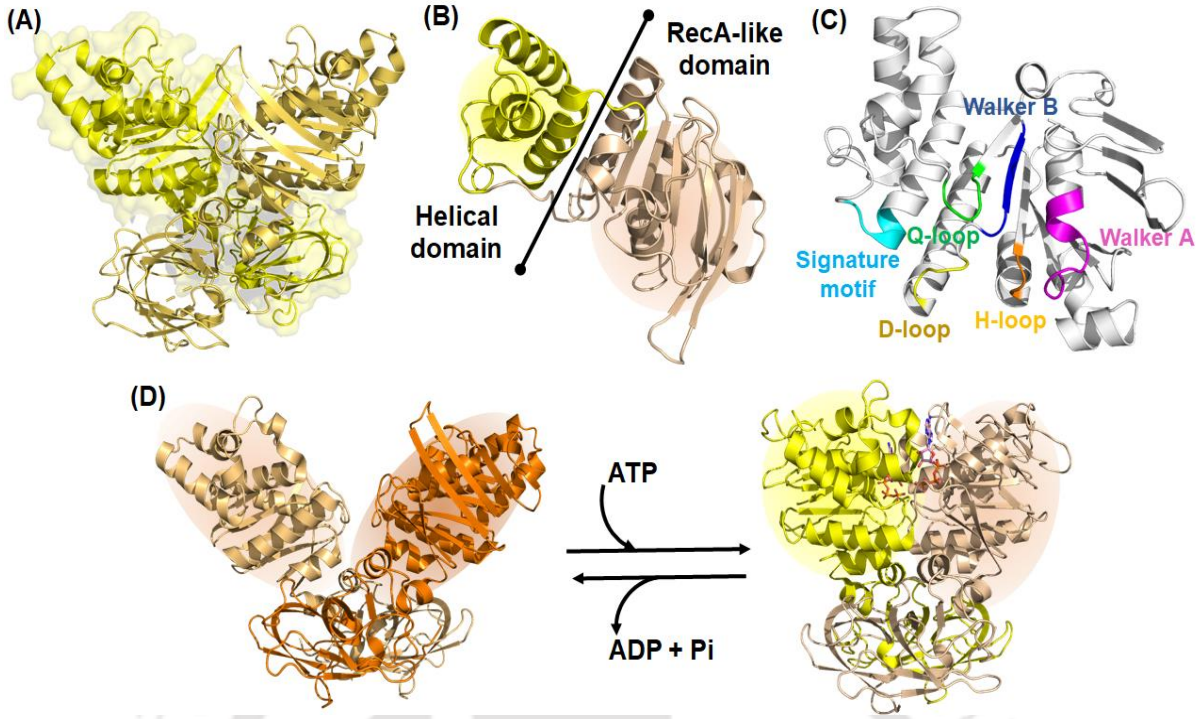


Figure 1.9. Architecture of NBD. *(A) Overview of the dimeric arrangement of NBD (MalK, PDB id: 4KHZ). (B) Domain arrangement of a monomer of MalK. The NBD can be divided into two distinct domains, helical (yellow) and RecA-like (wheat). (C) Organization of conserved motifs in NBD. Walker A motif, Walker B motif, signature motif, D-loop, H-loop and Q-loop are represented in magenta, blue, cyan, yellow, green and orange, respectively. (D) Dimeric conformation of MalK in an open state (nucleotide-free state; PDB id: 1Q1E) and closed state (ATP bound state; PDB id: 1Q12).*

NBDs can be divided into three main classes, I, II and III. Class I comprises systems with fused TMD and NBD. Class II comprises soluble NBD but no TMD. Class III comprises mostly importers with separate TMD and NBD (Table 1.2). These classes engulf 29 different families, which can be further divided into sub-families on the basis of similarities between the NBDs in order to distinguish the biological roles of systems. Phylogenetic tree analysis reveals distinct clustering of these families, indicating their evolutionary relationship (Dassa, 2011). As per reports, the MlaF protein belongs to the Mkl family of Class III. The Mkl family has been reported to be involved in the retrograde transport of lipids, organic solvent resistance and steroid uptake (Dassa,

2011). Furthermore, MlaF orthologs are reported to be involved in the intracellular spreading of bacteria, adherence and invasion of epithelial cells and safeguarding outer cell surface integrity (Dassa and Bouige, 2001; Dassa, 2011).

Table 1.2. Classification of NBDs.

Class	Characteristics	Proteins	Function
I	Fused NBD and TMD	MsbA	LipidA export
II	Only NBD present	UvrA	DNA repair and drug resistance
III	NBD and TMD separate	MlaF	Retrograde transport of lipids, organic solvent resistance, steroid uptake

1.4.5. Substrate (solute)-binding protein (SBP)

The SBPs are additional functional domains present in importers that are dedicated to the internalization of substrates from the extracellular milieu to the cytoplasm through the membrane bilayer. SBPs are critical for maintaining substrate specificity as well as selectivity and impart directionality to the process of transport. The substrates include a wide variety of nutrients or solute molecules such as carbohydrates, vitamins, amino acids, peptides, etc. which are captured by the SBPs and handed over to the TMDs (Scheepers et al., 2016). SBPs share low sequence similarity but possess a highly conserved architecture. They are composed of two α/β domains, viz. N- and C-terminal domains (NTD and CTD), which are linked by a flexible hinge region. In each domain, α-helices flank the β-sheets, which vary in number. The binding pocket of SBP is present in the form of a cleft between the two domains. The hinge region demonstrates a bending motion resulting in the movement of the NTD and CTD, thus, facilitating the entrapment of ligands (Figure 1.10A-1.10B)(Fukami-Kobayashi et al., 1999). During the process of ligand binding, the SBP transitions from open conformation (ligand-free state) to closed conformation (ligand-bound state). In Gram-negative bacteria, the SBPs remain confined to the periplasm and remain in a free-floating state. However,

in Gram-positive bacteria, SBPs are anchored to the cytoplasmic membrane via a lipid moiety located near their cognate TMD (Figure 1.10C) (Tanaka et al., 2018).

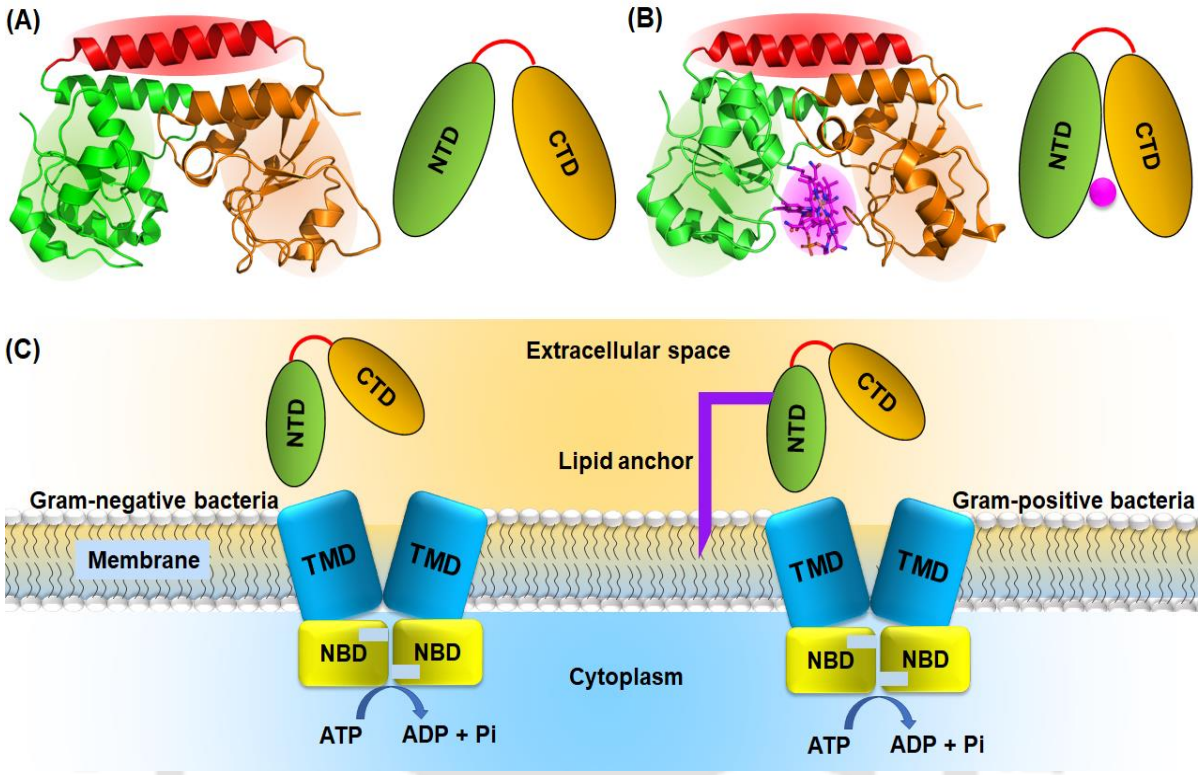


Figure 1.10. Overview of the SBP. *(A-B) Crystal structure of ligand-free (PDB id: 1N4D) and VitaminB_{12}-bound (PDB id: 1N4A) structure of BtuF from E. coli (left, right) along with their schematic representation highlighting the domain movement, respectively. (C) (Left, right) Schematic representation of SBP organization in Gram-negative and Gram-positive bacteria, respectively. CTD, C-terminal domain (orange); NBD, nucleotide-binding domain (yellow); NTD, N-terminal domain (green); SBP, substrate-binding protein (green-orange); TMD, transmembrane domain (cyan).*

1.4.5.1 Classification of SBPs

Based on the topological arrangement of the β-sheets in the core regions of NTD and CTD, SBPs have been categorized into three different types, I, II and III. In Type I and II, the β-sheet organization follows $\beta_2\beta_1\beta_3\beta_4\beta_5$ and $\beta_2\beta_1\beta_3\beta_n\beta_4$ (where n is the cross-over β-strand) arrangement. Both types possess connecting strands as the hinge region (Fukami-Kobayashi et al., 1999). However, in Type III, a single α-helix serves

as a hinge region that links the NTD and CTD of the SBP (Lee et al., 1999). Since this report, the structural data on SBP has increased substantially, which tends to exhibit immense diversity. In order to understand such diverse data, a new classification scheme of SBP was introduced, which was entirely based on the structural fold and substrate specificities. As per this scheme, SBPs can be classified into six distinct clusters (A-F) (Berntsson et al., 2010). This classification was further updated by taking into consideration the even larger number of SBP structures (504 unique SBP entries) that were deposited in PDB in due course of time. As per this scheme, the SBPs are classified into seven clusters (A-G) with additional subclusters (Scheepers et al., 2016). In recent times, the scheme of classification was again updated by taking into consideration a higher number of PDBs (764 unique SBP entries). As per the latest classification, SBPs can be divided into eight distinct clusters (A-H) with new subclusters (Figure 1.11A-1.11H and Table 1.3). Furthermore, the report also highlights the subdivision of NTD and CTD into subdomains (N1, N2, C1 and C2), giving rise to diverse topologies (Chandravanshi et al., 2021).

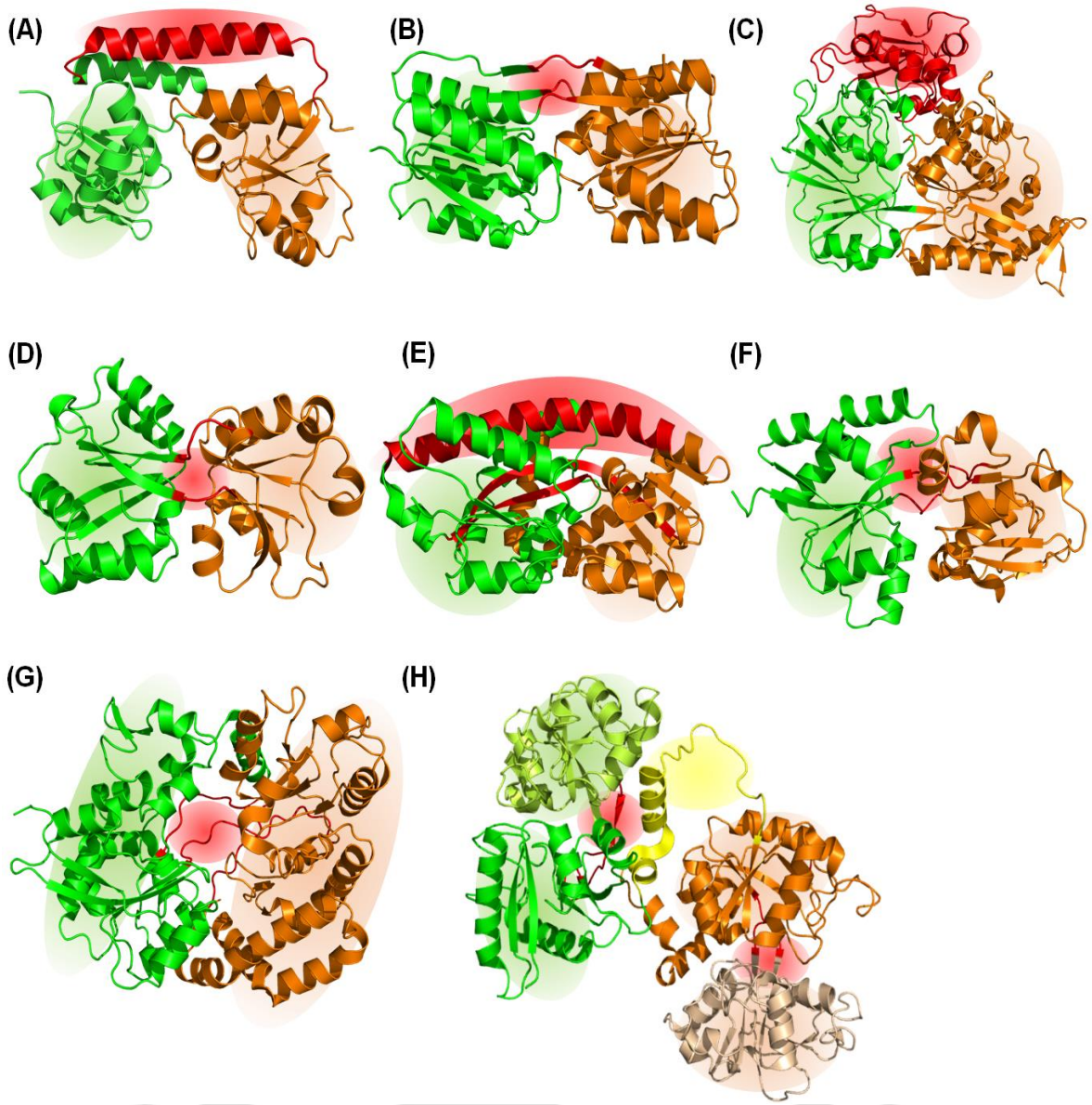


Figure 1.11. Structure-based classification of SBPs. *Based on the overall topology and structural features of hinge regions (highlighted in red) SBPs are classified into (A) cluster A (e.g. BtuF, PDB id: 1N2Z), (B) cluster B (e.g. Ribose-binding protein, PDB id: 1DRJ), (C) cluster C (e.g. OppA, PDB ID: 3DRF), (D) cluster D (e.g. ModA, PDB id: 1AMF), (E) cluster E (e.g. UehA, PDB id: 3FXB), (F) cluster F (e.g. MetQ, PDB id: 4GOT), (G) cluster G (e.g. AlgQ1, PDB id: 1Y3P) and (H) cluster H (e.g. Serotransferrin, PDB id: 1JNF). For each SBP, NTD and CTD are represented in green and orange, respectively. The short peptide connecting NTD and CTD in cluster H is represented in yellow.*

Table 1.3. An overview of the SBP classification.

Cluster	Subcluster	Types of ligands
A	A-I	Metal ions
	A-II	Metal ions, Fe^{3+} hydroxamate heme, Ferric vibriobactin, ferrichrome, Petrobactin, siderophore, Staphyloferrin A, Vitamine B12
B	B-I	Sugar, sugar alcohols, Autoinducer 2
	B-II	Amino acids
	B-III	Aromatic compounds
	B-IV	Natriuretic peptides
	B-V	Glutamate, L-glutamine
	B (unclassified)	n.d.
C	C-I	(Oligo)peptide, metal ion, sialic acid, glutathione, heme, Methanobactin
	C-II	Nickel, peptide, staphylopine
	C-III	(Di/oligo)peptides, heme, agropinic acid, mannopinic acid
	C-IV	(Oligo/Tetra)peptide, heme, raffinose
	C-V	(Chito)oligosaccharide, oligopeptide, phosphate
	C (unclassified)	Dipeptide
D	D-I	Maltose, maltotriose, molybdate, glucose, galactose
	D-II	Spermidine, putrescine, citrate, shikimate, phosphate, sodium azide,

		mannopine, agropine
	D-IIIa	Phosphate, tungstate, sulfate, molybdate, iron
	D-IIIb	Molybdate, tungstate
	D-IV	Iron, aromatic acids, polyamine
	D^a	Phosphate
	D (unclassified)	Maltose, acarbose, tungstate, thiamin, polyamine
E	E-I	Sialic acid, D-glucuronate, galacturonate, succinate, glycerol-3-phosphate, 4-hydroxybenzoate
	E-II	Glutamate, Aspartate, Oxoadipate, malate
F	F-I	Phosphite, bicarbonate, nitrate, sulfonate, phosphonate
	F-II	Methionine
	F-III	Glycine betaine, proline, ethanolamine,
	F-IV	Amino acids, kainate, octopine, sulfobetaine
	F (classified)	Nitrate, sulfonate, thiamine
G	G	Alginate
H	H	Iron

Despite being SBPs, both MlaC and MlaD have not been incorporated in the SBP classification scheme. Interestingly, both proteins have been recognized as lipid transfer proteins (LTPs). The LTPs are involved in the transfer of lipid components of

bilayers across the aqueous phase. All LPTs provide an ideal hydrophobic environment in which the lipids are more stable as compared to in acquires solution. This is due to the cavity present in the LTPs with a hydrophobic lining that accommodates the lipid molecule. Owing to its resemblance to a box with an internal cavity, MlaC is categorized as a box-like lipid shuttle. Such LTPs shuttle between donor and acceptor compartments following multiple steps, which include docking, lipid extraction, undocking, cytosolic diffusion and deposition of lipids. On the other hand, MlaD is categorized as bridge-like LTPs (Wong et al., 2019). Such LTPs have a continuous opening that extends along their lengths. Such an arrangement creates a tunnel-like structure through which lipids pass while the LTP remains stationary (Wong et al., 2017; Wong et al., 2019).

1.5. Mla system shows similarity to the Mce system

Interestingly, the Mla transporter shows similarity to the mammalian cell entry (Mce) system present in *Mycobacterium tuberculosis* and other members of the phylum actinobacteria. The Mce system is responsible for the acquisition of host cholesterol as well as fatty acids in *M. tuberculosis* and has been reported to contribute to virulence and pathogenesis (Klepp et al., 2022). The similarity between Mce and Mla systems is mainly due to the presence of the highly conserved Mce domain in MlaD and Mce proteins. The domain is an 81 amino-acid-long sequence with multiple conserved hydrophobic residues and is the defining feature of Mce proteins (Isom et al., 2017). The first gene (*mce1A*) encoding a Mce protein was identified when a fragment of the *M. tuberculosis* genome was introduced into a non-pathogenic *E. coli* strain. The transformed *E. coli* cells were able to invade and survive within the non-phagocytic HeLa cell line and macrophage (Arruda et al., 1993). Eventually, the complete genome sequencing of *M. tuberculosis* revealed that it contains four *mce* operons (*mce1-4*), each containing six *mce* genes (*mce1A-1F*, *mce2A-2F*, *mce3A-3F* and *mce4A-4F*). In addition to this, each operon in the upstream contains two *yrbE* genes (*yrbE1A-1B, yrbE2A-2B, yrbE3A-3B* and *yrbE4A-4B*) (Figure 1.12A). The *mce* genes encode 24 Mce proteins (Mce1A-4F), each containing a single copy of the Mce domain and are orthologs of the MlaD protein from *E. coli*. On the hand, the *yrbE* genes encode eight permeases (YrbE1A-4B) which are the orthologs of the MlaE protein from *E. coli* (Cole et al., 1998; Casali and Riley, 2007). Similar to MlaD, the Mce proteins remain associated with the IM of *M. tuberculosis* through their N-terminals (Fenn et al., 2020).

Additionally, the *mceG* gene, situated away from the *mce* operons, encodes an NDB, which is the ortholog of the MlaF protein from *E. coli* and remains associated with YrbE proteins. Although the basic architecture of *mce* operons consists of *mce* and *yrbE* genes, various other associated genes such as *mceR*, (orphaned) Mce-associated membrane (*omam*, *mam*) and lipid uptake coordinator A (*lucA*) genes have been reported to play a critical role in the proper functioning of the Mce system (Perkowski et al., 2016; Nazarova et al., 2017; Wilburn et al., 2018). Although the structural data of the Mce system from *M. tuberculosis* is not available to date, it has been proposed that Mce proteins from each operon would give rise to a hetero-hexameric arrangement with a central tunnel. The hexamer would rest on the dimer of YrbE permease, forming a continuous translocation pathway for the internalization of substrates acquired from the host (Ekiert et al., 2017; Perkowski et al., 2016; Wilburn et al., 2018). This proposed macromolecular arrangement is overall comparable to that of the Mla system (Figure 1.12B).

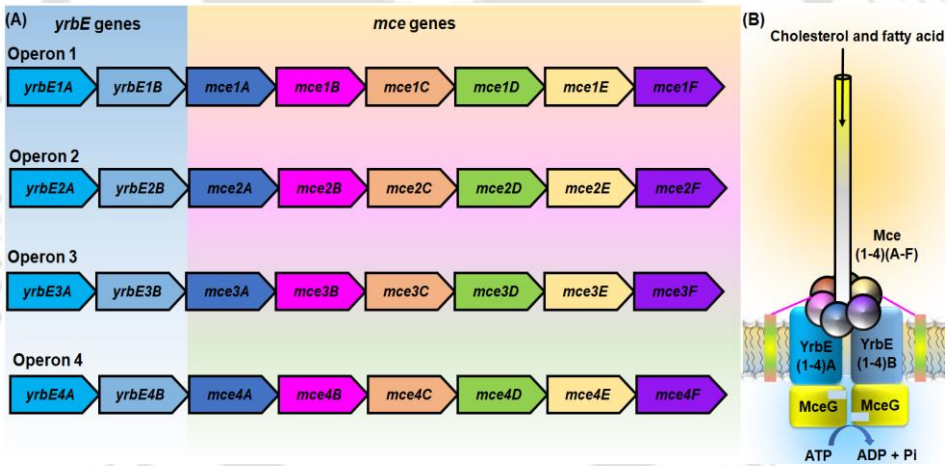


Figure 1.12. Organization of mce operon and Mce system in M. tuberculosis. *(A) Distribution of 4 mce operons (mce1-4) in M. tuberculosis. Each operon consists of six core mce genes (mce1A-1F, mce2A-2F, mce3A-3F and mce4A-4F) that encode membrane-anchored Mce proteins. These genes are preceded by two copies of yrbE genes (yrbE1A-1B, yrbE2A-2B, yrbE3A-3B and yrbE4A-4B) that encode integral membrane proteins. (B) Proposed macromolecular arrangement of Mce system in M. tuberculosis. The Mce proteins from each operon give rise to a hetero-hexameric ring with a central tunnel via which fatty acid and cholesterol molecules enter the bacterial*

cell. The energy required for the transport is obtained by the hydrolysis of ATP molecules performed by MceG (Ekiert et al., 2017).

1.6. Mechanism of substrate translocation by ABC importers

1.6.1. Ligand capture

During the process of ligand capture, the SBP engulfs the ligand and seizes it in the binding pocket between the NTD and CTD. Thus, SBP in the open unliganded state transitions to a closed liganded state after ligand capture. Despite the variation in structural folds and the nature of the transporting substrates, the SBPs capture ligands by following one of the four main mechanisms, (1) the “Venus Fly-trap” mechanism (synchronized movement of NTD and CTD), (2) the “asymmetric domain movement” mechanism (NTD exhibit a movement larger than that of the CTD), (3) “one domain movement” mechanism (NTD exhibits a prominent movement while the CTD remains static) and (4) “subdomain movement’’ mechanism (differential movement of the subdomain, N1, N2, C1, C2) (Figure 1.13A-1.13D) (Mao et al., 1982; Pandey et al., 2016; Chandravanshi et al., 2020a; Chandravanshi et al., 2020b).

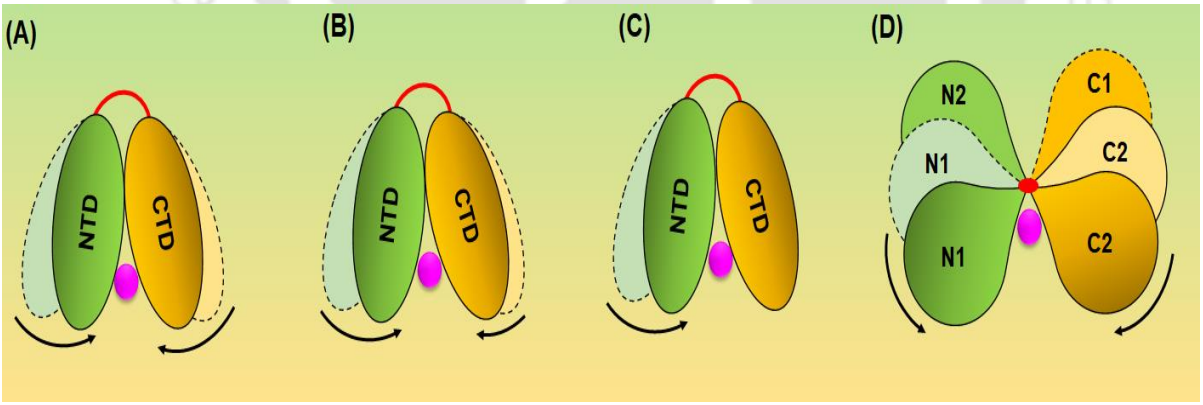


Figure 1.13. Overview of the different ligand binding mechanisms of SBP. *(A-D) ‘’Venus Fly-trap” mechanism, “asymmetric domain movement” mechanism, “one domain movement” mechanism and “subdomain movement’’ mechanism, respectively. The directionality of domain movements is shown by curved arrows. CTD, C-terminal*

domain (orange) and NTD, N-terminal domain (green). N1/N2 and C1/C2 represent the subdomains of NTD and CTD, respectively.

1.6.2. Internalization of ligand

The internalization of ligands involves induced conformational changes in the TMDs. After ligand capture, substrate-bound closed SBP interacts with its cognate TMDs leading to conformation changes in TMDs. Although the conformational landscape of TMDs can vary in different importers, the process, in general, involves the transition of the TMDs from an inward open to an outward open state (Rice et al., 2014). In the former state, NBD subunits remain in an open nucleotide-free state (Figure 1.14A). The interaction between ligand-bound SBP and TMDs causes the partial closure of NBDs (Figure 1.14B). ATP binding to NBD triggers their closure, conformational changes in TMDs leading to the narrowing of the translocation pathway and attainment of outward open conformation. Subsequently, the SBP releases the substrate molecule into the translocation passage (Figure 1.14C). Upon ATP hydrolysis, ADP and inorganic phosphate are released while TMDs return to their inward open conformation. The NBDs attain the nucleotide-free open state and the SBP departs the transport system (Figure 1.14D).

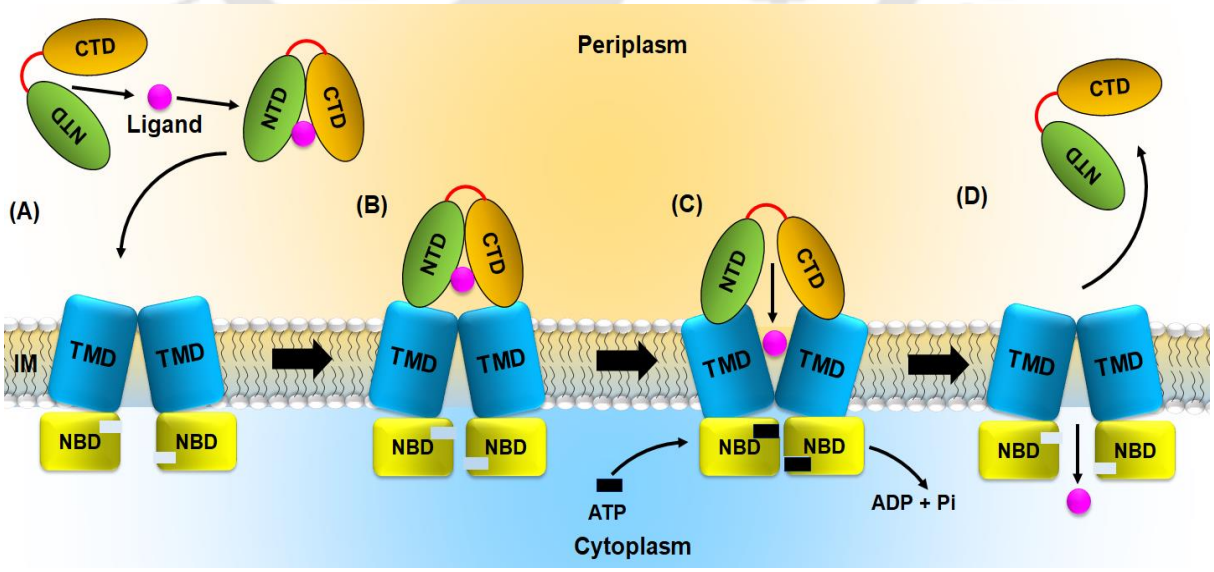


Figure 1.14. Schematic representation of substrate translocation by ABC importer. *(A) The Ligand-bound substrate binds to nucleotide-free inward-facing conformation of the transporter. (B) Partial closing of NBDs due to the interaction between SBP and TMD. (C) Closure of NBDs due to ATP binding, the collapse of translocation pathway, attainment of outward-open conformation and the release of ligand followed by undocking of SBP. (D) Hydrolysis of ATP resets the transporter to its original conformation. CTD, C-terminal domain (orange); NBD, nucleotide-binding domain (yellow); NTD, N-terminal domain (green); SBP, substrate-binding protein (green-orange); TMD, transmembrane domain (cyan). ATP and ligand are represented as rectangles (black) and circles (magenta), respectively.*

1.6.3. Mechanism of substrate translocation by ABC exporters

Based on the available structural data, ABC exporters have been reported to have different molecular mechanisms than ABC importers for ligand translocation. The working model for multidrug resistance (MDR) transporter is considered ideal for understanding the mechanism of ABC exporters. The exporters have a high affinity for ligands in the nucleotide-free state, which leads to ligand binding (Figure 1.15A). Ligand binding induces conformational changes in TMDs that result in increased affinity for ATP in the NBDs. ATP binding leads to the formation of a closed NBD dimer, which in turn causes a larger conformational change in the TMDs apt for ligand translocation (Figure 1.15B). This is followed by ATP hydrolysis, which initiates the dissolution of the closed NBD dimer and further conformational changes in TMDs (Figure 1.15C). After the release of ADP and Pi molecules, the transporter returns to the high-affinity state for the ligand (Figure 1.15D) (Linton et al., 2007).

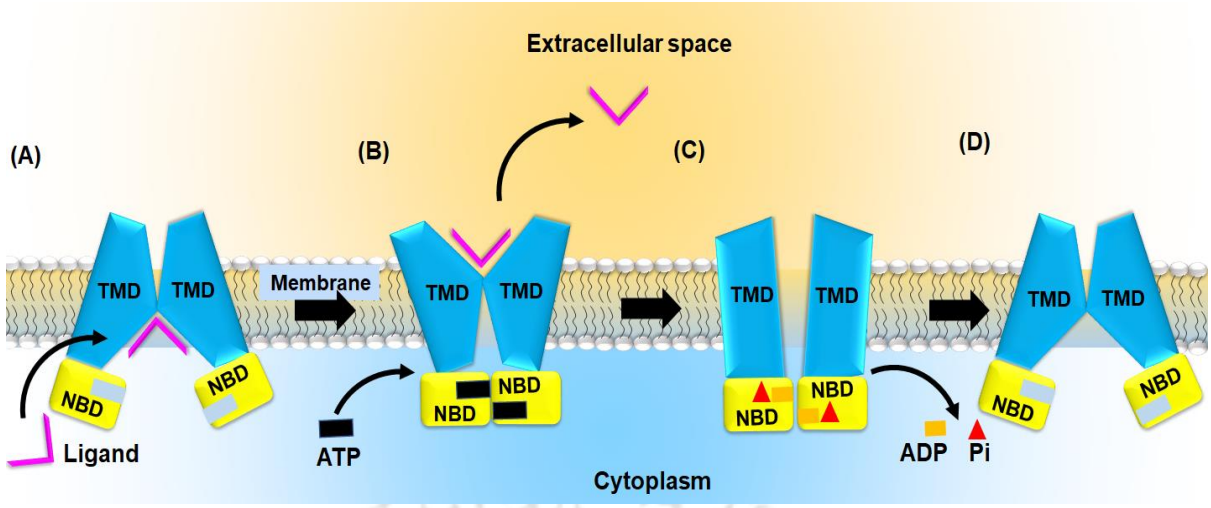


Figure 1.15. Schematic representation of substrate translocation by ABC exporter. *(A) ABC exporter in the nucleotide-free state has a high affinity for a ligand, which leads to binding. (B) Ligand binding increases the affinity of NBD for ATP binding, causing the closure of NBD dimers and conformational changes in TMD, leading to the expulsion of ligand. (C) ATP hydrolysis leads to the dissolution of NBDs and conformational changes in TMDs. (D) The release of ADP and Pi restores the transporter to its original state. CTD, C-terminal domain (orange); NBD, nucleotide-binding domain (yellow); NTD, N-terminal domain (green); SBP, substrate-binding protein (green-orange); TMD, transmembrane domain (cyan). ATP, ADP, Pi and ligand are represented as a rectangle (black), rectangle (orange), triangle (red) and half frame (magenta), respectively.*

1.7. Dilemma of directionality

Unlike typical ABC transporters that are dedicated to substrate translocation in a particular direction, the Mla system presents an unusual case. This is because the directionality of substrate transport by the Mla system still remains a debatable subject. As a result, a definite model describing the mechanism of PL transport is still lacking. Based on the available reports, the directionality of PL transport of the Mla system can be speculated from three major aspects, retrograde transport, anterograde transport and bidirectional/ambiguous transport.

1.7.1. Retrograde transport

Malinverni and Silhavy laid down the foundation stone for understanding the directionality of PL transport by the Mla system. They emphasized that the system would be involved in the retrograde movement of misplaced PL molecules from the OM to the IM (Malinverni and Silhavy, 2009). Based on biochemical assays, Chong and co-workers reported that either OmpC would flip the PLs, which would be removed by MlaA, or MlaA would directly extract PLs from the OM and hand them to MlaC (Chong et al., 2015). By conducting evolution experiments on LOS-deficient *A. baumannii*, it was reported that the Mla system contributes to the retrograde movement of PLs (Powers and Trent, 2018). Furthermore, mutational along with radioactive labeling studies have revealed that the Mla system does not have any demonstrable role in the anterograde PL transport (Powers et al., 2020). This finding was further supported by a proteoliposome-based study that highlighted the ATP-dependent retrograde transport of PLs (Tang et al., 2021). Crystal structures have also revealed that a major portion of MlaA remains OM embedded, which reaches out to the outer leaflet. As a result, only the misplaced PLs from the outer leaflet and not the PLs from the inner leaflet of the OM get access to the central channel of MlaA. The PLs from the outer leaflet would enter the channel and move towards the periplasm with the aid of MlaC (Abellón-Ruiz et al., 2017). Also, based on photocrosslinking studies, a retrograde model of PL transport by the Mla system was suggested. It was proposed that the Omp-MlaA complex would transfer PL to MlaC without the aid of any ATP hydrolysis. The high affinity of PL for MlaC would act as a driving force for such a transfer of substrates. However, this affinity is so strong that additional energy would be required for the delivery of the bound PL to the MlaFEDB complex (Ercan et al., 2018). Based on the available structural data of the MlaFEDB complex, Low and Chng, tried to trace the trajectory of PL movement and put forward a respectable model of PL retrograde transport coupled with ATP hydrolysis. As per this model, the PL-bound MlaC would dock on the MlaFEDB complex in a nucleotide-free state and hand over the PL (Figure 1.16A.1). The initial ATP binding would trigger conformational changes in the complex that would aid in the rearrangement of PL (Figure 1.16A.2). Eventually, the complete ATP binding would result in the collapse of the translocation pathway and extrusion of the PL into IM (Figure 1.16C.3). The final step would involve ATP hydrolysis so as to restore the MlaFEDB complex to resting state (Figure 1.16A.4) (Low and Chng, 2021).

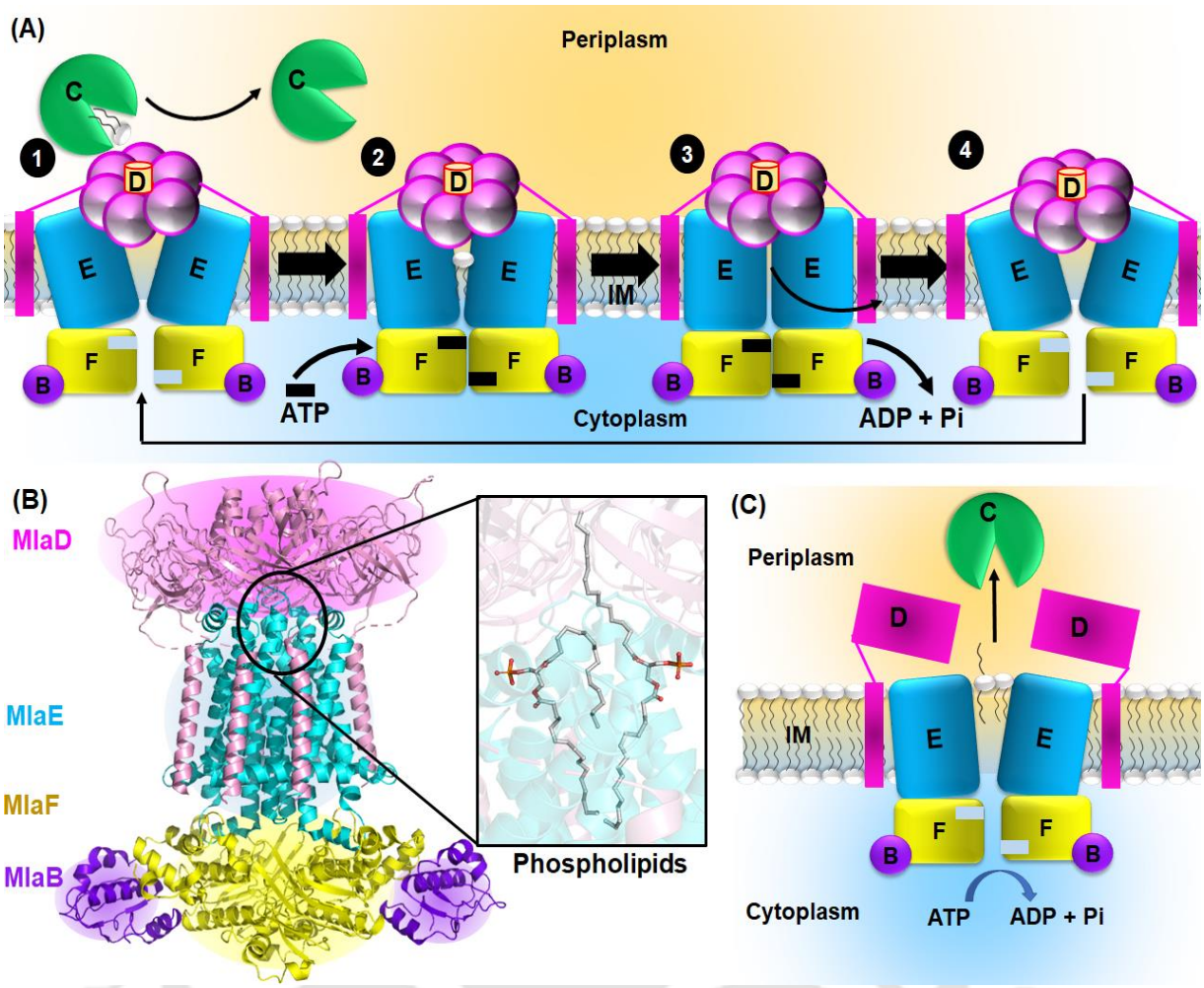


Figure 1.16. Proposed mechanism of PL transport by Mla system. *(A) Proposed mechanism of PL import by Mla system. (A) Docking of PL bound MlaC to nucleotide-free outward open IM Mla complex. (B) ATP binding causes conformational changes and re-orientation of PL. (C) The collapse of the translocation pathway and extrusion of PL into the IM. (4) Hydrolysis of ATP resets the IM complex to its original state (Low and Chng, 2021). (B) (Left) Cryo-EM structure of MlaFEDB complex from E. coli (PDB id: 6XBD). (Right) Two PLs bound to the MlaFEDB complex in two different states of export. (C) Proposed mechanism of PL export by Mla system. The MlaD hexamer would exhibit a slanting movement and MlaE would undergo conformation changes leading to the export of PL from the IM to MlaC (Coudray et al., 2020). The proteins MlaBCDEF are represented in purple, green, pink, cyan and yellow, respectively.*

1.7.2. Anterograde transport

One of the first reports to highlight the export function of the Mla system was made by Kamischke and co-workers on the pathogenic bacteria *A. baumannii* (Kamischke et al., 2019). Using stable ^{13}C-isotope labeling, they were able to monitor the movement of PL from IM to OM. This work was further supported by the findings of Hughes and co-workers, who utilized various biophysical and biochemical approaches to establish that the Mla system is involved in the anterograde movement of PLs in *E. coli*. PL transfer assays conducted by using PL free/bound *Ec*MlaC/MlaD clearly showed that PL transfer from *Ec*MlaD-PL to *Ec*MlaC(Apo) can occur but not the other way around. By utilizing the quartz crystal microbalance with dissipation monitoring (QCM–D) technique, they were able to demonstrate the transfer of PL from the MleFEDB complex to MlaC, thus, confirming the retrograde movement (Hughes et al., 2019). The anterograde movement was further supported by the structural studies on the MlaFEDB complex from *E. coli*. The cryo-EM structure of the MlaFEDB complex reveals a slanted confirmation of MlaD hexamer along with the presence of two PLs in different stages of export in the translocation pathway. As per the report, such a conformation is possible when the PLs are to be extracted from the IM complex to MlaC. This, in turn, suggests that the Mla system would be involved in the anterograde movement of PLs (Coudray et al., 2020). The anterograde transport model was further supported by Mann and co-workers, who suggested the central translocation pathway of the MlaFEDB complex to be extremely flexible and accommodating for PLs in *A. baumannii*. The conformational dynamicity would aid in the transport of PL to MlaC from the MlaFEDB complex (Mann et al., 2021).

1.7.3. Bidirectional/ambiguous transport

Interestingly, there are also reports that have either highlighted the bidirectional movement of PLs or have maintained ambiguity about the directionality of transport. Owing to the interaction of MlaC with MlaA-OmpF and MlaFEDB obtained through interferometry, it was suggested that the Mla system might be involved in the bidirectional movement of PL (Ekiert et al., 2017). The experiments involving ^{14}C-labeled PL tracing reveal that the transfer of PL from MlaC and to the MlaFEDB complex reconstituted into nanodiscs takes place spontaneously and can be enhanced with ATP hydrolysis. However, the findings also highlight that MlaC can receive PL from the MlaFEDB complex, though ATP hydrolysis disrupts such transport (Low et al., 2021). Based on extensive biochemical and structural studies, Chi and co-workers

reported that the MlaFEDB complex can release PL via an extrusion mechanism involving the conformational changes of the central translocation pathway. However, in the proposed mechanism, they maintained a diplomatic stand when it came to addressing the directionality of PL transport by the Mla system (Chi et al., 2020).

1.7.4. Auxiliary function

Interestingly, the Mla system in different Gram-negative bacteria is involved in auxiliary functions. The Mla system is critical for the formation of outer membrane vesicles (OMV) in *Haemophilus influenzae* and *Vibrio cholerae*. The disruption of the Mla system results in the increased production of OMV in both human pathogens. The OMV formation is crucial for relieving the selective pressure of the local defense mechanism during the initial colonization of pathogens in the host body, thereby facilitating proliferation (Roier et al., 2016). In addition to this, the ortholog of the Mla system can even participate in the acquisition of nutrients that are essential for survival. This is evidenced by the existence of the Lin transporter system in *Sphingobium japonicum*. Like the Mla system, the Lin system also contains an IM-associated periplasmic protein containing MlaD/Mce domain (LinM), an IM translocate (LinK), an ATPase (LinL) as well as a lipoprotein (LinN). However, the system lacks MlaC and MlaB counterparts. Unlike the Mla system, the Lin system is involved in the internalization of insecticide γ-hexachlorocyclohexane (γ-HCH) (Nagata et al., 2007).

1.8. Significance of the study

The asymmetric OM of Gram-negative bacteria serves as a robust permeability barrier for antimicrobial compounds to breach. As a result, antibiotics are unable to penetrate the bacterial envelope and enter the bacterial cell. This has contributed significantly to the emergence of multi-drug resistance (MDR) and the global health crisis. Thus, systemic disruption of the OM integrity would result in compromised barrier function and ensure proper penetration of antibiotics. As the Mla system is dedicated to safeguarding the PL asymmetry of the OM, it has been recognized as an excellent drug target. This is further supported by the findings that the mutation of Mla components results in OM defects, reduced permeability and increased susceptibility. An additional advantage of targeting the Mla system is its transmembrane localization. This would pose less of a challenge in terms of accessibility as compared to any cytoplasmic protein/system. Also, aiming the Mla system might enhance the efficacy

of previously ineffective antibiotics that attack cytoplasmic components but enable them to penetrate the OM barrier. In spite of being such a promising drug target, mechanistic insights into the functioning of the Mla system are still scarce. Added to that, unlike typical ABC transporters, most of the components of the system remain unplaced in various classification schemes owing to their atypical features. Hence, in order to characterize all the components and acquire the molecular picture of PL transport, in this study, the Mla system of *E. coli* was extensively studied. Structural data obtained from the study would be used to understand the architecture of the protein, especially the binding pocket that would lay down the foundation stone for structure-based drug designing for the development of inhibitor molecules. Thus, the outcomes would help in the development of targeted therapeutics and expand the horizon for combating infections caused by other Gram-negative bacteria. The mechanistic characterization of the proteins would help in the development of working models that would help understand the directionality of transport of the Mla system. Furthermore, the information obtained would also contribute to the understanding of ABC transporter-aided import and export mechanisms.

1.9. Objectives

To get insight into the PL transport mechanism of the Mla system, the following objectives were designed:

1. *In silico* characterization of Mla proteins from *E. coli*.

2. Molecular cloning, overexpression and purification of MlaC and MlaD proteins from *Escherichia coli* to homogeneity.

3. Crystallization of the purified proteins to elucidate their three-dimensional structures.

4. Biophysical and mechanistic characterization of the proteins.

5. Molecular docking and molecular dynamics (MD) simulation studies of *Ec*MlaC protein.

Chapter 2

Materials and methods

भारतीय प्रौद्योगिकी संस्थान गुवाहाटी
Indian Institute of Technology Guwahati

2. Materials and methods

The general outline of Chapter 2

This chapter describes the different methodologies and experimental designs used for the structural and functional characterization of the Mla proteins from *E. coli*. A general work plan representing the experimental outline is provided below in Figure 2.1.

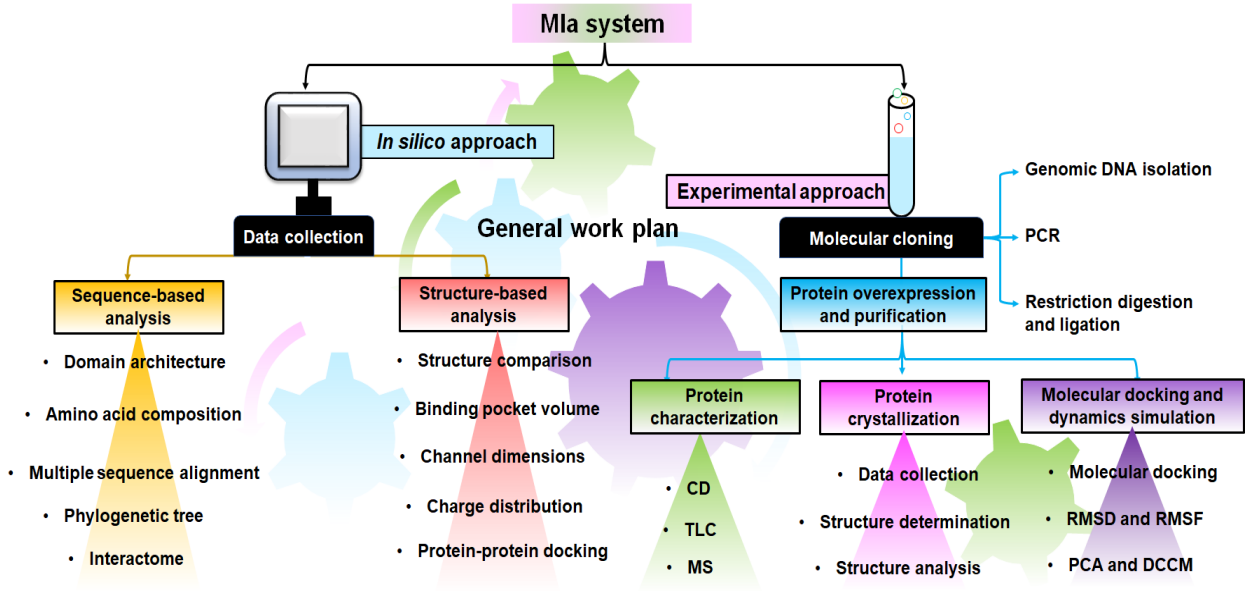


Figure 2.1. *A general work plan of the methodology used to accomplish the objectives.*

2.1. MATERIALS

2.1.1. Reagents

The various enzymes used for molecular cloning, such as restriction enzymes, shrimp alkaline phosphatase, T4 DNA ligase, etc., were obtained from New England Biolabs (NEB, USA). The different kits used for the plasmid isolation, PCR clean-up and gel extraction used for recombinant cloning were procured from QIAGEN (Germany), HiMedia (India) and GCC Biotech (India). The Ni^{2+}-NTA affinity resin and pierce centrifuge columns used for the purification of the proteins were procured from QIAGEN (Germany) and Thermo Fisher Scientific (USA), respectively. The Amicon Ultra-15 and Vivaspin Turbo 15 protein concentrators were purchased from Merck (Germany) and Sartorius (Germany), respectively, for the concentration of purified

proteins. The various crystallization buffers, microbatch and hanging-drop plates and other protein crystallization-related chemicals and tools were procured from Hampton Research (USA) and Molecular Dimensions (UK). In addition to this, various chemicals used for the preparation of buffers were obtained from Merck (Germany), HiMedia (India) and Sisco Research Laboratories (India).

2.1.2. Phospholipid (PL)

The PL, L-α-Phosphatidylethanolamine, dioleoyl was procured from Merck (Germany).

2.2. METHODS

Overall, the work was conducted via two main approaches, *in silico* and experimental.

2.2.1. *In silico* approach

2.2.1.1. Sequence-based analysis

2.2.1.1.1. Retrieval of sequences

The nucleotide sequences of the genes of interest were obtained from KEGG and the Gene database of the National Centre for Biotechnology (NCBI) (Kanehisa and Goto, 2000). The information related to the operonic arrangement was retrieved from the Gene database of NCBI, EcoCyc database and MycoBrowser (Kapopoulou et al., 2011; Keseler et al., 2017). The protein sequences and their homologs were retrieved from the UniProtKB database (The UniProt Consortium, 2023).

2.2.1.1.2. Analysis of domain architecture and amino acid composition

Domains are structural, functional and evolutionary units of proteins. Domain architecture is defined as the arrangement of its constituent domains in the protein sequence. In order to study the architectural diversity of the domains present in the Mla proteins and to classify them, two databases, viz. Pfam and InterPro were used (El-Gebali et al., 2019; Mitchell et al., 2019; Mistry et al., 2021). The protein domains often have unique chemical properties owing to their amino acid profiles which in turn determine their biological functions. Hence, the amino acid composition of the domains of interest belonging to different architectures was determined by using ProtParam (Gasteiger et al., 2005). The profiles obtained were visualized as heat maps generated using the statistical package RStudio (RStudio Team, 2015).

2.2.1.1.3. Sequence alignment

In order to study the conservation at the residue level, the retrieved sequences were subjected to pairwise as well as multiple sequence alignments (MSA). For pair-wise sequence alignment, the web tool BLAST with the default set of parameters was used (Altschul et al., 1990). The MSAs were performed using the program Clustal Omega with the default set of parameters (Sievers and Higgins, 2014). The program utilizes a combination of seeded guide trees and hidden Markov model (HMM) iterations to build alignments between the multiple sequences. For better visualization and clarity, the outputs were rendered using the web tool ESPript 3.0 (Gouet et al., 2003). The consensus motif(s) present was determined using the consensus sequence tool embedded in the web tool ESPript followed by its visualization using the web server WebLogo and the program ConSurf (Crooks et al., 2004; Ashkenazy et al., 2016).

2.2.1.1.4. Phylogenetic tree

To determine the evolutionary relationship among the homologous protein/domain sequences and to understand their ancestry, phylogenetic tree analysis was performed. The program MEGA version 7.0 was used to generate the phylogenetic trees (Molecular Evolution and Genetic Analysis, Kumar et al., 2016). Construction of the phylogenetic trees involved the following three steps: (1) Alignment of the retrieved sequences with the default set of parameters using the program ClustalW embedded in the software MEGA 7.0. (2) Generation of a tree using the Maximum-likelihood (ML) method. The significance of the tree was evaluated using the bootstrapping method, which defines the confidence level of similarity between the clusters formed. (3) The generated trees were further rendered for clear clustering and to obtain relevant information about the evolutionary relationship.

2.2.1.1.5. Interactome analysis

Mla proteins are involved in coordinated interaction among one another in order to perform the task of PL transport between the IM and the OM. These protein-protein interactions are highly specific transient physical contacts that are critical for PL transport. In order to identify such molecular interactions, including the identification of domain-domain interaction and proteins belonging to different systems yet aiding in PL transport, an interactome analysis was performed using the STRING database, followed by their visualization using the tool Cytoscape 3.8.1 (Shannon et al., 2003; Szklarczyk et al., 2017).

2.2.1.2. Structure-based analysis

2.2.1.2.1. Data retrieval and structural comparison

Crystal and cryo-EM structures of the proteins of interest were obtained from the Protein Data Bank (PDB) (Berman et al., 2000). The modeled three-dimensional structures of proteins were obtained from the AlphaFold database (Jumper et al., 2021; Varadi et al., 2022). The structural comparisons, including the estimation of root mean square deviation (RMSD) of the protein structures, were carried out using the software 3dSS (Sumathi et al., 2006) and PyMOL (The PyMOL Molecular Graphics System, Schrodinger, LLC).

2.2.1.2.2. Analysis of binding-pocket and charge distribution

The binding-pocket volumes of proteins of interest were calculated by using the program CASTp with the default probe radius of 1.4 Å (Tian et al., 2018). In the case of channel-forming proteins, the channel dimensions were determined using the plugin CAVER 3.0, available in PyMOL with the default set of parameters (Chovancova et al., 2012). In order to study the composition of binding pockets and protein surfaces, the YRB colour scheme was used. The script colors the structures at the atomic level wherein the carbon atoms not bound to nitrogen or oxygen atoms are highlighted in yellow, oxygen atoms carrying the negative charges in glutamate and aspartate residues are coloured in red and nitrogen atoms carrying the positive charges in lysine and arginine residues are coloured in blue. The remaining atoms are displayed in white (Hagemans et al., 2015). The YRB script is available in the public domain and runs using PyMOL.

2.2.1.2.3. Protein-protein docking

For protein-protein docking, the freely available online docking tool ClusPro was used (Kozakov et al., 2017). In ClusPro, one protein molecule was submitted as a receptor while another protein molecule was submitted as a ligand by keeping the default set of parameters provided by the program. From the multiple protein-protein complexes predicted, the balanced results (providing 10 clusters) were considered, which were further analyzed so as to obtain the best model. The atomic interactions between the complexes were identified using the webserver PDBsum (Joosten et al., 2010; Laskowski et al., 2018). The molecular interactions of docked ligands and other structural figures were prepared using the program PyMOL.

2.2.2. Experimental approach

2.2.2.1. Construction of recombinant protein overexpression

Molecular cloning is a collection of experimental procedures used to isolate and expand a DNA fragment of interest from a genome into a host organism in order to create many identical copies. In this study, the molecular cloning method was used for the creation of the recombinant construct. Hence, the genome of interest was first isolated and the target gene from it was amplified and placed into a proper vector for propagation. In the presence of the ligase enzyme, the vector and gene were combined and injected into the host organism, whereupon the planned construct underwent replication leading to the amplification of the desired recombinant clone (Figure 2.2). Details of each step of molecular cloning performed in this study are described below.

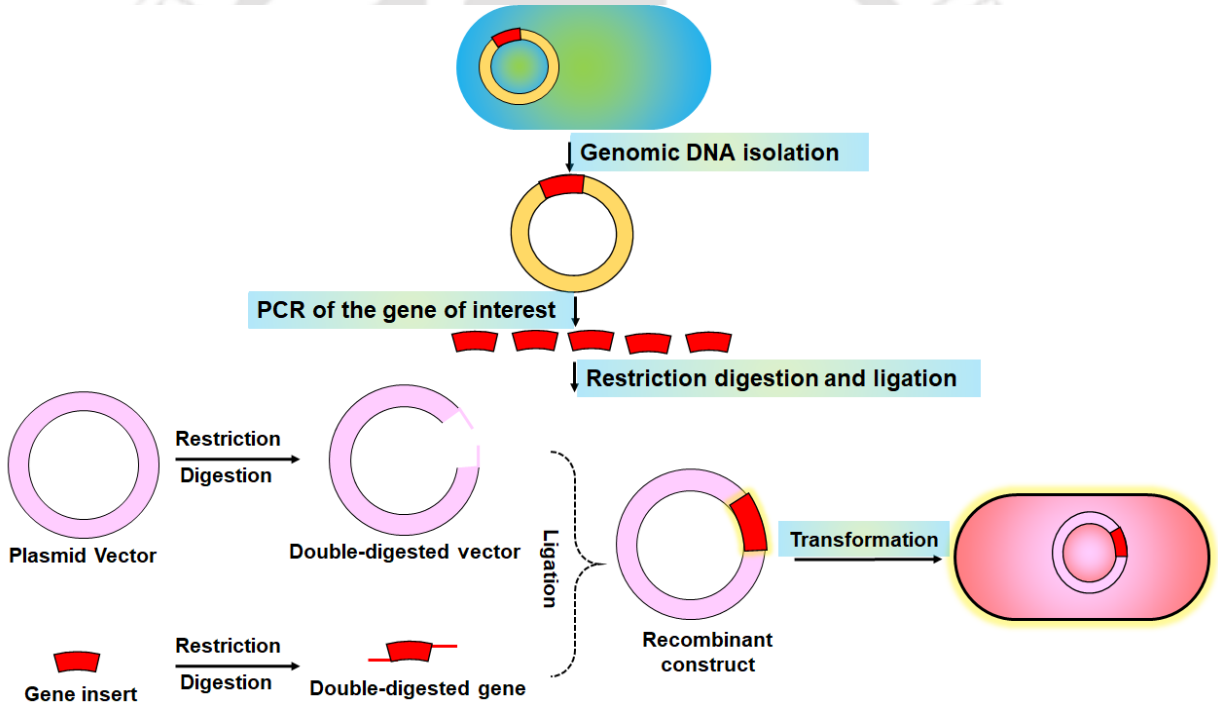


Figure 2.2. *Schematic representation for molecular cloning.*

2.2.2.2. Preparation of competent cells for gene cloning and protein overexpression

Using a calcium chloride ($CaCl_2$)-based chemical technique, the competent cells for several strains of *E. coli* required for molecular cloning and protein overexpression

were created. By re-suspending the bacterial cells in a 100 mM $CaCl_2$ solution during their log phase, followed by a 30-minute ice incubation at 4°C, the bacterial cells were made competent. The competent cells were then flash-frozen in liquid nitrogen, re-suspended in an autoclaved solution of 100 mM $CaCl_2$ and 15% glycerol and stored at -80°C. For increasing the cloned plasmid copy number, *E. coli* DH5α competent cells were used as they have high transformation efficiency. On the other hand, BL21(DE3) competent cells were used for the overexpression of target proteins. These cells possess a chromosomal copy of the T7 RNA polymerase for expression of the protein under the control of the T7 promoter. Overexpression of the targeted proteins was achieved by inducing the expression cells with an analog of lactose, i.e. Isopropyl β-D-1-thiogalactopyranoside (IPTG), which binds to the repressor.

2.2.2.3. Gene amplification

For the amplification of the genes of interest, gene-specific customized primers were created by taking into consideration factors like GC content, melting temperature (Tm), restriction digesting sites, etc. Furthermore, the 6x-His-tag required for Ni-NTA affinity chromatography-based protein purification at the N-terminal was incorporated in the primers. To evaluate and confirm all the parameters, the primers were examined using IDT Technologies' web application OligoAnalyzer. For amplification of the genes of interest, the genome from *E. coli* BL21(DE3) competent cells were isolated. Using this genome as the template and the gene specific primers, the genes were amplified by polymerase chain reaction (PCR). With the help of primers, any N-terminal signal peptide sequence or N-terminal membrane-embedded helical region was removed from the genes. The PCR reaction was performed in the following steps: 1) initial denaturation (95°C, 5-10 min), 2) final denaturation (95°C, 30-40 sec), 3) annealing (Tm of genes, 45 sec), 4) initial extension (72°C, 50 sec) and 5) final extension (72°C, 10 min).

2.2.2.4. Double digestion of the gene of interest and vector

In order to execute directional cloning, double digestion of the amplified genes was performed using the respective restriction enzymes before inserting them into the pET22b plasmid vector. The restriction enzymes *Nde*I and *Xho*I were chosen, which generate the 5' and 3' sticky ends in both the genes and the vector. For restriction digestion, the genes were added with the required restriction enzymes and allowed to incubate at 37°C for 3 hours. Simultaneously, the pET22b plasmid vector was

incubated with the same restriction enzymes and incubated at 37°C for 3 hours. After incubation, the enzyme alkaline phosphatase was added to the vector reaction mixture and further incubated for 1 hour at 37°C to remove the 5'-phosphate group. The double-digested genes and vector were purified using the PCR purification kit (QIAGEN) and gel extraction kit (QIAGEN), respectively.

2.2.2.5. Ligation of the gene of interest to vector and clone confirmation

In order to obtain the recombinant plasmid, target genes were inserted into a digested pET22b vector. The T4 DNA ligase enzyme, which creates a covalent link between the two DNA pieces, was used to accelerate the ligation reaction. For ligation, double digested vector and genes with the sticky ends were combined in a variety of ratios (1:1, 1:2, 1:3, 1:4, and 1:5) and incubated at 22°C for 4 hours with T4 DNA ligase. Each ligation mixture was first incubated and the *E. coli* DH5α competent cells were transformed with the mixture in order to propagate the ligated recombinant clone. By growing the transformed cells on Luria Bertani (LB) agar plates supplemented with ampicillin, positive colonies were identified. The fully-grown colonies were inoculated in LB broth enriched with ampicillin and then subjected to plasmid isolation for further confirmation. To validate the presence of the desired genes after plasmid isolation, the expected cloned plasmids were treated with the appropriate restriction enzymes and then incubated at 37°C for 3 hours. Following validation by double digestion, the clones were further confirmed by DNA sequencing.

2.2.2.6. Protein overexpression

The genes cloned in the pET22b vector were overexpressed by transforming *E. coli* BL21 (DE3) expression host cells with the recombinant plasmid. The overexpression was optimized at 37ºC for different time intervals (1 to 7 hrs) by using different concentrations of IPTG. Since a higher concentration of IPTG leads to cell death or unproductive protein formation, the IPTG concentrations (0.1, 0.5 and 1.0 mM) were also optimized. Once the expression host, optimum temperature, IPTG concentration and expression time were optimized, the overexpressed proteins were checked for solubility.

2.2.2.7. Solubility of proteins

The cells containing the overexpressed proteins of interest were harvested by centrifugation, followed by their resuspension in a lysis buffer for homogenization. In

order to disrupt the cells, the homogenized suspension was then sonicated at 33% amplitude for 15 min, maintaining the sonication cycle at 2 s on and 10 s off. The cell lysate was then centrifuged at a maximum speed of 15,644*g* for 40 min to segregate the pellet and supernatant fractions. The solubility of the protein was analyzed by running the obtained pellet and supernatant fractions in 12% SDS-PAGE. Once the solubilized proteins were identified in the supernatant fraction, they were further subjected to protein purification by using chromatography techniques.

2.2.2.8. Affinity chromatography for protein purification

In order to purify the proteins of interest, the immobilized metal affinity chromatography (IMAC) technique was used. The N-terminals of the recombinant proteins were tagged with a 6X-His because IMAC follows the precise coordination geometry between a transition metal ion (Co^{2+}, Ni^{2+}, Cu^{2+}, or Zn^{2+}) and polyhistidine residues, which coordinate with immobilized Ni^{2+} ions on a matrix. The piercing centrifuge columns were filled with Ni-NTA resins which are equilibrated using the required buffers. The packed and equilibrated Ni-NTA resins were subjected to a 2 h incubation period with the supernatant fraction containing the protein of interest. In order to remove the weakly bound non-specific proteins, the column was washed with buffers containing different concentrations of imidazole (10-20 mM). In order to reduce the non-specific binding of host proteins, imidazole was combined with various additives such as salt (KCl), a reducing agent (β-mercaptoethanol), and a protease inhibitor, phenylmethanesulphonyl fluoride (PMSF). In order to elute the protein of interest, a higher concentration of imidazole (250-400 mM) was used in the elution buffer. This was done so as to break the coordination bond between the histidine residues (tagged in the recombinant protein) and the Ni^{2+} ions (present in the resin). The imidazole molecules competitively bind to the Ni^{2+} ion and dislodge the tagged protein bound to the resin. The purity of the eluted fraction was then evaluated by running the samples in 12% SDS-PAGE. To completely remove the imidazole, gradient dialysis was performed after pooling up the purified elutes. The dialyzed proteins were then concentrated using protein concentrators up to the desired concentration so as to perform biophysical and biochemical experiments.

2.2.2.9. Protein characterization

2.2.2.9.1. Circular dichroism (CD) spectroscopy

In order to determine the secondary structural content and the thermal stability of the proteins, the biophysical method of CD spectroscopy was used. In the wavelength range of 260 to 190 nm, CD offers the secondary structural data of proteins, including α-helix, parallel and anti-parallel β-sheet, turns and random coil. In this study, secondary structure makeup and unfolding thermodynamics of the chosen proteins were observed. As a function of temperature, the secondary structural composition of proteins was examined for α-helices and β-sheets. The melting temperature (Tm) of the targeted SBPs is calculated using the thermodynamics of protein unfolding as a function of temperature. For the chosen proteins, thermal stability experiments were conducted in a temperature range of 20 to 90ºC.

2.2.2.9.2. Thin layer chromatography (TLC)

TLC is a chromatography technique used for the separation of lipids from a mixture based on the polarity of the components. In order to observe the presence of any bound PL, a lipid extraction assay of the protein of interest was first performed following the Bligh-Dyer method (Bligh and Dyer, 1959, Thong et al., 2016). For lipid extraction, the purified protein sample was mixed with chloroform:methanol (1:2 vol/vol). The mixture was vortexed and sonicated for 30 s three times, followed by centrifugation at 21,000*g* for 5 min. The supernatant was collected and an appropriate amount of chloroform and buffer were added to form two-phase mixture of chloroform:methanol:water. In order to separate the organic and aqueous phases, the mixture was then centrifuged at 4,000*g* for 5 min. After centrifugation, the organic phase was gently removed and the organic solvent in it was evaporated at room temperature. The dried lipids that remained were resuspended in chloroform:methanol (4:1 v/v). The sample was spotted in a TLC plate which was placed in the solvent system (65:25:4 chloroform:methanol:water) designed to separate PLs based on head group polarity. The plate was allowed to remain in the solvent system until the solvent font moved to a required distance. After that, the plate was dried at room temperature and stained with iodine vapor.

2.2.2.9.3. Mass spectrometric analysis

Matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) was used for the determination of the molecular weight of protein. The purified protein samples were mixed with the matrix in a (protein:matrix) ratio of 1:2, 1:3 and 1:5. As the theoretical molecular weight of the selected proteins are more than 10 kDa, sinapinic acid was

used for analysis. The matrix was prepared in acetonitrile and 0.1% trifluoroacetic acid in a ratio of 30:70, respectively, to a concentration of 10 mg ml^{-1}. Sinapinic acid was dissolved in this solvent by sonicating the solution for 30 minutes and centrifuging it at 12000*g* for 10 minutes to get rid of the extra matrix crystals. The samples were also prepared with varying concentrations of protein (1, 5 and 10 mg ml^{-1}). Spots of 2 µL of the protein-matrix mixture were made on the ground steel MALDI plate grooves and air-dried. After drying, the drops were ionized in a Bruker autoflex speed and the data were plotted using an in-built Flex control analysis software.

2.2.2.10. Protein crystallization

2.2.2.10.1. Crystallization of the selected proteins

For the determination of the three-dimensional structures of the proteins, the technique of X-ray crystallography was used. The purified protein samples were first crystallized and the crystals were then exposed to an X-ray beam to obtain the diffractions. In order to obtain diffractable crystals, the protein solutions were screened with a variety of crystallization screens (buffers) to get an initial crystal hit followed by various optimizations. The process of protein crystallization progresses through four zones, viz. undersaturated, metastable, labile and precipitation. Crystal formation occurs at a supersaturated stage in a metastable zone (Figure 2.3). This supersaturated state of protein crystallization can be reached by vapor diffusion. This can be done by using two major approaches, microbatch-under-oil and hanging-drop vapor diffusion. For the microbatch-under-oil method, a mixture of the protein and precipitant is submerged under the oil to slow down the evaporation. In this study, a combination of paraffin and silicon oils was used in order to control the rate of diffusion. On the other hand, in the hanging drop method, a drop containing a mixture of the protein solution and precipitant was placed on a coverslip and subjected to equilibration against the reservoir solution in an inverted position. The crystal screening was initially performed using the microbatch-under-oil method at 4 and 20ºC. Once the initial crystal hits were obtained, crystal setup was done using the hanging-drop vapor diffusion method. Additionally, further optimization was done using various strategies such as varying the concentration of protein as well as crystallization conditions (salt, precipitant, etc.), incorporation of additives, microseeding, etc.

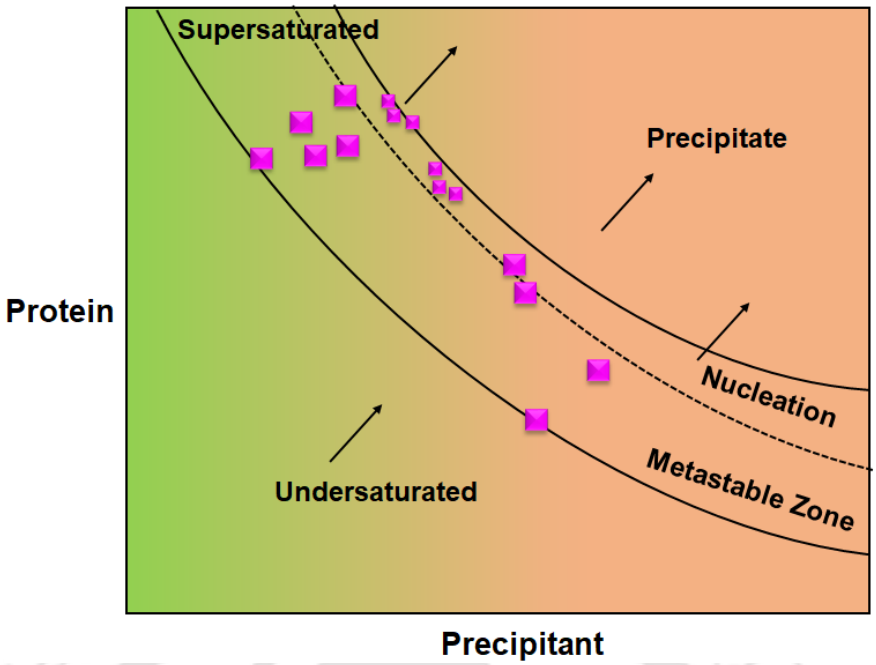


Figure 2.3. *Schematic representation for the phase diagram of protein crystallization. (Image is redrawn from Shaw Stewart et al., 2011).*

2.2.2.10.2. Data collection

After the successful crystallization of the proteins, suitable protein crystals were subjected to a monochromatic beam of X-ray to get the diffraction patterns. Owing to the periodic arrangement of protein molecules in a crystal, exposure to X-rays would result in scattering either in a constructive or destructive way. Bragg`s law defines the relationship between constructive interference and the geometry of crystal parameters. According to Bragg`s law, the X-ray scattering depends on the incident beam wavelength (λ), the glancing angle (2θ) and a positive integer (n).

$$2d\sin\theta = n\lambda$$

In this study, the X-ray intensity data were collected at a fixed wavelength (λ) of the incident X-ray beam (1.5418 Å) and recorded on an image plate (IP) detector. In order to collect data with all possible diffraction peaks, the angle of oscillation for the crystals was kept at 1.0° (1.0 degree per image) during data collection. The desired numbers of images or frames was collected for the complete data set depending on the space group of the crystals. Since the data were collected at a fixed wavelength (λ), other

parameters such as the crystal-to-detector distance (d) in a range of 80-200 mm and exposure time (300-600 s) were varied so as to improve the data collection process.

2.2.2.10.3. Data processing

During data processing, the raw X-ray data is transformed into Bragg reflections with measured intensities. This involves four steps: scaling, integration, merging and calibrating. Different algorithms are derived for each stage and can be found in software packages. In this study, the program iMosflm was used to process the raw diffraction pictures for initial processing and Aimless for scaling and merging (Battye et al., 2011; Evans and Murshudov, 2013). The CCP4 suite includes the embedded software iMosflm. The program iMosflm produces an index for each reflection and preliminary data on unit cell dimensions in a crystallographic symmetry. Data sets were scaled and merged using the program Aimless in order to determine the potential point and space groups. The module ctruncate distributed in the CCP4 package was used to convert the obtained intensities into structure factors (Winn et al., 2011).

2.2.2.10.4. Structure determination

The intensities obtained from the X-ray diffraction methods provide only information about structure amplitudes and do not give the phase angle information. But for the calculation of the electron density, information on phase angles is essential (Taylor, 2010). Various methods can be used to obtain the initial phase information. These include (1) molecular replacement (MR), (2) single- and multiple-wavelength anomalous dispersion (SAD and MAD) and (3) molecular isomorphous replacement (MIR). In this study, the MR method was used (Evans and McCoy, 2008). As per this method, a known molecular structure of the homologous protein is used to solve the structure of a new target protein using the program Phaser (McCoy et al., 2007). The program uses multivariate statistics and maximum likelihood probability theory. A significant similarity (>25%) between the known homologous protein and the targeted unknown protein is required in order to obtain the initial phase by MR. The obtained information from the model structure was further utilized to calculate the phases from the target protein intensities.

2.2.2.10.5. Model building and structure refinement

After obtaining initial phase angles from the target protein reflections, an electron density map was calculated. This density map helps to build a model structure by

tracing and modeling the secondary structural elements using the program Coot (Crystallographic Object-Oriented Toolkit, Emsley et al., 2010). During the process of model building, various physicochemical parameters, such as bond lengths, bond angles, torsion angles,atomic overlaps, etc. of the model were analyzed using the programs such as PDB Goodies (Hussain et al., 2002) and PSAP (Protein Structure Analysis Package, Balamurugan et al., 2007).

In this study, the program Refmac5 was used for model refinement (Vagin et al., 2004). The program utilizes Bayesian statistics and maximum likelihood probability. Model refinement results in the improvement of the agreement between the calculated structure factors from the model and observed structure factors from the experiment.

Following each cycle of refinement, an agreement between the calculated structure factors (F_{calc}) and the measured structure factors (F_{obs}) was evaluated. This agreement is expressed as:

$$R = (\Sigma(F_{obs} - F_{calc})/(\Sigma F_{obs})$$

Following refinement, the R-factor values and electron-density maps (*2Fo-Fc* and *Fo-Fc*) contoured at 1.0 σ and 3.0 σ, respectively, were examined to confirm the improvement in the model.

2.2.2.10.6. Cross-validation

The quality of the model was assessed using statistical cross-validation and the R-factor (Brunger, 1992). The pre-refinement diffraction data were split into two groups for cross-validation. While 5-10% of the total data set was discarded as a small complementary test set, the remaining 90-95% of the diffraction data were employed as a sizable working set during the refinement process. The accuracy of the predicted model was determined by calculating the R-free value using the remaining 10% test data set.

$$R_{free} = \frac{\sum_{hkl \subset T} ||F_{obs}| - k|F_{calc}||}{\sum_{hkl \subset T} |F_{obs}|}$$

Where hkl⊂T represents all reflections belonging to test set *T* of unique reflections.

2.2.2.10.7. Structure validation

The reliability of the three-dimensional structures of the proteins was assessed by the structure validation process. A structural validation procedure typically consists of three steps: 1) validity of the experimental measurements, 2) consistency in the atomic model and 3) consistency of the known chemical and physical properties of the atomic model. For structure validation of the selected proteins, the following aspects have been considered in this study: (1) geometrical parameters and (2) root mean square deviation (RMSD) of the refined structures. The assessment of the structures was done using the online servers PROCHECK (Laskowski et al., 1993) and MolProbity (Chen et al., 2010).

2.2.2.10.8. PROCHECK and MolProbity

The stereochemical quality of the model structure was evaluated using PROCHECK after each refinement cycle (Laskowski et al., 1993). The program verifies the geometry of the model structure residue-by-residue. The following details are provided by PROCHECK- Ramachandran plot, glycine & proline Ramachandran and Chi1-Chi2 plots, main-chain and side-chain parameters, residue properties, main-chain bond length and angle distributions and RMS distances from planarity and distorted geometry plots. The geometry of the chosen proteins was estimated using the Ramachandran plot.

The web server MolProbity is utilized for structural checking (Davis et al., 2007). This program assesses the three-dimensional structures of proteins both at the global and local levels. MolProbity optimizes both polar and nonpolar hydrogens, all-atom contact analysis, covalent geometry, and torsion angles of the protein for a given crystal structure. It also enhances the local chemical environment due to the resolution of structures, which include Ramachandran outliers and flipped-branched protein side chains. In this study, MolProbity was utilized to find the Ramachandran and torsion-angle outliers.

2.2.2.10.9. Structure analysis

Crystal and cryo-EM structures of the proteins of interest were obtained from PDB. The modeled three-dimensional structures of proteins were obtained from the AlphaFold database. The structure-based homology search for proteins was performed using the web server Dali (Holm, 2020). The structural comparisons, including the determination of root mean square deviation (RMSD) of the protein structures, were carried out using

the software 3dSS and PyMOL. The secondary structural contents of the analyzed structures were determined using the web servers DSSP and topology diagrams were generated using TopDraw embedded in the CCP4 package (Kabsch and Sander, 1983; Bond, 2003). The domain arrangement in protein structures and their ancestral analysis were performed using the CATH Protein Structure Classification Database and Structural Classification of Proteins (SCOP) database, respectively (Dawson et al., 2017; Andreeva et al., 2020). The binding-pocket volumes of proteins of interest were calculated by using the program CASTp with a default probe radius of 1.4 Å. In the case of channel-forming proteins, the channel dimensions were determined using the plugin CAVER 3.0, available in PyMOL with the default set of parameters. The surface areas and analysis of oligomeric interfaces were performed using the web server PDBePISA with the default set of parameters (Krissinel and Henrick, 2007). In order to study the composition of binding pockets and protein surfaces, the YRB colour scheme was used. In order to perform the sequence and structure-based alignments to generate informative multiple alignments that are consistent with both sequence and structural information, the online program PROMALS3D was used (Pei et al., 2008). The orientation of proteins was determined by using the webserver PPM v2.0 and TMHMM (Krogh et al., 2001; Lomize et al., 2012). The programs Coot and PDBsum were used to investigate the intermolecular interactions between ligands and proteins. The programs PyMol, Chimera and Visual molecular dynamics (VMD) were used for the visualization of the three-dimensional protein structures (Humphrey et al., 1996; Pettersen et al., 2004). These programs were also used for the generation of molecular graphic depictions.

2.2.2.11. Molecular docking and dynamics simulation

2.2.2.11.1. Molecular docking

Molecular docking is a computer-based method that is utilized to make predictions about preferred orientations, affinities and interactions of a ligand with the binding site of a protein (Tiwari and Singh, 2022)**.** The binding affinities of different ligands to the protein of interest were estimated by molecular docking analysis performed using the freely available program AutoDock v4.2 (Morris et al., 2009). The three-dimensional atomic coordinates of the ligands were retrieved from their protein-bound structures available at PDB. Where applicable, the hydrogen atoms were added to the protein

and ligand prior to each docking experiment using the AutoDock module. The Gasteiger charge algorithm was then used to give the partial charges to the protein atoms (Gasteiger and Marsili, 1980). To perform the blind docking of the ligands with the protein, a grid size of 126x126x126 with a grid spacing of 0.375 Å was generated by taking the center of mass of the targeted protein as the grid center. By maintaining the flexibility of the ligand atoms relative to the rotatable bonds while maintaining the rigidity of the protein atoms, rigid molecular docking was achieved. A total of 2000 runs of the Lamarckian genetic algorithm (LGA) were set up for each molecular docking experiment. The docked ligand conformations were clustered with an RMSD cut-off of 2.0 Å. The atomic interactions between the protein and the docked ligand(s) were identified using the program Coot and PDBsum. The program PyMOL was used to prepare the molecular interactions of docked ligands and other structural figures.

2.2.2.11.2. Molecular dynamics (MD) simulation

MD is a computer-based simulation method used for the identification and analysis of protein motions and other biomolecules over time. The simulations capture the physical behavior of proteins in extensive atomic detail which help understand their conformational landscape and dynamicity (Hollingsworth and Dror, 2018). It is based on the principles of classical mechanics and thermodynamics and involves simulating the motions of the individual atoms and molecules in a system as they interact with each other and with their environment. The method incorporates a series of steps, which involves the preparation of a molecular system where the topological parameters of a protein are generated from the PDB file. This system is then put into a definitive box where water or any other solvent is added. The system is then energetically minimized to remove any steric clashes and reduce the potential energy. This is followed by the equilibration of the system in order to reach a stable state at constant temperature and pressure. Once it is equilibrated, the production phase or MD run begins for a longer period of time to gather data and study the behaviour of the system (GROMACS Documentation, 2021.1). In this study, MD simulations were performed using the package GROMACS 2021.1 (Abraham et al., 2015). The experimental setup can be divided into two different categories. These included simulations of the systems in the (A) ligand-free and (B) ligand-bound states. The simulations of both systems were performed using the force field Amber ff99SB embedded in GROMACS (Ponder and Case, 2003; Hornak et al., 2006). The prepared systems were solvated with the TIP3P water model, followed by the addition of chloride and sodium ions to neutralize

the overall charge of the systems (Price and Brooks, 2004). The production run for the systems was 1 µs. MD trajectories were analyzed using in-build GROMACS modules and Bio3D library in R-packages. Structural visualization of simulation trajectories was analyzed using VMD software and PyMOL.

Chapter 3

In silico characterization of Mla proteins from *Escherichia coli*

A part of this chapter has been published as:

1. **Dutta A**, Chandravanshi M and Kanaujia SP (2021). Conserved features of the MlaD domain aid the trafficking of hydrophobic molecules. *Proteins* 89(11), 1473-1488.

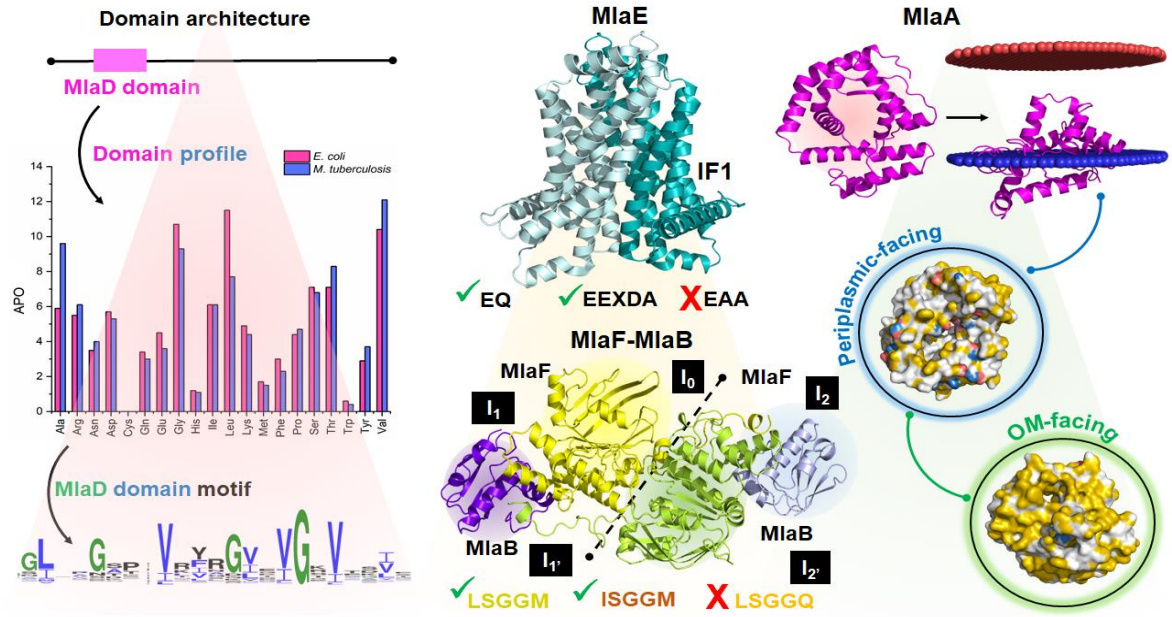
2. **Dutta A** and Kanaujia SP. *In silico* characterization of Mla proteins. ***Under preparation***.

Abstract

In Gram-negative bacteria, the maintenance of the lipid asymmetry (Mla) system is involved in the transport of phospholipids (PLs) between the inner (IM) and outer (OM) membranes. The system represents a unique ATP-binding cassette (ABC) transporter owing to the presence of an auxiliary protein. The system utilizes seven proteins that give rise to three sub-cellular complexes – lipoprotein MlaA-OmpC/F (OM), MlaC (periplasmic) and MlaFEDB complex (IM). Among them, MlaD presents a unique case as it is a unique IM-associated periplasmic solute-binding protein (SBP), which possesses a conserved domain, the MlaD domain. While proteins carrying the MlaD domain are known to be primarily involved in the trafficking of hydrophobic molecules, not much is known about this domain itself. On the other hand, MlaB, which remains associated with MlaF, contains a Sulphate Transporter and AntiSigma factor antagonist (STAS) domain. Such an arrangement is not observed in the case of other ABC transporters. Furthermore, ABC transporters are not dependent on OM lipoprotein, as observed in the case of the Mla system. In addition to that, the remaining components of the Mla system remain poorly characterized. Thus, in this study, all the components of the Mla system from *E. coli* have been characterized using *in silico* approaches. The profiling of the MlaD domain of different architectures reveals the abundance of glycine and hydrophobic residues and the lack of cysteine residues. The domain possesses a conserved N-terminal region and a well-preserved glycine residue that constitutes a consensus motif across different architectures. Phylogenetic analysis shows that the MlaD domain archetypes are evolutionarily closer and marked by the conservation of a functionally crucial pore loop located at the C-terminal region. The study also establishes the critical role of the domain-associated permeases and the driving forces governing the transport of hydrophobic molecules. This sheds sufficient light on the structure-function-evolutionary relationship of the MlaD domain. The

hexameric interface analysis reveals that the MlaD domain itself is not the sole player in the oligomerization of the proteins. Further, an operonic and interactome map analysis reveals that the Mla and the Mce systems are dependent on the structural homologs of the nuclear transport factor 2 (NTF2) superfamily. The study also reveals that MlaE and its orthologues possess atypical TMD characteristics with unique ancestry along with the presence of an EQ loop. The same can be said about MlaF and its orthologues which tend to have originated from a separate lineage and consist of variable signature motifs. The study also shows that the functional association between MlaF and MlaB gives rise to additional interfaces which are critical for the proper functioning of the Mla system. On the other hand, the structural peculiarities of MlaA aid in the transport of PLs through it. The protein defies typical OM integral protein as well as lipoprotein structures. The study highlights that MlaA is extremely flexible, which would provide the force for PL transport without the utilization of ATP hydrolysis. The central channel of the protein is extremely dynamic, which would provide the ideal ambience for the accommodation of transporting PL. Also, MlaA would remain localized in the OM via interactions with OmpC/F. The orientation of MlaA would also be affected by the presence of the C-terminal helical region. Altogether, the analysis firmly establishes the Mla system to be a unique ABC transporter with atypical characteristics.

Graphical abstract



3.1. Introduction

Gram-negative bacteria are generally more resilient in comparison to Gram-positive bacteria due to the presence of an outer membrane (OM) (Delcour, 2009). The OM possesses a sophisticated organization with lipopolysaccharides (LPSs) and phospholipids (PLs) in its outer and inner leaflets, respectively. The asymmetric nature of the OM acts as a protective shield against the assaults of a number of toxic chemicals, such as antibiotics and detergents (Clifton et al., 2013). However, the PLs present in the inner leaflet of the OM has the tendency to flip and accumulate in its outer leaflet leading to the disruption of the membrane integrity (Henderson et al., 2016). In order to restore and maintain the asymmetry of the OM, bacteria have developed a number of strategies (Henderson et al., 2016; Bishop, 2008); one such strategy involves the utilization of a multi-component machinery known as the maintenance of lipid asymmetry (Mla) system (Henderson et al., 2016).

The Mla system was first reported in *Escherichia coli* and is a highly conserved intermembrane machinery that consists of three components, an OM-associated lipoprotein (MlaA), a free-floating periplasmic protein (MlaC) and an inner membrane

(IM) ATP-binding cassette (ABC) transporter complex (MlaFEDB) (Malinverni and Silhavy, 2009). Recent studies have suggested that MlaA would remain associated with OmpC/F forming a stable complex (Chong et al., 2015). It has been proposed that the PLs accumulated in the OM are supposedly extracted by MlaA and are handed over to MlaC via a unique retrograde movement and are further handed over to the IM complex supposedly for internalization (Henderson et al., 2016; Malinverni and Silhavy, 2009; Ekiert et al., 2017). However, recent studies have suggested that the Mla system might be involved in an anterograde movement where the PLs are transported to the OM from the IM (Hughes et al., 2019).

A typical ABC importer comprises a single free-floating periplasmic solute-binding protein (SBP) (Moussatova et al., 2008; Rice et al., 2014). On the other hand, the Mla system consists of two SBPs (a free-floating periplasmic protein MlaC and a unique IM-associated periplasmic protein MlaD)(Malinverni and Silhavy, 2009). The protein MlaD contains a highly conserved domain named after itself, that is, the MlaD domain, which is also referred to as the mammalian cell entry (Mce) domain (Ekiert et al., 2017). This is due to its first identification in the Mce proteins of the human pathogen *Mycobacterium tuberculosis* and is reportedly critical for host cell entry (Ekiert et al., 2017; Isom et al., 2017). The Mce proteins are involved in the transport of hydrophobic molecules (cholesterol and fatty acids) and the virulence of *M. tuberculosis* (Gioffré et al., 2005; Pandey and Sassetti, 2008; Nazarova et al., 2017). In addition to *E. coli* and *M. tuberculosis*, the MlaD domains have been identified in various proteins from actinobacteria, Gram-negative bacteria and even plants (Awai et al., 2006; Casali and Riley, 2007; Nakayama et al., 2017). Owing to the conserved MlaD domain architecture and the associated function, the MlaD protein of *E. coli* is considered an ortholog of the Mce proteins of *M. tuberculosis* and Trigalactosyldiacylglycerol 2 (TGD2) protein of *Arabidopsis thaliana* (Malinverni and Silhavy, 2009).

Apart from the protein MlaD, two additional proteins of *E. coli,* viz. PqiB and YebT contain the MlaD domain in multiple copies (Ekiert et al., 2017; Moussatova et al., 2008). These two proteins also mediate the trafficking of PLs in order to maintain the OM asymmetry (Nakayama et al., 2017). Regardless of different organisms and protein types, the MlaD domain is rich in hydrophobic residues and conserves the topology constituted by β-sheets with interconnecting loops (Ekiert et al., 2017; Moussatova et

al., 2008). However, architecture types of the MlaD domain are found to be diverse across different proteins and organisms as well. Although the MlaD domain is a conserved feature, a proper characterization of the domain is still lacking.

The usual methods of sequence-based study of the MlaD domain are not adequate due to the diversity of domain architecture and lack of sufficient reports. Hence, the characterization of the MlaD domain becomes necessary in order to deduce its function. In addition to MlaD, the other components of the Mla system remain poorly characterized. The functional association of Sulphate Transporter and AntiSigma factor antagonist (STAS) domain-containing MlaB protein with MlaF (Nucleotide-binding domain, NBD) or a dependency upon OM lipoprotein (MlaA) is not observed in the case of other ABC transporter. Additionally, the MlaE (transmembrane domain, TMD), along with MlaF, remains an enigma that still requires in-depth study. In order to eradicate this dearth of knowledge about the Mla system, we have used various computational approaches to characterize all the components of the Mla system from *E. coli*. Using these strategies, we have studied the MlaD domain from the MlaD protein of *E. coli*, its archetypes and those belonging to different architectures. Furthermore, the MlaD domain-containing intermembrane systems of *E. coli*, *M. tuberculosis* and *A. thaliana* were investigated in order to gain a better understanding of the domain function. The study also reports the unique evolutionary lineages of MlaE as well as MlaF, along with their unusual structural features. The work further highlights the dynamic nature of MlaA, which would aid in the transport of PLs. Altogether, the *in silico* study conducted provides significant insight into the features of Mla components from *E. coli*.

3.2. Materials and methods

3.2.1. Data collection

The amino acid sequences of the proteins MlaD, PqiB and YebT from *E. coli*, Mce from *M. tuberculosis* and TDG2 from *A. thaliana* were extracted from the UniProtKB database (The UniProt Consortium, 2021). In order to study the architectural diversity of the MlaD domain, two databases, Pfam and InterPro, were searched by using the keyword “MlaD’’ (El-Gebali et al., 2019; Mitchell et al., 2019). Subsequently, the search hits (Pfam; Accession No.: PF02470) and “Mce/MlaD” (InterPro; Accession No.: IPR003399) containing the keyword “MlaD” were chosen. The domain architectures

from each database were shortlisted by scanning the architecture/organization section. The final list of architectures was prepared by including only the unique MlaD domain architectures shortlisted from both databases. One representative from each architecture type was randomly selected for further analysis. The protein sequence of each representative was collected from the UniProtKB database.

3.2.2. Sequence analysis

The amino acid sequences of the MlaD domain were trimmed from the respective protein sequences considering the domain boundaries provided in the databases and literature (Isom et al., 2020). To determine the sequence similarities among the MlaD domains, pairwise sequence alignment was performed using the web tool BLAST with the default set of parameters (Altschul et al., 1990). In order to investigate the features with respect to amino acid composition, their percentage of occurrences (POs) in the MlaD domain was calculated using the web tool ProtParam (Gasteiger et al., 2005). The average percentage of occurrence (APO) of each amino acid over all the MlaD domains was calculated by taking the mean of their individual PO values. The amino acids with APOs greater and lesser than the median value were categorized as “abundant” and “limited,” respectively. The MlaD domain profiles obtained from the PO values were visualized as heat maps generated using the statistical package RStudio (RStudio Team, 2015). To obtain the conserved regions of the MlaD domain, multiple sequence alignment (MSA) was performed using the program Clustal Omega followed by visualization of the results using the web tool ESPript (Gouet et al., 2003; Sievers et al., 2014). The consensus motif(s) present in the domains of different architecture types was determined using the consensus sequence tool embedded in the web tool ESPript with a default global score of 0.7, followed by its visualization using the web server WebLogo and the program ConSurf (Crooks et al., 2004; Ashkenazy et al., 2016). To understand the evolutionary relationships among the MlaD domains, the sequences were aligned by using the program ClustalW, followed by the construction of a phylogenetic tree using the program MEGA7 employing the maximum likelihood method with the default set of parameters (Kumar et al., 2016). Subsequently, the phylogenetic tree was validated by performing 1000 bootstrap iterations. The organization of the genes of Mla, Mce and TGD transport systems was analyzed from the available literature and the Gene database of the National Center for Biotechnology Information (NCBI) as well as MycoBrowser (Kapopoulou et al., 2011). To identify the interactions of the MlaD domain-containing proteins, the protein-protein interaction

analysis was performed using the database STRING v10.5 with the default set of parameters (Szklarczyk et al., 2017). Subsequently, the interactome networks were visualized using the tool Cytoscape 3.8.1 (Shannon et al., 2003). Wherever applicable, the MlaD domain from the protein MlaD of *E. coli* was taken as the reference in addition to the MlaD domains of the proteins Mce and TGD2 from *M. tuberculosis* and *A. thaliana*, respectively. Analysis of signal peptides of lipoproteins was performed using the web server DOLOP (Madan Babu & Sankaran, 2002).

3.2.3. Structure analysis

The crystal/cryo-EM structures of the proteins of interest were obtained from RCSB Protein Data Bank (PDB) (Berman et al., 2000). The modeled three-dimensional structures of the proteins for which crystal/cryo-EM structures are not available were obtained from Protein Structure Database (Jumper et al., 2021; Varadi et al., 2022). The structural comparisons between the MlaD domains were performed using the web server 3dSS (Sumathi et al., 2006) and the results were visualized using PyMOL (The PyMOL Molecular Graphics System, Schrodinger, LLC). The analysis of protein interfaces was performed using the web server PDBePISA with the default set of parameters (Krissinel and Henrick, 2007). The hydrophobicity of the proteins, including the MlaD domains, was visualized using the YRB color scheme in the program PyMOL (Hagemans et al., 2015). In the YRB scheme, the carbon atoms not bound to nitrogen and oxygen atoms are highlighted in yellow, nitrogen atoms in the side chains of lysine and arginine are in blue, oxygen atoms in the side chains of glutamate and aspartate are in red and all remaining atoms are in white. The program PyMOL was used to visualize protein structures and generate figures. The dynamics of the selected protein were performed by using the web server anisotropic network model 2.0 (ANM 2.0) from where the PyMOL script was obtained and executed in a local system (Eyal et al., 2015). Structure-based positioning of proteins in the membrane was determined by using the online server PPM v2.0 with the default sets of parameters (Lomize et al., 2012). The dimensions of the central channels were estimated using the plugin CAVER 3.0, available in PyMOL with the default set of parameters (Chovancova et al., 2012). Protein-protein interactions were studied using the webserver PDBsum (Laskowski et al., 2018).

3.2.3.1. Protein-ligand docking

For the docking experiment, modeled three-dimensional structure of MlaA from *E. coli* (*Ec*MlaA$_m$) was obtained from the AlphaFold database. The binding affinity of PL (Ligand id: PEF; extracted from MlaC from *Ralstonia solanacearum*, PDB id: 2QGU) to *Ec*MlaA$_m$ (with/without C-terminal helix) was estimated by molecular docking analysis performed using the freely available program AutoDock v4.2 (Morris et al., 2009). Wherever applicable, before each docking experiment, the hydrogen atoms were added to the protein and ligand using the module available in the program AutoDock. Subsequently, the partial charges were assigned to the protein atoms using the Gasteiger charge algorithm (Gasteiger and Marsili, 1980). To perform the blind docking of the ligands to the protein, a grid size of 126x126x126 with a grid spacing of 0.375 Å was generated by taking the center of mass of the targeted protein as the grid center. The rigid molecular docking was performed by keeping the protein atoms rigid and the ligand atoms flexible, along with the rotatable bonds. For each molecular docking experiment, a total of 2000 runs of the Lamarckian genetic algorithm (LGA) were set up. The docked ligand conformations were clustered with a root mean square deviation (RMSD) cut-off of 2.0 Å. The protein-ligand complexes were observed using the program PyMOL.

3.2.3.2. Protein-protein docking

For protein-protein docking, the freely available web tool ClusPro was used (Kozakov et al., 2017). The proteins OmpC (*Ec*OmpC, PDB id: 2J1N) and OmpF from *E.coli* (*Ec*OmpF, PDB id: 5NUQ) were considered as receptors, while *Ec*MlaA$_m$ (with/without C-terminal helix) was considered as a ligand. The calculations were performed with the default set of parameters while the models under the balanced results sections were considered. The complexes obtained were studied using the program PyMOL.

3.3. Results

3.3.1. The MlaD domains are organized in different domain architecture types

A search for the MlaD domain architecture led to the collection of 155 MlaD domains. Based on their unique architecture types, the domains were organized into 20 different (sub)groups (A–E; Figure 3.1 and Table A1.1). Out of 20, architecture group A contains only MlaD domain(s), while the remaining groups contain additional functional domains (Table A1.1). Interestingly, the architecture of the MlaD domain varies with

different proteins from the same organism; for example, in *E. coli*, three proteins, viz. MlaD, PqiB and YebT contain 1, 3 and 7 copies of the MlaD domains, respectively (referred to as *Ec*MlA_MlaD1, *Ec*PqA_MlaD1-3 and *Ec*YeA_MlaD1-7; where the organism and protein names are encoded in first and second two letters followed by the architecture group between one to five letters and the number of MlaD domains are provided at the last digits), belonging to the architecture group A.

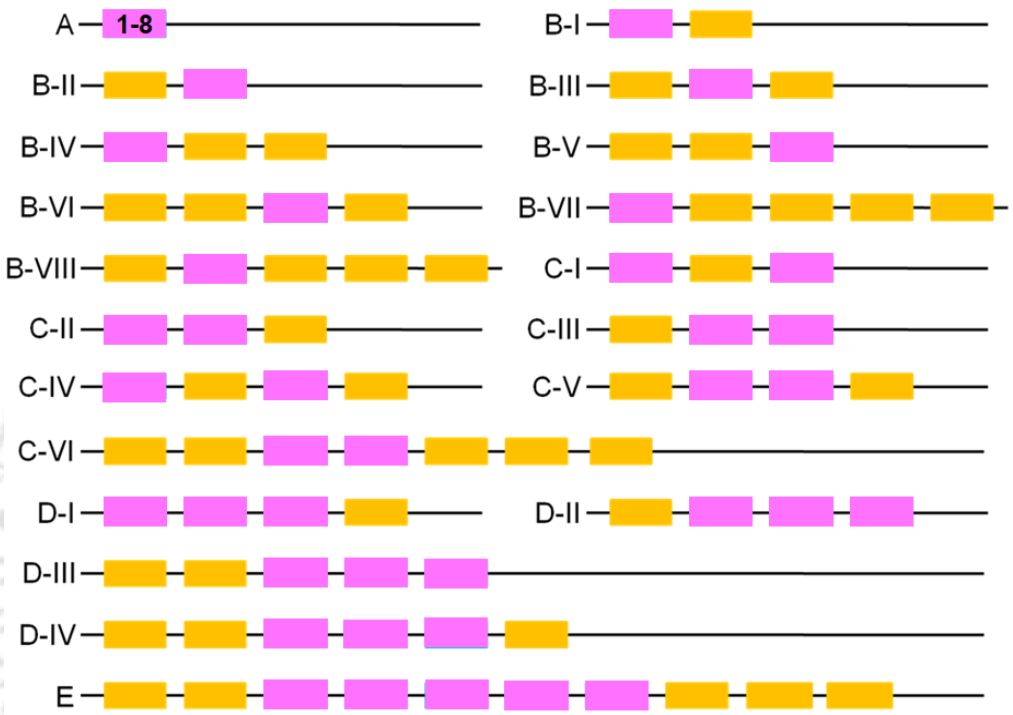


Figure 3.1. Schematic representation of the MlaD domain architectures. *A schematic depiction of 20 different architecture types of the MlaD domains organized in different proteins collected from Pfam and InterPro databases. These architectures are categorized into different groups as labeled. The proteins having only the MlaD domains ranging from 1 to 8 copies are kept in group "A." The MlaD domain is represented as pink rectangles, while other domains (proteins) are orange rectangles. The details of these architectures are provided in Table A1.1.*

Akin to the protein MlaD, 24 orthologous Mce proteins from *M. tuberculosis* H37Rv also contain the MlaD domain and are grouped into two different architecture types, viz. A (contains a single copy of the MlaD domain; *Mt*1BA_MlaD1, *Mt*1CA_MlaD1, *Mt*1EA_MlaD1, *Mt*1FA_MlaD1, *Mt*2BA_MlaD

1, *Mt*2CA_MlaD1, *Mt*2DA_MlaD1, *Mt*2EA_MlaD1, *Mt*3BA_MlaD1, *Mt*3CA_MlaD1, *Mt* 3DA_MlaD1, *Mt*3EA_MlaD1, *Mt*3FA_MlaD1, *Mt*4BA_MlaD1, *Mt*4CA_MlaD1, *Mt*4EA_ MlaD1 and *Mt*4FA_MlaD1) and B-I (a single copy of MlaD domain in association with cholesterol uptake porter domain, viz. Mce4_CUP1; *Mt*1ABI_MlaD1, *Mt*1DBI_MlaD1, *Mt*2ABI_MlaD1, *Mt*2FBI_MlaD1, *Mt*3ABI_MlaD1, *Mt* 4ABI_MlaD1 and *Mt*4DBI_MlaD1). Similarly, in *A. thaliana*, the protein TGD2 contains a single copy of the MlaD domain (*At*TgA_MlaD1) and belongs to the architecture group A (Figure A1.1 and Table A2.1-A2.2).

Furthermore, the analysis of the collected MlaD domain architectures suggests that the MlaD domains are mostly found in proteins from Gram-negative bacteria. However, they are also identified to be present among the various proteins from Gram-positive bacteria (*Lr*PlBI_MlaD1), archaea (*Ea*MlA_MlaD1), arthropoda (*Ga*MlBII_MlaD1, *Ap*UnCV_MlaD1-2 and *Ln*PiCVI_MlaD1-2), nematoda (*Tt*PrBVIII_MlaD1 and *Tt*McDIV_MlaD1-3), fungi (*Bb*YeE_MlaD1-5), cyanobacteria (*Oa*AbBI_MlaD1 and *Cm*MlBI_MlaD1) and placozoa (*Ta*AdBIII_MlaD1). In some proteins, the MlaD domain is found in certain unclassified sources (*Cc*UnBI_MlaD1, *Cc*UnBIV_MlaD1, *Ms*UnBII_MlaD1, *Sc*UnBIII_MlaD1, *Ch*BlBIII_Mla D1, *Cr*RaBIV_MlaD1 and *Ub*McBI_MlaD1; Table A2.1-A2.2).

3.3.2. The MlaD domains share a similar amino acid profile across different architectures

As the MlaD domains are ordered in different architectural groups, it becomes necessary to find out whether they differ at the sequence level as well. For this, the MlaD domains present in the proteins MlaD, PqiB and YebT of *E. coli* were aligned pairwise. The results reveal that in spite of being a conserved evolutionary unit, the MlaD domains share a highly diverse amino acid composition (Table A1.2).

With respect to other MlaD domains of *E. coli*, *Ec*MlA_MlaD1 shows identities, similarities and query coverages in the range of 25–55%, 36–65% and 6–67%, respectively. In fact, *Ec*MlA_MlaD1 does not show any significant alignment with *Ec*YeA_MlaD4 and *Ec*YeA_MlaD5. Furthermore, *Ec*MlA_MlaD1 shares the highest sequence identity (55%) with *Ec*YeA_MlaD1 but has a query coverage of only 30% in spite of being a conserved domain. Within the protein PqiB, *Ec*PqA_MlaD1 shows the highest identity at 75% but at a query coverage of a mere 3%

with *Ec*PqA_MlaD3. Surprisingly, *Ec*PqA_MlaD1 does not show any significant alignment with *Ec*PqA_MlaD2, while the latter shows an identity of 55% (similarity: 63%, query coverage: 20%) with *Ec*PqA_MlaD3. Such variations are also seen among the domains of the protein YebT where identities, similarities and query coverages are in the range of 23–47%, 47–82% and 14–100%, respectively. As observed in PqiB, some of the domains in YebT do not show significant alignments. Furthermore, *Ec*PqA_MlaD3 and *Ec*YeA_MlaD6 share no significant alignments with many domains, while *Ec*YeA_MlaD7 shows reasonable alignments with all the domains (Table A1.2).

Since the pairwise sequence alignment between the MlaD domains could not disclose their conservational details, the amino acid compositions of these domains were calculated by their sequence profiles. The sequence profiles of the MlaD domains of proteins from *E. coli* and *M. tuberculosis* reveal that 10 amino acids are abundantly present with APOs greater than the median APO value (Figure 3.2A and Table A2.2). In the MlaD domains of proteins from *E. coli*, the amino acids leucine, glycine, valine, threonine, serine, isoleucine, alanine, aspartate, arginine and lysine (in decreasing order of their APOs, the highest being 11.5) are abundant. Similarly, in the MlaD domains of proteins from *M. tuberculosis*, the amino acids valine, alanine, glycine, threonine, leucine, serine, arginine, isoleucine, aspartate and proline (in decreasing order of their APOs, the highest being 12.1) are abundantly present. Notably, the amino acids tryptophan (APO of 0.6 and 0.4 in *E. coli* and *M. tuberculosis*, respectively) and cysteine (APO of 0 in both the organisms) show their least presence in the MlaD domains (Figure 3.2A and Table A2.2). These results were also confirmed using the heat map profiles of the MlaD domains of proteins from *E. coli* and *M. tuberculosis* (Figure 3.2B-3.2D). A similar amino acid profile was also observed for the MlaD domain of the protein TGD2 from *A. thaliana*, where 10 amino acids are abundantly present with APOs greater than the median APO value of 4.5. Moreover, this profiling suggests that the amino acids cysteine, histidine, phenylalanine and tyrosine (each have an APO of 1.3) and tryptophan (APO of 0) are the least abundant (Table A2.2).

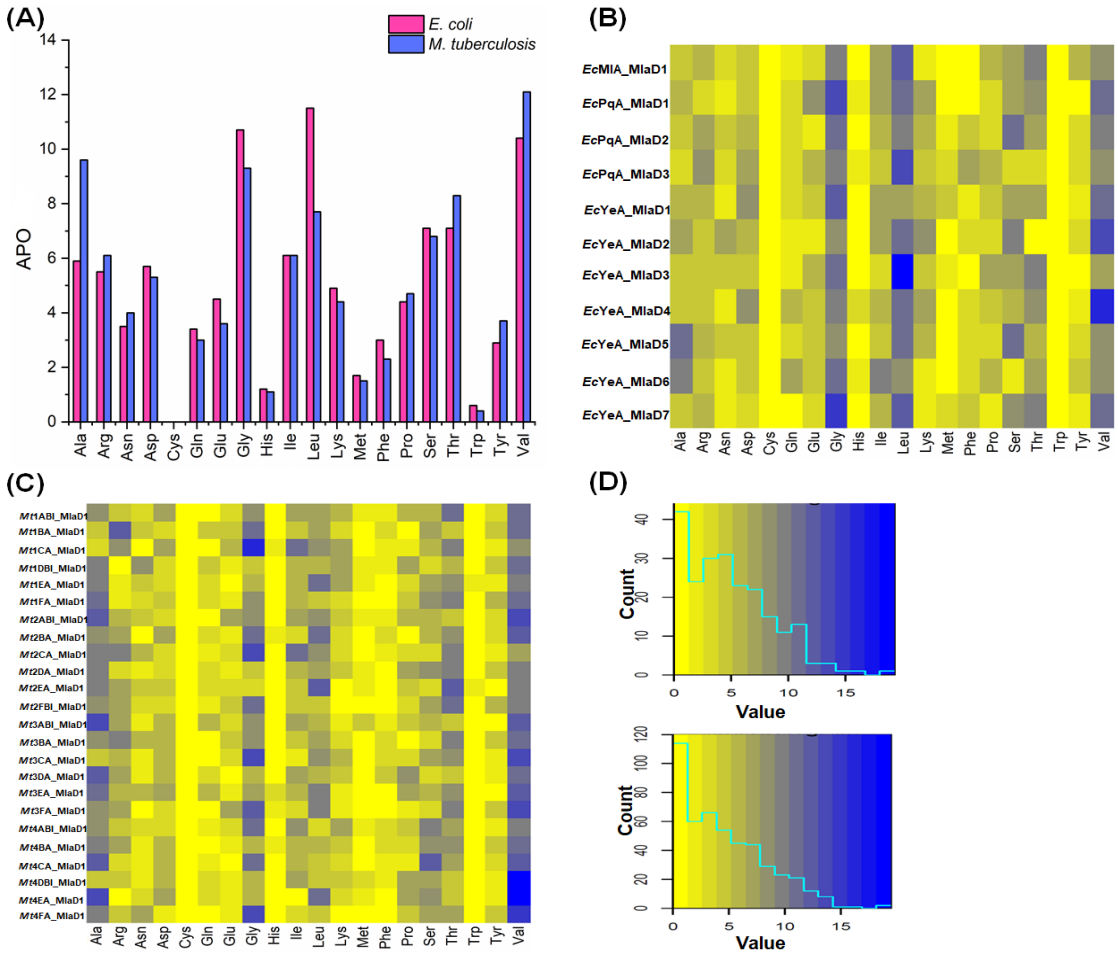


Figure 3.2. The distribution of amino acids in the MlaD domains of the proteins from E. coli and M. tuberculosis. *(A) Graph depicting the APOs of amino acids present in the MlaD domains of proteins from E. coli and M. tuberculosis. The amino acids and their respective APOs are represented along the X- and Y-axes, respectively. (B-C) A heat map showing the distribution of the 20 amino acids in the MlaD domains of proteins from E. coli and M. tuberculosis, respectively. In both the heat map diagrams, the amino acids are represented in columns and the MlaD domains are represented in rows. (D) The color key and histogram of the heat map for E. coli (upper panel) and M. tuberculosis (lower panel), respectively. The X-axis in the color key represents the number of cells and the Y-axis represents the values in each cell. The increasing gradient of blue signifies the increase in the PO values. The histogram within the color key indicates the number of cells in the heat map carrying a PO.*

The amino acid profiling hints that despite the low sequence identity, the MlaD domain conserves the overall amino acid composition. To further corroborate the result, an attempt was made to study the conserved compositional features across the collected 155 MlaD domains from 20 different domain architecture types. The result reveals that the amino acids valine, glycine, leucine, serine, isoleucine, alanine, threonine, aspartate, arginine and lysine (in decreasing order of their APOs, the highest being 11.3 and the median APO being 4.7) are abundantly present. Expectedly, the amino acid cysteine is found to be the least occurring (APO of 0.1) among the MlaD domains (Figure A1.2A). These results were further corroborated using the heatmap profiles of the collected MlaD domains (Figure A1.2B).

3.3.3. The N-terminal region of the MlaD domain is conserved

To further identify the significance of the position-specific conservation of amino acids in the MlaD domains, MSA of all the collected MlaD domains was performed systematically. The domain *Ec*MlA_MlaD1 present in the protein MlaD extends from 38th to 117th amino acid. For the purpose of visualization and analysis, the domain boundary is labeled from 1 to 75. Results suggest that in *E. coli*, the MlaD domains possess multiple conserved glycines (Gly13, Gly25, Gly29 and Gly68) and hydrophobic residues (Leu5, Ala7, Iso12, Leu15, Val21, Ile23, Val26, Val28, Val31, Ile34, Leu36, Val45, Leu47, Ile49, Ile56 and Leu75) that are mostly situated at the N-terminal region. Among them, the residue Gly29 is the most conserved (Figure 3.3A). Similarly, in *M. tuberculosis*, the MlaD domains also possess multiple conserved glycines (Gly25, Gly29 and Gly71) and hydrophobic residues (Ala7, Leu15, Val21, Ile23, Val26, Val28, Val31, Iso34, Val45, Iso49, Iso56, Leu62, Iso64, Leu69, Leu70, Leu75 and Leu77; Figure 3.3B). Among them, the residues Val21 and Gly29 are the most conserved (Figure 3.3B). Akin to *E. coli* and *M. tuberculosis*, the MlaD domain of *A. thaliana* also possesses multiple conserved glycines as well as hydrophobic residues (Figure A1.3A).

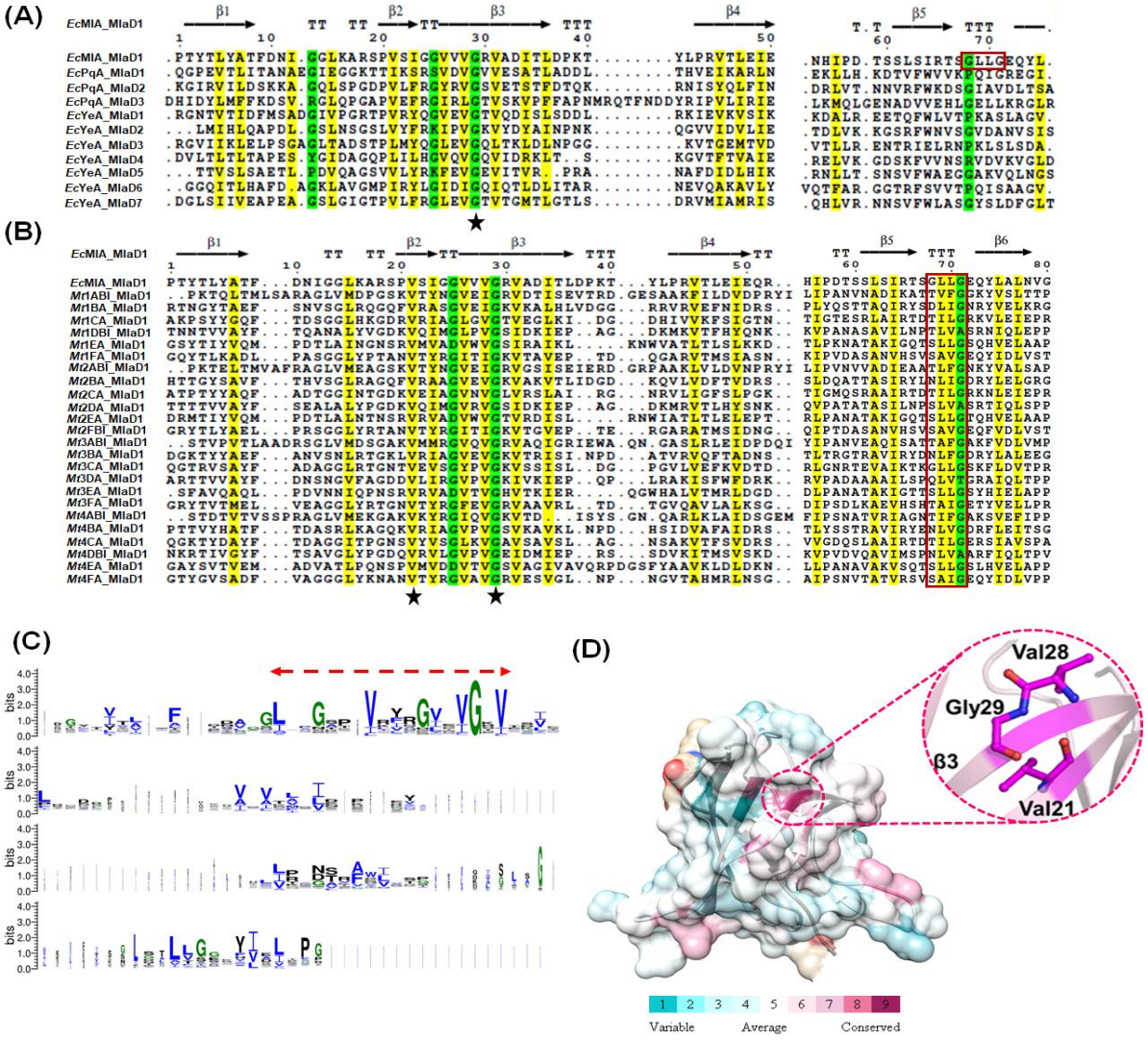


Figure 3.3. Conserved features of different MlaD domains. *(A-B) MSA of the MlaD domains of proteins from E. coli and M. tuberculosis, respectively. Similar hydrophobic and conserved residues are highlighted in yellow and green, respectively. Highly conserved residues Val21 and Gly29 are marked with a star. The conserved PLP region is enclosed in a red box. The domain boundary of EcMlA_MlaD1 is 38 to 117. For the purpose of visualization and analysis, it is labeled from 1 to 75. (C) Sequence logo representation of the conservation of the amino acids in the MlaD domains across different domain architecture groups. Glycine and hydrophobic residues are displayed in green and blue, respectively, while the remaining residues are displayed in black. The motif L-X_2-G-X_3-V-X_3-G-X_2-V-G-X-V has been marked with a double-headed dotted arrow in red. (D) The conservation coloring profile of the MlaD domain obtained from the web server ConSurf mapped onto the surface representation of the EcMlA_MlaD1 (PDB id: 5UW2) (Ashkenazy et al., 2016). The conservation scores 1–9 represent highly variable to highly conserved residues and the respective colors*

are shown below the domain. A magnified view of the most conserved residue Gly29 is provided on the right-hand side of the model.

In addition to the position-specific conservation, the MSA results also reveal a unique characteristic of the pore loop (PLP). The PLP is a motif that extends out of the core MlaD domain and lines the central pores of the proteins MlaD, PqiB and YebT (Ekiert et al., 2017; Liu et al., 2020). The results indicate that the PLP region of *Ec*MlA_MlaD1 is not conserved among the MlaD domains of the proteins PqiB and YebT from *E. coli* (Figure 3.3A). However, the region is found to be conserved in the MlaD domains from *M. tuberculosis* and *A. thaliana* (Figures 3.3B and A1.3A).

To further affirm the conservation of the N-terminal region of the MlaD domains, MSA of the MlaD domains from different domain architecture groups was performed. The result asserts that irrespective of different architectures, the N-terminal region contains similar amino acid residues. Along with the highly conserved residue Gly29, other residues such as Leu15, Val21, Gly25, Val28 and Val31 are also preserved in the majority of the MlaD domains (Figure 3.3C and Figure A2.1). Interestingly, these conserved amino acid residues from 15 to 31 constitute a motif L-X_2-G-X_3-V-X_3-G-X_2-V-G-X-V (where X denotes any amino acid) that is overall present in the MlaD domains across the varying architecture. Compared to the N-terminal region, the C-terminal region of the MlaD domain is highly variable (Figure 3.3C). At the tertiary structure level, the conserved residues are mostly located at the β3 strand of the domain (Figure 3.3D).

3.3.4. The *Ec*MlA_MlaD1 domain is evolutionarily closer to the MlaD domains of *M. tuberculosis* and *A. thaliana* than those of *E. coli*

To understand the evolutionary relationship of *Ec*MlA_MlaD1, a phylogenetic analysis of the domains *Ec*PqA_MlaD1-3, *Ec*YebA_MlaD1-7, *Mt*1ABI_MlaD1-*Mt*4FA_MlaD1 and *At*TgA_MlaD1 was performed. The phylogenetic tree can be divided into two distinct clusters that have been named Cluster 1 and 2 (Figure 3.4). Interestingly, *Ec*MlA_MlaD1 groups together with the *Mt*1ABI_MlaD1-*Mt*4FA_MlaD1 and *At*TgA_MlaD1 in Cluster 1, while *Ec*PqA_MlaD1-3 and *Ec*YebA_MlaD1-7 group together in Cluster 2 (Figure 3.4A). To probe the reasons for this, the domain architectures belonging to these two clusters were compared. These domains either

belong to the architecture group A or B–I. Cluster 1, in which *Ec*MlA_MlaD1 is placed, contains 26 MlaD domains that either belong to the architecture group A (19 domains) or BI (7 domains). All the domains of this cluster occur in a single copy with the conserved PLP region. Interestingly, the seven domains belonging to architecture group BI are a part of proteins (Mce1A, Mce1D, Mce2A, Mce2F, Mce3A, Mce4A and Mce4D) that carry a single copy of an additional domain, viz. Mce4_CUP1. Furthermore, the domains of Cluster 1 are from highly diverse source organisms. On the other hand, Cluster 2 is constituted of 10 MlaD domains that are from *E. coli* and exclusively belong to the architecture group A. Interestingly, all these domains occur in multiple copies, unlike Cluster 1. Furthermore, all the members of Cluster 1 are components of ABC transporter systems, while members of Cluster 2 are components of non-ABC transporter systems.

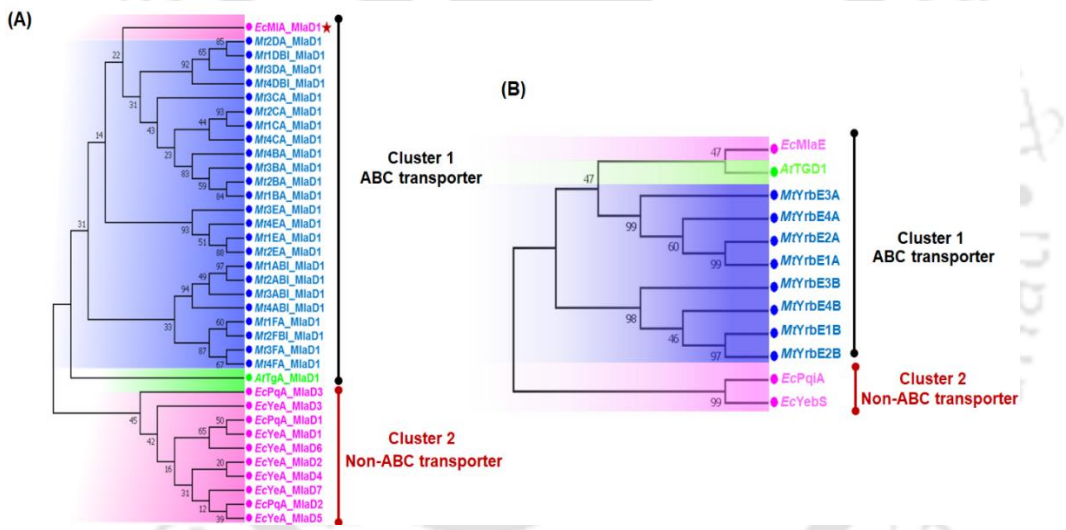


Figure 3.4. Phylogenetic trees displaying the evolutionary relationship among the MlaD domains and their associated TMDs of E. coli, M. tuberculosis and A. thaliana. *(A) Evolutionary relationship among the MlaD domains of E. coli, M. tuberculosis and A. thaliana. The tree was generated for 36 MlaD domains, including EcMlA_MlaD1 (red star). (B) Evolutionary relationship among the MlaD domains associated TMDs of E. coli, M. tuberculosis and A. thaliana. The tree was generated for 12 TMDs and details are provided in Table A1.3. Two clusters are labeled and the clustering of each domain/TMD of E. coli, M. tuberculosis and A. thaliana is highlighted in pink, blue and green, respectively. Numbers at each internal node of the phylogenetic tree signify the probability of its occurrence (in percentage) out of 1000 bootstrap replicates.*

As MlaD domains perform the function of transport in association with multipass transmembrane proteins, a phylogenetic analysis was performed using their respective permeases. For the analysis, the proteins MlaE, PqiA and YebS from *E. coli* (*Ec*MlaE, *Ec*PqiA and *Ec*YebS), YrbE1A-4B from *M. tuberculosis* (*Mt*YrbE1A-4B) and TGD1 from *A. thaliana* (*At*TGD1) were taken into consideration. The phylogenetic tree forms two distinct clusters named Cluster 1 and 2. *Ec*MlaE groups together with *Mt*YrbE1A-4B and *At*TGD1 in Cluster 1, while, *Ec*PqiA and *Ec*YebS group together in Cluster 2 (Figure 3.4B). This clustering corresponds to the clustering shown by the MlaD domains. Furthermore, the members of Cluster 1 belong to the MlaE family (Pfam id: PF02405) and are ABC transporter permease. On the other hand, the members of Cluster 2 belong to the PqiA family (Pfam id: PF04403) and are non-ABC transporter permeases. This is suggestive of ATP-dependent and independent mechanisms of PL transport.

3.3.5. The MlaD domain adopts a conserved structure with varying loop regions

As the *Ec*MlA_MlaD1 domain shares a low sequence similarity (~44%) with the other MlaD domains of *E. coli*, an attempt was made to understand their (dis)similarity at the tertiary structure level. The structural details of the MlaD domains of the proteins MlaD (PDB id: 5UW2), PqiB (PDB id: 5UVN) and YebT (PDB ids: 6KZ3 and 6KZ4) from *E. coli* and the protein MlaD from *Acinetobacter baumannii* (*Ab*MlA_MlaD1, PDB id: 6IC4) are available in the literature. Comparison of these structures reveals that *Ec*MlA_MlaD1, *Ec*PqA_MlaD1-3, *Ec*YeA_MlaD1-7 and *Ab*MlA_MlaD1 share a similar topology that majorly consists of β-sheets and have RMSDs in the range of 1.0-1.5 Å (Figure 3.5A-3.5K). However, the PLP regions vary in length and organization. The structural superimposition of these 12 MlaD domains clearly shows the preservation of the core β-sheet-constituted regions with extremely variable loop regions (Figure 3.5). The *Ab*MlA_MlaD1, which also is one of the longest MlaD domains, possesses an insertion between the β4 and β5 strands in the form of α helices (Figure 3.5K). Interestingly, the PLP region in the *Ab*MlA_MlaD1 domain is conserved at the sequence as well as at the structure level (Figures 3.5K and Figure A1.3B).

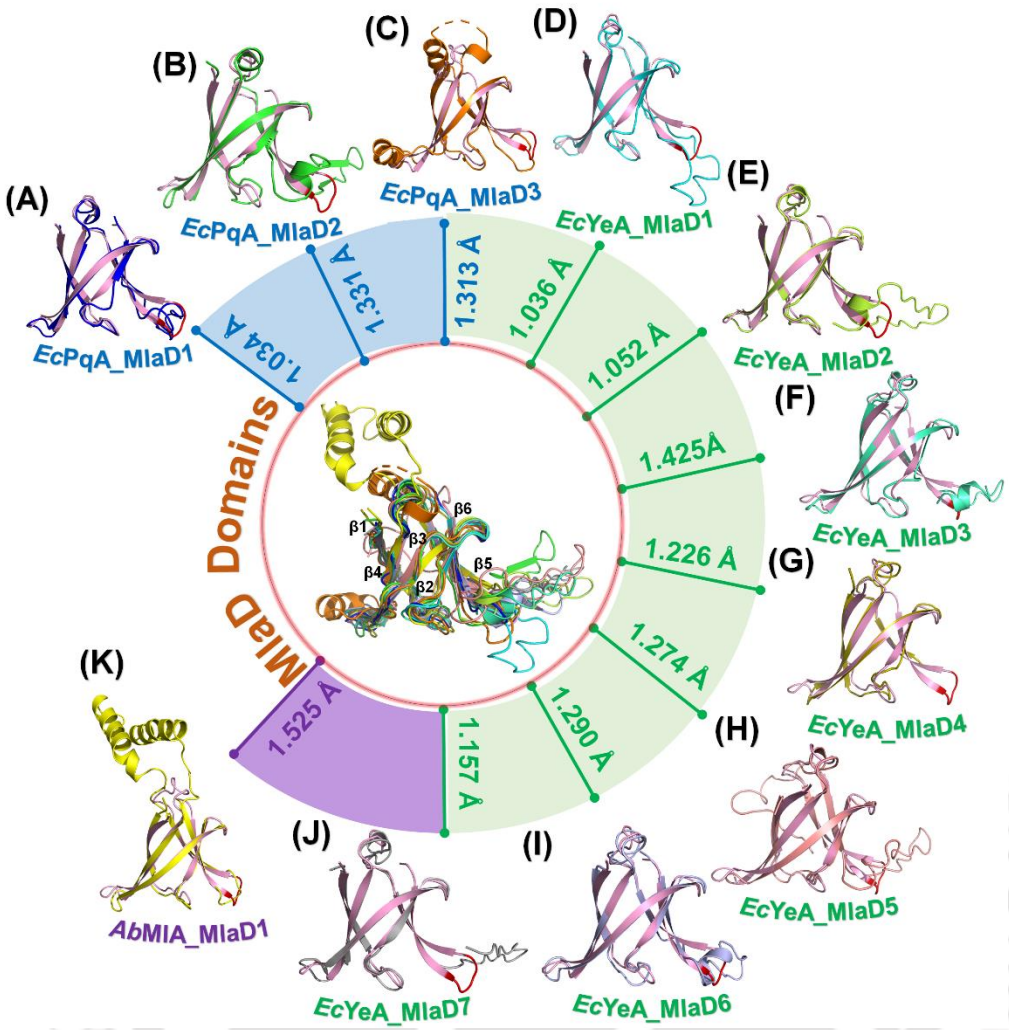


Figure 3.5. Overall structural comparison of EcMlA_MlaD1 with various MlaD domains. *Structural superimposition of EcMlA_MlaD1 (pink) and (A–C) EcPqA_MlaD1 (blue), EcPqA_MlaD2 (green) and EcPqA_MlaD3 (orange), respectively, (D–J) EcYeA_MlaD1 (cyan), EcYeA_MlaD2 (limon), EcYeA_MlaD3 (green cyan), EcYeA_MlaD4 (olive), EcYeA_MlaD5 (salmon), EcYeA_MlaD6 (light blue), EcYeA_MlaD7 (gray), respectively and (K) AbMlA_MlaD1 (yellow). The overall structural superimposition of the 12 MlaD domains is shown at the center. The individual superimpositions of EcMlA_MlaD1 with EcPqA_MlaD1-3, EcYeA_MlaD1-7 and AbMlA_MlaD1 are labeled and highlighted in blue, green and purple, respectively. The respective RMSDs are mentioned along these lines and the PLP region of EcMlA_MlaD1 is highlighted in red.*

3.3.6. The MlaD domain forms a hydrophobic pore

In *E. coli*, the proteins MlaD, PqiB and YebT containing the MlaD domains oligomerize to form homo-hexameric rings with a central pore (Ekiert et al., 2017). To obtain insights into the functional involvement of the MlaD domain, the spatial position of its amino acid residues was evaluated. Analysis reveals that the residues of the MlaD domains are spatially located at the inner line of the hydrophobic pore, with the PLPs lining the central channel of the homo-hexameric ring (Figure 3.6A). In the protein MlaD from *E. coli*, the *Ec*MlA_MlaD1 domain generates a predominant hydrophobic pore with a diameter of 16.3 Å at the opening site (Figure 3.6B). In the protein PqiB from *E. coli*, the domains *Ec*PqA_MlaD1-3 make similar hexameric arrangements but with varying pore diameters of 39.5, 37.5 and 19.9 Å, respectively (Figure 3.6C-3.6E). A similar observation is made for *Ec*YeA_MlaD1-7 of the protein YebT (Figure A1.4A-A1.4G) except for the *Ec*YeA_MlaD4 domain (Figure A1.4D), where the electron density for the PLP region was poorly resolved (Liu et al., 2020). For this reason, the PLP region was obtained from available literature (Isom et al., 2020). The length of the PLP region of each domain, along with the diameter of the central pores at the opening sites, are provided in Table A1.4. The results clearly demonstrate that the diameter of the pores is not correlated with the lengths of the PLP region. The variations in the sequences, lengths and amino acid compositions of the PLP regions of the MlaD domains of *E. coli* were further visualized through an MSA (Figure A1.5). The analysis again indicates that such differences would give rise to hydrophobic pores with varying diameters.

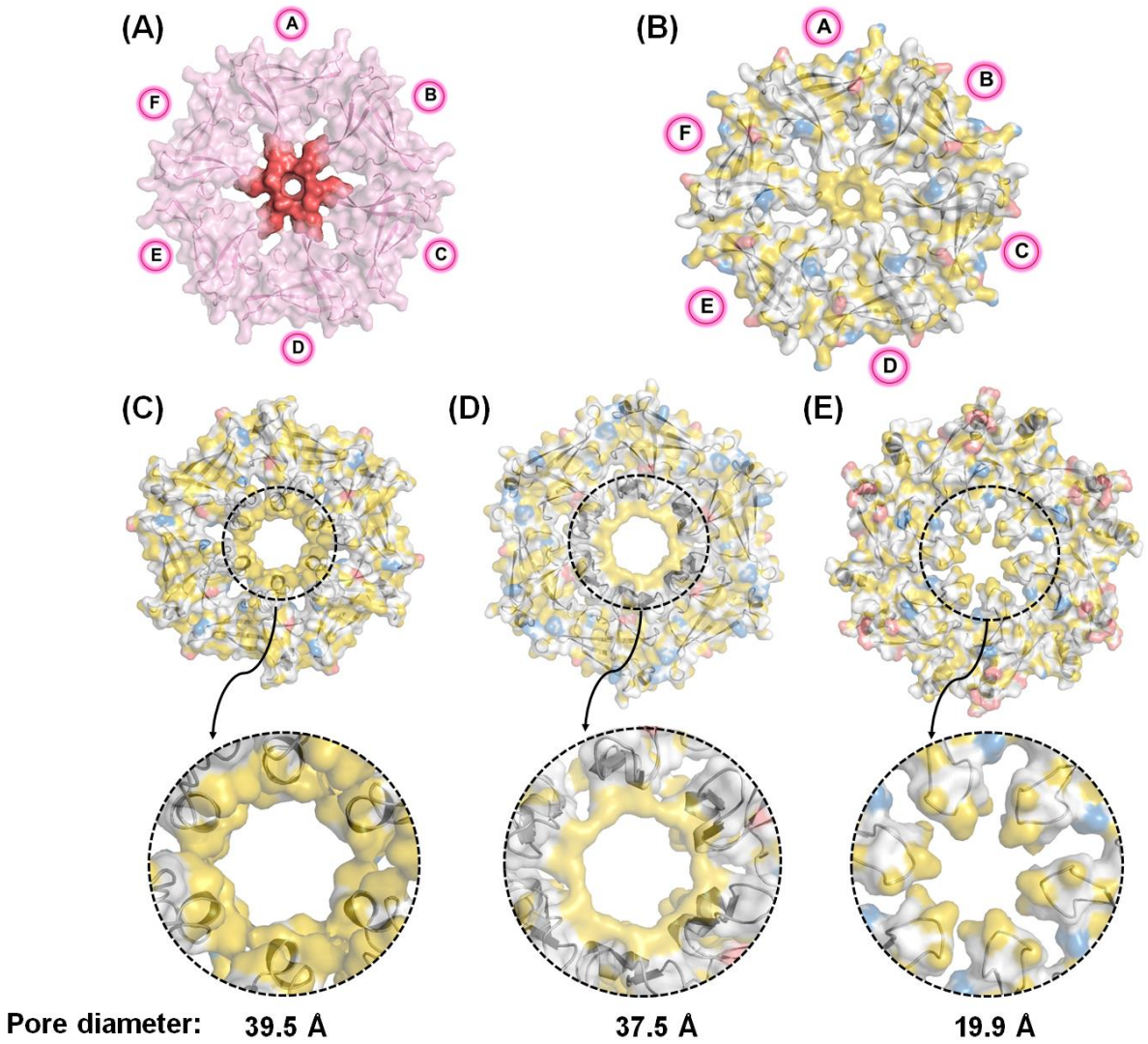


Figure 3.6. Overview of the hexameric assemblies of the MlaD domains and their respective pores in E. coli. *(A) An overview of the hexameric assemblies of the MlaD domains with the PLP region (dark pink) lining the central pore of the protein MlaD (PDB id: 5UW2). (B) Surface YRB representation of EcMlA_MlaD1 hexamer of the protein MlaD (PDB id: 5UW2). (C–E) Surface YRB representation of EcPqA_MlaD1-3 hexamers of the protein PqiB (PDB id: 5UVN), respectively. The hexameric assemblies of the domains are shown at the top and the magnified views of their respective central pores lined by the PLP regions, along with the pore diameters, are shown at the bottom.*

The MlaD domains in the three proteins (MlaD, PqiB and YebT) show a stacking arrangement giving rise to extended hydrophobic channels (Figure 3.7A-3.7C). In

order to get further insights into the role of PLP in such macromolecular arrangements, the cryo-EM structure of the MlaFEDB complex from *E. coli* (PDB id: 6XBD) was studied (Figure A1.6A). The structure reveals the presence of two bound PLs in two different conformations. The MlaD ring deviates from the expected 6-fold symmetry due to asymmetric interaction with PLs. The Arg97 residue from each MlaE subunit forms a salt bridge with the phosphate head group. At the same time, Leu107 from the MlaD protein makes hydrophobic interactions with one of the acyl chains of the PL (Figure A1.6B). It was proposed that the MlaD pore would be involved in the re-orientation of PL prior to translocation (Coudray et al., 2020). Interestingly, Leu107 corresponds to Leu70 in *Ec*MlA_MlaD1 and is located in the PLP. This signifies that the PLP region is also crucial for the proper re-orientation of the PLs. Considering that the region is conserved among the members of Cluster 1 and works in association with permeases belonging to the MlaE family, it seems likely that the PLP would modulate the pore diameter as well as substrate conformation in all the 26 MlaD domains. However, the PLPs among the members of Cluster 2 are extremely variable. Furthermore, the domains remain associated with permeases belonging to the PqiA family, indicating that the PL transport would be independent of ATP hydrolysis. This again corroborates the findings of the evolutionary analysis. Considering the available information, it can be proposed that MlaD domains have developed structural attributes, especially the PLPs, based on the mechanism of PL transport and association with permeases.

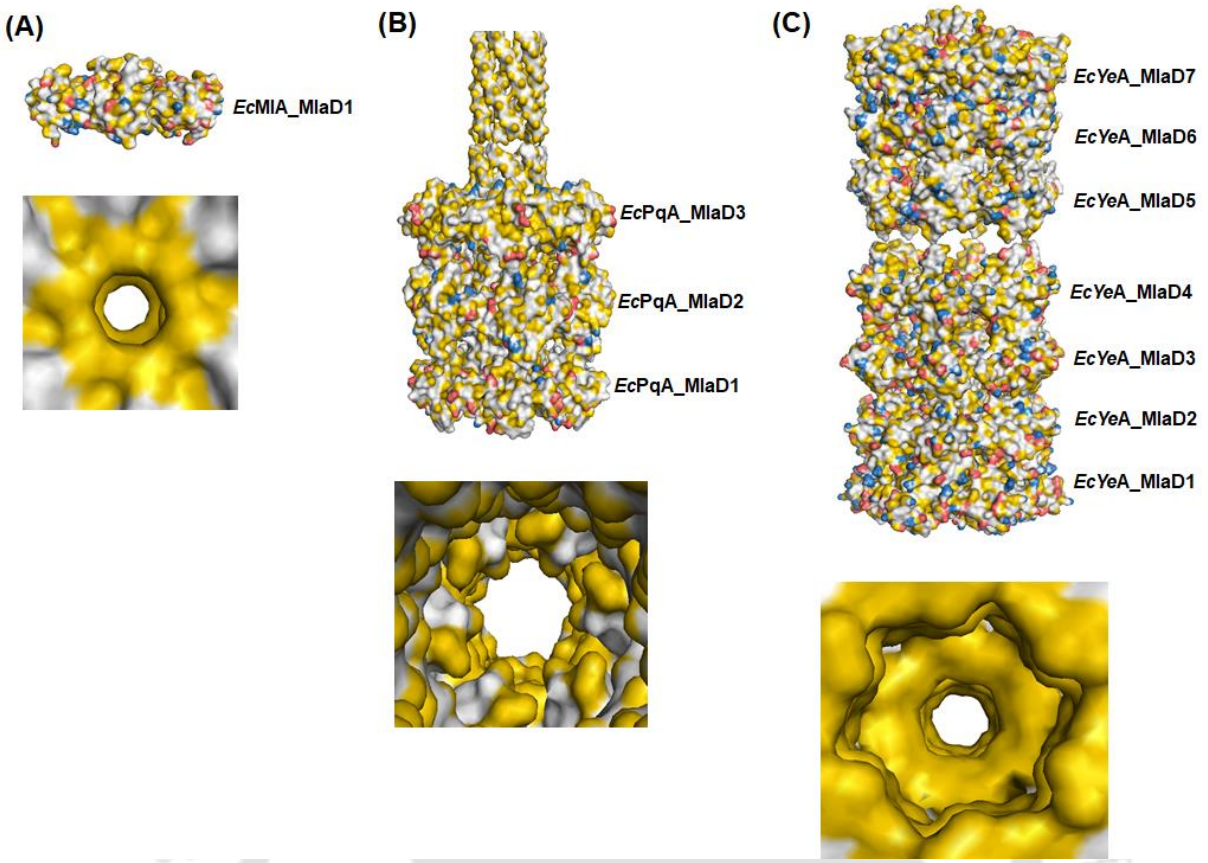


Figure 3.7. An overview of the organization of the MlaD domains in the proteins MlaD, PqiB and YebT from E. coli. *(A-C) (top) The side view of the MlaD hexamer (PDB id: 5UW2), PqiB hexamer (PDB id: 5UVN) and YebT hexamer (PDB ids: 6KZ3 and 6KZ4) highlighting one, three and seven layers of EcMlA_MlaD1, EcPqA_MlaD1-3 and EcYeA_MlaD1-7, respectively. (Bottom) The bottom view of the hydrophobic channel present in the center of these proteins, respectively.*

3.3.7. The MlaD domains are not involved in determining the oligomeric state of the proteins

As the MlaD domains are conserved units and assemble in hexameric rings in *E. coli*, it becomes imperative to find out whether they will determine the oligomerization of the proteins MlaD, PqiB and YebT. For this, the amino acid residues present at the hexameric interface of the protein MlaD were investigated, anticipating their absolute conservation if they are critical for oligomerization. The residues present at the dimeric interface (Figure 3.8A) of the homo-hexamer were identified using the web server PDBePISA and the details are provided in Table A1.5. Each dimeric interface can be divided into three regions, viz. head, body and base (Figure 3.8B-3.8D). A total of 50

residues are present at the dimeric interface of the protein MlaD, of which about 70% are from the MlaD domain. Among them, the residues Ile12, Gly13, Gly25, Val26, Val28, Ile34, Leu36, Ile56, Ser67 and Gly68 are almost conserved in the MlaD domains of proteins PqiB and YebT from *E. coli* (Figure 3.3A). With respect to *Ec*MlA_MlaD1, in the MlaD domains of Mce proteins, the residues Ile23, Gly25, Val26, Val28, Ile34, Ile56, Ile64, Leu69 and Leu70 show conservation (Figure 3.3B). In *At*TgA_MlaD1 of the protein TGD2 from *A. thaliana*, the residues Ile12, Gly14, Ile23, Gly25, Val26, Val28, Ile34, Thr35, Lys39, Thr40, Ile56, Ile64, Thr66, Ser67, Gly68, Leu70, Gly68, Leu69 and Gln73 are conserved (Figure A1.3A). MSA of all the collected MlaD domains shows that out of 50 residues, only a few residues Ile12, Gly25, Val26, Val28, Ile34, Ile56, Ile64 and Leu69 are conserved (Figure A2.1). The variation in the conservation of interfacial residues within the domain indicates that MlaD domains do not govern the process of oligomerization in proteins.

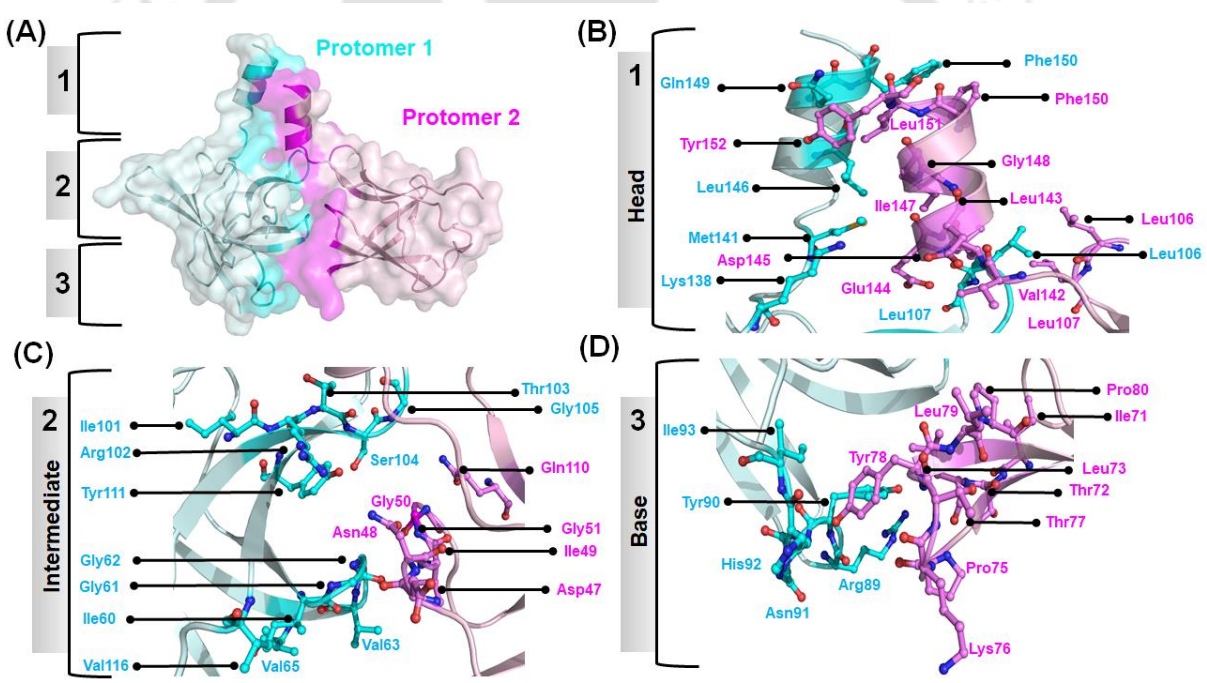


Figure 3.8. The residues present at the dimeric interface of the protein MlaD. *(A) Surface representation of a dimeric interface where two protomers of protein MlaD are shown in cyan and pink. (B-D) The residues of each protomer present at the interface belonging to the head, body and base parts, respectively.*

3.3.8. Nuclear transport factor 2 (NTF2) dependent Mla system might work in coordination with the Lpt system

It is interesting to note that the MlaD domain-containing systems of *E. coli*, *M. tuberculosis* and *A. thaliana* are involved in the transport of hydrophobic molecules (Ekiert et al., 2017). However, the organization and number of the MlaD domains in these systems vary. Thus, to draw a comparative relationship between the transport of hydrophobic molecules and the associated metabolic processes of these systems, genetic organizations were studied. The operonic arrangement of the Mla system shows that the gene encoding the protein MlaC is present upstream to that of the gene encoding the protein MlaD (Figure 3.9A). The Mla system is immensely dependent on MlaC as it is involved in the trafficking of PLs between the OM and IM. During this process, MlaC interacts with OM-associated protein MlaA and IM-associated protein MlaD (Figure A1.7A). Interestingly, the protein MlaC belongs to the NTF2 superfamily, unlike other solute-binding proteins (SBPs) (Wong et al., 2017). In contrast, the operonic arrangement of PqiABC and YebST systems lacks gene(s) encoding protein(s) that is structurally or functionally similar to MlaC (Figure 3.9A). In fact, both systems are not dependent on any periplasmic protein as PqiB and YebT directly reach out to the OM (Figure A1.7A). The operonic arrangement of the Mce system in *M. tuberculosis*, shows the presence of genes encoding proteins that have been reported to show structural similarity to the NTF2 superfamily. These include a total of 13 genes, 8 of which encode Mam (Mce-associated membrane) proteins (Mam 1A, 1B, 1C, 1D, 3A, 3B, 4A and 4B) and the remaining five genes encode for Omam (Orphaned mam) proteins (Perkowski et al., 2016). The *mam* genes are distributed in the operons *mce1*, *mce3* and *mce4*, while the *omam* genes are situated away from the *mce* operons (Figure 3.9A) (Forslund et al., 2019). Both Mam and Omam proteins have been reported to stabilize the Mce protein complex, thus suggesting potential interaction (Figure A1.7B). Furthermore, the genes encoding other associated proteins, such as MceG (ATPase) and LucA, which facilitate the uptake of fatty acids and cholesterol by stabilizing the protein subunits of the Mce system, are also situated away from the *mce* operons (Figure 3.9A). On the other hand, in the TGD system of *A. thaliana*, no protein has been reported that is structurally similar to the NTF2 superfamily. The protein TGD5, which is a functional analog of the protein MlaC, is not free-floating but is membrane-associated (Figure A1.7C). Moreover, genes encoding for proteins TGD1-5 are not located in an operon but are dispersed (Figure 3.9A). The genes *tgd1*, *tgd3* and *tgd5* are found on chromosome 1, while the

genes *tgd2* and *tgd4* at chromosome 3. These observations suggest that only the Mla and Mce systems are dependent on the NTF2 proteins for the transport of hydrophobic molecules, while Pqi, Yeb and TGD systems transport the hydrophobic molecules without the aid of the NTF2 proteins.

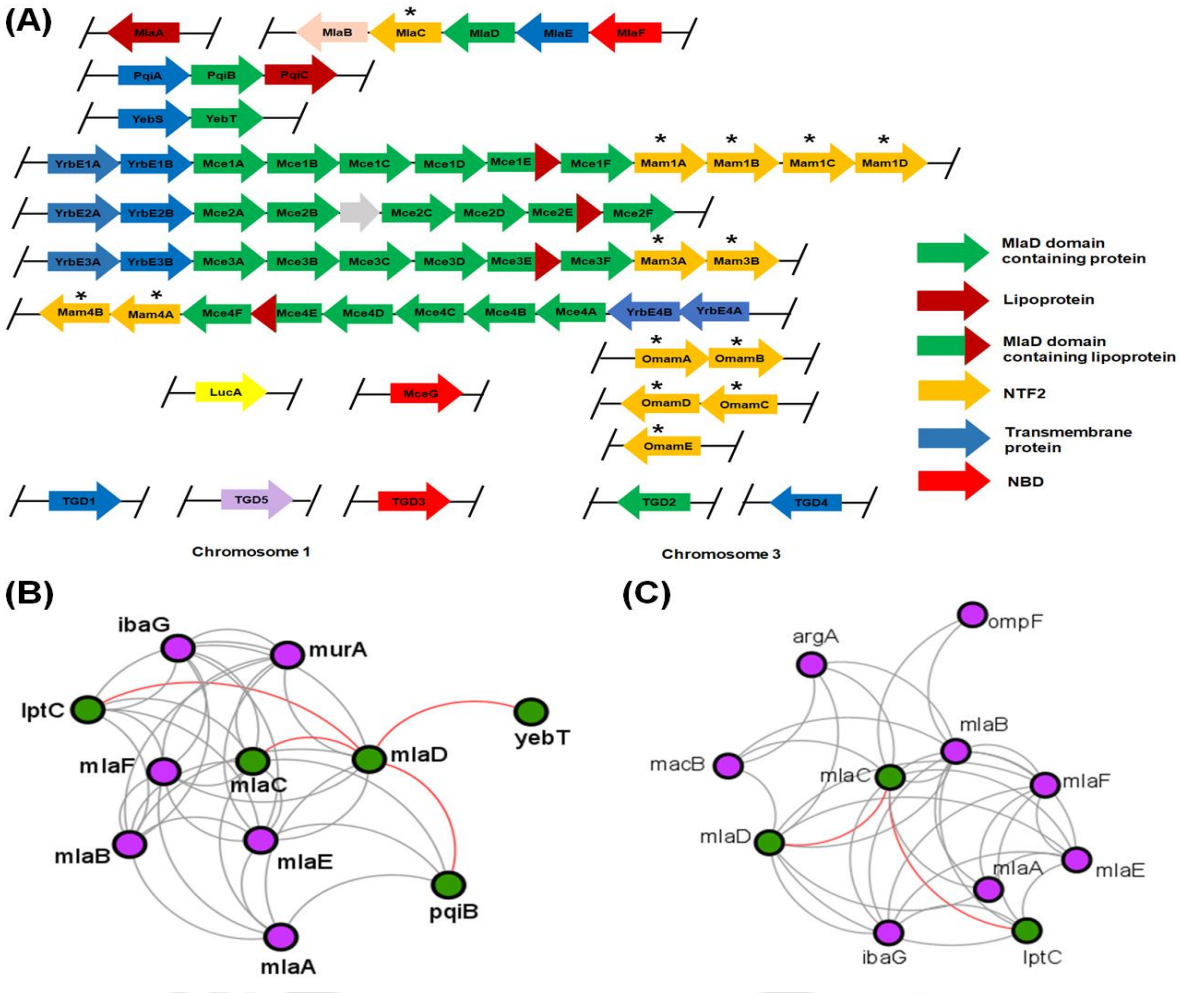


Figure 3.9. Genetic organization of the genes encoding proteins containing the MlaD domain and the functionally associated components in E. coli, M. tuberculosis and A. thaliana. *(A) Schematic representation of the genes encoding the proteins of Mla, Pqi and Yeb systems in E. coli, Mce in M. tuberculosis and TGD in A. thaliana. Each gene is represented by an arrow (with the encoded protein name) indicating its direction of transcription. The pseudogene region is shown as a gray arrow. The structural homologs of the NTF2 superfamily are highlighted with asterisks. (B-C) A protein-protein network map showing the interaction of proteins MlaD and MlaC, respectively. Each node represents a protein and the connections between them represent a predicted interaction. The interactions of interest are highlighted with green*

nodes and red lines. At the same time, the rest are highlighted with purple nodes and gray lines. Further details of the interacting proteins are provided in Table A1.6.

To further explore the biological interaction(s) of the MlaD protein possessing *Ec*MlA_MlaD1, a network analysis was performed. The interactome analysis shows that the MlaD protein interacts with PqiB and YebT along with MlaC (Figure 3.9B). Interestingly, the analysis also suggests an interaction with the protein LptC, a component of the Lpt system. In *E. coli*, the Lpt system is involved in the transport of LPS from the IM through the periplasm to the OM. The protein LptC is an IM-associated periplasmic protein that plays a critical role in the transport of LPS molecules. Surprisingly, the network analysis of the protein MlaC also shows an interaction with the protein LptC (Figure 3.9C). In summary, the interaction analysis indicates that the MlaC, MlaD and LptC proteins could interact with each other and work in synergy.

3.3.9. *Ec*MlaE and its orthologues have evolved separately from other TMDs

The interactome analysis clearly shows that MlaD interacts with MlaE. Furthermore, the cryo-EM structure of the MlaFEDB complex suggests that MlaD is in constant association with MlaE (PDB id: 6XBD; Coudray et al., 2020). Although detailed structural information about the Mce complex from *M. tuberculosis* is not available, the cryo-EM structures of the Mce complex from *M. smegmatis* indicate similar macromolecular arrangement (PDB ids: 8FED, 8FEE and 8FEF; Chen et al., 2022). The comparable organization has also been reported in other MlaD domain-containing systems (Lu and Benning, 2009; Ekiert et al., 2017). However, such an arrangement is not observed in any typical ABC transporters. In order to find out if such organizational differences led to any significant variation in the TMD, a phylogenetic study was performed by taking into consideration the TMDs of Mla, PqiB and Yeb systems of *E. coli*, Mce systems of *M. tuberculosis* and *M. smegmatis,* TGD system of *A. thaliana* and other ABC transporters of *E. coli* (Lpt and Mal transporters) (Table A1.7). Interestingly, analysis shows three distinct clusters – I, II and III. The protein *Ec*MlaE belongs to Cluster I, constituted by the TMDs of Mce systems of *M. tuberculosis* and *M. smegmatis*, as well as the TGD system of *A. thaliana.* On the other hand, Cluster II is constituted by TMDs from other ABC transporters of *E. coli*. Surprisingly, Cluster III is formed by the TMDs of the Pqi and Yeb systems. Thus, the

analysis shows that TMDs associated with MlaD domain-containing proteins form two separate clusters (I and III). Furthermore, *Ec*MlaE is evolutionarily closer to TMDs of *M. tuberculosis*, *M. smeagmatis* and *A. thaliana* (Figure 3.10.A).

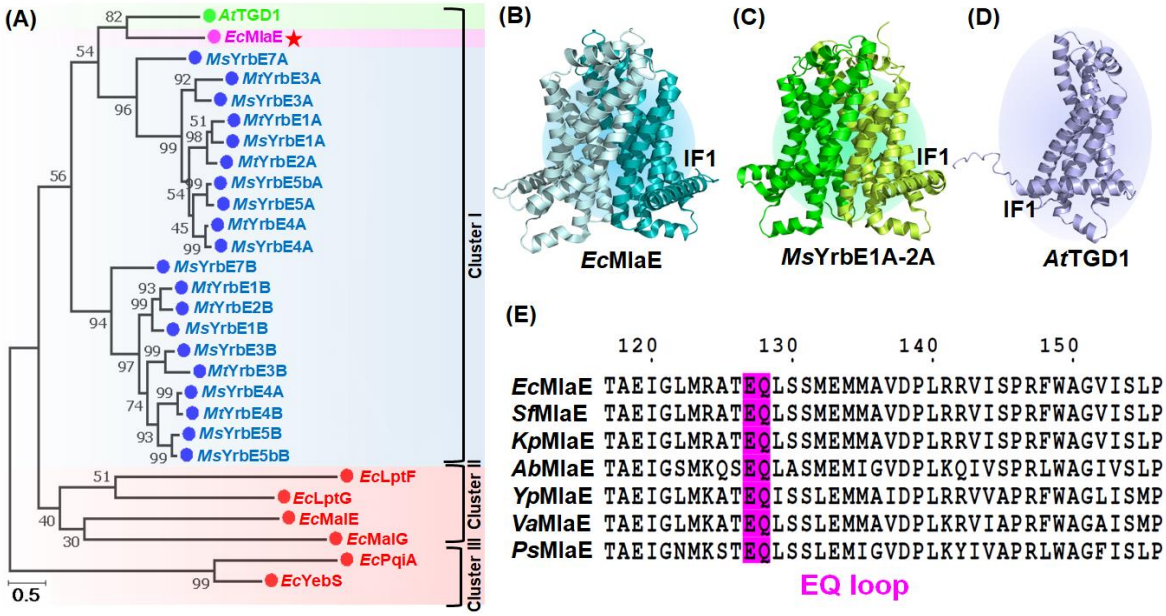


Figure 3.10. Sequence and structure analyses of EcMlaE. *(A) Phylogenetic tree analysis of EcMlaE, its orthologues and other TMDs. Three clusters (I, II and III) are labeled. The TMDs from Mla (E. coli), TGD (A. thaliana) and (M. tuberculosis/ M. smegmatis) are highlighted in magenta, green and blue. The remaining TMDs are highlighted in red. Further details are provided in Table A1.7. Numbers at each internal node of the phylogenetic tree signify the probability of its occurrence (in percentage) out of 1000 bootstrap replicates. (B-D) Overall structures of EcMlaE (PDB id: 6XBD, palecyan-cyan), MsYrbE1A-2A (PDB id: 8FEF, green and limon, respectively) and AtTGD1 (modeled, lightblue), respectively. (E) MSA of EcMlaE and its orthologues from Gram-negative bacteria. The conservation of the EQ loop is highlighted in magenta.*

3.3.10. *Ec*MlaE and its orthologue possess N-terminal interfacial helix

After getting an insight into the evolutionary relationship among the TMDs, a structural study of the members of the three clusters was made. For this, the crystal/cryo-EM (PDB ids: 6XBD (*Ec*MlaE), 6MHZ (*Ec*LptFG), 2R6G (*Ec*MalFG) and 8FEF(*Ms*YrbE1A-1B)) modeled structures (*Ec*PqiA, *Ec*YebS, *Mt*YrbE1A-4B and *At*TGD1) were obtained

from PDB and AlphaFold database, respectively. The study shows that *Ec*MlaE, *Ms*YrbE1A-7B, *Mt*YrbE1A-1B and *At*TGD1 (Cluster I) contain N-terminal interfacial helix (IF1) (Figure 3.10.B and Figure A1.8A-F). This IF1 helix is important for interaction with the helical region of MlaD domain-containing proteins (Coudray et al., 2020). On the other hand, cluster II members do not possess the IF helix. Surprisingly, the helix is absent in members of cluster III even though they are functionally associated with MlaD domain-containing proteins. In fact, both PqiA and YebS are similar in structure but significantly different from any TMD associated with ABC transporters.

3.3.11. *Ec*MlaE and its orthologue show variation in coupling helix

It has been previously reported that the TMDs of ABC transporters possess a highly conserved EAA region/motif present in the coupling helix that interacts with their cognate ATPase (Mourez et al., 1997; Casali and Riley, 2007). In order to find out whether the motif is conserved, the TMD sequences of *Ec*MalFG proteins were taken as references as they are well-studied (Mourez et al., 1997). Accordingly, the MSAs of the collected sequences were systematically performed. The MSA of *Ec*MlaE and its orthologues from Gram-negative bacteria shows the presence of a conserved EQ loop instead of an EAA loop (Table A1.7 and Figure 3.10.E). The MSA of *Ec*MlaE, *At*TGD1, *Ec*MalF and *Ec*MalG clearly shows that the EAA motif is conserved in the latter two sequences. However, in the former two sequences, the EQ loop is conserved, which is situated before the EAA loop (Figure A1.8G). Interestingly, in *Mt*YrbE1A-4A and *Ms*YrbE1A-7B, the EEXDA loop is conserved (Figure A1.8H). However, in *Ec*PqiA and *Ec*YebS, there is no presence of either EAA or EQ or EEXDA loop, indicating that these two proteins do not have any cognate ATPase.

3.3.12. *Ec*MlaF and its orthologues have evolved separately from other NBDs

The NBDs are the signature feature of ABC transporters (Dassa, 2011). The cryo-EM structure of the MlaFEDB complex from *E. coli* reveals that the NBD MlaB remains associated with MlaF in the cytoplasmic side (Coudray et al., 2020). However, such interactions of the auxiliary protein with NBDs have not been observed in the case of other ABC transporters. Additionally, MlaF and its orthologue from the Mce system of *M. tuberculosis*, MceG, belong to the Mkl family. As per the available classification scheme, MlaF is involved in the import of PL from the OM to IM (Dassa and Bouige P, 2001; Dassa, 2011). In order to probe the evolutionary relationship of MlaF with other Mkl and non-Mkl proteins, a phylogenetic tree analysis was performed. For the

analysis, *Ec*MlaF orthologues from Gram-negative bacteria, Actinomycetales, as well as plants were taken into consideration. Also, various other non-Mkl NBD sequences from *E. coli* were taken into account (Table A1.8). The phylogenetic tree forms two distinct clusters – I and II. Cluster I is constituted by *Ec*MlaF and other Mkl proteins. This clearly indicates that the Mkl NBDs evolved separately from other NBDs of ABC transporters. On the other hand, Cluster II is formed by non-Mkl proteins. Also, three distinct sub-clusters are observed in Cluster I (Ia, Ib and Ic). The sub-clusters Ia, Ib and Ic are formed by *Ec*MlaF orthologues from Gram-negative bacteria, Actinomycetales and plants, respectively (Figure 3.11A).

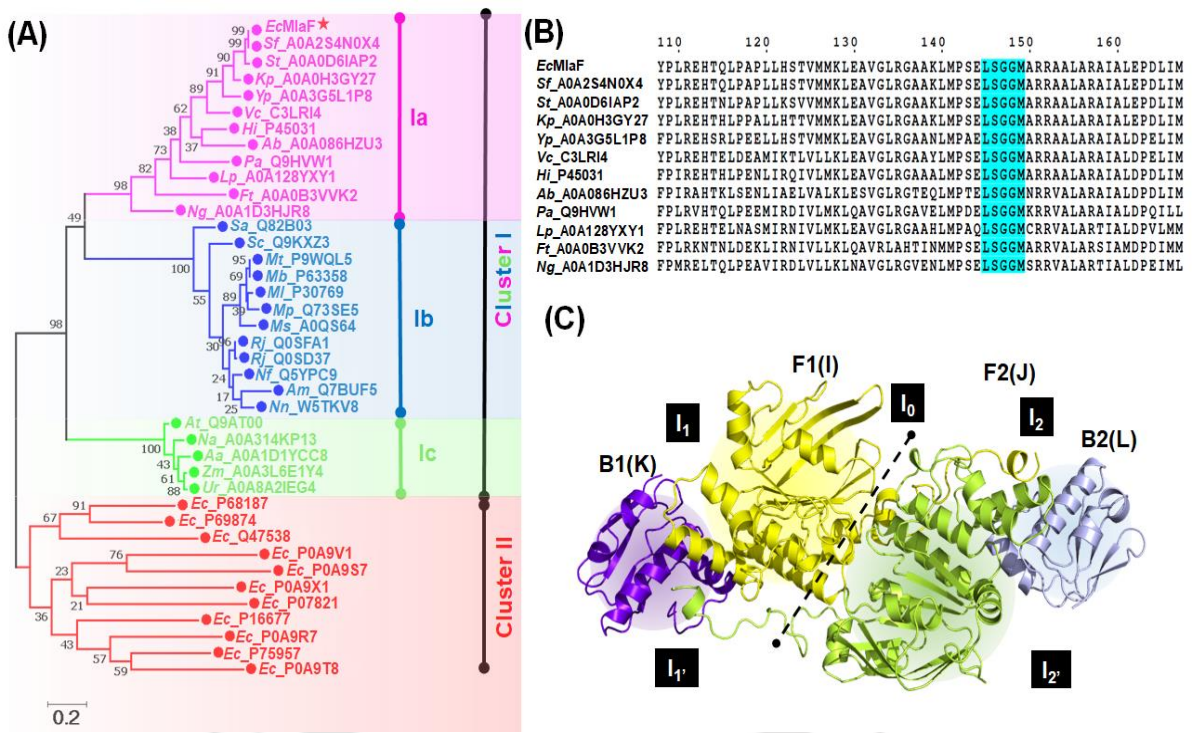


Figure 3.11. Sequence and structure analysis of EcMlaF. *(A) Phylogenetic tree analysis of Mkl NBDs, including EcMlaF and its orthologues from Gram-negative bacteria (highlighted in magenta), Actinomycetales (highlighted in blue) and plants (highlighted in green) along with other non-Mkl NBDs (highlighted in red). Two clusters (I and II) are labeled. Further details are provided in Table A1.8. Cluster I is further divided into three sub-clusters (Ia, Ib and Ic). EcMlaF is marked by a red star. Numbers at each internal node of the phylogenetic tree signify the probability of its occurrence (in percentage) out of 1000 bootstrap replicates. (B) MSA of EcMlaF and its orthologues from Gram-negative bacteria. The LSGGM motif is highlighted in cyan. (C) Interface analysis of EcMlaF-EcMlaB complex (PDB id: 6XBD). The complex can be*

divided into two halves (shown with a dotted line), each containing one unit of EcMlaF (F1, yellow, chain I; F2, limon, chain J) and EcMlaB (B1, purpleblue, chain K; B2, lightblue, chain L). The interfaces of F1-F2, F1-B1, F1-B2, F2-B2 and F2-B1 are designated as I_0, I_1, $I_{1'}$, I_2 and $I_{2'}$, respectively. Further details are provided in Table A1.9.

3.3.13. *Ec*MlaF and its orthologues show variation in signature motifs

The members of the (sub)clusters were systematically analyzed. NDBs possess the conserved motif LSGGQ (Signature motif) (Kolich et al., 2020). However, the MSA of *Ec*MlaF and its orthologs from Gram-negative bacteria show the presence of the LSGGM motif (Cluster Ia). Furthermore, this motif is present in the case of the Mkl proteins from plants (Cluster Ic). Interestingly, in Actinomycetales, the motif ISGGM is present (Cluster Ib). In the case of other Gram-negative bacteria, the LSGGQ motif is overall conserved, though the motif varies in some NBDs (Figure 3.11B and Figure A1.9A-C).

3.3.14. The association of *Ec*MlaB gives rise to four additional interfaces

The protein of MlaB is a cytoplasmic component that remains associated with MlaF and contains a STAS domain. Interestingly, the STAS domain is found in bacterial anti-sigma factor antagonists (ASA) and the C-terminal region of SLC26 (SulP) anion transporters. The ASA of *Bacillus,* viz. SpoIIAA works in association with Anti-Sigma Factor SpoIIAB and together, they regulate the gene expression during the process of sporulation. Similar to MlaB, SpoIIAA has also been implicated in having NTPase activity (Campbell et al., 2002). As per the available structural information (PDB ids: 6XBD, 6XGY and 6XGZ), MlaF units (F1 and F2) form a dimer giving rise to one interface (I_0). Interestingly, each MlaF unit (F1 and F2) remains associated with one copy of MlaB (B1 and B2). As a result, two new interfaces are formed, which we have named I_1 (F1-B1) and I_2 (F2-B2). However, each MlaF unit has a unique C-terminal extension (CTE), with the help of which it reaches out to the MlaB unit of the other MlaF unit. As a result, again, two new interfaces are formed, which we have named $I_{1'}$ (F1-B2) and $I_{2'}$ (F2-B1) (Figure 3.11C and Figure A1.10A-E). Thus, in addition to the dimeric interface, as seen in typical NBDs, the MlaF and MlaB give rise to four additional interfaces. A similar arrangement is observed in the cases of *A. baumannii* and *Pseudomonas aeruginosa* (PDB ids: 7D06 and 7CH9, respectively) (Table A1.9).

3.3.15. *Ec*MlaB possesses a DSSG motif

Additionally, STAS domains have been reported to bind to NTP (Nucleotide triphosphate), wherein the DSSG motif (second S) undergoes phosphorylation (Dever et al., 1987; Najafi et al., 1996). In order to probe whether *Ec*MlaB (*Ec*_P64602) can bind to ATP molecule or possess any such motif, the amino acid sequences of *Ec*MlaB and various STAS-domain-containing proteins from other organisms, were collected from the UniProtKB database and analyzed at the sequence level (Table A1.10). For the analysis, the STAS domain of SpoIIAA protein from *B. subtilis* (*Bs*_P10727), which has been reported to have NTPase activity, was taken as the reference. The MSA of the collected sequences clearly shows the conservation of the DSSG motif, which is associated with GTP binding. Also, the sequence GLGVILGR present in *Bs*_P10727 is partially conserved in the other STAS domains, including *Ec*MlaB (Figure 3.12A). The sequence is comparable to the conserved Walker A motif, GXXXXGK, that is, critical for NTP binding among the NBDs (Najafi et al., 1996). Also, the study shows that the DSSG, as well as GLGVILGR, overlap each other. In the DSSG motif, the serine residue present in the second position is the site for phosphorylation and is critical for the functioning of *Bs*_P10727. In the case of *Ec*_P64602, the serine residue is replaced by threonine residue, which also can act as a site of phosphorylation and can be critical for the maintenance of the proper function of the Mla system. This hypothesis is supported by the finding that the mutation of threonine residue to alanine residue in the MlaFEB complex results in a decrease in ATPase activity (Thong et al., 2016). Among the collected sequences, three proteins, viz. *Mt*_P9WGF7, *Ec*_P0AFR2 and *Hs*_P40879 do not have the serine/threonine residue. This is of indicative that the function of these three proteins might not be dependent on the process of phosphorylation. To further investigate the presence of DSSG as well as the Walker A-like motif, the MlaB proteins from different Gram-negative bacteria were analyzed. The MSA clearly indicates the general conservation of motifs in all the selected MlaB proteins, along with the preservation of the phosphorylation site (Figure 3.12B). Based on this sequence-based analysis, it can be proposed that *Ec*_P64602, as well as its orthologues from different Gram-negative organisms, might bind to ATP.

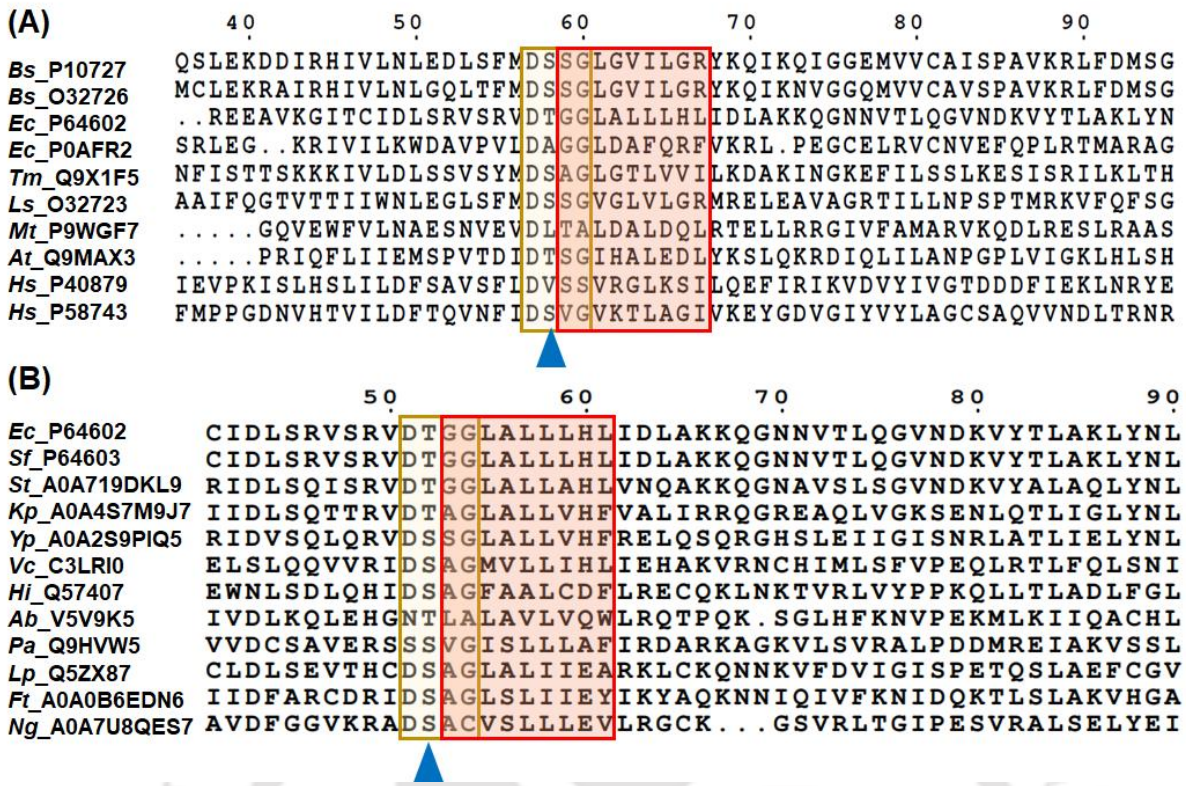


Figure 3.12. Sequence analysis of STAS domain-containing proteins. *(A) MSA of EcMlaB (Ec_P64602) and other STAS-domain containing proteins from different source organisms. (B) MSA of Ec_P64602 and its orthologues from other Gram-negative bacteria. Further details provided in Table A1.10. Conservation of DSSG and Walker A-like motifs are highlighted in yellow and red boxes, respectively. The conservation of Serine residue (site of phosphorylation) is highlighted by a blue upward arrow.*

3.3.16. *Ec*MlaA possesses a doughnut-shaped structure

The protein *Ec*MlaA is situated in the OM and is mainly responsible for the rearrangement of the PLs in the membrane without the aid of ATP hydrolysis. *Ec*MlaA forms a complex with *Ec*OmpC and the complex aids in the maintenance of OM asymmetry (Malinverni and Silhavy, 2009; Chong et al., 2015). The crystal structures of MlaA from *Klebsiella pneumoniae* and *Serratia marcescens* (*Kp*MlaA, *Sm*MlaA; PDB ids: 5NUQ, 5NUP, 5NUO and 5NUR) are available. As per these structures, the MlaA protein forms a stable complex with *Ec*OmpF/C (Abellón-Ruiz et al., 2017). However, structural information on the *Ec*MlaA-Omp complex is not available. In order to get insights, a preliminary sequence-based analysis was performed wherein

Kp/*Ec*/*Sm*MlaA along with *Mt*Mce(1-4)E along with *Ec*PqiC were taken into account. Typical lipoproteins possess a positively charged *n*-region, hydrophobic/uncharged h-region followed by a consensus lipobox sequence ([LV][ASTVI][GAS][C]) (Madan Babu and Sankaran, 2002). The MSA results clearly show the presence of these three distinct regions (Figure A1.11A). As the crystal structure of *Ec*MlaA is not available, the three-dimensional model of *Ec*MlaA (*Ec*MlaA$_m$) was obtained from the AlphaFold database. At the sequence level, *Ec*MlaC shares decent similarity with *Kp*MlaA (Identity: 88%, similarity: 94%, query coverage: 100%) and *Sm*MlaA (Identity: 78%, similarity: 87%, query coverage: 100%). As per the model, *Ec*MlaA$_m$ comprises three distinct regions – the N-terminal α-helix (containing the lipobox), the core region and the C-terminal helix (CTH). Surprisingly, the CTH region is absent in the crystal structures of *Kp*MlaA and *Sm*MlaA even though the clone construct contains the region (Figure 3.13A). However, the CTH regions are present in the AlphaFold predicted structures of *Kp*MlaA and *Sm*MlaA (Figure 3.13B). For the purpose of analysis, the core region of *Ec*MlaA$_m$, *Kp*MlaA and *Sm*MlaA were considered. As per *Ec*MlaA$_m$, the core region of the protein comprises three distinct components – seven helices (H1, H2, H3, H4, H5a-b, H6 and H7), a pore loop (between H4 and H5a) and a ridge (from H5a to N-terminal of H6). This is similar to the crystal structures of *Kp*MlaA and *Sm*MlaA. It has a ring-shaped structure resembling a doughnut and is predominantly α-helical in nature (Figure 3.13C).

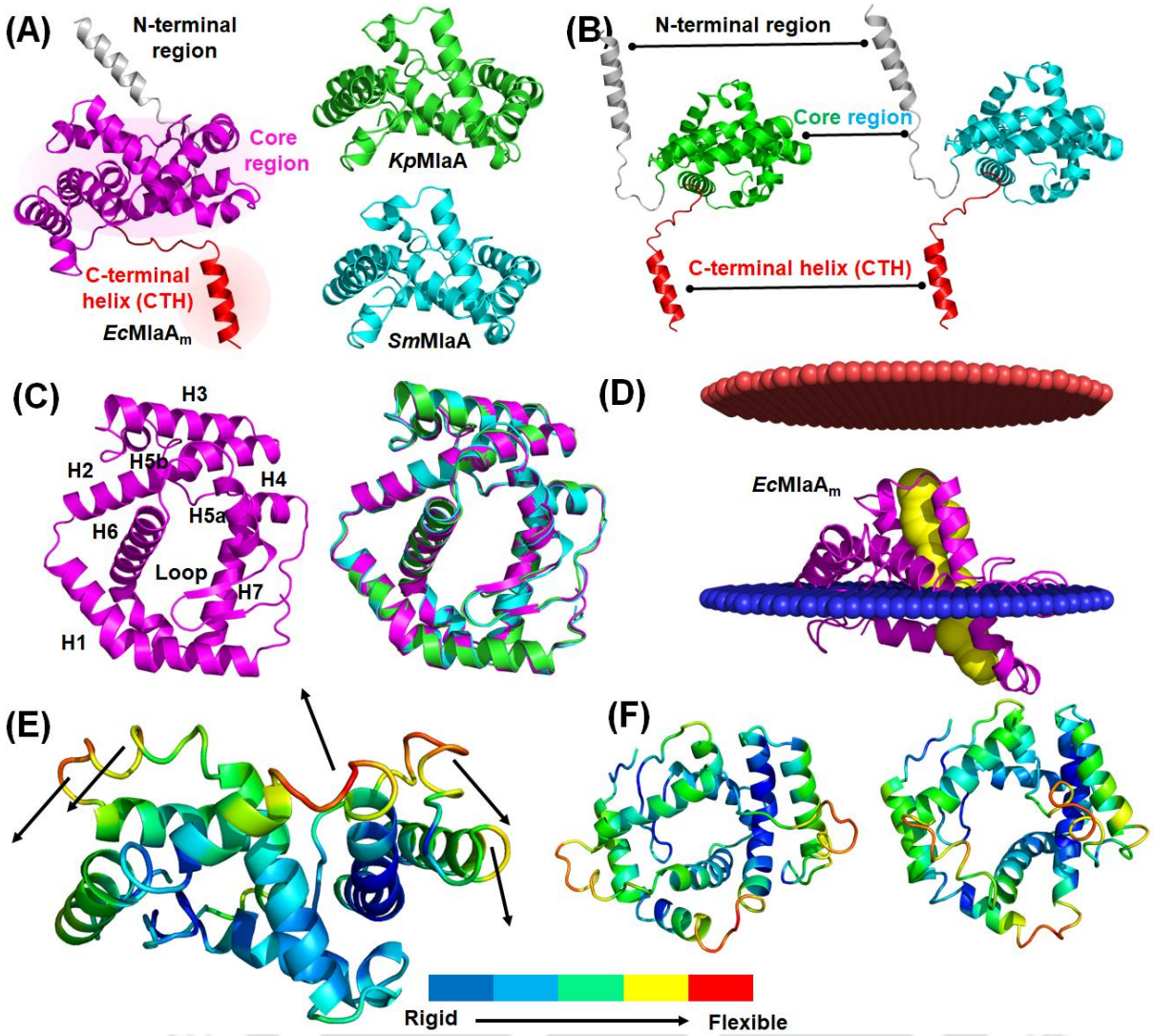


Figure 3.13. Structure analysis of EcMlaA. *(A) (Left) Overall structure of modeled EcMlaA (EcMlaA$_m$) obtained from AlphaFold database. The protein is divided into three distinct regions, the N-terminal region (grey), the core region (magenta) and the C-terminal helix (red). Crystal structure (Right, top) of KpMlaA (PDB id: 5NUP, green) and (Right, bottom) SmMlaA (PDB id: 5NUQ, cyan). (B) (Left, right) Modeled structures of KpMlaA and SmMlaA, respectively, highlighting the three distinct regions. (C) Structural comparison of EcMlaA$_m$. (Left) The overall structure of EcMlaA$_m$ (magenta) with its constituent helices (H1-H7) and loop. (Right) Structural superimposition of EcMlaA$_m$ (magenta) upon KpMlaA (PDB id: 5NUP, green) and SmMlaA (PDB id: 5NUQ, cyan). (D) Orientation analysis of EcMlaA$_m$ (magenta). The outer and inner leaflets of OM are represented by red and blue planes, respectively. The central tunnel is depicted as a yellow surface. (E) ANM analysis of EcMlaA$_m$. Protein dynamics observed via ANM. The arrows indicate the direction in which the parts of the protein*

are moving. (F) Different conformations (Left, open; right, close) of EcMlaA$_m$ observed via ANM analysis. The increasing shade from blue to red signifies an increase in the flexibility of protein regions (Figures E and F).

3.3.17. *A* major portion of *Ec*MlaA remains embedded in the OM and gives rise to a dynamic central channel

In order to understand the localization of MlaA, the available crystal structures, along with *Ec*MlaA$_m$, were subjected to orientation analysis. Interestingly, in the crystal structures, *Kp*MlaA and *Sm*MlaA were found to be present in the complex with Omp (PDB ids: 5NUQ (*Sm*MlaA-*Ec*OmpF), 5NUP (*Kp*MlaA-*Kp*OmpC) and 5NUO/5NUR (*Kp*MlaA-*Ec*OmpF)) (Table A1.11). Hence, in order to get better insights, the complexes were also included in the analysis. As anticipated, the major portions of the Omp remain embedded in the membrane, with the interconnecting loops protruding out of the membrane. Surprisingly, a major portion of *Kp*MlaA and *Sm*MlaA remain embedded in the membrane suggesting that MlaA is an unusual integral membrane protein (Figure A1.11B-A1.11E). Analysis performed using only *Kp*MlaA and *Sm*MlaA (after extracting them from the complex) as well as *Ec*MlaA clearly shows that H6 is perpendicular to the membrane. This gives rise to a ridge-like arrangement that leads to the outer leaflet. On the other hand, only a small portion protrudes out of the membrane constituting the other end (Figure 3.13D and Figure A1.11F-A1.11G).

After getting an insight into the orientation, a study was made to find out the distribution of charge in the central channel of the protein. For this, charge, as well as channel mapping of *Kp*/*Sm*/*Ec*MlaA, was performed. The study shows that the periplasmic-facing side of MlaA is hydrophilic in the region with an amphipathic charge distribution. However, the major of the protein is hydrophobic in nature, helping it to remain embedded in the membrane. Surprisingly, the channel is not entirely hydrophobic even though it serves as a passage for PL movement (Figure A1.12A-A1.12C). Furthermore, channel mapping reveals that MlaA would give rise to an uneven tunnel-like arrangement that is not straight but convoluted. The tunnel would stretch from the periplasmic-facing side and extent to the outer leaflet via the ridge (Figure 3.13D and Figure A1.11F-A1.11G). As per the analysis using CASTp, the overall volumes of the tunnels in *Kp*/*Sm*/*Ec*MlaA are similar (1202, 1521 and 1236 Å^3, respectively). Furthermore, in order to understand the conformational dynamicity of *Ec*MlaA$_m$, an

ANM analysis was performed. The analysis clearly shows that loop regions connecting the helices would demonstrate maximum fluctuations (Figure 3.13E). Such movements modulate the dimensions of the central channel leading to the expansion and contraction of protein, resulting in the opening and closing of the *Ec*MlaA$_m$ (Figure 3.13F).

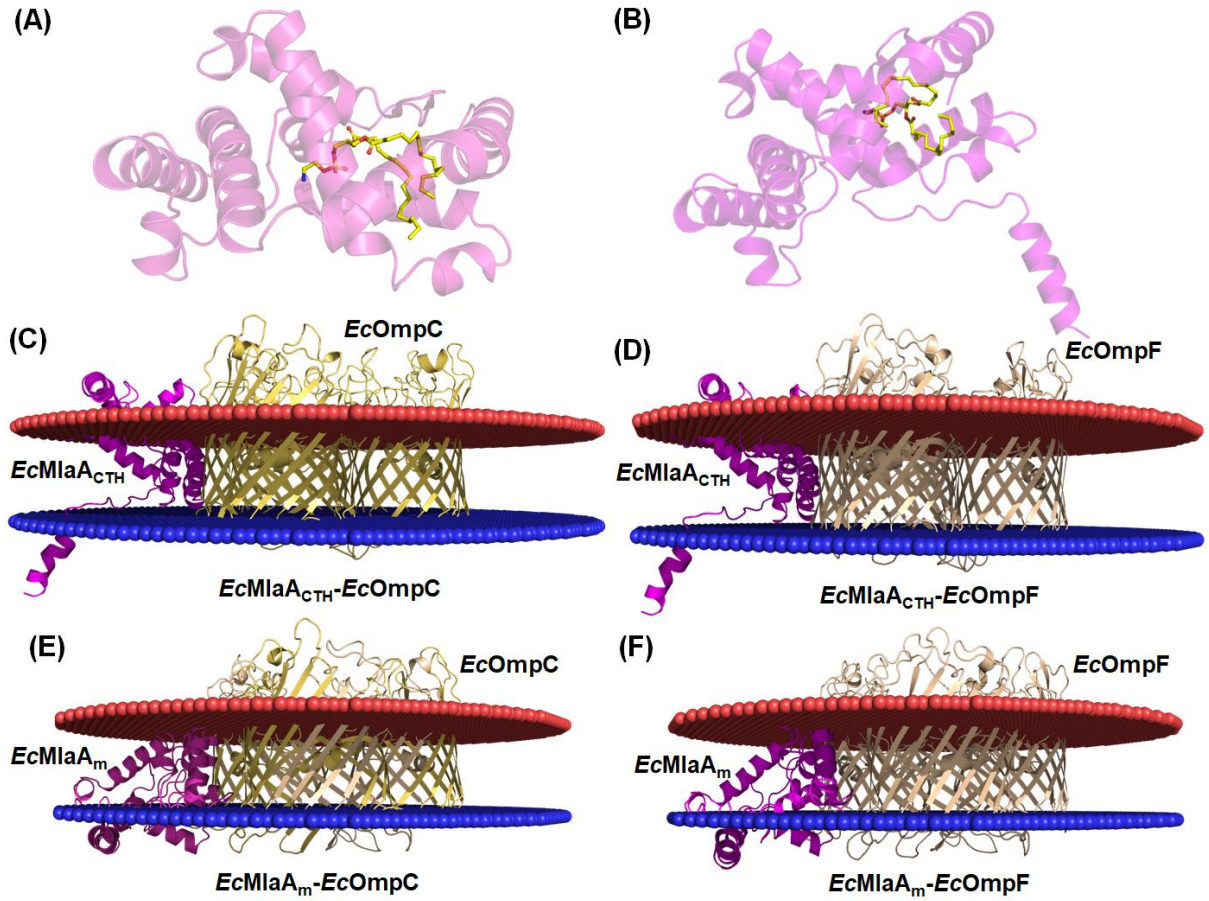


Figure 3.14. Analysis of protein-ligand and protein-protein docking. *(A-B) Molecular docking of EcMlaA$_m$ (magenta) and EcMlaA$_{CTH}$ (magenta) with PEF (yellow), respectively. Protein-protein docking of EcMlaA$_{CTH}$ (magenta) with (C) EcOmpC (yelloworange) and (D) EcOmpF (wheat). Protein-protein docking of EcMlaA$_m$ with (E) EcOmpC (yelloworange) and (F) EcOmpF (wheat). The outer and inner leaflets of OM are represented by red and blue planes, respectively.*

3.3.18. The central channel of *Ec*MlaA can bind PL, while the CTH region might determine the orientation of the protein

To date, the crystal structure of MlaA in complex with PL is not available. In order to find out whether MlaA can bind PL, a molecular docking study was made wherein the PL molecule was docked with *Ec*MlaA$_m$ (no CTH) and *Ec*MlaA$_{CTH}$ (with CTH). This was done so as to find out whether the CTH region has any negative impact on PL binding. The results show that in both cases, the PL molecule can bind to the central channel of *Ec*MlaA$_{m/CTH}$ with similar binding energies (-4.11 and -3.32 kcal mol^{-1}, respectively). This indicates that the CTH region does not hamper the binding of PL to MlaA (Figure 3.14A-3.14B).

The crystal structures of *Kp*/*Sm*MlaA in complex with OmpC/F are available, which provide insight into the macromolecule arrangement. However, the structural information is not available for *Ec*MlaC. Hence, a study was made to construct the complex and understand the organization. The analysis of the crystal structure reveals *Kp*/*Sm*MlaA-OmpC/F complex is mainly stabilized by salt bridges, hydrogen bonds and non-bonded contact. Also, irrespective of the Omp, MlaA proteins bind to the interface of two dimers of Omp trimers (Figure A1.13A-A1.13D). Interactome analysis reveals that *Ec*MlaA would interact with *Ec*OmpC/F. Hence, protein-protein docking experiments were performed for *Ec*MlaA$_{m/CTH}$ (ligand) and *Ec*OmpC/F (receptor). The results indicate that in both cases, *Ec*MlaA$_{m/CTH}$ would bind to the dimeric interface of the *Ec*OmpC/F trimer. However, the orientations of *Ec*MlaA$_{m/CTH}$ in the complex are quite different from that of available crystal structures and from each other (Figure 3.14C-3.14F). This is suggestive that *Ec*MlaA$_{m/CTH}$ can undergo changes in the orientation owing to its dynamic nature. It can also be speculated that the CTH might play a critical role in determining the orientation of *Ec*MlaA$_{CTH}$.

3.4. Discussion

In *E. coli*, the Mla system plays a significant role in maintaining the asymmetry of OM by transporting PLs in between the OM and IM. The Mla system belongs to the ABC transporter superfamily and contains six proteins, MlaABCDEF. The IM-associated periplasmic protein MlaD possesses a highly conserved MlaD/Mce domain. The domain is found in proteins of Gram-negative bacteria, actinobacteria as well as plant chloroplast and is often associated with the trafficking of hydrophobic molecules like

PLs, cholesterol and fatty acids (Ekiert et al., 2017). However, detailed analysis and characterization of the MlaD domain have not been reported in the literature. In this study, a sequence- and structure-based characterization of the MlaD domain, in particular from the Mla system of *E. coli*, has been performed. Additionally, an attempt has been made to provide insight into the MlaD domains of various architectures and the associated components of the MlaD domain systems.

3.4.1. The MlaD domain possesses a diverse domain architecture but a conserved amino acid profile

The analysis reveals that the MlaD domain is present across different architectural types and organisms. This suggests that the domain could have undergone recombination and might participate in auxiliary functions other than PL transport (Perkowski et al., 2016). For instance, the LinABC transport system (LinKLMN) in *Sphingobium japonicum,* which comprises a membrane-bound periplasmic protein LinM possessing a single MlaD domain, is involved in the import of synthetic chlorinated pesticide γ-hexachlorocyclohexane (γ-HCH) for degradation (Lal et al., 2010; Nagata et al., 2011). The study shows that the MlaD domains from *E. coli* have diverse sequences but comparable amino acid profiles. Results demonstrate that the hydrophobic amino acids valine and leucine, as well as the amino acid glycine, are abundantly present in the MlaD domains. The amino acid glycine is known to adapt dynamic conformations, while valine has been reported to be a strong β-strand former that enhances the flexibility of the β-regions of proteins (Kočа et al., 1994). Hence, it can be speculated that the abundant presence of valine and glycine residues would help impart dynamic conformations to the MlaD domain. Also, the amino acid cysteine is almost completely absent in the MlaD domains indicating the absence of disulfide bonds. This would, in turn, confer conformational plasticity to the MlaD domain (Khoo and Norton, 2011). As a whole, the amino acid profile analysis suggests that the MlaD domain is predominantly hydrophobic and flexible.

3.4.2. The MlaD domain possesses a conserved N-terminal region with a well-preserved motif

The study also reveals that most of the conserved residues are mainly confined to the N-terminal region of the MlaD domain. Residue Gly29, situated in the midst of the β3 strand and the part of the motif L-X_2-G-X_3-V-X_3-G-X_2-V-G-X-V (where X stands for any

amino acid), is preserved across the 20 different domain architecture types. We propose this motif be referred to as the MlaD domain motif. Although the rationale for such conservation could not be proposed, it can be speculated that the motif might be associated with the functions of PL trafficking or domain movement. This is because such conservation of the glycine residue has been previously reported to be crucial for the movement and conformational changes of neighbouring amino acids in the LolA protein (Takeda et al., 2003).

3.4.3. *Ec*MlA_MlaD1 is not evolutionarily closer to the Mla domains from *E. coli*

In addition to the motif, 26 out of 36 MlaD domains taken as references possess a conserved PLP region. These include the MlaD domains of the proteins MlaD, Mce and TGD2 from *E. coli*, *M. tuberculosis* and *A. thaliana*, respectively. In contrast, the remaining 10 domains of the proteins PqiB and YebT from *E. coli* vary in lengths and amino acid compositions. The evolutionary analysis of the selected domains reveals that the MlaD domains cluster according to their architecture types, that is, domains occurring in single copy cluster together irrespective of the source organisms (e.g. *Ec*MlA_MlaD1, *Mt*1ABI_MlaD1-*Mt*4FA_MlaD1 and *At*TgA_MlaD1). It can, thus, be hypothesized that the proteins MlaD, Mce and TGD2 from *E. coli*, *M. tuberculosis* and *A. thaliana*, respectively, have diverged from a common ancestor. Interestingly, some of the Mce proteins (Mce1A, 1D, 2A, 2F, 3A, 4A and 4D) have adapted an additional domain, Mce4_CUP1. The Mce4_CUP1 domain belongs to the “family of putative Mce4 transporters of cholesterol” and is always observed in association with the MlaD domain. Considering the association of the Mce4_CUP1 domain with cholesterol transport and the MlaD family, it might play a critical role in the process of cholesterol import in *M. tuberculosis*. Also, the distinct clustering of the MlaD domains occurring in multiple copies hints that they might have developed separately from a common ancestor during the course of evolution. Based on this, it can be speculated that the Mla system having a single copy of the MlaD domain might have evolved initially, followed by the Pqi and Yeb systems having multiple copies of the MlaD domain for PLs trafficking. This is also reflected through their variation in the organization of the MlaD domains. The rationale behind the presence of three PL transporters in *E. coli* might be associated with ligand size as well as the timely maintenance of OM asymmetry. For the Mla system, the PL is bound in the interior region of the protein MlaC (Ekiert et al., 2017). A large molecule like PL requires a higher surface area to bind to the protein with multiple contact sites (Reynolds et al.,

2008; Smith et al., 2012). Eventually, the protein MlaC requires to release the bound PL molecule and the entire process of transport would be time-consuming. However, there is no such requirement for the binding and release of PLs for proteins PqiB and YebT due to their tunnel-like structures. Thus, Pqi and Yeb systems might allow the movement of a wider range of PLs and also within a short time. The phylogenetic analysis also establishes the relationship between the domains and the permeases based on the driving force of PL transport. The MlaD domains of Mla, Mce and TGD systems transport PLs at the expense of ATP hydrolysis with the aid of TMDs belonging to the MlaE family. However, the driving force responsible for the transport of PLs by the MlaD domains of PqiB and YebT is still elusive. It has been speculated that the proton motive force or facilitated passive equilibration of PLs might play a significant role in the process of transportation (Ekiert et al., 2017; Isom et al., 2020).

3.4.4. MlaD domain has a conserved structure with variable loops

Although *Ec*MlA_MlaD1 is overall superimposable on other MlaD domains, it shows highly variable loop regions. The modeled structures of the proteins Mce1A and TGD2 from *M. tuberculosis* and *A. thaliana*, respectively, contain a core structure of the domain that is similar to that of the *Ec*MlA_MlaD1 (Lu and Benning, 2009; Khan et al., 2018). In the case of the largest MlaD domain present in *Ab*MlA_MlaD1, there is the presence of insertion of two α helices between β4 and β5 strands (Kamischke et al., 2019). In spite of that, the overall topology of the MlaD domain is similar to that of *Ec*MlA_MlaD1. This indicates that the overall structure of the MlaD domain comprising β-sheets is conserved across different domain architectures and organisms.

3.4.5. MlaD domain imparts flexibility to the protein

A dominant presence of hydrophobic amino acids and the PLP region in the MlaD domain create a hydrophobic central pore. The variation in amino acid compositions, lengths as well as organizations of the PLP regions results in varying diameters of the central pore. Considering the conformational dynamicity of the MlaD domain as well as the PLP region, it seems likely that the protein might modulate the diameter of its hydrophobic channel in order to accommodate a wider range of PLs. Such a mechanism has been suggested in the literature as well (Isom et al., 2020). This could be critical for providing an ideal environment for PL trafficking as they tend to show high flexibility and movement when free of the membrane arrangement (Surovtsev and

Dzuba, 2014). Thus, the plasticity of the channel would help hold the PLs and provide directionality to their movements. Furthermore, unlike the protein MlaD, the proteins PqiB and YebT require the presence of the repeated MlaD domain to traverse the periplasm. In the case of the protein MlaD, the lack of multiple numbers of MlaD domains is compensated by the involvement of the protein MlaC that freely moves in the periplasmic space. Interestingly, the PLP region of *Ec*MlA_MlaD1, along with the MlaE subunit, has been reported to play a critical role in re-orienting the PLs for their binding to MlaC. In fact, the conformational changes transduced by the MlaE subunit produce the conformation in MlaD that is necessary for PL extraction by MlaC (Coudray et al., 2020). Owing to the conservation of this PLP region in the MlaD domains of Mce proteins from *M. tuberculosis* and TGD2 protein from *A. thaliana*, along with their association with ABC transporter permease MlaE, it can be speculated that the region would have a similar function of re-orientating substrates in the three systems. On the other hand, the variations in the PLP regions of the MlaD domains of PqiA and YebT might be associated with the ATP-independent transport of PLs. Furthermore, the PLPs have been reported to possess different conformational dynamics that might be modulated by the TMDs or other components yet to be identified (Isom et al., 2020). It is likely that the MlaD domains of PqiB and YebT have evolved separate mechanisms of PL transport that are independent of ATP hydrolysis. This is suggestive that MlaD domains might have developed the structural attributes, especially the PLPs, based on the mechanisms of PL transport. This observation, for the first time, provides insights into the structure-function-evolutionary relationship of the MlaD domain. Even though the overall topology of the MlaD domain is preserved, its presence does not necessarily lead to the oligomerization of the protein. This is evidenced by the involvement of the residues from within and outside the domain at dimeric interfaces in the hexameric assembly of the MlaD domain. Moreover, only a few interfacial residues are conserved that too for varying degrees.

3.4.6. MlaD domain-containing systems have diverse interactomes

In *E. coli*, the protein MlaD utilizes a structural homolog of the NTF2 superfamily, that is, protein MlaC for PL trafficking, whereas no such component is required for the proteins PqiB and YebT. This could again be attributed to the presence of multiple copies of the MlaD domains in the latter two proteins, which traverse through the periplasm to reach the OM. Interestingly, the Mce proteins containing a single copy of the MlaD domain are also dependent on NTF2-like proteins. However, unlike the

protein MlaD, a total of 13 membrane-bound proteins that are structural homologs of the NTF2 superfamily, known as Mam and Omam, are supposedly associated with the core Mce protein complex. Although a clear functional understanding of these proteins is not known, it has been proposed that they are involved in imparting stability to the Mce complex. Hence, it can be speculated that the sophisticated organization of the Mce transporter, as well as the different cell envelope architecture of *M. tuberculosis* require such an elaborate arrangement. The involvement of the NTF2 protein(s) also hints toward the similarity in the origin and mode of action of both the Mla and Mce systems. In *A. thaliana*, a functional analog of the protein MlaC is found as a small glycine-rich protein TGD5, which is not the structural homolog of the NTF2 superfamily (Fan et al., 2015). It is plausible that during the course of evolution, the chloroplast lipid trafficking system might have adapted a different structure in order to aid lipid trafficking. Thus, it seems that the involvement of the homologs of the NTF2 superfamily in lipid trafficking might be confined to prokaryotes, thus, complementing the previous report (Wong et al., 2017). Further, the role of MlaC in the case of the Mla system might not just be confined to PL transport. It might also act as a mediator between the Mla and Lpt systems. This could, in turn, regulate the movement of PLs and LPSs between the membranes leading to the maintenance of OM asymmetry.

3.4.7. *Ec*MlaE has a unique evolutionary lineage

The analysis of *Ec*MlaE reveals that it possesses an atypical TMD construct, not observed in the usual TMDs. This includes the presence of the IF1 helix, which has been reported to interact with the helical region of *Ec*MlaD (Coudray et al., 2020). This unique feature can be attributed to the separate line of evolution from typical TMDs, which might have led to the origin of *Ec*MlaE and its orthologues. Interestingly, the *Ec*MlaE orthologues also form sub-clusters. However, surprising was the separate cluster formed by *Ec*PqiA and *Ec*YebS. It could possibly be because both these TMDs are not associated with any NBD, as the Pqi and Yeb systems are non-ABC transporter systems. This is also apparent from the absence of the EAA/EQ/EEXDA loop. Additionally, the structural peculiarities of *Ec*MlaE, along with its association with *Ec*MlaD, might aid in the bidirectional movement of PLs. As a result, in spite of possessing an exporter fold, the Mla system might be able to transport PLs in both directions, as observed in recently studied ABC transporters (Low et al., 2021).

3.4.8. *Ec*MlaF-*Ec*MlaB gives rise to a unique functional association and macromolecular arrangement

The association between *Ec*MlaF and *Ec*MlaB is critical for the proper functioning of the Mla system (Thong et al., 2016). The presence of CTE in *Ec*MlaF, along with the interaction with *Ec*MlaB, results in new interfaces, which in turn stabilizes the MlaFEDB complex (Thong et al., 2016; Kolich et al., 2020). Such a dynamic arrangement is not observed in typical ABC transporters and hints towards a unique substrate transport mechanism. The presence of the DSSG motif in *Ec*MlaB and its orthologues suggests possible ATP binding and phosphorylation. However, further experimental approaches are required for proper validation. It can be further speculated that this MlaF-MlaB association is important for the bidirectional movement of PLs.

3.4.9. The structural features of *Ec*MlaA might aid in the ATP-independent transfer of PLs

The bacterial OM proteins have a β-barrel architecture that remains embedded in the membrane, whereas the lipoproteins have the protein component protruding out into the periplasm (Rollauer et al., 2015; El Rayes et al., 2021). However, *Ec*MlaA presents a unique case as it has an α-helical structure and remains embedded in the OM in spite of being a lipoprotein. This association with the OM is mainly aided by the interaction of OmpC/F. This also contributes to the ideal orientation of the central channel that reaches out to the outer leaflet. The seven helices create a suitable passage for the movement of the PLs. In addition to this, the flexible nature of *Ec*MlaA would provide the necessary force for the transport of PLs without the utilization of ATP. Thus, it can be speculated that *Ec*MlaA would exhibit a wide conformational space in order to aid PL movement. This would be critical as PLs, when not present in membrane arrangement, demonstrate ‘lipid gymnastics’ (Neumann et al., 2017). This is further supported by the conformational state of PL observed in our docking experiment. The dynamic movements of PL would be accommodated by the flexible *Ec*MlaA. This flexibility would also be aided by the CTH region, which would help in orienting the protein during the process of transport.

3.5. Conclusion

In the current study, an extensive characterization of all the components of the Mla system from *E. coli* by utilizing various bioinformatics-based approaches. The study highlights the diverse architectural types of the MlaD domain, establishing its association with different proteins from various organisms. This sheds sufficient light on the possibility of the multi-functionality of the domain. The substantial presence of glycyl and hydrophobic residues suggests the conformational dynamicity of the domain and its potential interaction(s) with hydrophobic molecule(s). This would immensely contribute to the characterization of the proteins possessing MlaD domains. The reportage of the conserved Gly29 residue, along with the consensus motif at the N-terminal region of the domain, highlights the critical role of the region. The motif can be used for the identification of Mce proteins and is an excellent candidate for mutational analysis that would further aid in the functional characterization of the domain. The findings from the phylogenetic tree establish for the first time the common ancestry of the MlaD domain of the proteins MlaD from *E. coli*, Mce from *M. tuberculosis* and TGD2 from *A. thaliana*. This contributes to the fact that the evolution of the MlaD domain is based on the domain architecture. This relatedness is reflected through the conservation of PLP regions and the association of the domains with ABC transporter systems. On the other hand, the MlaD domains of the proteins PqiB and YebT from *E. coli* are evolutionarily closer, although they lack the conservation of PLP regions. Unlike the former group, the domains are components of non-ABC transporter systems and are not dependent on ATP hydrolysis for transporting PLs. This led to the learning that the MlaD domains belonging to the two clusters would participate in two distinct mechanisms of PL transport.

The structural studies combined with the evolutionary analysis shed light on the relationship between the MlaD domains, especially the PLPs and permeases based on driving forces. This led to the understanding that the Mla, Mce and TGD systems would mechanistically function in a similar way. The hydrophobic mapping of MlaD domains, conducted for the first time, provides a detailed view of the central pore formed by domain hexamers. The study clearly highlights the critical role of MlaD domains in providing an ideal hydrophobic environment for PLs. Interestingly, the MlaD domains are found to be not responsible for protein oligomerization. The findings also led to the learning that functionally associated protein MlaC of the Mla system from *E. coli* and Mam and Omam of the Mce system from *M. tuberculosis* are structural homologs of proteins belonging to the NTF2 superfamily. This highlights the presence

of noncanonical SBPs and additional components that might be critical for the functioning of ABC transporters. The study, for the first time, establishes the functional association between Mla and Lpt systems, especially the role of MlaC as a mediator. This led to the understanding that both the systems synergistically transport PLs and LPSs in order to maintain OM asymmetry. The work also presents the unusual features of the remaining components of the Mla system. This includes the evolutionary lineage of MlaE and MlaF, along with the variation in their coupling helix and signature motifs, respectively. This has also resulted in the identification of additional interfaces in the MlaF-MlaB complex. Furthermore, the study also highlights the dynamic as well as the flexible nature of MlaA, which would aid in the transport of PL. As a whole, these findings shed a sufficient amount of light on the various aspects of the Mla system.

Appendix A. Supplementary data

Supplementary Tables, A1 (A1.1-A1.11) and A2 (A2.1-A2.2).

Supplementary Figures, A1 (A1.1-A1.13) and A2 (A2.1).

Chapter 4

Structural and mechanistic characterization of MlaC protein from *Escherichia coli*

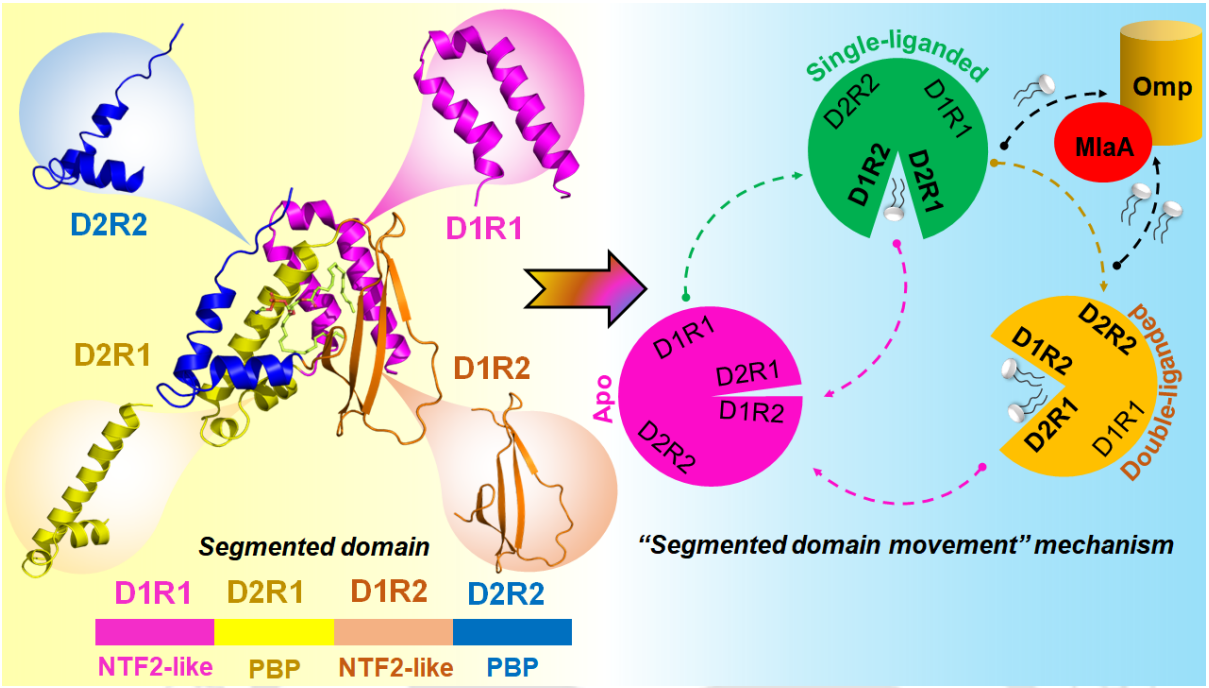
The part of this chapter has been published as:

1. **Dutta A** and Kanaujia SP (2022). MlaC belongs to a unique class of non-canonical substrate-binding proteins and follows a novel phospholipid-binding mechanism. *J. Struct. Biol.* 214(4), 107896.

Abstract

The outer membrane (OM) of Gram-negative bacteria acts as a formidable barrier against a plethora of detrimental compounds owing to its asymmetric nature. This is because the OM possesses lipopolysaccharides (LPSs) in the outer leaflet and phospholipids (PLs) in the inner leaflet. The maintenance of the lipid asymmetry (Mla) system is involved in preserving the distribution of PLs in OM. The periplasmic component of the system MlaC serves as the substrate-binding protein (SBP) that shuttles PLs between the inner and outer membranes. However, an in-depth report highlighting its mechanism of ligand binding is still lacking. This study reports the crystal structure of MlaC from *Escherichia coli* (*Ec*MlaC) at a resolution of 2.5 Å in a quasi-open state, complexed with PL. The structural analysis reveals that *Ec*MlaC and orthologs comprise two major domains, viz. nuclear transport factor 2-like (NTF2-like) and phospholipid-binding protein (PBP). Each domain can be further divided into two subdomains arranged in a discontinuous fashion. This study further reveals that *Ec*MlaC is polyspecific in nature and follows a reverse mechanism of the opening of the substrate-binding site during the ligand binding. Furthermore, MlaC can bind two PLs by forming subsites in the binding pocket. These findings, altogether, have led to the proposition of the unique “segmented domain movement” mechanism of PL binding, not reported for any known SBP to date. Further, unlike typical SBPs, MlaC has originated from a cystatin-like fold. Overall, this study establishes MlaC to be a non-canonical SBP with a unique ligand-binding mechanism.

Graphical abstract



4.1. Introduction

Gram-negative bacteria are generally more resilient than Gram-positive bacteria due to the presence of a unique outer membrane (OM) (Delcour, 2009). The OM possesses an asymmetric organization of lipopolysaccharides (LPSs) in the outer leaflet and phospholipids (PLs) in the inner leaflet. The asymmetric nature of the OM acts as a protective shield against the assaults of a plethora of toxic chemicals such as antibiotics, detergents, etc. (Clifton et al., 2013). However, PLs located in the inner leaflet have the tendency to flip and accumulate in its outer leaflet leading to the disruption of the permeability barrier of the OM (Henderson et al., 2016). In order to restore OM integrity, Gram-negative bacteria have evolved two major systems, viz. PldA and PagP to eliminate the excess PLs by modifying them (Bishop, 2008). A recently-identified third system involves a highly conserved multi-component intermembrane machinery, viz. maintenance of lipid asymmetry (Mla) system (Malinverni and Silhavy, 2009). The Mla system consists of three major components - an OM lipoprotein-Osmoporin (Omp) complex (MlaA-OmpC/F), a free-floating periplasmic protein (MlaC), and an inner membrane (IM) ATP-binding cassette (ABC)

transporter complex (MlaFEDB) (Malinverni and Silhavy, 2009; Henderson et al., 2016; Coudray et al., 2020). The Mla system was initially proposed to serve as an importer, which extracts the excess PLs accumulated in the OM. The OM-associated MlaA extracts PLs and hands them over to MlaC that ultimately transfers to the IM complex (MlaFEDB) for internalization (Malinverni and Silhavy, 2009; Ekiert et al., 2017). However, recent findings have suggested that the Mla system adopts an ABC exporter fold and can export PLs to the OM (Hughes et al., 2019; Coudray et al., 2020). Furthermore, MlaFEDB can spontaneously transfer PLs to MlaC in the absence of ATP (Low et al., 2021). This firmly establishes that MlaC can transport PLs in both directions.

The basic architecture of both ABC importers and exporters consists of transmembrane- and nucleotide-binding domains (TMDs and NBDs) (Rees et al., 2009; Wilkens, 2015). However, ABC importers require an additional component termed substrate (or solute)-binding protein (SBP). SBPs capture the substrates and bring them to the TMDs for subsequent translocation, thus, rendering specificity and directionality to the importers (Maqbool et al., 2015). In the Mla system, proteins MlaC, MlaE, and MlaF serve as SBP, TMD, and NBD components, respectively. However, the system is also dependent on three additional proteins, viz. MlaA (OM-associated), MlaB (IM-associated), and MlaD (IM-associated) (Malinverni and Silhavy, 2009; Coudray et al., 2020; Ekiert et al., 2017; Abellón-Ruiz et al., 2017; Kolich et al., 2020). MlaA has a doughnut shape with a central pore and works in association with osmoporins (Omps) (Abellón-Ruiz et al., 2017). The protein MlaD possesses a highly conserved Mce or MlaD domain and forms a hexameric ring with a central pore that rests on MlaE (Ekiert et al., 2017). The pores in the proteins (MlaA and MlaD) have been recognized as the translocation passages for PLs that are mobilized through the periplasm with the help of free-floating MlaC (Coudray et al., 2020).

Although MlaC serves as the SBP component for the Mla system, it surprisingly does not share any structural similarity with the SBP-like fold (Yero et al., 2021). Although diverse at the primary structure level, canonical SBPs possess a highly conserved structural fold with two globular α/β N- and C-terminal domains (NTD and CTD) and a central β-sheet flanked by α-helices. The NTD and CTD are connected by a flexible hinge region which allows their movement for ligand capturing. The substrate-

binding site is a cleft at the interface of the two domains (Scheepers et al., 2016). The hinge-bending motion facilitates the transition from an open to a closed state upon substrate binding. Based on the topological arrangement of the β-sheets in the core regions of NTD and CTD, SBPs have been categorized into two different types, I and II. In Type I, both the domains share $\beta_2\beta_1\beta_3\beta_4\beta_5$ arrangement, while in Type II, the two domains share $\beta_2\beta_1\beta_3\beta_n\beta_4$ arrangement where n is the cross-over β-strand (Fukami-Kobayashi et al., 1999). In addition to this, a third type, III, has been reported, wherein a single α-helix serves as a linker between the two domains (Lee et al., 1999). On the basis of the structural fold and substrate specificities, SBPs have been classified into eight clusters, A-H (Berntsson et al., 2010; Scheepers et al., 2016; Chandravanshi et al., 2021). Structural studies on MlaC suggest that it lacks the typical NTD and CTD arrangement (Ekiert et al., 2017; Hughes et al., 2019). Instead, it has been suggested to have a non-contiguous domain organization and cannot be placed in any of the previously recognized clusters (Yero et al., 2021). Furthermore, MlaC has been categorized as a box-like lipid transfer protein (LTP) without any lid (Wong et al., 2019). Moreover, available three-dimensional structures of MlaC show a variation in the number of bound ligands, thereby hinting at a distinct ligand-binding mechanism (Ekiert et al., 2017; Hughes et al., 2019; Yero et al., 2021). MlaC has also been highlighted as a promising drug target for the development of effective therapeutics against pathogenic Gram-negative bacteria (Huang et al., 2016; Huang et al., 2019). However, there is a lack of an in-depth study of MlaC addressing the structural aspects of PL binding and transport. Also, a proper reportage on non-canonical SBPs such as MlaC and its ancestry is unavailable to date.

In this study, we report, for the first time, the three-dimensional structure of MlaC from *E. coli* (*Ec*MlaC) in a quasi-open state, complexed with an endogenously bound ligand. The structural study of *Ec*MlaC enabled us to postulate the presence of segmented domains and to hypothesize a unique mechanism of binding-site opening followed by ligand binding and re-orientation. To the best of our knowledge, such a mechanism has not been reported for any known SBP to date. This study also enriches the current knowledge on the structural diversity of non-canonical SBPs and sheds a sufficient amount of light on their ancestries.

4.2. Materials and methods

4.2.1. Cloning, overexpression, and purification

The genomic DNA was isolated from *Escherichia coli* BL21(DE3) competent cells and the gene fragment corresponding to the signal peptide-cleaved mature *Ec*MlaC protein (residues 22–211) was amplified by polymerase chain reaction (PCR) from the genome using the forward primer 5′-GCGC**CATATG**CATCATCATCATCATCATGCAGACCAGACCAATCCG-3′ containing *Nde*I restriction site (bold) followed by a 6xHis-tag (underlined) and the reverse primer 5′-CCG**CTCGAG**TTATTTTTTCTCTTCCAGAGTG-3′ containing *Xho*I restriction site (bold). The amplified gene was inserted into pET22b(+) vector excised with the same set of restriction enzymes. The resultant recombinant plasmid was confirmed by double digestion and plasmid DNA sequencing (Figure 4.1A). For overexpression, *E. coli* BL21(DE3) competent cells were transformed by using the recombinant plasmid. The transformed cells were grown at 37 ºC in a Luria-Bertani (LB) medium supplemented with 100 µg mL^{-1} ampicillin until the culture attained an optical density (OD) of 0.6 at 600 nm. Subsequently, the secondary cell culture was induced for protein expression by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) at a concentration of 0.1 mM. The culture was further grown for 7 h and at each hour, cells were harvested in order to obtain the ideal induction time (Figure 4.1B). Based on this experiment, the secondary cell culture was induced for 4h at 37 ºC. The cells were harvested by centrifugation at 3,836*g* for 10 min and resuspended in lysis buffer composed of 20 mM Tris-HCl pH 7.5, 150 mM KCl, 5 mM imidazole, 5% glycerol, 1 mM phenylmethylsulfonyl fluoride (PMSF) and 3 mM β-mercaptoethanol (β-ME). The resuspended cells were lysed by sonication (2 s on and 10 s off at 33% amplitude) followed by the removal of insoluble debris by centrifugation at 15,644*g* for 40 min at 4 °C (Figure 4.1C). The supernatant was then applied to a pierce centrifuge column (Thermo Fisher Scientific, USA) packed with Ni^{2+}-affinity resin (Qiagen, Germany) pre-equilibrated with binding buffer (20 mM Tris-HCl pH 7.5, 150 mM KCl, 5 mM imidazole, 5% glycerol, 1 mM PMSF and 3 mM β-ME) and incubated for 2 h. The column was washed with five column volume of each wash buffer A (20 mM Tris-HCl pH 7.5, 200 mM KCl, 10 mM imidazole, 5% glycerol, 1 mM PMSF and 3 mM β-ME) and wash buffer B (20 mM Tris-HCl pH 7.5, 200 mM KCl, 20 mM imidazole, 5% glycerol, 1 mM PMSF and 3 mM β-ME). The bound protein was

eluted with 400 mM imidazole in the wash buffer (Figure 4.1D). The eluted fractions were collected and step-wise dialysis was performed against 20 mM Tris-HCl pH 7.5 and 200 mM KCl in order to remove the imidazole. The dialyzed protein was concentrated to a final concentration of 10 mg ml^{-1} using a Vivaspin turbo 15 (Sartorius, Germany) centrifugal filter unit with a molecular weight cut-off of 10 kDa (Figure 4.1E).

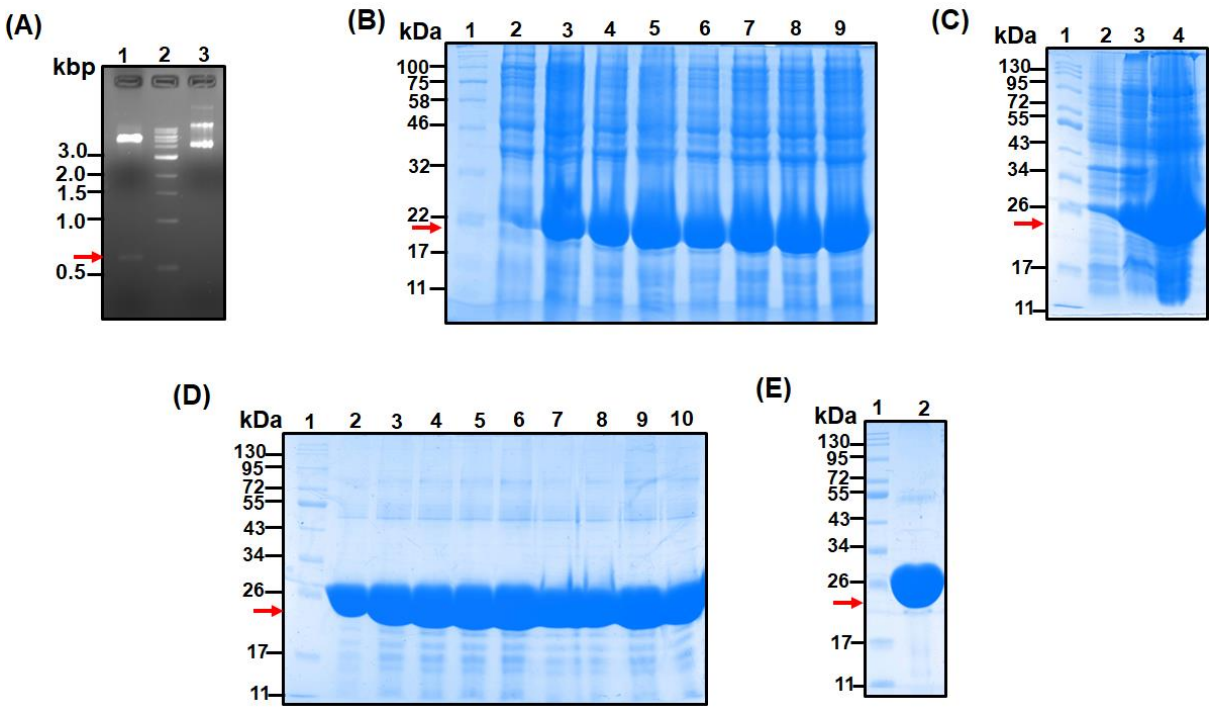


Figure 4.1. Cloning, overexpression, solubility and purification of EcMlaC. *(A) Clone confirmation by double digestion of the vector pET22b carrying the truncated gene mlaC by NdeI and XhoI restriction enzymes (lane 1: insert pop out after double digestion of the plasmid pET22b, lane 2: DNA ladder and lane 3: pET22b wild type). (B) Overexpression of the protein EcMlaC in BL21 (DE3) at 0.1 mM IPTG at a different time interval (lane 1: protein marker, lane 2: uninduced or 0 hrs, lane 3-9: 1-7 hr, respectively) after induction. (C) SDS-PAGE analysis for optimized protein solubility at 37°C for 4 hrs (lane 1: protein marker, lane 2: uninduced sample, lane 3: pellet fraction and lane 4: supernatant). (D) SDS-PAGE analysis for the eluted fractions of the protein using Ni-NTA metal-affinity chromatography (lane1: protein marker, lanes 2-10: eluted fractions of the protein). (E) SDS-PAGE analysis for the concentrated protein (lane 1: protein marker and lane 2: concentrated protein). The gene and the protein of interest are marked by a red arrow.*

4.2.2. Circular dichroism (CD) experiments

The Far-UV CD spectrum of *Ec*MlaD was recorded between 190 and 260 nm wavelengths with a 1 mm optical path length quartz cuvette using a Jasco J-1500 spectrometer (Jasco, Germany). The protein prepared in a buffer containing 20 mM Tris-HCl pH 7.5 and 200 mM KCl was diluted in water to a final concentration of 5 µM. The CD parameters response, sensitivity and scan speed were fixed at 2 s, 100 millidegrees and 100 nm min^{-1}, respectively, for spectra measurements. The final spectrum was obtained by averaging three scans and subtracting from the blank that contained the buffer as a reference. The spectrum was analyzed by using Spectra Manager software provided by JASCO. In order to study the thermal stability of *Ec*MlaD, the sample was subjected to a temperature range of 20-90 ºC with a heating rate of 5 ºC min^{-1}. The spectra were recorded from 190 to 260 nm wavelengths, and the denaturation profiles were plotted as a function of temperature and wavelength using the software Origin v9.6.

4.2.3. Lipid extraction and thin-layer chromatography

In order to find out whether the purified protein is bound to any PL, a lipid extraction assay of *Ec*MlaC was performed following the Bligh-Dyer method, as previously reported (Bligh and Dyer, 1959; Thong et al., 2016). For lipid extraction, 1 ml of 2.5 mg ml^{-1} of purified *Ec*MlaC was mixed with 3.75 volumes of chloroform:methanol (1:2 (v/v)). The mixture was vortexed and sonicated for 30 s three times, followed by centrifugation at 21,000*g* for 5 min. The supernatant was collected and an appropriate amount of chloroform and buffer were added to form two-phase mixture chloroform:methanol:water. In order to separate the organic and aqueous phases, the mixture was then centrifuged at 4,000*g* for 5 min. After centrifugation, the organic phase was gently removed and the organic solvent in it was evaporated at room temperature. The dried lipids that remained were resuspended in 50 µL chloroform:methanol (4:1 (v/v)). A TLC Silica gel 60 F_{254} plate (Merck) was cut to a length of 11 cm and a width of 4 cm. A total of 20 µL of the sample was spotted on the plate, about 1 cm from the bottom of the plate. The plate was then placed in the solvent system (65:25:4 chloroform: methanol: water) designed to separate PLs based on head group polarity. The plate was allowed to remain in the solvent system until the solvent font was 1 cm from the top of the plate. After that, the plate was dried at room temperature and stained with iodine vapor.

4.2.4. Crystallization, data collection, and structure determination

The initial screening of the purified protein was performed using microbatch-under-oil technique by mixing 2 µL of protein and 2 µL of crystallization buffer available in the crystallization screens from Hampton Research at 4 ºC and 20 ºC temperatures. Crystal hits were obtained in the buffer containing 0.7 M sodium citrate tribasic dihydrate and 0.1 M Bis-Tris propane pH 7.0 at 20 ºC within 1-2 months (Figure 4.2A). In order to get diffraction-quality crystals, optimization was performed using the hanging-drop vapor-diffusion method at 20 ºC (Figure 4.2B). Optimization was also performed in order to reduce the incubation period for protein crystallization. We were able to obtain *Ec*MlaC crystals in a condition containing 0.2 M Lithium sulfate, 0.05 M Zinc acetate dihydrate, 0.1 M Bis-Tris 7.5, 22.5% (v/v) polyethylene glycol (PEG) Smear Broad (a mixture of PEG compounds with a broad range of molecular weights) at 20 ºC within 48 hours of incubation (4.2C). However, the crystals failed to produce quality diffraction. X-ray intensity diffraction data were collected at −173 ºC using the home source Rigaku MicroMax-007 HF diffractometer (operated at 40 kV and 30 mA) and R-Axis IV++ image-plate detector with rotating anode X-ray tube at wavelength 1.5418 Å available at Central Instruments Facility (CIF) of Indian Institute of Technology Guwahati, India. Data was collected by maintaining the crystal-to-detector distance at 150 mm with an oscillation of 1° and an exposure time of 300 s per degree. At the time of data collection, crystals were cryoprotected using 20% glycerol and flash cooled under the stream of liquid nitrogen. The collected data set was processed and scaled using the programs iMosflm and Aimless embedded in the CCP4 package (Battye et al., 2011; Winn et al., 2011; Evans and Murshudov, 2013). The X-ray intensities were converted to structure factors using the program ctruncate embedded in the CCP4 package. The three-dimensional structure of *Ec*MlaC was determined by the molecular replacement method using the program Phaser using the crystal structure of MlaC from *E. coli* (PDB id: 5UWA) as the search model (McCoy et al., 2007). In order to calculate the R_{free}, a total of 5% of the total reflections was kept aside as a test dataset (Brünger, 1992). The atomic model building and the iterative cycles of structural parameters refinement were carried out by using the programs Coot and Refmac5, respectively (Vagin et al., 2004; Emsley et al., 2010). The first cycle of structure refinement showed a clear electron density for the endogenously bound ligand phosphatidylethanolamine (PEF) in the binding pocket. However, the protein atoms and water molecules were first fitted in the electron density map. Subsequently,

the ligand molecule PEF was fitted to the electron density map. The quality of the refined model was validated by using the programs Procheck and MolProbity (Laskowski et al., 1993; Chen et al., 2010). The refinement and validation statistics of the model are provided in Table B1. The three-dimensional atomic coordinates and the structure factors have been deposited in RCSB Protein Data Bank with the accession code 7VR6 (Berman et al., 2000).

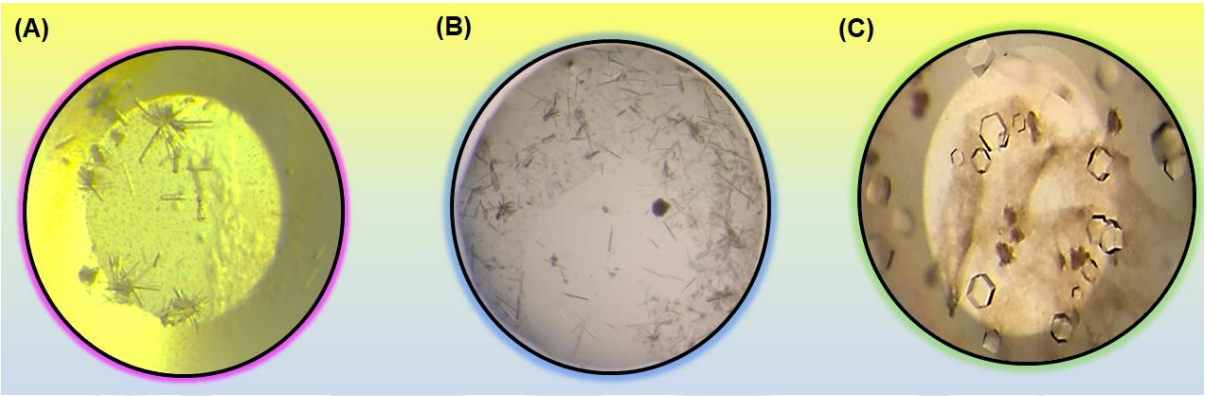


Figure 4.2. Crystallization of EcMlaC. *(A-C) Protein crystals of EcMlaC.*

4.2.5. Mass spectrometry analysis

In order to determine the molecular mass of ligand-bound *Ec*MlaC, the matrix-assisted laser desorption/ionization time of flight (MALDI-TOF) mass spectrometry method was employed. The homogeneously purified *Ec*MlaC was mixed with the matrix in protein: matrix ratios of 1:2, 1:3, and 1:5. The sinapinic acid matrix was prepared at a concentration of 10 mg ml^{-1} in a solvent containing acetonitrile and 0.1% trifluoroacetic acid in a ratio of 30:70, respectively. The matrix was dissolved in the solvent by sonicating it for 30 min, followed by centrifugation at 12,000*g* for 10 min to get rid of the excessive matrix particles. The prepared matrix was mixed in the above-mentioned ratios with 2 mg ml^{-1} concentration of *Ec*MlaC. Spots of 1 µL protein:matrix mixure were made on the ground-steel MALDI plate and were allowed to dry. After drying, the drops were ionized in a Bruker autoflex speed and the obtained data were analyzed using the in-built flexControl analysis software. The graphs were plotted using the software Origin version 8.5 for final representation.

4.2.6. Molecular docking

The binding affinities of the different PLs (Ligand ids: PEF, 8ND, H3T and GOT) to *Ec*MlaC were estimated by molecular docking analysis performed using the freely available program AutoDock v4.2 (Morris et al., 2009). The apo form of the protein was obtained by manually removing the bound ligand PEF from the binding pocket using the program Coot. The three-dimensional atomic coordinates of the ligands were extracted from the available structures of MlaC-ligand complexes (PDB ids: 5UWA and 6HSY) which were retrieved from the RCSB Protein Data Bank**.** Wherever applicable, before each docking experiment, the hydrogen atoms were added to the protein and ligand using the module available in the program AutoDock. Subsequently, the partial charges were assigned to the protein atoms using the Gasteiger charge algorithm (Gasteiger and Marsili, 1980). To perform the blind docking of the ligands to the protein, a grid size of 126x126x126 with a grid spacing of 0.375 Å was generated by taking the center of mass of the targeted protein as the grid center. The rigid molecular docking was performed by keeping the protein atoms rigid and the ligand atoms flexible, along with the rotatable bonds. For each molecular docking experiment, a total of 2000 runs of the Lamarckian genetic algorithm (LGA) were set up. The docked ligand conformations were clustered with a root mean square deviation (RMSD) cut-off of 2.0 Å. The atomic interactions between the protein and the docked ligand(s) were identified using the web server PDBsum and validated using the program Coot (Laskowski et al., 2018). The interactions were mainly divided into polar and non-polar, with the maximum distance cut-off of 3.5 and 4 Å, respectively. The molecular interactions of docked ligands and other structural figures were prepared using the program PyMOL v2.4.1.

4.2.7. Sequence and structure analyses

The protein sequences and the atomic coordinates of structures analyzed in the study were retrieved from the UniProtKB database and Protein Data Bank, respectively (The UniProt Consortium, 2023). The structure-based homology search for *Ec*MlaC was performed using the web server Dali (Holm, 2020). Domain organization of *Ec*MlaC was obtained using the CATH Protein Structure Classification Database (Dawson et al., 2017). In order to determine the sequence similarities among the *Ec*MlaC orthologs, pair-wise sequence alignment was performed using the web tool BLAST with the default set of parameters (Altschul et al., 1990). Following the previously reported structural parameters, the PL molecule has been divided into two parts – head

and tail regions (Cevc, 1993). The C1-C3 carbon atoms of the glycerol moiety were first identified. The C1-C3 carbon atoms, along with the moiety attached to C1 carbon, constitute the head region. Furthermore, the oxygen atoms present in the PL is part of the head region. The acyl chains attached to the C2 and C3 carbons of the glycerol moiety have been designed as β- and γ-chains, respectively. In order to determine the conformation of PLs, the crystal structures of *Ec*MlaC were fixed in C-terminal to N-terminal orientation. A PL was considered to possess β-γ conformation if the β chain appeared first, followed by the γ chain. When *vice versa*, the PL was considered to possess γ-β conformation. Furthermore, in order to study the orientation of the PL in the binding pocket, we have introduced the concept of the "binding plane". It is an imaginary plane that passes through the PL molecule. To determine the subdomain boundaries, structure-based multiple sequence alignment was performed using the program PROMALS3D (Pie et al., 2008). The volumes of the ligand-binding pockets were calculated using the program CASTp with a default probe radius of 1.4 Å (Tian et al., 2018). The protein-ligand interactions were analyzed using the web server PDBsum and validated using the program Coot. The interactions were mainly divided into polar and non-polar, with the maximum distance cut-off of 3.5 and 4 Å, respectively. In order to determine the translational movement of the individual β-strand, the crystal structures of *Ec*MlaC were superimposed using the Secondary Structure Matching (SSM) tool of Coot (Krissinel and Henrick, 2004). Following this, the Cα distances of each amino acid residue in each β-strand were measured with respect to the apo form. From this, the average Cα distance was calculated, which was designed as the average translational movement of the strand. The greater the value of the average translational movement, the larger the movement of the β-sheet. Helix curvatures were analyzed using the tool Bendix embedded in the program Visual Molecular Dynamics (VMD) (Humphrey et al., 1996; Dahl et al., 2012). For the comparison of kinks, the kink residue was identified and numbered as 0. The flanking residues around the kink residue were numbered from −6 to + 6 for the analysis (Wilman et al., 2014). In order to identify conserved residues, multiple sequence alignment of protein sequences was performed using the program Clustal Omega, and the results were rendered using the web tool ESPript 3.0 (Gouet et al., 2003; Sievers and Higgins, 2017). The residue conservation was further studied with the help of the sequence logo generated using the web server Weblogo (Crooks et al., 2004). The evolutionary conservation of residues in the crystal structure of *Ec*MlaC was studied

using the web tool ConSurf (Ashkenazy et al., 2016). The ancestral analysis of SBPs was performed using the Structural Classification of Proteins (SCOP) database (Andreeva et al., 2020). The operonic arrangement of the genes encoding SBPs was studied using the Gene database of the National Centre for Biotechnology Information (NCBI) and the EcoCyc database (Keseler et al., 2017). The visualization of crystal structures and the molecular graphic figures were generated using the program PyMOL (The PyMOL Molecular Graphics System, Schrodinger, LLC). The hydrophobicity mapping of the proteins was performed by using the YRB color scheme in the program PyMOL (Hagemans et al., 2015). As per this scheme, the carbon atoms not bound to nitrogen and oxygen atoms are highlighted in yellow, nitrogen atoms in the side chains of lysine and arginine are in blue, oxygen atoms in the side chains of glutamate and aspartate are in red, and all remaining atoms are in white.

4.3. Results

4.3.1. The protein is folded properly and co-purified with phospholipid

In order to get an insight into the secondary structural elements, purified MlaC from *E. coli* (*Ec*MlaC) was subjected to CD spectroscopy. The CD spectrum of MlaC shows a positive peak at 193.8 nm along with two negative peaks at 210 nm and 220 nm. This clearly suggests the presence of α and β secondary structural elements (Figure 4.3A). In order to probe the thermal stability of the protein, a thermal ramping experiment was performed by heating the protein sample from 20 to 90 ºC. It was observed that at around 70 ºC, the protein starts losing its structural integrity (Figure 4.3B). The further increase in the temperature leads to the further loss of secondary structure. The previously reported structure of *Ec*MlaC (PDB id: 5UWA) highlights the presence of an endogenously bound phosphatidic acid molecule (Ligand id: 8ND) in the binding pocket (Ekiert et al., 2017). Based on TLC experiments, *Ec*MlaC has also been reported to bind three different PLs viz, phosphatidylglycerol, phosphatidylethanolamine and cardiolipin (CL) (Hughes et al., 2019). Hence, to investigate the presence of any bound ligand, TLC of *Ec*MlaC extract was performed. The results reveal that *Ec*MlaC has been co-purified with three different species of PL (Figure B1A).

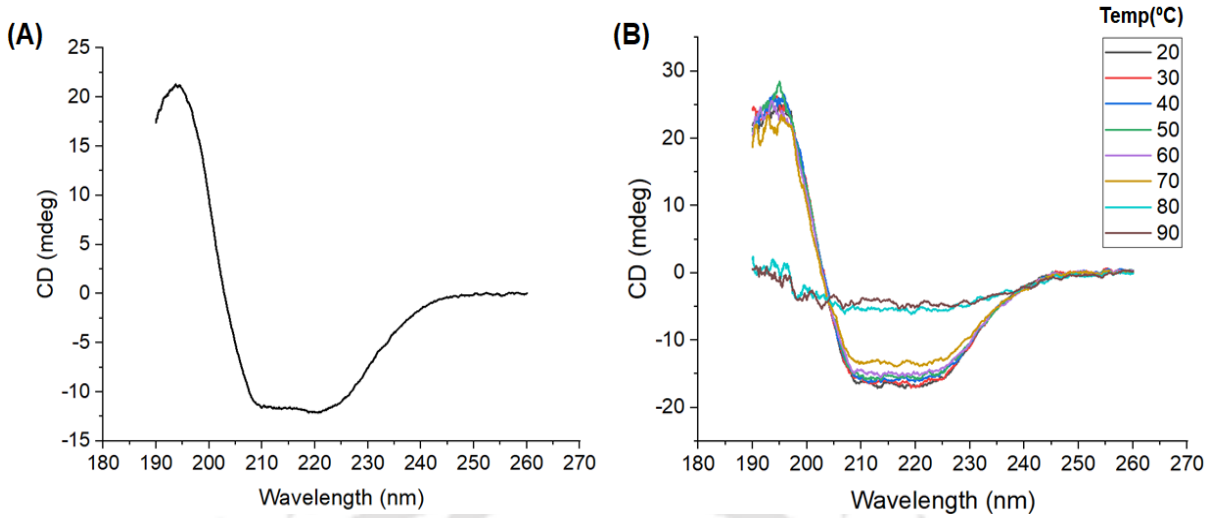


Figure 4.3. Biophysical characterization of EcMlaC. *(A) CD spectrum of EcMlaC measured in the wavelength range of 190-260 nm. (B) Thermal melting plots depicting the change in secondary structure elements of EcMlaC at different temperatures in the range of 20-90 ºC.*

4.3.2. The overall structure of *Ec*MlaC

In order to get further insights, the crystal structure of *Ec*MlaC was determined at a resolution of 2.5 Å in the space group *H*3 containing one protomer in the asymmetric unit (PDB id: 7VR6, Table B1). The structure reveals that *Ec*MlaC adopts an α+β fold consisting of four anti-parallel β-strands and eight α-helices (Figure 4.4A). This is in accordance with the previously reported structure of *Ec*MlaC (PDB id: 5UWA). Interestingly, an additional electron density other than that for the amino acid residues was clearly observed in the binding pocket of *Ec*MlaC. The crystal structure of MlaC ortholog from *Ralstonia solanacearum* (PDB id: 2QGU) submitted by Northeast Structural Genomics (NESG) was crystallized with ligand modeled as phosphatidylethanolamine (Ligand id: PEF), which contains a head group (Figure B1B and Table B2). However, this additional electron density for the head group in the endogenously bound ligand (Ligand id: 8ND) was not observed in the previously reported single-liganded state of MlaC (PDB id: 5UWA, Figure B1B and Table B2). The electron densities of the ligands bound to the MlaC orthologs *Pseudomonas putida* and *Pseudomonas aeruginosa* were also analyzed (PDB ids: 5UWB and 6HSY,

respectively; Figure B1B and Table B2). In our structure, the electron density of the ligand corresponded to that of phosphatidylethanolamine (Ligand id: PEF) molecule (Figure 4.4A). Hence, a PEF molecule was identified to be endogenously bound to the binding pocket. In order to affirm the presence of this ligand, a MALDI-TOF analysis of *Ec*MlaC was performed. It was anticipated that a distinct peak corresponding to the molecular weight of the holo form of *Ec*MlaC bound to PEF (22686.78 + 691.96 = 23378.74 Da) would be observed. Surprisingly, the data showed a prominent peak at ~ 22662.51 Da, which corresponds to the theoretical molecular weight of the apo form of MlaC (22686.78 Da) (Figure B1C). Furthermore, split peaks of lower intensities but of higher molecular weights were observed. At closer inspection, a peak corresponding to the molecular weight of the holo form of *Ec*MlaC (23378.183 Da) with a decent intensity (1688 a.u.) was observed. This suggests that the sample contains *Ec*MlaC bound to PEF. However, owing to the presence of different species of *Ec*MlaC with different PLs, peak splitting was observed. A comparable mass spectrum has been previously reported for insulin and arg-insulin mixture (Bilati et al., 2005). Furthermore, the spectrum also hints at the possible fragmentation of PLs, which has been previously reported to occur during MALDI-TOF experiments (Schiller et al., 1999; Al-Saad et al., 2003; Rujoi et al., 2004; Woods et al., 2004; Fuchs et al., 2010; Engel et al., 2022). However, among the different species, *Ec*MlaC bound to PEF was favored for crystallization. Such an observation has been made in the case of the *Ec*MlaC ortholog from *Pseudomonas aeruginosa*, which can bind different species of PLs but crystallized with only two endogenously bound PLs (Ligand ids: GOT and H3T) (Yero et al., 2021).

The phospholipid phosphatidylethanolamine (PEF) endogenously bound to *Ec*MlaC obtained in this study adopts the β-γ conformation with its two acyl tails deeply buried and comparatively spread out. The head group of PEF remains solvent-exposed, which forms hydrogen bonds with the residue Tyr100 and two water molecules (Figure B1D). Furthermore, in total, there are 43 interactions that stabilize the PEF molecule at the binding site of the protein (Table B3). Also, the shape of the binding cavity resembles that of a decanter with a narrow neck and a broad base, unlike a cleft buried between NTD and CTD, as observed in the canonical SBPs.

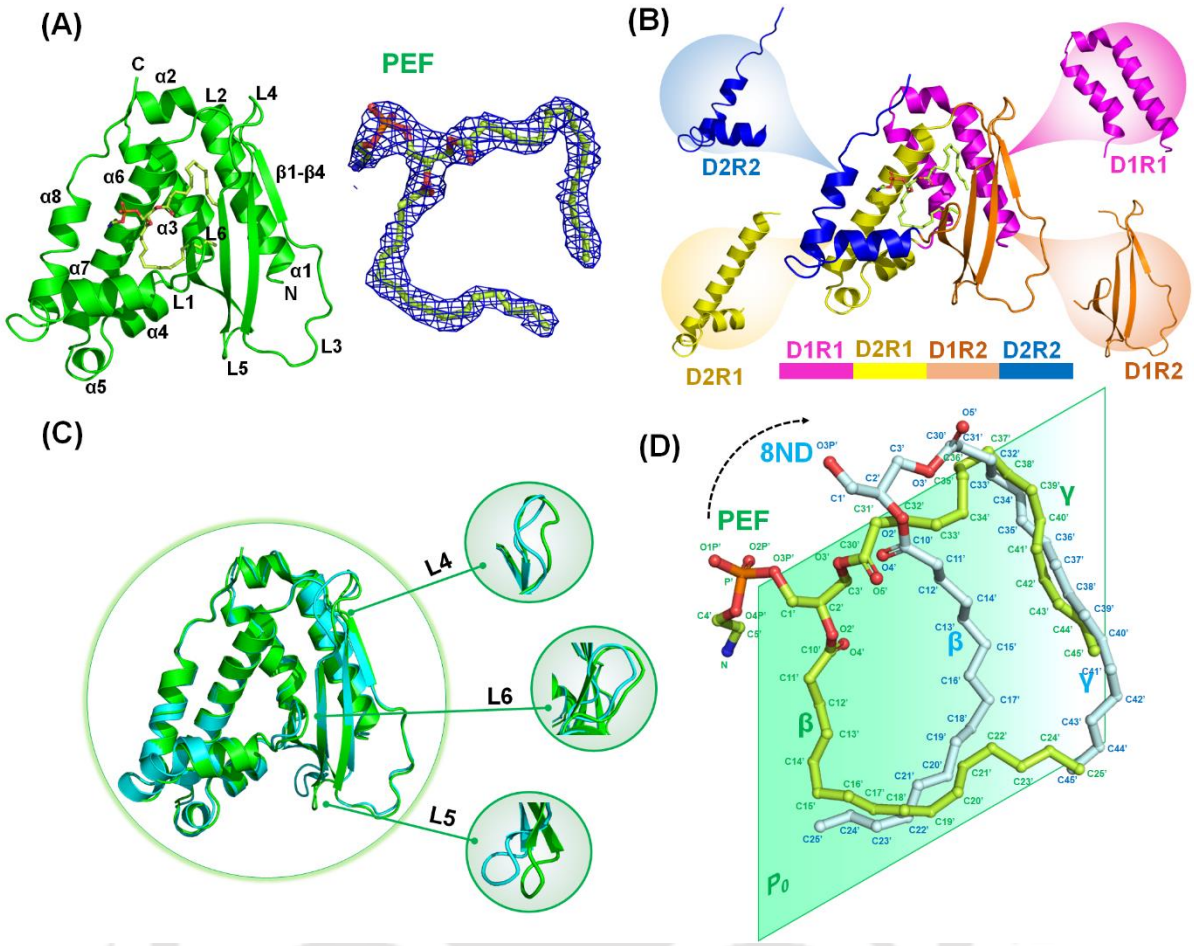


Figure 4.4. The overall structure of EcMlaC. *(A) (Left) The three-dimensional structure of EcMlaC in complex with PEF (PDB id: 7VR6). (Right) The composite omit map of bound PEF contoured at 1.0 σ is displayed in a blue mesh. (B) The protein EcMlaC consists of two main domains, NTF2-like (D1) and PBP (D2). Each of the two domains can be divided into two subdomains (D1R1, D1R2; D2R1, D2R2) that are arranged in a discontinuous fashion. The schematic representation of the discontinuous subdomain arrangement is shown at the bottom. (C) Structural superimposition of EcMlaC_P2$_1$2$_1$2$_1$_holo1 (PDB id: 5UWA, cyan) and EcMlaC_H3_holo1 (PDB id: 7VR6, green) along with the magnified views of the structural superimpositions of L4-L6 loops. (D) Superimposition of the endogenously bound phospholipids 8ND and PEF. Both the ligands occupy the same binding plane, P_0 (green parallelogram). The acyl chains in 8ND are comparatively closer to each other than those of PEF. The rightward shift of the head group of 8ND is denoted by a dotted arrow.*

4.3.3. *Ec*MlaC possesses a unique segmented domain arrangement

A structural comparison of *Ec*MlaC with available SBPs revealed that *Ec*MlaC possesses an arrangement that has not been observed in any canonical SBP. A structural homology search of *Ec*MlaC shows that it shares similarities with the nuclear transport factor 2 (NTF2) superfamily rather than to any canonical SBP (Table B4). A search in the CATH database against *Ec*MlaC shows that it contains a single domain belonging to the class: Alpha Beta (3), architecture: Roll (3.10), topology: nuclear transport factor 2; Chain: A (NTF2) (3.10.450) and superfamily: Tgt2/MlaC (3.10.450.710). Surprisingly, a structural search of the MlaC ortholog from *Ralstonia solanacearum* (*Rs*MlaC, PDB id: 2QGU) in the CATH database revealed the presence of two different domains (D1 and D2). Each domain is discontinuous and is segmentally arranged as subdomains, referred to as D1R1 (29–73), D2R1 (74–116), D1R2 (117–172), and D2R2 (173–207). The D1R1 and D1R2 subdomains belong to an unnamed (UN) superfamily (CATH id: 3.10.450.50). On the other hand, D2R1 and D2R2 subdomains belong to the phospholipid-binding protein (PBP) superfamily (CATH id: 1.10.10.640) (Figure B2A). The UN superfamily and domain D1 have previously been referred to as NTF2-like superfamily and NTF2-like domain, respectively (Yero et al., 2021). Based on the superfamily, the domain D2 is referred to as the PBP domain in this study. As per the CATH database, the UN superfamily is one of the largest superfamilies under the NTF2 topology. On the other hand, the PBP superfamily belongs to the class: Mainly Alpha (1), architecture: Orthogonal Bundle (1.10), and topology: Arc Repressor Mutant, subunit A (1.10.10). This suggests that *Rs*MlaC represents a unique fusion of two domains belonging to two different superfamilies. This analysis indicates that *Ec*MlaC and *Rs*MlaC (sequence identity: 29%) possess different (sub)domain arrangements. The former has a single domain, while the latter has two domains with a segmented arrangement. However, in a recent study, a combination of the two structural domains in the proteins Tgt2/MlaC has been highlighted. Based on an in-depth structural search, this study further proposes the subdomain D2 to be associated with the small alpha (or C) domain of AAA^+ (ATPase Associated with diverse cellular Activities) proteins, referring to it as AAA helical-bundle domain. Furthermore, Tgt2/MlaC to which *Ec*MlaC belongs consists of the *Ec*MlaC ortholog from *Pseudomonas putida*. This affirms that the superfamily Tgt2/MlaC is derived from the combination of the NTF2-like and PBP superfamilies. Thus, *Ec*MlaC

and its orthologs possess a unique segmented domain arrangement consisting of NTF2-like (D1R1, D1R2) and PBP (D2R1, D2R2) subdomains. A structure-based multiple sequence alignment (MSA) of *Ec*MlaC orthologs revealed the boundaries of the four subdomains D1R1, D1R2, D2R1, and D2R2 (Figure 4.4B, Figure B2B and Table B5). Furthermore, in *Ec*MlaC and its orthologs, the binding site remains sandwiched between the subdomains D1R1-D2R1 and D1R2-D2R2 (Figure 4.4B).

4.3.4. *Ec*MlaC shows variation in ligand interaction

Since the crystal structure of *Ec*MlaC has already been published, a comparative analysis was done between the two structures (PDB ids: 5UWA and 7VR6). The previously published *Ec*MlaC was crystallized in the space group $P2_12_12_1$ (with cell dimensions, a = 33.63 Å, b = 115.89 Å, c = 133.22 Å, α = β = γ = 90º). In this study, *Ec*MlaC was crystallized in the space group *H*3 (with cell dimensions, $a = b$ = 114.94 Å, c = 46.10 Å, α = β = 90º, γ = 120º). For comparison, the two structures are referred to as *Ec*MlaC_$P2_12_12_1$_holo1 and *Ec*MlaC_H3_holo1, respectively. Both the structures are similar (RMSD: 0.7 Å) for all Cα atoms, except for L4-L6 and β1-β4 regions. These regions exhibit significant structural differences (Figure 4.4C). In comparison to *Ec*MlaC_H3_holo1, the three loops (L4, L5, and L6) in *Ec*MlaC_$P2_12_12_1$_holo1 show inward movements towards the binding site (distance: 1.9, 2.6, and 1.1 Å, respectively). Further, *Ec*MlaC_$P2_12_12_1$_holo1 has a comparatively larger binding-site cavity (381.515 $Å^3$) than that of *Ec*MlaC_H3_holo1 (304.629 $Å^3$). Interestingly, both the structures were solved with endogenously bound phospholipids 8ND and PEF, respectively. Structural superimposition reveals that both ligands are present in the single binding plane, referred to as P_0. Even though both the PLs adopt β-γ conformations, the tails of the two molecules adopt different conformations. Additionally, both PLs differ in their head groups (Figure 4.4D). The head group of 8ND in *Ec*MlaC_$P2_12_12_1$_holo1 is closer to the loop L6, while that of PEF is closer to the helix α6. This is due to the formation of hydrogen bonds between the head group of PEF and the residue Tyr100 and water molecules. On the other hand, 8ND is hydrogen-bonded to Gln115 and two water molecules. This results in a rightward shift of the 8ND molecule (Figure 4.4D and Figure B1D). Overall, PEF in *Ec*MlaC_H3_holo1 shows 43 contacts, whereas 8ND in *Ec*MlaC_$P2_12_12_1$_holo1 has 22 contacts. Among the interacting residues, Phe39, Tyr105, Ala108, Leu109, Met111, Gln115, Phe150,

and Met173 coordinate the ligands in both the structures (Figure B3A-B3B and Table B3). These findings, altogether, signify that ligands bound to *Ec*MlaC can exhibit dynamic conformations and can have varying interactions.

4.3.5. Ligand binding and orientation dictate the volumetric expansion of the binding pocket of MlaC

In order to gain further insights into ligand binding, a comparative analysis of the structures of *Ec*MlaC available in different liganded forms and its orthologs was performed. These included structures of MlaC from *E. coli* (*Ec*MlaC_P$3_1$21_apo, PDB id: 6GKI; *Ec*MlaC_P$2_12_12_1$_holo1, PDB id: 5UWA; *Ec*MlaC_H3_holo1, PDB id: 7VR6), *Ralstonia solanacearum* (*Rs*MlaC_holo1, PDB id: 2QGU), *P. putida* (*Pp*MlaC_holo2, PDB id: 5UWB) and *P. aeruginosa* (*Pa*MlaC_holo2, PDB id: 6HSY); digit along with the term holo denotes the number of bound ligands (Table B2). Although *Ec*MlaC orthologs share diverse sequence similarities, their tertiary structures are similar, with RMSD values in the range of 0.7 to 3.9 Å for all Cα atoms (Figure B4A-S4F and Table B6). In-depth structure analysis reveals that the structures bound to two PLs show higher differences, primarily due to the outward bending of the α8 helix (Figure B4D-B4E).

The binding pocket of *Ec*MlaC is highly hydrophobic in nature (Figure 4.5A). A comparison of binding-pocket volumes reveals that *Ec*MlaC_P$3_1$21_apo has a smaller cavity (59.117 Å^3) than that of the *Ec*MlaC_P$2_12_12_1$_holo1 (381.515 Å^3) and *Rs*MlaC_holo1 (383.932 Å^3). Notably, *Pa*MlaC_holo2 and *Pp*MlaC_holo2 have ~ 19-fold larger binding-pocket volumes (1114.763 Å^3 and 1047.451 Å^3, respectively) owing to the binding of two PLs (Figure 4.5B). This clearly hints that ligand binding results in the opening of MlaC. This is contrary to canonical SBPs, which undergo conformational changes to attain a closed state upon ligand binding. The study also suggests that the binding-pocket volume of MlaC expands with the increasing number of ligands in the binding site. For further analysis, the MlaC structures were categorized into three groups, viz. Group I, II, and III, corresponding to apo, single-liganded, and double-liganded states, respectively. As per this scheme of classification, the *Ec*MlaC_H3_holo1 structure belongs to Group II. Surprisingly, its binding-pocket volume is significantly smaller (304.629 Å^3) than that of other Group II members (~380 Å^3). A previous molecular dynamics (MD) simulation

of *Ec*MlaC_P2$_1$2$_1$2$_1$_holo1 showed a transition from an open (ligand-bound) state to a closed (apo) state in the absence of ligand with a sequential decrease in the binding-pocket volume (Hughes et al., 2019). These observations suggest that the *Ec*MlaC_H3_holo1 structure represents a quasi-open transition state between the open (ligand-bound, *Ec*MlaC_P2$_1$2$_1$2$_1$_holo1) and closed (apo, *Ec*MlaC_P3$_1$21_apo) states.

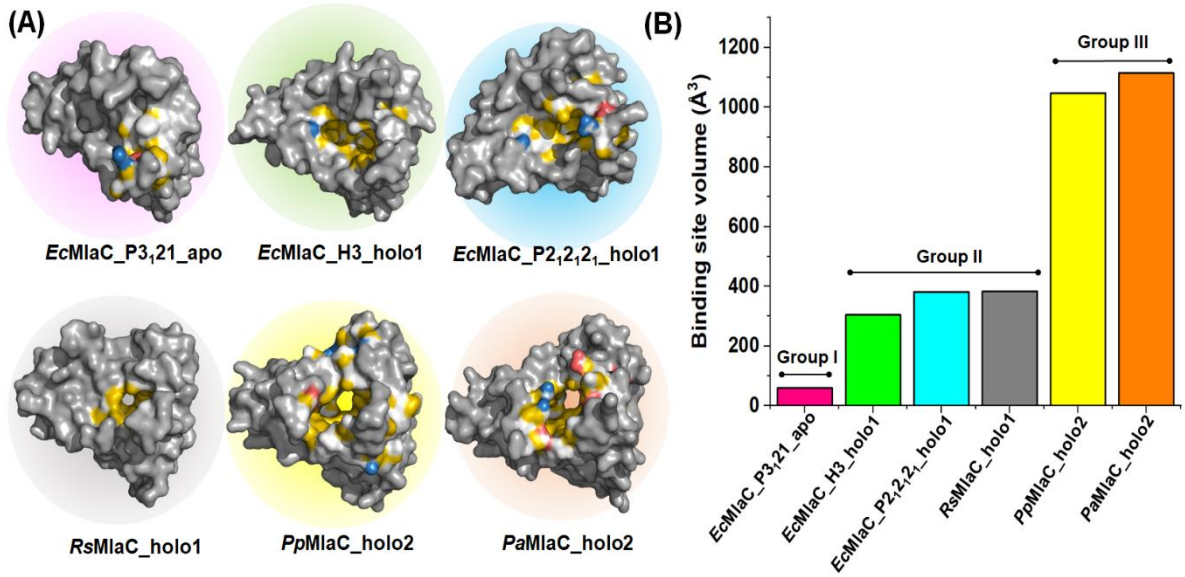


Figure 4.5. Comparison of the binding-pocket volumes. *(A) Hydrophobicity mapping of the binding pockets of EcMlaC_H3_holo1 and its orthologs. The binding sites have been computed by using the program CASTp, and the hydrophobicity mapping was performed by using the YRB scheme. (B) Histogram depicting the binding-pocket volumes of different EcMlaC_H3_holo1 and its orthologs. The apo, single-liganded and double-liganded MlaC species have been designated as Group I, II and III, respectively.*

After identifying the variations in the binding-pocket volumes, the conformations and orientations of the PLs bound to Groups II and III MlaC structures were compared (Figure B5A). All the Group II members (*Ec*MlaC_P2$_1$2$_1$2$_1$_holo1, *Ec*MlaC_H3_holo1, and *Rs*MlaC_holo1) are bound to the ligands (8ND, PEF and PEF', respectively) in β-γ conformation in a single binding plane, P_0 (Figure B5B). On the other hand, all the Group III members are bound to ligands in mixed conformations in two non-overlapping

planes. In *Pp*MlaC_holo2, two molecules of PEF (referred to as PEF_1 and PEF_2) are present in β-γ and γ-β conformations in two different non-overlapping planes referred to as P_1 and P_2, respectively (Figure B5C). Similarly, in *Pa*MlaC_holo2, the two molecules of diacyl glycerophospholipids H3T and GOT adopt β-γ and γ-β conformations, respectively, in two different non-overlapping planes (referred to as $P_{1'}$ and $P_{2'}$, respectively) (Figure B5D). These observations signify that during double ligation, MlaC accommodates ligands in mixed conformations and in two non-overlapping planes. Further investigation of these planes reveals that the set of P_1-P_2 planes is almost perpendicular to those of $P_{1'}$-$P_{2'}$. Altogether, these planes resemble the four walls of a rectangular box (Figure B5E). Interestingly, a comparison of the ligand-binding planes of the ligands of Groups II and III revealed that plane P_0 almost diagonally touches P_1-P_2 and $P_{1'}$-$P_{2'}$ (Figure B5F). These observations signify that plane P_0 is at a relaxed state as compared to planes P_1-P_2 and $P_{1'}$-$P_{2'}$. Based on this, it can be proposed that the binding of one molecule to the protein is relaxed compared to the two molecules, which needs to be restrained. In fact, owing to this restraint, two distinct subsites are formed in Group III members.

4.3.6. The role of the D1R1 subdomain is ambiguous

The individual residues of each of these four subdomains from all the *Ec*MlaC_H3_holo1 orthologs were studied in order to identify their contribution to the overall conformational change of the subdomains upon ligand binding. In *Ec*MlaC_H3_holo1, the bound phospholipid PEF establishes seven contacts (Figure 4.6A and Table B3). Among them, the interaction with the residues Met31 and Phe39 possibly helps in stabilizing the wider tails of the phospholipid PEF. However, the ligand interaction in *Ec*MlaC_$P2_12_12_1$_holo1 is not identical to that of *Ec*MlaC_H3_holo1 (Figure 4.6A and Table B3). This signifies that in both the *Ec*MlaC_holo structures, the subdomains have varying interactions with the bound PLs that lead to their conformational dynamicity. Furthermore, with respect to *Ec*MlaC_$P3_121$_apo, the aromatic ring of the residue Phe39 in *Ec*MlaC_holo structures demonstrate rotameric conformations (Figure 4.6A). Furthermore, with respect to *Ec*MlaC_$P3_121$_apo, the residues of the D1R1 subdomain tend to exhibit outward movements. In spite of this, the analysis of these three structures does not reveal any drastic movement of the D1R1 subdomain. The variation in ligand interacting residues is also observed in *Rs*MlaC_holo1 (Figure 4.6B and Table B3). Also, with respect to *Ec*MlaC_H3_holo1,

the corresponding residues in *Rs*MlaC_holo1 tend to exhibit outward movements suggesting an increase in the binding pocket volume.

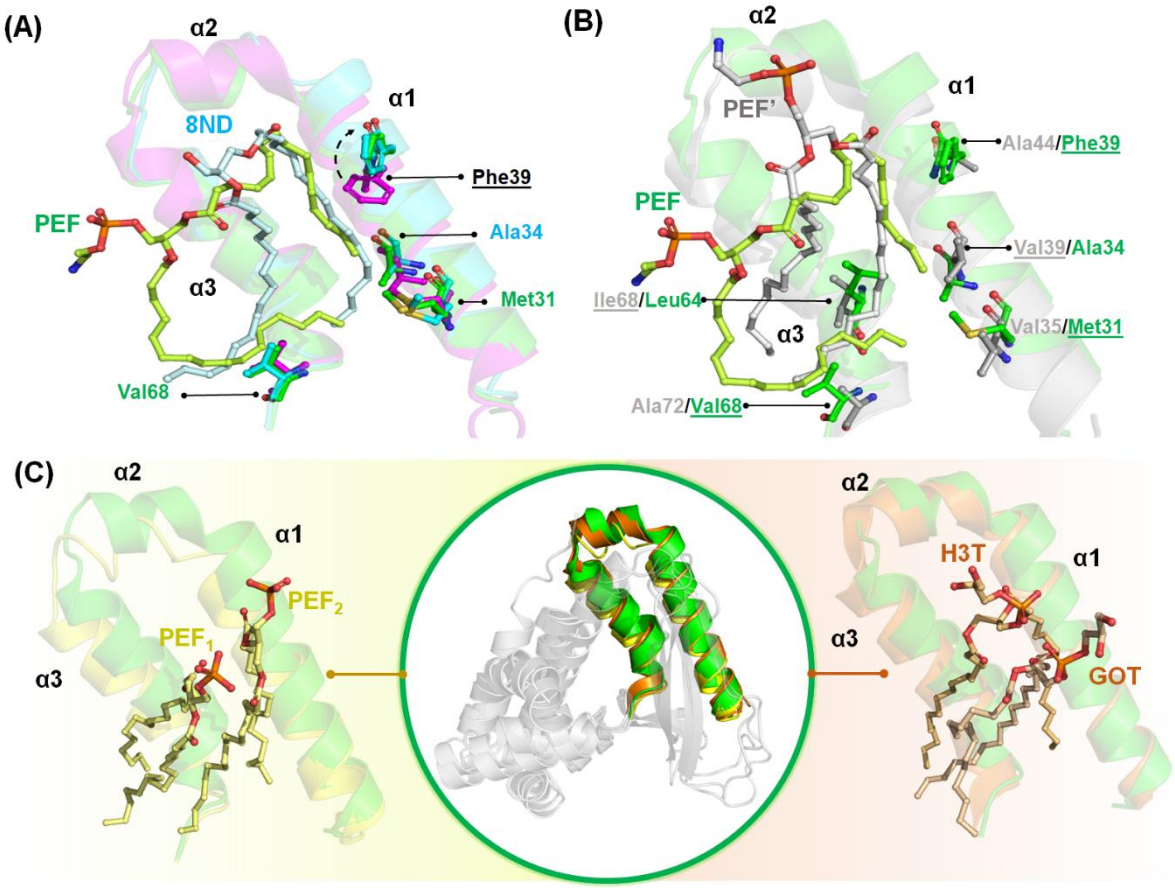


Figure 4.6. Structural details of the ligand interaction with the D1R1 subdomain. *Interaction of D1R1 subdomain residues of (A) EcMlaC_H3_holo1 (green) with PEF (limon) & EcMlaC_P2$_1$2$_1$2$_1$_holo1 (cyan) with 8ND (palecyan) and (B) RsMlaC_holo1 (grey) with PEF' (grey). The corresponding residues of EcMlaC_P3$_1$21_apo (magenta) and EcMlaC_H3_holo1 (green) are shown in (A) and (B), respectively. The change in orientation of the side chain of Phe39 is denoted by the dotted black arrow. (C) (Center) Superimposition of EcMlaC_H3_holo1 (green), PpMlaC_holo2 (yellow) and PaMlaC_holo2 (orange). For the purpose of visualization, only the D1R1 subdomain is highlighted. (Left) Orientation of the ligands PEF$_1$ and PEF$_2$ (light yellow) in PpMlaC_holo2. (Right) Orientation of the ligands GOT (wheat) and H3T (light orange) in PaMlaC_holo2. The ligands and the residues of interest are represented as ball-and-stick models.*

Unlike Group II, the members of Group III MlaC have segregated binding planes. As a consequence, in *Pp*MlaC_holo2, the plane P_1 of PEF_1 is closer to the α3 helix, while the plane P_2 of PEF_2 is closer to the α1 helix of D1R1 (Figure 4.6C). As a result, PEF_1 establishes nine contacts through its γ-chain and PEF_2 makes two contacts through its β-chain with D1R1. On the other hand, in *Pa*MlaC_holo2, the phospholipid H3T is situated closer to the D1R1 subdomain and, thus, is able to make five contacts through both chains. But GOT makes a single contact solely through its γ- chain (Figure B6A-B6B and Table B3). Thus, these variations in interactions of the ligands are contributed by the orientation of the binding planes. Moreover, α1 and α3 helices display slight outward shifts, which contribute to the increase in the binding pocket volume. However, the lack of substantial movement indicates that the subdomain as a whole is not flexible enough. Owing to the absence of conformational dynamicity, the exact role of the D1R1 subdomain remains ambiguous.

4.3.7. The D2R1 subdomain modulates the binding-pocket volume

As per previous reports, the α6 helix (D2R1 subdomain) does not establish any polar interaction(s) with the bound PL (Ekiert et al., 2017; Yero et al., 2021). However, in *Ec*MlaC_H3_holo1, the head group of the bound molecule PEF is pulled towards the D2R1 subdomain due to a hydrogen-bond interaction with the residue Try100 from the α6 helix. In contrast, the head region of the endogenously bound phospholipid 8ND in *Ec*MlaC_$P2_12_12_1$_holo1 does not establish any polar interaction(s) with the D2R1 subdomain (Figure B1D). Furthermore, the phospholipid PEF establishes 22 contacts with the D2R1 subdomain in *Ec*MlaC_H3_holo1. On the other hand, phospholipid 8ND in *Ec*MlaC_$P2_12_12_1$_holo1 makes only seven non-polar interactions with the D2R1 subdomain (Figure 4.7A-4.7B and Table B3). Thus, similar to the D1R1 subdomain, the D2R1 subdomain shows variation in the ligand interaction. Furthermore, a structural comparison of the α6 helix of *Ec*MlaC_$P3_121$_apo with *Ec*MlaC_H3_holo1 and *Ec*MlaC_$P2_12_12_1$_holo1 revealed an angular movement of 6.7º. This is mainly contributed by the conformational shifts of the residues Tyr105, Met111, and Tyr112 by 3.3 and 4.7 Å, 6.3 and 9.6 Å, and 4.2 and 4.5 Å, respectively (Figure B7A). Interestingly, in *Rs*MlaC_holo1, the ligand PEF' is oriented towards the right, compelling its γ-chain to be distant from the D2R1 subdomain. As a consequence, the interactions are confined only to its head and β-chain (Figure 4.7C and Table B3).

Notably, the C-terminal region (residues 105–116) of the α6 helix in *Rs*MlaC_holo1 exhibits a kink which results in the widening of the distance between the α6 helix and β4-strand (Figure 4.7D). This causes the binding pocket of *Rs*MlaC_holo1 to open up further. Overall, the analysis again suggests that in the single liganded states of MlaC, the bound PL can have varying conformations as well as interactions.

Notably, both the Group III members of MlaC possess the bend at the C-terminal (residues 105–112) region of the α6 helix, as observed in the case of *Rs*MlaC_holo1 (Figure B7B). As seen in the case of the D1R1 subdomain, in the Group III members, the ligand interaction is again dependent on the orientation of the binding planes. In *Pp*MlaC_holo2, the plane P_1 of PEF_1 is closer to D2R1 compared to that of plane P_2 of PEF_2. However, in *Pa*MlaC_holo2, both the planes $P_{1'}$ of GOT and $P_{2'}$ of H3T are close to the D2R1 subdomain. The difference in the arrangement of binding planes leads to variations in the number of ligand interactions (Figure B7C-B7D and Table B3).

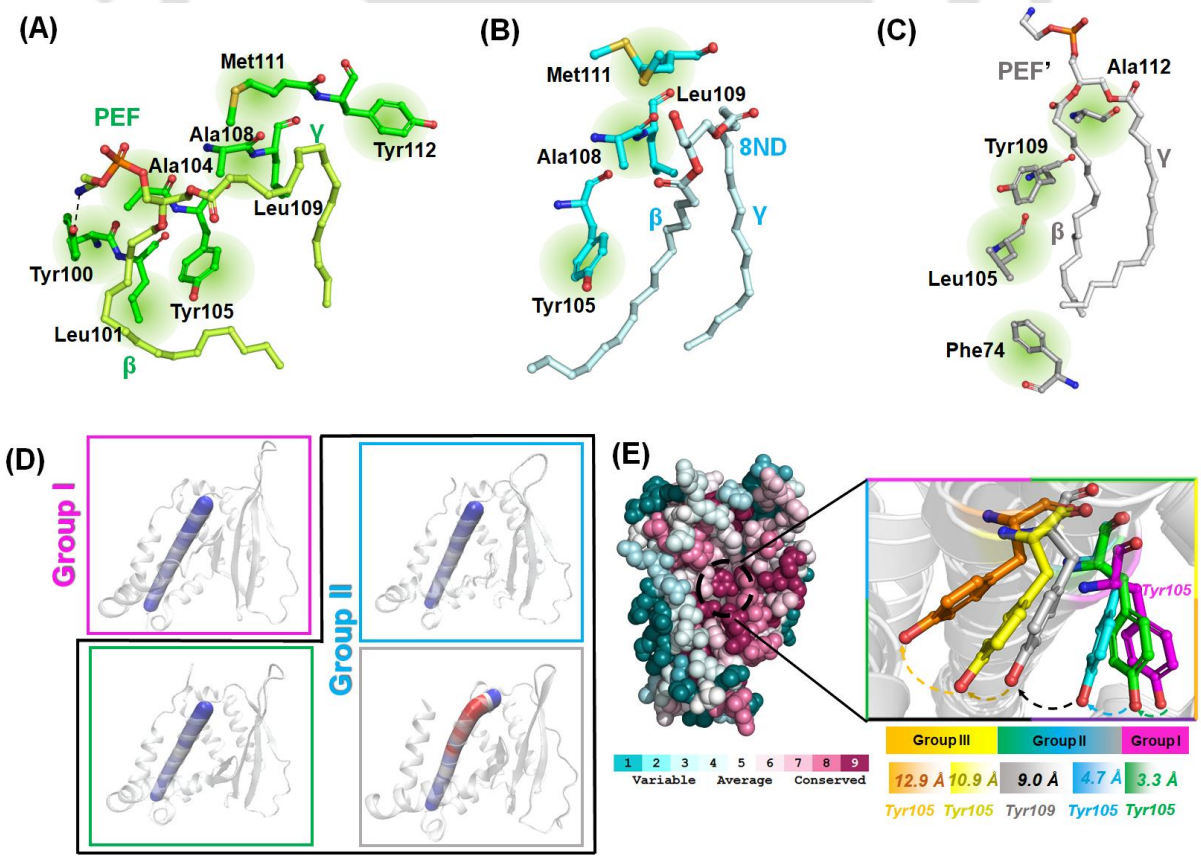


Figure 4.7. Structural details of the ligand interaction with the D2R1 subdomain. *Interaction of D2R1 subdomain residues of (A) EcMlaC_H3_holo1 (green) with PEF (limon), (B) EcMlaC_P2$_1$2$_1$2$_1$_holo1 (cyan) with 8ND (palecyan) and (C) RsMlaC_holo1 (grey) with PEF' (grey). The ligands and the residues of interest are represented as ball-and-stick models. (D) Analysis of helix curvatures of MlaC orthologs belonging to Groups I and II. The curvature has been color-coded using the blue-white-red scheme. As per the scheme, blue, white, and red color signify straight, curvature and angle, respectively. EcMlaC_P3121_apo, EcMlaC_P2$_1$2$_1$2$_1$_holo1, EcMlaC_H3_holo1, and RsMlaC_holo1 are enclosed in magenta, cyan, green and grey boxes, respectively. (E) Evolutionary conservation of the residue Tyr105 (Left) The conservation profile of MlaC is presented using a space-filled model. The amino acids are colored as per their conservation grades using the color-coding bar, with turquoise-through-maroon indicating variable-through-conserved. (Right) The distance moved by the residue Try105 in Groups II and III has been measured with respect to Group I.*

Notably, the residue Tyr105 that bisects the binding pocket shows substantial movement in all the *Ec*MlaC_H3_holo1 orthologs. Sequence comparison of these members revealed the conservation of the residue Tyr105 in the *Ec*MlaC_H3_holo1 orthologs. In comparison to *Ec*MlaC_P3$_1$21_apo, the residue Tyr105 demonstrates an outward movement of 3.3, 4.7, 9.0, 10.9 and 12.9 Å in *Ec*MlaC_H3_holo1, *Ec*MlaC_P2$_1$2$_1$2$_1$_holo1, *Rs*MlaC_holo1, *Pp*MlaC_holo2 and *Pa*MlaC_holo2, respectively (Figure 4.7E). Thus, the movement systematically increases and becomes the largest among the Group III members. It can be correlated that as the movement of the residue Tyr105 increases, the binding pocket of MlaC opens up. Altogether, the analysis affirms that conformational shifts in the residues of the D2R1 subdomain contribute to the increase in the binding pocket volume. However, in both Group II and III members, the bound ligands tend to show varying interactions, as observed in the case of other subdomains.

4.3.8. The D1R2 subdomain exhibits a change in β-sheet curvature upon ligand binding

The D1R2 subdomain is the only part of MlaC that comprises β-strands. The subdomain is constituted by four twisted β-strands (β1-β4) that are anti-parallel to each other and have interconnecting loops. The structural superimposition of *Ec*MlaC_H3_holo1 and its orthologs reveals a distinct change in the curvatures of β1-β4 strands upon ligand binding (Figure 4.8A). Due to this, the loops L5 and L6 show inward and outward movements, respectively. The distance between the α6 helix & β4-strand (in general) and the changes in the distances between the residues Ala112 (α6) & Ser169 (L6) and Val116 (α6) & Ser169 (L6) in *Rs*MlaC_holo1 have been reported as the principal determinants of the opening and closing of the binding pocket (Huang et al., 2016; Hughes et al., 2019). The distances between the corresponding residues Ala108 & Glu169 and Tyr112 & Glu169 are 10.2 Å and 13.9 Å in *Ec*MlaC_P3$_1$21_apo, 13.8 Å and 16.7 Å in *Ec*MlaC_P2$_1$2$_1$2$_1$_holo1 and 14.8 Å and 17.8 Å in *Ec*MlaC_H3_holo1, respectively (Figure 4.8B). Thus, the analysis shows that with respect to *Ec*MlaC_P3$_1$21_apo, distances between the D2R1 and D1R2 subdomains in the *Ec*MlaC_H3_holo1 and its orthologs have increased after ligand binding. A similar observation is made in the case of other *Ec*MlaC_H3_holo1 orthologs (Figure B8A-B8B). This change in distance results in the opening up of MlaC.

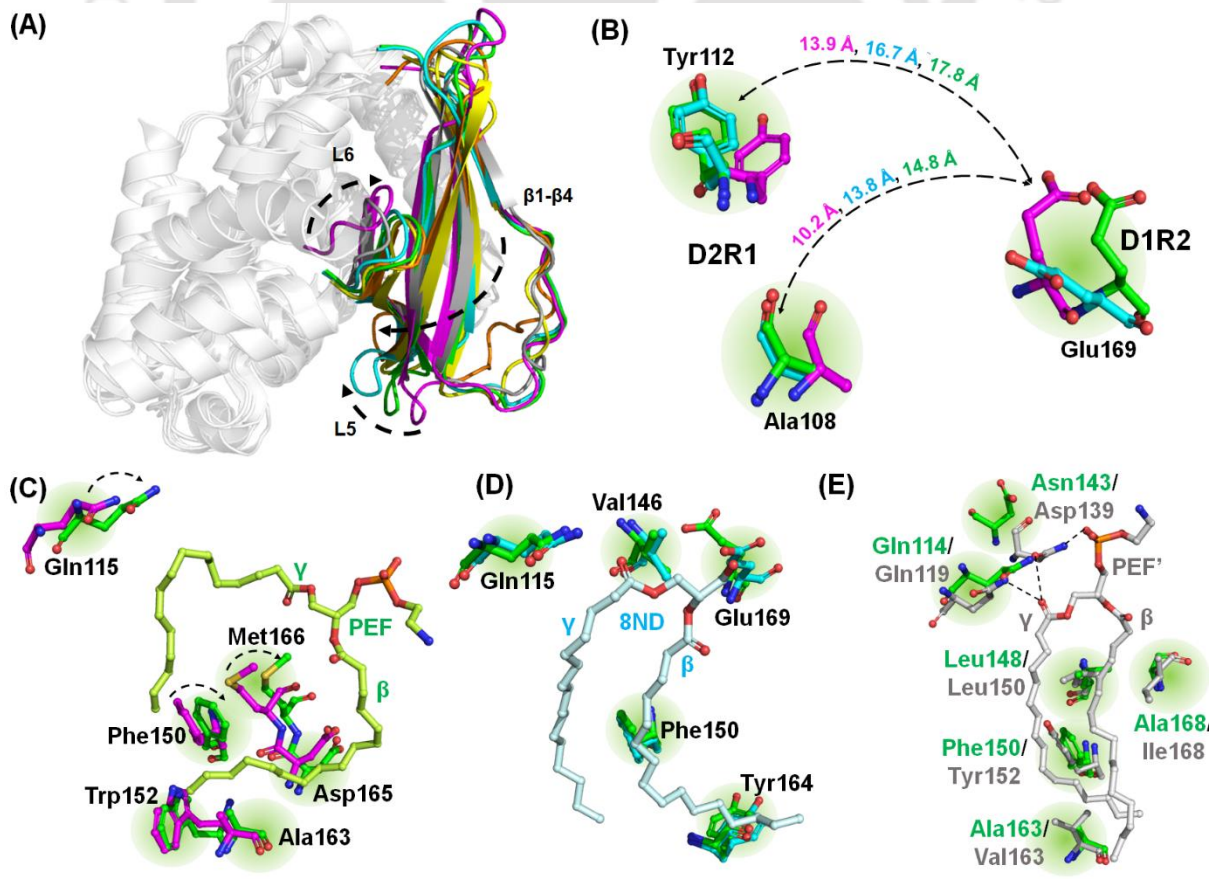


Figure 4.8. Structural details of the ligand interaction with the D1R2 subdomain. *Structural superimposition of EcMlaC_P3$_1$21_apo (magenta), EcMlaC_P2$_1$2$_1$2$_1$_holo1 (cyan), EcMlaC_H3_holo1 (green), RsMlaC_holo1 (grey), PpMlaC_holo2 (yellow), and PaMlaC_holo2 (orange). Upon ligand binding, β1-β4 strands undergo a change in curvatures resulting in the inward and outward movement of the L5 and L6 loops, respectively. (B) Movement of the residues from the D2R1 and D1R2 subdomains modulating the opening and closing of MlaC. The distances between the residues Ala108 & Glu169 and Tyr112 & Glu169 in EcMlaC_P3$_1$21_apo (magenta), EcMlaC_P2$_1$2$_1$2$_1$_holo1 (cyan), and EcMlaC_H3_holo1 (green) are mentioned and denoted by dotted double-headed arrows. Interaction of D2R1 subdomain residues of (C) EcMlaC_H3_holo1 (green) with PEF (limon), (D) EcMlaC_P2$_1$2$_1$2$_1$_holo1 (cyan) with 8ND (palecyan) and (E) RsMlaC_holo1 (grey) with PEF' (grey). The corresponding residues of EcMlaC_P3$_1$21_apo (magenta) and EcMlaC_H3_holo1 (green) are shown in (C) and (D-E), respectively. The ligands and the residues of interest are represented as ball-and-stick models. Hydrogen bonds and the residue movements are denoted by straight and curved dotted lines, respectively.*

Surprisingly, the distance between helix α6 and loop L6 in *Ec*MlaC_P2$_1$2$_1$2$_1$_holo1 and *Ec*MlaC_H3_holo1 is comparable. However, the binding-pocket volume of the two structures is 381.515 Å^3 and 304.629 Å^3, respectively. This difference of ~77 Å^3 in the binding-pocket volume is mainly contributed by the difference in the translational movements of β1-β4 strands owing to changed curvatures. Hence, in order to get further insight, the average translational movements of each β1-β4 strand from *Ec*MlaC_P2$_1$2$_1$2$_1$_holo1 and *Ec*MlaC_H3_holo1 were measured with respect to *Ec*MlaC_P3$_1$21_apo. With respect to *Ec*MlaC_P3$_1$21_apo, in *Ec*MlaC_H3_holo1, the β1-β4 strands exhibit average translations of 1.2, 1.2, 1.6, and 1.9 Å, respectively. Whereas, in *Ec*MlaC_P2$_1$2$_1$2$_1$_holo1, the β1-β4 strands exhibit translations of 2.3, 2.9, 3.0, and 2.7 Å, respectively. The results clearly indicate that β1-β4 strands from *Ec*MlaC_H3_holo1 have undergone less translational movement than *Ec*MlaC_P2$_1$2$_1$2$_1$_holo1. This contributes to the reduced binding pocket volume in *Ec*MlaC_H3_holo1, giving rise to a quasi-open state.

At the residue level, the interaction of the D1R2 subdomain with the ligands in MlaC shows a similar variation as observed in the case of the D1R1 and D2R1 subdomains. In *Ec*MlaC_H3_holo1, the phospholipid PEF makes ten contacts, whereas, in *Ec*MlaC_P$2_12_12_1$_holo1, the phospholipid 8ND forms eight contacts with the D1R2 subdomain (Figure 4.8C-4.8D and Table B3). In comparison to *Ec*MlaC_P3_121_apo, these interacting residues exhibit significant conformational shifts (Figure 4.8C). On the other hand, in *Rs*MlaC_holo1, the phospholipid PEF' establishes 17 contacts with the D1R2 subdomain (Figure 4.8E and Table B3). Thus, the findings again highlight that the single liganded states of *Ec*MlaC, including the ortholog, establish different contacts with the D2R1 subdomain.

As anticipated, in Group III members, the interactions of the two bound ligands with the D1R2 subdomain are again dependent on the orientation of their respective binding planes. In *Pp*MlaC_holo2, P_1 of PEF_1 is closer to the D2R1 subdomain, whereas P_2 of PEF_2 is closer to the D1R2 subdomain. On the other hand, in *Pa*MlaC_holo2, both the planes $P_{1'}$ of GOT and $P_{2'}$ of H3T are close to the D1R2 subdomain. Accordingly, the interactions of the ligands with the D1R2 subdomain change (Figure B9A-B9B and Table B3). Also, compared to *Ec*MlaC_H3_holo1, all these interacting residues exhibit outward movements, which contribute to the increase in the binding pocket volume of Group III members. Altogether, the findings signify that upon ligand binding, the curvatures of the β strands in the D1R2 subdomain would change, facilitating the opening of the binding pocket.

4.3.9. The D2R2 subdomain plays a significant role in accommodating two phospholipid molecules

The most striking feature of the D2R2 subdomain is the presence of a kink in the α8 helix of the Group IIII members. This results in an outward bending of the α8 helix. With respect to *Ec*MlaC_H3_holo1, the α8 helix of *Pp*MlaC_holo2 and *Pa*MlaC_holo2 makes kink angles of 68º and 73º, respectively (Figure 4.9A). Due to this drastic change in the α8 helix conformation, the binding-pocket volume of Group III members is increased by ~19-fold as compared to Group I members. In the Group III members, *Pp*MlaC_holo2 and *Pa*MlaC_holo2, kink occurs at the residues Gly194 and Gly195, respectively. The low helix-forming propensity of glycine is suggested to be responsible for the kink formation, whereas the conserved Trp196 residue is proposed

to stabilize the kink (Yero et al., 2021). However, in *Ec*MlaC and *Rs*MlaC, the residue glycine has been substituted by the residues Glu196 and Phe197, respectively. However, to further probe whether the kink formation is exclusive to Group II members, the α8 helix of Group III members and *Ec*MlaC_H3_holo1 (Group II) were compared.

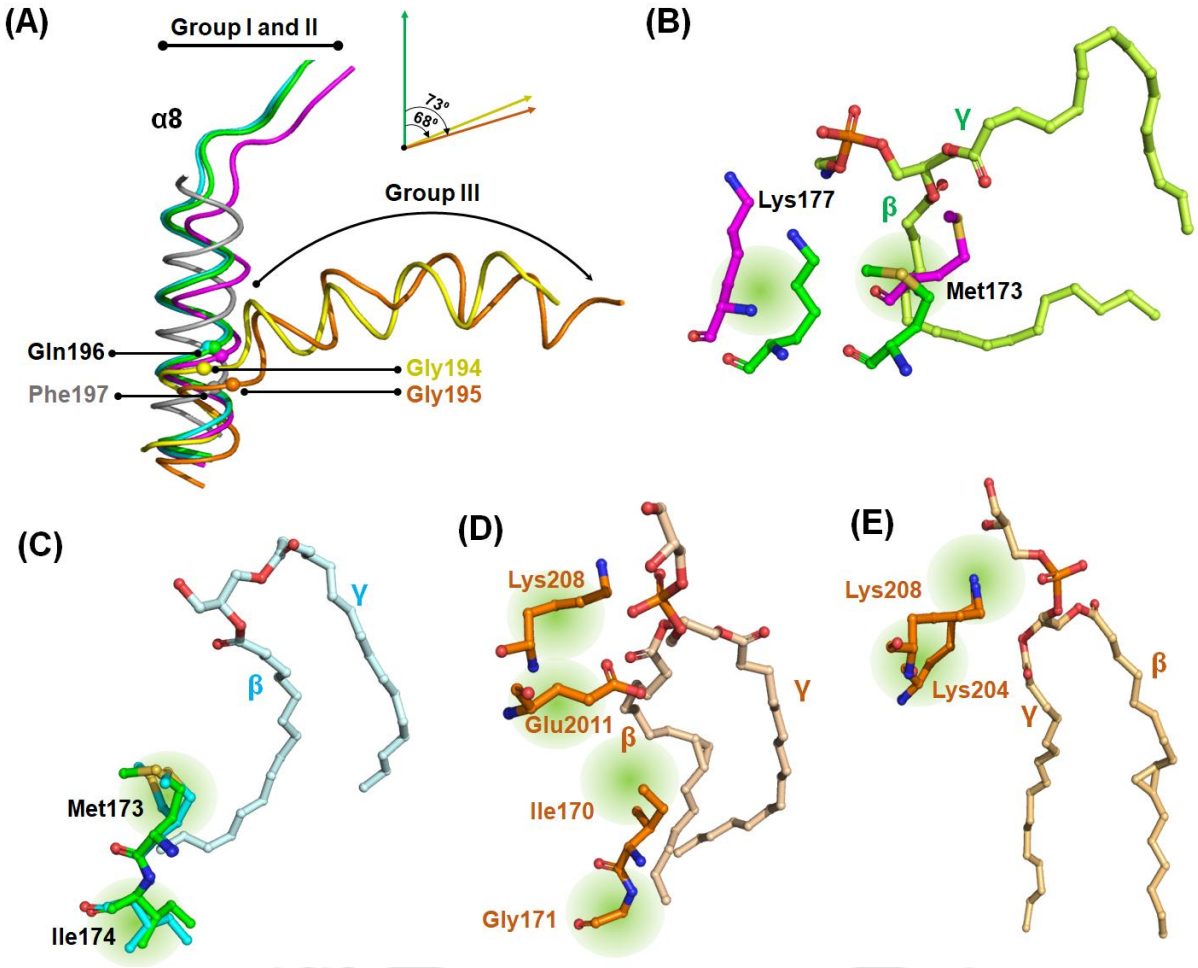


Figure 4.9. Structural details of the ligand interaction with the D2R2 subdomain. *(A) In Group III, the α8 helix forms a kink at Gly194/195. In Group I and II, the α8 helix remains straight, and the glycine residues are replaced by Gln196 or Phe197. The schematic representation of the kink angles of the Group III members is shown at the top right. Interaction of the D2R2 subdomain residues of (B) EcMlaC_H3_holo1 (green) with PEF (limon), (C) EcMlaC_P2$_1$2$_1$2$_1$_holo1 (cyan) with 8ND (palecyan) and PaMlaC_holo2 (orange) with (D) GOT(wheat) and (E) H3T (lightorange). The corresponding residues of EcMlaC_P3$_1$21_apo (magenta) and EcMlaC_H3_holo1 (green) are shown in (B) and (C), respectively.*

For the analysis, positions −6 to + 6 around the kink residue were considered. The results revealed that most of the amino acid residues around the kink are hydrophobic in nature. In *Pp*MlaC and *Pa*MlaC, the + 1 position is occupied by a tryptophan residue, whereas in *Ec*MlaC, it is a leucine residue (Figure B10A-B10C). An MSA analysis of Group III members showed that neither Gly195 at kink (at position 0) nor Trp196 (at position + 1) is conserved (Figure B10D-B10E and Table B7). Thus, in Group III members, since the residues Gly195 and Trp196 were hypothesized to form kink, their substitution suggests that these two residues might not be the sole determinants for the α8 helix bending. Furthermore, *Ec*MlaC has been reported to bind CL, which would require undergoing volumetric expansion (Henderson et al., 2016; Hughes et al., 2019). The AAA + proteins have been reported to bind CL through positively charged arginine and lysine (Makise et al., 2001). The D2R1 and D2R2 subdomains resemble the C-domain of AAA + proteins and possess arginine and lysine residues in all *Ec*MlaC_holo orthologs (Figure B11). Thus, it is tempting to speculate that, akin to Group III, Group II members can also bind two PLs by expanding the binding-pocket volume.

Furthermore, in *Ec*MlaC_H3_holo1, the phospholipid PEF makes four contacts, whereas, in *Ec*MlaC_$P2_12_12_1$_holo1, the phospholipid 8ND makes three contacts with the D2R2 subdomain (Figure 4.9B-4.9C). A comparison of *Ec*MlaC_$P3_121$_apo and *Ec*MlaC_$P2_12_12_1$_holo1 revealed that the α7 helix also undergoes conformational changes upon ligand binding, causing the interaction between protein (Lys177) and the bound PL (Figure B12). However, in *Rs*MlaC_holo1, this conformation change of the α7 helix is comparatively subtle. Furthermore, owing to the rightward shift of the ligand, no significant interaction is established between the protein and PEF' molecule.

Also, in Group III members, both the helices α7 and α8 show conformational changes (Figure 4.9A and Figure B12). In *Pp*MlaC_holo2, the planes P_1 of PEF_1 and P_2 of PEF_2 are situated away from the D2R2 subdomain and hence, do not establish any interaction. In *Pa*MlaC_holo2, the planes $P_{1'}$ of GOT and $P_{2'}$ of H3T are closer to the D2R2 subdomain, which enables the PLs to establish interactions (Figure 4.9D-4.9E). Akin to other subdomains, the orientation of the binding planes again dictates the interaction of the PLs with the D2R2 subdomain. Thus, the systematic study of the four subdomains clearly shows that Group II and III members do not establish identical

interactions with the bound PLs. Also, the binding pocket of MlaC is extremely flexible and can accommodate PLs in varying conformations.

4.3.10. The phospholipid head groups show variations in the polar interactions

In all the liganded structures of MlaC, the head groups of the bound PLs are solvent-exposed, which are involved in polar interactions. In order to probe whether there exist variations among them, the polar interactions established between the ligands and MlaC orthologs were studied. In *Ec*MlaC_H3_holo1 and *Ec*MlaC_P$2_1 2_1 2_1$_holo1, the head groups of the bound PLs form 3 hydrogen bonds with *Ec*MlaC and water molecules (Figure B1D and Table B3). Surprisingly, in *Rs*MlaC_holo1, the head group of PEF' makes 14 hydrogen bonds with the residues Gln119 and Asn143 and nine water molecules (Figure 4.10A). In Group III members, there are no water molecules involved in the interaction with the head group of the bound PLs. In *Pp*MlaC_holo2, although PEF_1 does not show any interaction, the head group of PEF_2 forms five hydrogen bonds (Figure 4.10B and Table B3). Similarly, in *Pa*MlaC_holo2, the phospholipids GOT and H3T form two and one hydrogen bond, respectively (Figure 4.10C and Table B3). However, in Group III members, the PL headgroups, in spite of being solvent-exposed, do not interact with water molecules. To sum up, both Group II and Group III members show variations in polar interactions. It can be speculated that this variation is due to head group specificities and might play a critical role in determining the orientation of PLs.

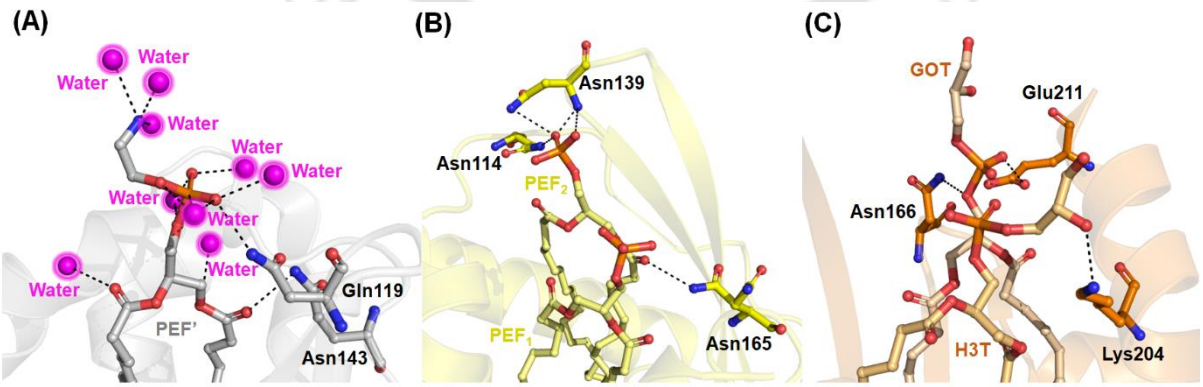


Figure 4.10. Phospholipid head group interactions with EcMlaC_H3_holo1 orthologs. *Polar interactions of (A) PEF' with RsMlaC_holo1, (B) PEF2 with PpMlaC_holo2 and (C) GOT and H3T with PaMlaC_holo2. The interacting residues,*

water molecules, and hydrogen bonds are depicted as ball-and-stick models, magenta spheres, and black dotted lines, respectively.

4.3.11. MlaC demonstrates polyspecificity towards phospholipids

The structural analysis firmly establishes that MlaC from *E. coli* and its orthologs can have varying interactions with PLs. Hence, it becomes imperative to find out whether *Ec*MlaC_H3_holo1 can bind different ligands (8ND, GOT and H3T). Therefore, to estimate the binding affinities, molecular docking experiments of *Ec*MlaC_H3_holo1 with these ligands along with PEF were performed. The docking results demonstrate that PEF molecule binds to *Ec*MlaC_H3_holo1 in β-γ conformation with a binding energy of −4.0 kcal mol^{-1}. The molecule is held by 17 residues while its head group remains solvent-exposed, which establishes hydrogen bonds with Ala104, Gln107 and Met111 (Figure 4.11A and Table B8). On the other hand, the phospholipids 8ND, GOT and H3T bind to *Ec*MlaC_H3_holo1 with a binding energy of −6.91, −3.13 and −2.78 kcal mol^{-1}, respectively. The 8ND and H3T molecules also bind in β-γ conformations, whereas GOT molecule docks in the γ-β conformation. Furthermore, both 8ND and GOT molecules interact with 20 residues, whereas the H3T molecule interacts with 22 residues. Also, among these three PLs, the head groups of GOT and H3T establish hydrogen bonds (Figure 4.11B-4.11D and Table B8). However, in the case of 8ND, the head remains shielded from the solvent owing to its reduced size. On the other hand, in the case of phospholipids PEF, GOT and H3T, the head groups remain inclined towards the D2R1 subdomain owing to the interactions. This indicates that the head group might play a critical role in determining the orientation of the ligand in MlaC. All four ligands have overlapping binding planes with varying conformations and interact with all four subdomains (Figure B13A-B13E). This signifies that during single ligation, the ligand can bind in any conformation and orientation as the binding pocket of MlaC does not impose any restraint on the ligand.

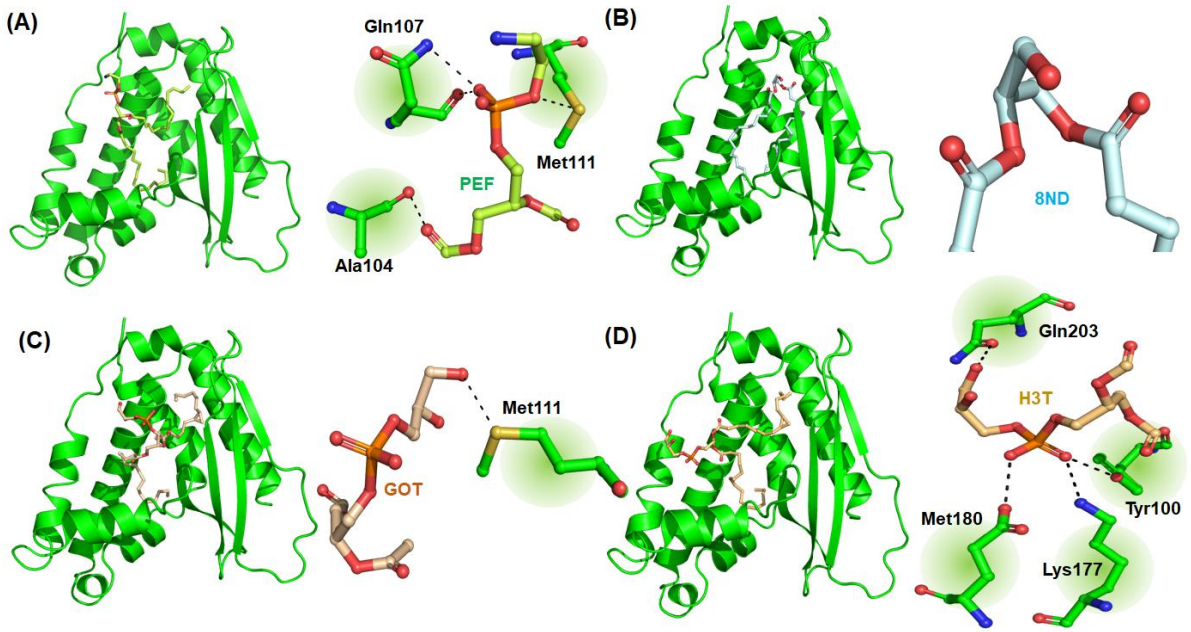


Figure 4.11. Molecular docking of EcMlaC_H3_holo1 with phospholipids. *Molecular docking of EcMlaC_H3_holo1 with (A-D) (Left) PEF, 8ND, GOT and H3T. (Right) Polar interactions (dotted lines) between EcMlaC_H3_holo1 and the head groups of the respective PLs. The interacting residues and the ligands are represented as ball-and-stick models, while the hydrogen bonds are represented as dotted lines. Further details of the interactions are provided in Table B8.*

4.3.12. MlaC has a unique ligand-binding mechanism involving ligand re-orientation

MlaC has been suggested to be involved in the bidirectional transfer of PLs (Low et al., 2021). Previously, the export function of PL through MlaFEDB from *E. coli* (PDB id: 6XBD) has been highlighted by considering the substrate conformations (Coudray et al., 2020). The cryo-EM structure reveals the presence of two PLs in two different conformations trapped in the process of being transferred to the protein MlaD. Based on this, it was proposed that the MlaFEDB complex would extract the PLs from the IM and re-orient them to “extended” or “tails-up” conformation. Furthermore, due to the presence of two PLs in the MlaFEDB, it was speculated that MlaC might transport two molecules of PL per ATP-hydrolysis cycle. Although the proposed PL transport model is overall insightful, it does not include the role of segmented domain arrangement, binding-pocket volume, and ligand interactions of MlaC.

Structural analysis of Groups II and III members reveals that MlaC in its apo (closed) state would undergo a volumetric expansion at the binding site in order to accommodate a PL, thus, transitioning to a holo (open) state. The results of molecular docking experiments clearly indicate that the tails of PLs would first enter the binding pocket of MlaC. It can be proposed that unliganded MlaC, after diffusing through the periplasm, would dock to MlaD of IM (donor membrane). This would be followed by the initial expansion of the binding site brought about by an outward movement of the D2R1 and D1R2 subdomains. This would involve a change in the curvatures of β1-β4 strands, leading to the opening of MlaC, in turn, capturing the PL at the binding site. The head group of the PL would establish interactions for the proper orientation of the PL molecule that facilitate its transfer to the MlaA-Omp complex (Figure 4.12A). In the case of two PL binding, the outward movement of the α8 helix results in a further increase of the binding-pocket volume to accommodate the second ligand (Figure 4.12A). After ligation, MlaC would undock from the IM and diffuse through the periplasm to dock to the MlaA-Omp complex of the OM (Acceptor membrane) for the transfer of PL. Our analysis reveals that the process of ligand binding involves a coordinated sequence of structural changes in the subdomains. Thus, we propose the term “segmented domain movement” mechanism in order to describe the binding of ligands to MlaC.

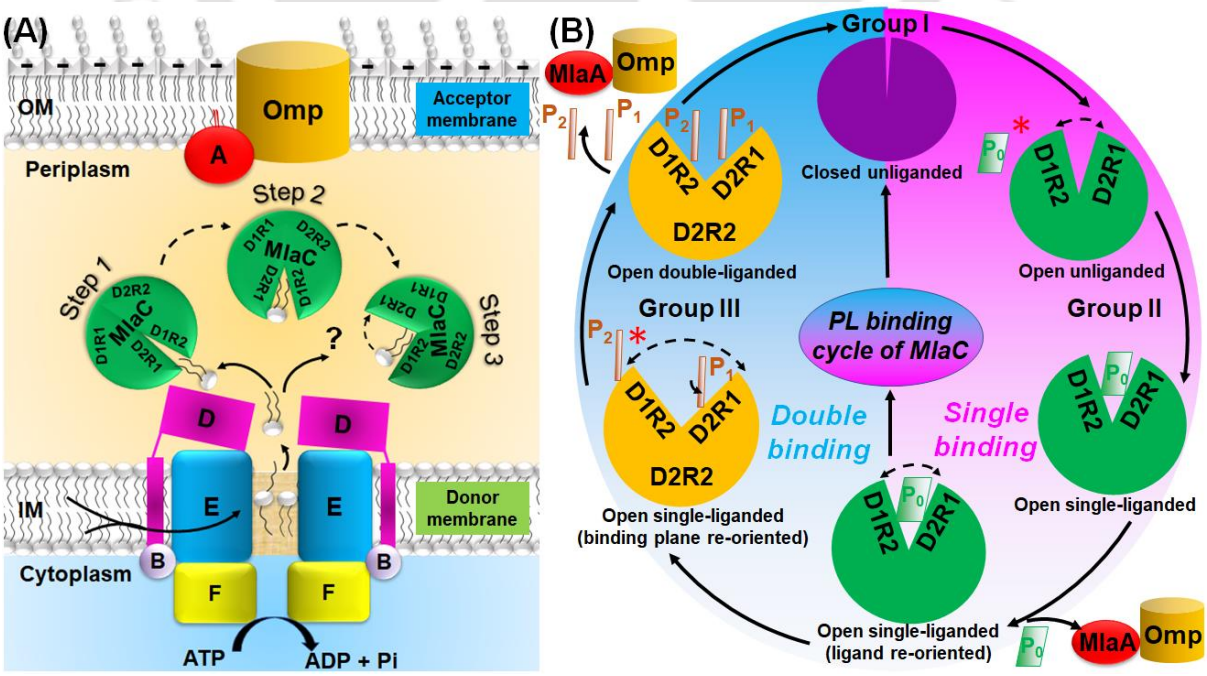


Figure 4.12. Overview of phospholipid transport by the Mla system. *(A) The proposed model of PL transport. The initial extraction of PLs from IM, followed by their binding to MlaC. This would involve three steps: (1) opening of the binding site due to the outward movement of the D1R2 and D2R1 subdomains, (2) initial docking of the PL molecule, and (3) re-orientation of PL in the binding site. For two PL binding, the α8 helix would move outward in order to increase the binding pocket volume. (B) PL binding cycle of MlaC. During single PL binding, the apo (close) MlaC would open up owing to the movement of the D1R2 and D2R1 subdomains, followed by PL binding at the binding plane P_0. After ligand re-orientation, the PL would be transferred to the MlaA-Omp complex. In the case of two PL binding, the α8 helix of the D2R2 subdomain would move outward, and the binding plane P_0 would re-orient to P_1, giving rise to two subsites in the binding pocket. This would be followed by the binding of the second PL at the binding plane P_2. Subsequently, the two PLs would be transferred to the MlaA-Omp complex. The binding plane P_0 is represented as a green parallelogram, while the binding planes $P_{1\text{-}2}$ are presented as orange rectangles. The transient states during the PL binding cycle are denoted by asterisks.*

An in-depth structural analysis of MlaC reveals that in the case of double-liganded states, two PLs are oriented in two different (β-γ and γ-β) conformations with non-overlapping binding planes ($P_1/P_{1'}$ and $P_2/P_{2'}$). It seems that the protein binds the first PL in a random conformation at plane P_0. Considering the observations in the available MlaC structures, it is tempting to propose that in the apo state, the protein remains in a closed state (Group I). Upon ligand binding, the protein opens up by regulating the D1R2 and D2R1 subdomains and traps the PL in a particular conformation at the plane P_0, leading to a single-liganded open state (Group II). Following ligand re-orientation, the protein MlaC transfers the PL to MlaA, attaining an apo state (Group I). However, in the case of two PLs binding to the protein MlaC, the α8 helix of the D2R2 subdomain moves outward and the plane P_0 re-orients to $P_1/P_{1'}$, leading to the formation of two subsites. This allows the binding of the second PL at the plane $P_2/P_{2'}$, giving rise to a double-liganded open state (Group III). Subsequently, the two PLs are transferred to the MlaA-Omp complex, resulting in an apo state of the protein MlaC (Group I) (Figure 4.12B).

4.3.13. MlaC constitutes a novel class of non-canonical SBPs

The structural analysis of MlaC establishes it to be an atypical SBP with a unique domain arrangement and ligand binding mechanism. Typical SBPs- Type I, II and III (referred to as SBP-I, II and III, respectively) possess the NTD-CTD arrangement and belong to the protein class α/β (SCOP id: 1000002). Furthermore, SBP-I, II and III possess Rossmann(2x3)oid (SCOP id: 2000016), SBP2HA-like (SCOP id: 2000060) and chelatase-like (SCOP id: 2001032) fold, respectively, and belong to Type 1 solute binding protein-like (SCOP id: 3000125), Type 2 solute binding protein-like (SCOP id: 3000083) and metal receptor (SCOP id: 3001679) superfamily, respectively (Figure B14A-B14C). Owing to such structural and evolutionary similarities, SBP-I, II and III follow similar ligand binding mechanisms involving NTD-CTD movement.

However, a scan in the SCOP database reveals that MlaC belongs to the MlaC-like family (SCOP id: 4005548), which comes under the NTF2-like superfamily (SCOP id: 3000472) and possesses the cystatin-like fold (SCOP id: 2000326) and belongs to protein class α + β (SCOP id: 1000003) (Figure 4.13A). The functional analogs of MlaC viz., LolA (LolABCDE transporter) and LptA (LptABCDEFG transporter) also have different ancestries. Similar to MlaC, both these proteins connect the OM and the IM, do not possess NTD-CTD arrangement and are involved in the transport of lipoprotein and LPS, respectively. Both the proteins LolA and LptA belong to the β-protein class (SCOP id: 1000001) (Figure 4.13B). On the other hand, the protein LolA belongs to the LolA-like superfamily (SCOP id: 3001964), possessing a LolA-like fold (SCOP id: 2001203). Whereas LptA belongs to the OstA-like beta-folder (SCOP id: 3002196) superfamily and LptA-like family (SCOP id: 4004381), possessing beta-folder fold (SCOP id: 2001334) (Figure 4.13B). Owing to such structural and evolutionary dissimilarities, the three proteins follow different ligand binding mechanisms. Thus, based on the study, we propose a new class of SBP named non-canonical SBPs constituted by MlaC, LolA and LptA. The members of this class do not possess NTD-CTD arrangement and would have different binding mechanisms.

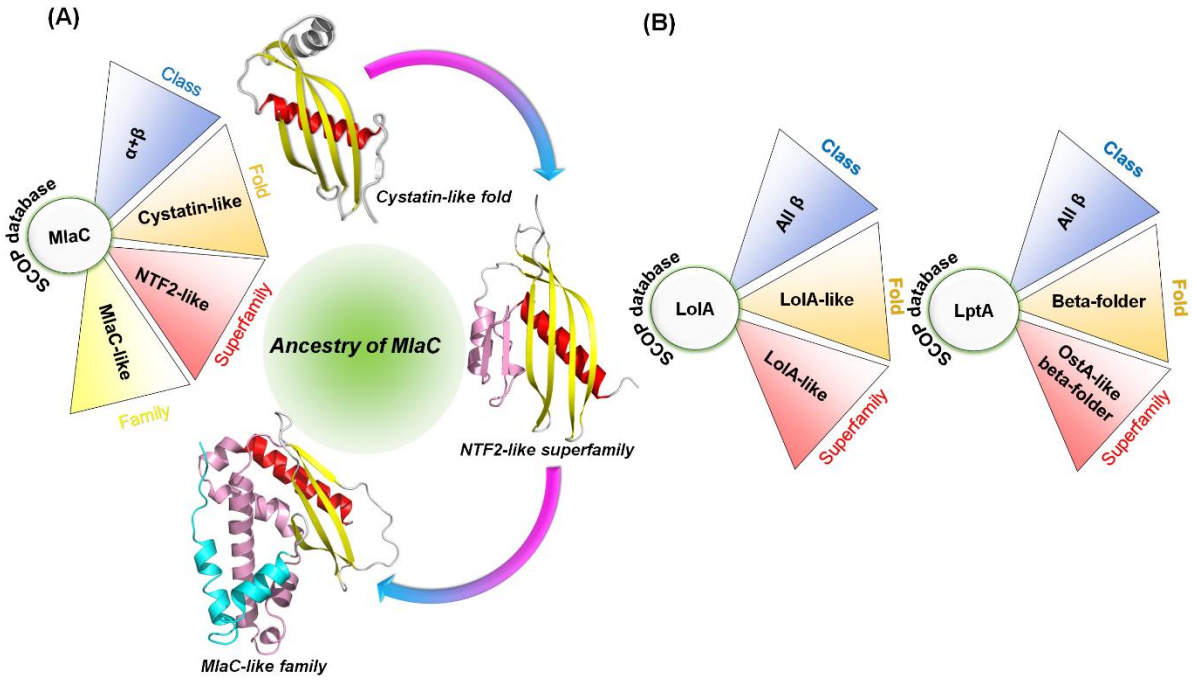


Figure 4.13. Evolutionary analysis of MlaC. *(A) Ancestral study of MlaC. (Left) Classification for MlaC provided by the SCOP database: α + β (Class), Cystatin-like (Fold), NTF2-like (Superfamily), and MlaC-like (Family). (Right) Comparison of MlaC-like family with the cystatin-like fold and NTF2-like superfamily. The cystatin-like fold is marked by a single α-helix (red) that is packed against coiled anti-parallel β-sheets (yellow). The signature feature of the NTF2-like superfamily is the insertion of β-α(2)-β (pink). The MlaC-like family is marked by the insertion of α(5) (pink) insertions and elongation of α(2) (cyan) following the main helix and β-strands, respectively. (B) Classification provided by the SCOP database for LolA: All β (Class), LolA-like (Fold) and LolA-like (Superfamily) and LptA: All β (Class), Beta-folder (Fold) and OstA-like beta-folder (Superfamily).*

4.4. Discussion

The protein MlaC of the Mla system is functionally analogous to the SBP component of ABC importers and ferries PLs between the OM and IM (Malinverni and Silhavy, 2009; Ekiert et al., 2017). The protein possesses a non-canonical structure and has been reported to be involved in the bidirectional transport of PLs (Ekiert et al., 2017; Hughes et al., 2019; Coudray et al., 2020; Yero et al., 2021). However, an in-

depth investigation providing mechanistic insights into the function of MlaC is lacking in the literature. This study reports the crystal structure of *Ec*MlaC in a quasi-open state with an endogenously bound phospholipid PEF and proposes a unique mechanism of ligand binding that has not been reported to date.

4.4.1. A unique binding pocket that is extremely plastic

The crystal structure of *Ec*MlaC in complex with phospholipid 8ND has previously been reported. However, our study reports the crystal structure of *Ec*MlaC in complex with phospholipid PEF in an unusual spread-out conformation. This clearly hints that *Ec*MlaC can accommodate PLs in varying conformations. Owing to the difference in ligand conformations, both structures show significant variations in ligand interactions and binding pocket volumes. Similar to the phospholipid 8ND, the head group of PEF is solvent-exposed but establishes hydrogen bonds with *Ec*MlaC and water molecules while the acyl chains remain buried in the binding pocket. Thus, the binding pocket of *Ec*MlaC resembles a box-like cavity without any lid. The archetypal form of such type of LPTs possesses a cavity sufficient for accommodating one lipid molecule (Wong et al., 2019). However, through extensive structural analysis, our work for the first time establishes that some *Ec*MlaC orthologs are able to bind to two PLs by splitting the binding pocket into two sub-sites. This is done by accommodating the PLs in two different planes leading to a substantial increase in the binding pocket volume. This extreme plasticity of the binding pocket can be attributed to the unique subdomain arrangement of *Ec*MlaC. Such an arrangement gives rise to segmental mobilities, which leads to a flexible binding pocket. These findings provide new mechanistic insights into the process of multiple ligations that have not been reported before and firmly establish the binding pocket to be enormously plastic.

4.4.2. Two subdomains mainly execute the opening and closing of the binding pocket

The previous structural studies on *Ec*MlaC have not considered its unique subdomain arrangement and hence, the individual contributions of the subdomains have been totally neglected. Such an arrangement gives rise to segmental mobilities leading to conformational dynamicity. Such changes have also been observed for other LTPs (Tremblay et al., 1996; Yoder et al., 2001; Li et al., 2014). Our analysis clearly shows the D2R1 and D1R2 subdomains primarily regulate the opening and closing of *Ec*MlaC

by modulating the distance between them. Upon ligand-binding, the anti-parallel β-strands located in the D2R1 subdomain exhibit translational movements, which alter the β-sheet curvature. Such a change in curvature has been associated with the opening of the binding pocket in proteins possessing the NTF2-like fold and is mainly to accommodate the hydrophobic tails of PLs (Fujiwara et al., 2015; Basanta et al., 2020). Additionally, in *Ec*MlaC, this change in curvature is further complemented by the conformational changes (particularly in the side chains of the amino acid residues) in the α6 helix located in the D1R2 subdomain. The conserved residue Tyr105, which bisects the binding pocket of *Ec*MlaC and its orthologs, exhibits the most distinct outward movement on ligand binding. This can be correlated to the higher propensity and mobility of conserved tyrosine residue in the binding sites of proteins (Villar and Kauvar, 1994; Khazanov and Carlson, 2013; Li et al., 2014). In *Ec*MlaC orthologs, the movement of Tyr105 increases with the systematic increase in the volume of the binding pocket. This suggests that the mobility of Tyr105 is critical for the opening of the protein. Interestingly, for the accommodation of two PLs, the α8 helix located in the D2R2 subdomain is required to exhibit an outward movement, thereby increasing the binding pocket volume. Our study reveals that both Group II and III members possess the required attributes to interact with two PLs. This hints that *Ec*MlaC might be able to bind to two PLs, thereby supporting the previous report of the protein binding to CL (Hughes et al., 2019).

Structural analysis and docking studies firmly establish that PLs form both polar and non-polar bonds with *Ec*MlaC. However, hydrophobic interactions between the acyl chains of PLs and the binding pocket of *Ec*MlaC mainly stabilize the *Ec*MlaC-ligand complex. This has also been previously observed for various other LTPs also (Yoder et al., 2001; Hoh et al., 2005). Depending upon the conformation of the tails of the bound ligands, the number of interactions stabilizing the MlaC-ligand complex varies. Surprisingly, both the Group II and III members show significant variation in the number of hydrogen bonds.

4.4.3. A unique mechanism of ligand binding not observed among canonical SBPs

Canonical SBPs with NTD and CTD arrangement follow one of the four major mechanisms of ligand binding, (1) Venus Flytrap (Mao et al., 1982), (2) asymmetric

domain movement (Pandey et al., 2016), (3) one domain movement (Chandravanshi et al., 2020a) and (4) subdomain movement (Chandravanshi et al., 2020b). (Figure 4.14A-4.14D). As the overall structure of *Ec*MlaC does not possess the typical SBP characteristic, the protein would not follow the previously reported ligand-binding mechanism. Added to that, to obtain mechanistic insights into a protein, the molecular details of its conformational changes in the intermediate states are required. However, these intermediate states being transient, are difficult to trap. The crystal structure of *Ec*MlaC obtained in this study represents the quasi-open state and, thus, bridges the gap between apo and single-liganded states. Furthermore, an extensive comparative structural analysis of the apo, single- and double-liganded states of *Ec*MlaC and its orthologs revealed their different conformational landscapes of ligand binding. Since the subdomain movements exhibited by *Ec*MlaC and its orthologs do not fall into any of the previously reported mechanisms of ligand binding, it is tempting to propose a novel derived mechanism, viz. “segmented domain movement”, which involves the reverse mechanism of binding-site opening (Figure 4.14E). The mechanism involves the extraction of PLs from the IM complex by MlaC aid by the segmental movements of the D1R2, D2R1 and D2R2 subdomains. The tails of the PLs first enter the binding pocket of MlaC and docks, while the head groups remain solvent-exposed. The solvent accessibility of the head groups of the PLs might help in the proper presentation of the PLs to the acceptor membrane and help in the PL transfer process. Such a mechanism of substrate presentation has been previously reported for another LTP, viz. phosphatidylinositol transfer protein (PITP) (Yoder et al., 2001). Furthermore, such a unique mechanism of ligand binding has been observed in the functional analogs of MlaC, LolA and LptA (Murahari et al., 2013; Kaplan et al., 2018; Wong et al., 2019). Additionally, these three proteins have diverse ancestries. Hence, considering their atypical characteristics, we have proposed a new class of SBPs named non-canonical SBPs that consist of MlaC, LolA and LptA.

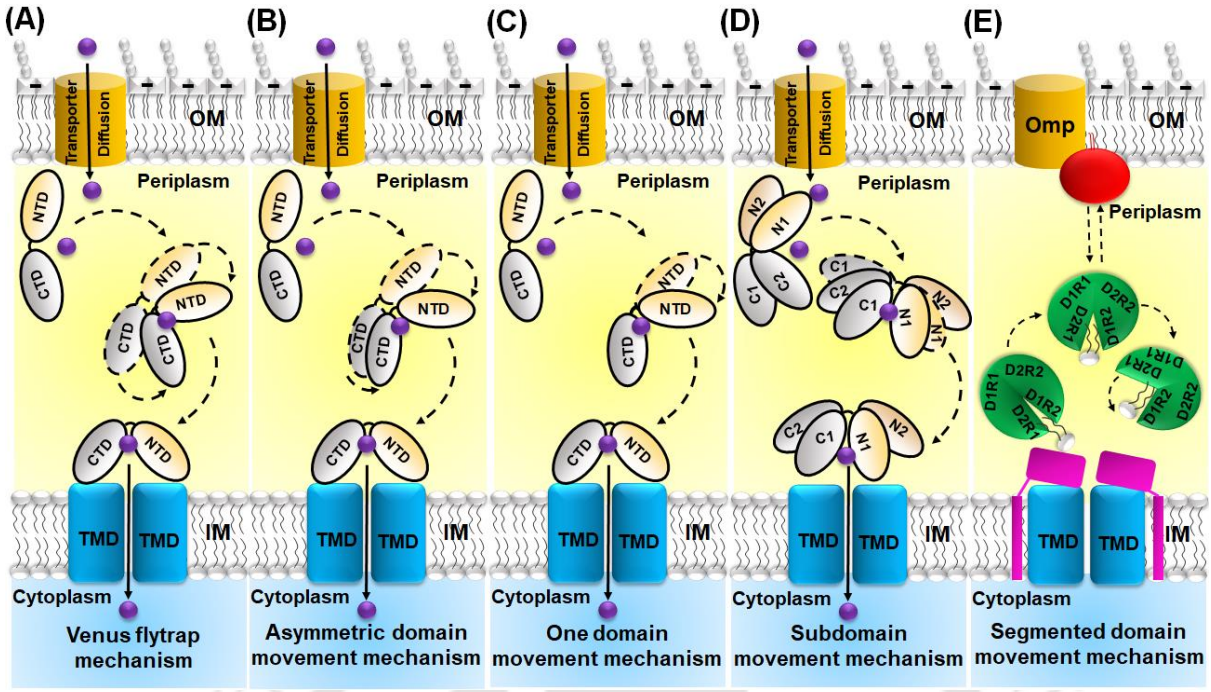


Figure 4.14. Schematic representation of the major models proposed for ligand-binding mechanisms. *(A) The "Venus Flytrap" mechanism (Mao et al., 1982). (B) The "asymmetric domain movement" mechanism (Pandey et al., 2016). (C) The "one domain movement" mechanism (Chandravanshi et al., 2020a). (D) The "subdomain movement" mechanism (Chandravanshi et al., 2020b). (E) The "segmented domain movement" mechanism (this study). NTD, CTD, and the associated ligands of canonical SBPs are depicted in grey, yellow and purple, respectively. Omp/OM proteins, TMD, MlaD, MlaC and MlaA, are shown in orange, cyan, pink, green, and red, respectively. The conformational changes of the domains(subdomains) upon ligand binding are depicted with dotted lines.*

4.5. Conclusion

This study provides significant insights into the structural and mechanistic aspects of *Ec*MlaC. To the best of our knowledge, it is the first comprehensive analysis of *Ec*MlaC and its orthologs. The extensive structural study establishes the unique segmented domain arrangement in *Ec*MlaC and how the respective subdomains interact with PLs. This led to the proposition of "segmented domain movement" that would involve a reverse mechanism of the binding-site opening. Such a mechanism has not been observed in any known SBP. However, the proposed mechanism should

be verified by further studies. The work also proposes a new class of SBPs named non-canonical SBPs, which are constituted by MlaC and other SBPs that are devoid of typical NTD-CTD arrangement as well as follow unique ligand binding mechanisms.

Appendix B. Supplementary data

Supplementary figures, Figure B1-B14.

Supplementary tables, Table B1-B8.

Chapter 5

Structural and mechanistic characterization of MlaD protein from *Escherichia coli*

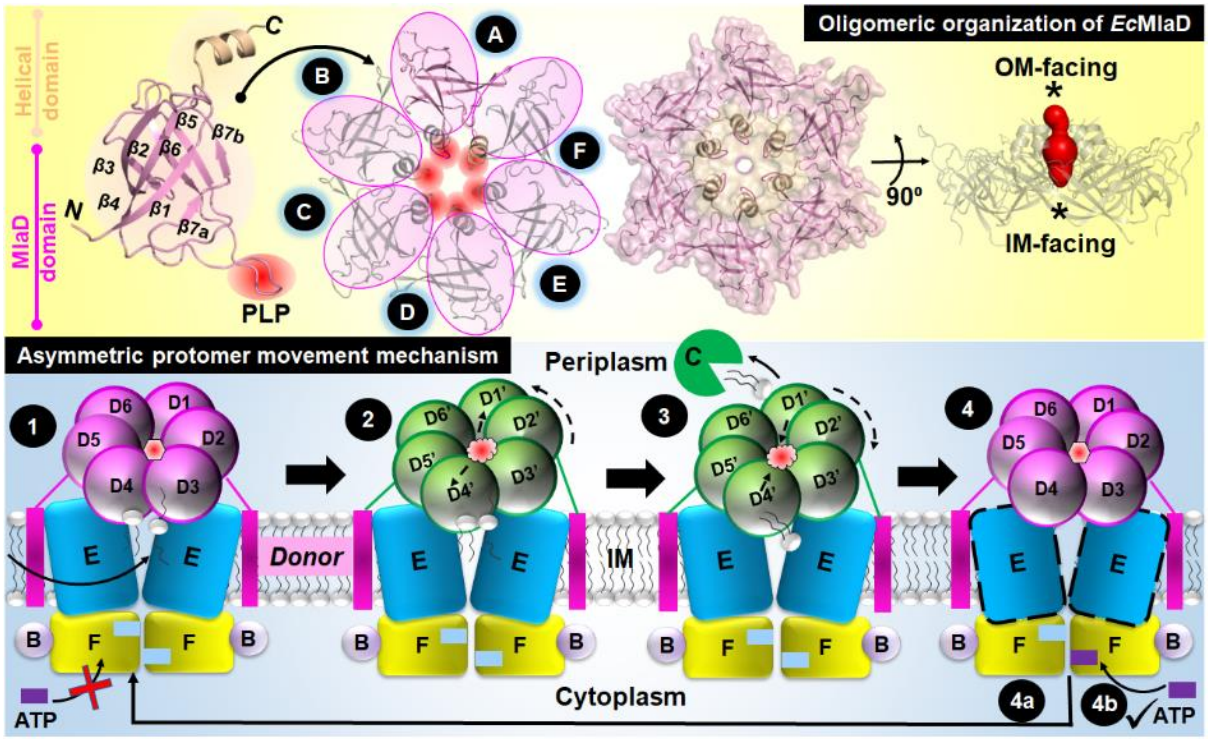
A part of this chapter is under communication for publication:

1. **Dutta A** and Kanaujia SP. The structural features of MlaD illuminate its unique ligand-transporting mechanism and ancestry. ***Under communication*.**

Abstract

The membrane-associated solute-binding protein (SBP) MlaD of the maintenance of lipid asymmetry (Mla) system has been reported to help the transport of phospholipids (PLs) between the outer and inner membranes of Gram-negative bacteria. However, the molecular mechanism underlying the transport of PLs and the ancestry of the protein MlaD remain unclear, owing primarily to its unique characteristics that are different from canonical SBPs. In this study, we report the crystal structures of MlaD from *Escherichia coli* (*Ec*MlaD) at the resolution range of 2.3-3.2 Å. The *Ec*MlaD protomer consists of two distinct regions, viz. N-terminal β-barrel fold consisting of seven strands (referred to as MlaD domain) and C-terminal α-helical domain (HD). The protein *Ec*MlaD oligomerizes to give rise to a homo-hexameric ring with a central channel that is hydrophobic and continuous with variable diameter. Interestingly, the structural analysis revealed that the HD, instead of the MlaD domain, plays a critical role in determining the oligomeric state of the protein. Based on the structural analysis, we propose a novel mechanism of PL transport by MlaD, viz. “asymmetric protomer movement (APM)”. Wherein half of the MlaD hexamer would rise in the periplasmic side along with an outward movement of pore loops, resulting in the change of the central channel geometry. Furthermore, this study highlights that, unlike typical SBPs, *Ec*MlaD possesses a fold similar to EF/AMT-type beta(6)-barrel and a unique ancestry. Altogether, the findings firmly establish *Ec*MlaD to be a non-canonical SBP with a unique ligand-transport mechanism.

Graphical abstract



5.1. Introduction

The outer membrane (OM) in Gram-negative bacteria acts as a formidable barrier against a plethora of harmful compounds such as toxins, drugs, etc. It is asymmetric in nature, possessing lipopolysaccharides (LPSs) in the outer leaflet and phospholipids (PLs) in the inner leaflet (Nikaido, 2003; Delcour, 2009). However, PLs have the tendency to accumulate in the outer leaflet of the OM, resulting in the disruption of the barrier function (Henderson et al., 2016). During extracellular stress, the restoration of the OM asymmetry is executed by two OM β-barrel enzymes, viz. PldA and PagP, which remove the excess PLs in the outer leaflet by cleaving or transferring their acyl chains (Bishop, 2008). On the other hand, in physiological conditions, a highly conserved multi-component intermembrane ABC transporter, viz. maintenance of lipid asymmetry (Mla) system, is used to safeguard the OM asymmetric distribution of PLs (Malinverni & Silhavy, 2009). The Mla system comprises three major components, viz. an OM lipoprotein-Osmoporin complex (MlaA-OmpC/F), a periplasmic protein (MlaC) and an inner membrane ABC transporter complex (MlaFEDB). The Mla system was

initially proposed to be an importer that would extract the excess PLs from the OM and ferry them to the inner membrane (IM) complex MlaFEDB for probable internalization (Malinverni & Silhavy, 2009). However, recent findings show that the Mla system adopts an exporter fold and can transport PLs in both directions (Hughes et al., 2019; Coudray et al., 2020; Low et al., 2021; Ekiert et al., 2022).

The general architecture of ABC importers and exporters comprises two copies, each of the transmembrane domain (TMD) and nucleotide-binding domain (NBD) (Rees et al., 2009; Wilkens, 2015). However, ABC importers are dependent on an additional domain, referred to as substrate (or solute)-binding protein (SBP). The SBPs render specificity and directionality to the importers by capturing the substrates and bringing them toward the TMDs for subsequent translocation (Maqbool et al., 2015). Although SBPs share low sequence similarities among themselves, they possess a highly conserved architecture consisting of two α/β globular domains in which α-helices flank the β-sheets, referred to as N- and C-terminal domains (NTD and CTD). The substrate-binding site is located at a cleft between the NTD and CTD, linked by a flexible hinge region that facilitates the movement of the two domains for substrate capture (Fukami-Kobayashi et al., 1999). However, the topological arrangement of β-sheets in the core regions of NTD and CTD, as well as the linker regions, show variations (Lee et al., 1999). Based on substrate specificities and structural folds, SBPs have been classified into eight clusters, A-H (Berntsson et al., 2010; Scheepers et al., 2016; Chandravanshi et al., 2021). Recently, based on the presence or absence of NTD-CTD arrangement, SBPs have been classified as canonical and non-canonical SBPs (Dutta & Kanaujia, 2022).

Interestingly, unlike canonical ABC importers, the Mla system possesses two SBPs, viz. MlaC (free-floating) and MlaD (IM-associated). Both MlaC and MlaD have been categorized as lipid-transfer proteins (LTPs) that are involved in non-vesicular lipid transport (Wong et al., 2019). Owing to its unique domain organization and mechanism of ligand binding, MlaC has been categorized as a non-canonical SBP (Dutta & Kanaujia, 2022). However, in spite of belonging to the same PL-transporting system and possessing a unique domain architecture, the protein MlaD has not been considered under any SBP classification scheme to date. Although the protein MlaD serves as the SBP for the Mla system, it does not share similar characteristics with that

of the canonical SBPs. Instead, the protein MlaD possesses a highly conserved mammalian cell entry (Mce) domain (also known as the MlaD domain), initially identified in the Mce proteins from *Mycobacterium tuberculosis* (Arruda et al., 1993; Ekiert et al., 2017). Previously, we reported the multiple architectures of the MlaD domain and its role in PL transport (Dutta et al., 2021). Besides MlaD, the proteins PqiB and YebT also possess the MlaD domain(s), reported to be involved in PLs transport between the IM and OM (Ekiert et al., 2017; Isom et al., 2020). In spite of a significant amount of information available in the literature, the molecular mechanism of PL transport by the protein MlaD is still not clear. In this study, we report the crystal structures of the ectodomain of the protein MlaD from *E. coli* (*Ec*MlaD). Through extensive structural analysis, we establish MlaD to be a non-canonical SBP, propose a novel mechanism of PL transport and report its probable evolutionary ancestry.

5.2. Materials and methods

5.2.1. Cloning, overexpression and protein purification

The genomic DNA was isolated from *Escherichia coli* BL21(DE3) competent cells and the gene fragment corresponding to the truncated form of MlaD (28-183, referred to as *Ec*MlaD) was amplified by polymerase chain reaction (PCR) from the genome using the forward primer 5'-GCCG**CATATG**CACCATCATCATCATCATGCGAACGTGACGTCCATACGTAC-3' containing *Nde*I restriction site (bold) followed by a 6xHis-tag (underlined) and the reverse primer 5'-CATT**CTCGAG**TTATTTCGTTGTACCCACAGGTTCAGTG-3' containing *Xho*I restriction site (bold). The amplified gene was inserted into pET22b(+) vector excised with the same set of restriction enzymes. The resultant recombinant plasmid was confirmed by double digestion followed by plasmid DNA sequencing (Figure 5.1A). For overexpression of the protein *Ec*MlaD, *E. coli* BL21(DE3) competent cells were transformed with the recombinant plasmid. The transformed cells were grown at 37 ºC in a Luria-Bertani (LB) medium supplemented with 100 µg mL^{-1} ampicillin until an optical density (OD) of 0.6 at 600 nm was attained. Subsequently, the secondary cell culture was induced for protein overexpression by the addition of 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). The secondary culture was grown for 7 h and at each hour, samples were collected so as to determine the ideal incubation time (Figure 5.1B). Based on this experiment, the secondary culture was

grown for 4 h at 37 ºC after IPTG induction. The grown cells were harvested by centrifugation at 3,836*g* for 10 min at 4 ºC. The cell pellets were resuspended in a lysis buffer containing 20 mM Tris-HCl pH 7.5, 150 mM KCl, 5 mM imidazole, 5% glycerol, 1 mM phenylmethylsulfonyl fluoride (PMSF) and 3 mM β-mercaptoethanol (β-ME). The resuspended cells were lysed by sonication (2 s on and 10 s off at 33% amplitude) followed by centrifugation at 15,644*g* for 40 min at 4 °C in order to remove insoluble debris (Figure 5.1C). The supernatant fraction was then applied to a pierce centrifuge column (Thermo Fisher Scientific, USA) packed with Ni^{2+}-NTA affinity resin (Qiagen, Germany) pre-equilibrated with a binding buffer (20 mM Tris-HCl pH 7.5, 150 mM KCl, 5 mM imidazole, 5% glycerol, 1 mM PMSF and 3 mM β-ME) and incubated for 2 h. The column was washed with five column volume of each of wash buffer A (20 mM Tris-HCl pH 7.5, 200 mM KCl, 10 mM imidazole, 5% glycerol, 1 mM PMSF and 3 mM β-ME) and wash buffer B (20 mM Tris-HCl pH 7.5, 200 mM KCl, 20 mM imidazole, 5% glycerol, 1 mM PMSF and 3 mM β-ME). The bound protein was eluted with 400 mM imidazole in the wash buffer (Figure 5.1D). The eluted fractions were pooled together and subjected to gradient dialysis against 20 mM Tris-HCl pH 7.5 and 200 mM KCl in order to remove the imidazole. The dialyzed protein was concentrated to a final concentration of 10 mg mL^{-1} using a Vivaspin turbo 15 (Sartorius, Germany) centrifugal filter unit with a molecular weight cut-off of 10 kDa (Figure 5.1E).

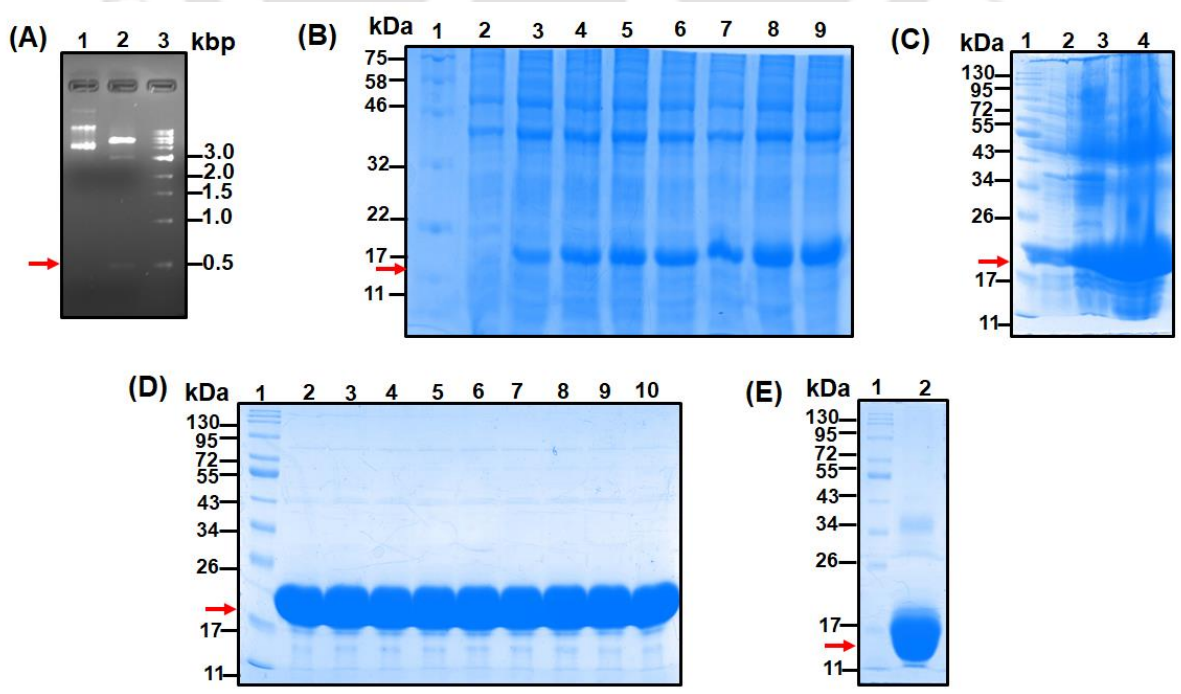


Figure 5.1. Cloning, overexpression, solubility and purification of EcMlaD. *(A) Clone confirmation by double digestion of the vector pET22b carrying the truncated gene mlaD by NdeI and XhoI restriction enzymes (lane 1: pET22b wild type, lane 2: insert pop out after double digestion of the plasmid pET22b and lane 3: DNA ladder). (B) Overexpression of the protein EcMlaD in BL21 (DE3) at 0.5 mM IPTG at a different time interval (lane 1: protein marker, lane 2: uninduced or 0 hrs, lane 3-9: 1-7 hr, respectively) after induction. (C) SDS-PAGE analysis for optimized protein solubility at 37°C for 4 hrs (lane 1: protein marker, lane 2: uninduced sample, lane 3: pellet fraction and lane 4: supernatant). (D) SDS-PAGE analysis for the eluted fractions of the protein using Ni-NTA metal-affinity chromatography (lane1: protein marker, lanes 2-10: eluted fractions of the protein). (E) SDS-PAGE analysis for the concentrated protein (lane 1: protein marker and lane 2: concentrated protein). The gene and the protein of interest are marked by a red arrow.*

5.2.2. Circular dichroism (CD) experiments

The Far-UV CD spectrum of *Ec*MlaD was recorded between 190 and 260 nm wavelengths with a 1 mm optical path length quartz cuvette using a Jasco J-1500 spectrometer (Jasco, Germany). The protein prepared in a buffer containing 20 mM Tris-HCl pH 7.5 and 200 mM KCl was diluted in water to a final concentration of 5 µM. The CD parameters response, sensitivity and scan speed were fixed at 2 s, 100 millidegrees and 100 nm min^{-1}, respectively, for spectra measurements. The final spectrum was obtained by averaging three scans and subtracting from the blank that contained the buffer as a reference. The spectrum was analyzed by using Spectra Manager software provided by JASCO. In order to study the thermal stability of *Ec*MlaD, the sample was subjected to a temperature range of 20-90 ºC with a heating rate of 5 ºC min^{-1}. The spectra were recorded from 190 to 260 nm wavelengths, and the denaturation profiles were plotted as a function of temperature and wavelength using the software Origin v9.6.

5.2.3. Crystallization, data collection and structure determination

The initial screening of purified *Ec*MlaD (12 mg mL^{-1}) was performed using the microbatch-under-oil technique by mixing 2 µL of protein and 2 µL of crystallization buffer available in the crystallization screens from Hampton Research at 4 ºC and 20

ºC. Initial crystal hits of the protein were obtained at 20 ºC within 1-2 months in a condition containing 0.2 M magnesium chloride hexahydrate and 20% (w/v) polyethylene glycol 3,350. In order to get diffraction-quality crystals, optimization was performed using the hanging-drop vapor-diffusion method at 20 ºC. However, the obtained crystals diffracted at extremely poor resolutions (Figure 5.2A-5.2B). Although other crystal hits were obtained, they failed to produce quality diffraction (Figure 5.2C-5.2H). Crystallization of protein-ligand complexes has been previously reported to produce better diffraction (Hassell et al., 2007). Hence, a crystallization setup for *Ec*MlaD with a cognate ligand L-α-phosphatidylethanolamine dioleoyl (PED) was made. For this, a hybrid system was prepared where *Ec*MlaD (200 µM) was incubated with the ligand (2000 µM; dissolved in the co-solvents dimethyl sulfoxide (DMSO) and ethanol) overnight at 4 ºC. This was followed by screening using the microbatch-under-oil technique by mixing 2 µL of ligand-incubated protein sample and 2 µL of crystallization buffer available in the crystallization screens from Hampton Research at 4 ºC and 20 ºC. Crystal hits for *Ec*MlaD were obtained in three different conditions, (1) 0.01 M nickel(II) chloride hexahydrate, 0.1 M tris pH 8.5, 20% w/v polyethylene glycol monomethyl ether 2,000, (2) 1.8 M sodium phosphate monobasic monohydrate, potassium phosphate dibasic pH 5.0 and (3) 0.2 M magnesium chloride hexahydrate, 0.1 M sodium citrate tribasic dihydrate pH 5.0, 10% (w/v) polyethylene glycol 20,000 (Figure 5.2I-5.2K, respectively) at 20 ºC within 1-2 months. X-ray intensity diffraction data were collected at -173 ºC using the home source Rigaku MicroMax-007 HF diffractometer (operated at 40 kV and 30 mA) and R-Axis IV++ image-plate detector with rotating anode X-ray tube at wavelength 1.5418 Å available at Central Instruments Facility (CIF) of Indian Institute of Technology Guwahati, India. Data were collected by maintaining the crystal-to-detector distance at 180 to 200 mm with an oscillation of 1° and an exposure time of 300 s. Before mounting, crystals were flash cooled under the stream of liquid nitrogen. The collected data set was processed and scaled using the programs iMosflm and Aimless embedded in the CCP4 package (Battye et al., 2011; Winn et al., 2011; Evans et al., 2013). The three-dimensional structure of *Ec*MlaD was determined by the molecular replacement method using the program Phaser and the previously-determined crystal structure of MlaD from *E. coli* (PDB id: 5UWA) as the search model (McCoy et al., 2007). In order to calculate the R_{free}, a total of ~5% of the total reflections was kept aside as a test dataset (Brünger, 1992). The atomic model building and the iterative cycles of structural parameters refinement were carried out

by using the programs Coot and Refmac5, respectively (Vagin et al., 2004; Emsley et al., 2010). The first cycle of structure refinement showed no clear electron density for any bound ligand. Thus, the protein atoms and water molecules were first fitted in the electron density map. Finally, even after fitting all the protein and water molecules, no electron density for the ligand could be observed at the binding site. The quality of the refined model was validated by using the programs Procheck and MolProbity (Laskowski et al., 1993; Chen et al., 2010). The refinement and validation statistics of the models are provided in Table C1. The three-dimensional atomic coordinates and the structure factors have been deposited in RCSB Protein Data Bank with the accession codes 8HQA, 8HPZ and 8HQ9 (Berman et al., 2000).

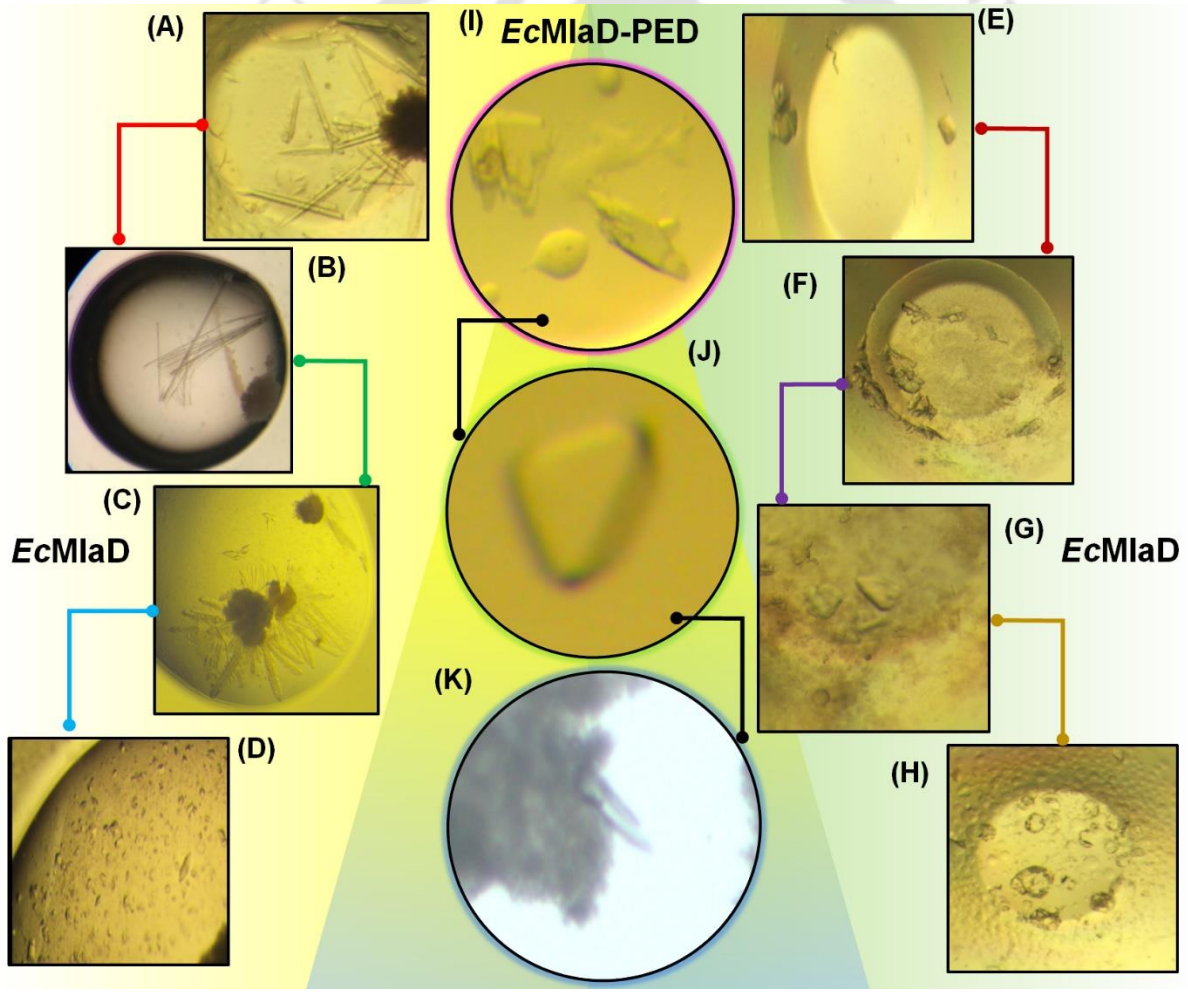


Figure 5.2. Crystallization of EcMlaD. *(A-H) Protein crystals obtained when crystal setup was done with EcMlaD. (I-K) Protein crystals obtained when crystal setup was done with EcMlaD and PED.*

5.2.4. Sequence and structure analyses

The protein sequences and the atomic coordinates of structures analyzed in the study were retrieved from the UniProtKB database and Protein Data Bank, respectively (The UniProt Consortium, 2023). In order to determine the sequence similarities among the MlaD orthologs, a pairwise sequence alignment was performed using the web tool BLAST with the default set of parameters (Altschul et al., 1990). In order to identify conserved residues, multiple sequence alignment (MSA) of protein sequences was performed using the program Clustal Omega and the results were rendered using the web tool ESPript 3.0 (Gouet et al., 2003; Sievers & Higgins, 2014). To understand the evolutionary relationships of proteins, a sequence-base phylogenetic tree analysis was performed by aligning the protein sequences using the program ClustalW, followed by the construction of a phylogenetic tree using the program MEGA7 employing the maximum likelihood method with the default set of parameters (Kumar et al., 2016). Subsequently, the phylogenetic tree was validated by performing 1000 bootstrap iterations. Sequence- and structure-based positioning of proteins in the membrane were determined by using the online server TMHMM and PPM v2.0, respectively, with the default sets of parameters. Identification of lipoproteins was performed using the web server DOLOP (Krogh et al., 2001; Madan Babu & Sankaran, 2002; Lomize et al., 2012).

The prediction of oligomeric states of *Ec*MlaD, their surface areas and analysis of oligomeric interfaces were performed using the web server PDBePISA with the default set of parameters (Krissinel & Henrick, 2007). The predicted three-dimensional structures of the proteins YrbE1A-4B and Mce1A-4F from *M. tuberculosis* and TGD1-2 from *A. thaliana* were obtained from the AlphaFold Protein Structure Database (Jumper et al., 2021; Varadi et al., 2022). The secondary structural contents of the analyzed structures were determined using the web servers DSSP and PDBsum (Kabsch & Sander, 1983; Joosten et al., 2010; Laskowski et al., 2018). The protein topology cartoons were generated by using the program TopDraw embedded in the CCP4 package (Bond, 2003). Structural comparisons were performed by using the programs PyMOL (The PyMOL Molecular Graphics System, Schrodinger, LLC) and UCSF Chimera (Pettersen et al., 2004). The dimensions of the central channels were estimated by using the plugin CAVER 3.0, available in PyMOL with the default set of parameters (Chovancova et al., 2012). The hydrophobicity mapping of the proteins

was performed by using the YRB color scheme in the program PyMOL (Hagemans et al., 2015). As per this scheme, the carbon atoms not bound to nitrogen and oxygen atoms are highlighted in yellow, nitrogen atoms in the side chains of lysine and arginine are in blue, oxygen atoms in the side chains of glutamate and aspartate are in red and all remaining atoms are in white. The visualization of crystal and modeled structures and the molecular graphic figures were generated using the program PyMOL and UCSF Chimera. The structure-based homology search for *Ec*MlaD was performed using the web server Dali (Holm, 2022). Domain organization in proteins was studied using the databases Pfam, InterPro and CATH (Dawson et al., 2017; Blum et al., 2021; Mistry et al., 2021). The ancestral analysis of SBPs was performed using the Structural Classification of Proteins (SCOP) database (Andreeva et al., 2020).

5.3. Results

5.3.1. Overall structure of *Ec*MlaD

The CD spectrum of *Ec*MlaD shows a single negative peak at ~208 nm, suggesting that the protein contains β-strands predominantly (Figure C1A). In order to probe the thermal stability of *Ec*MlaD, the thermal ramping (20-90 ºC) experiment was performed. The result revealed that the CD signal shifts from 208 to 200 nm at ~60 ºC (Figure C1B). This continues with the further increase in the temperature. This is suggestive that *Ec*MlaD might be present in different oligomeric forms and follow a two-state folding process. In order to further confirm this, a native-PAGE analysis of *Ec*MlaD was performed; bovine serum albumin (BSA) was considered as the reference. As per the previous report (Shukolyukov, 2009), BSA exists in three different forms, evident from three distinct bands corresponding to monomer (66 kDa), dimer (132 kDa) and trimer (198 kDa) (Figure C1C). Similarly, in the case of *Ec*MlaD, three distinct bands, ~140 kDa, ~200 kDa and ≥400 kDa, were observed (Figure C1C). This suggests that, in solution, *Ec*MlaD exists in three oligomeric forms.

The crystal structure of *Ec*MlaD was determined in three different space groups, viz. $P4_12_12$ (PDB id: 8HQA, a=b=64.28 Å, c=231.06 Å), I222 (PDB id: 8HPZ, a=60.28 Å, b=107.29 Å, c=117.92 Å) and $P2_122_1$ (PDB id: 8HQ9, a=60.42 Å, b=106.95 Å, c=117.12 Å) at resolutions 3.20 Å, 2.30 Å and 2.70 Å, respectively (Figure 5.3A, Figure

C2A-C2C and Table C1). Interestingly, in the crystal structure belonging to the space group $P4_12_12$, electron densities can be observed for the residues 35-152, whereas, in the crystal structures belonging to the space group $I222$ and $P2_122_1$, electron densities are available only for the residues 37-141 and 37-144, respectively. Hence, in all three structures, the electron density of the complete construct (28-183), especially the C-terminal region, was missing. This indicates that the C-terminal region of *Ec*MlaD is extremely flexible. This can be seen in the model predicted by the program AlphaFold as well, wherein the C-terminal region (156-183) shows a low per-residue confidence score (pLDDT). Thus, among the three structures of *Ec*MlaD, the one belonging to the space group $P4_12_12$ (*Ec*MlaD_$P4_12_12$) covers the MlaD domain (35-140, *Ec*MlaD_MlaD1^{P41212}) as well as the C-terminal region (141-152). On the other hand, structures belonging to space groups $I222$ and $P2_122_1$ cover only the MlaD domains (*Ec*MlaD_MlaD1^{I222} and *Ec*MlaD_MlaD1^{P21221}, respectively).

In *Ec*MlaD_$P4_12_12$, a single MlaD protomer consists of seven β-strands adopting a β-barrel fold which constitutes the highly conserved MlaD domain (*Ec*MlaD_MlaD1^{P41212}). Thus, as observed in the CD experiments, the protomer predominantly consists of β-strands. The β-strands are interconnected through extremely variable loop regions and their weak densities in certain regions indicate that they are flexible. The β3 strand is the longest, whereas the β7 strand is discontinuous and can be divided into two segments (β7a and β7b). The β7b strand gives rise to a C-terminal α-helical domain (HD). Thus, the periplasmic component of *Ec*MlaD consists of two main parts, viz. *Ec*MlaD_MlaD1^{P41212} and HD. The topology analysis indicates that in *Ec*MlaD_MlaD1^{P41212}, the β-strands are present in the order 143265 (Figure 5.3B). This corroborates the previously reported structure of *Ec*MlaD crystallized in the space group C2 (*Ec*MlaD_C2; a=122.36 Å, b=63.82 Å, c=91.15 Å, β=130.0º; Figure C2C; PDB id: 5UW2) (Ekiert et al., 2017). In *Ec*MlaD_ MlaD1^{I222} and *Ec*MlaD_ MlaD1^{P21221}, the overall core of the *Ec*MlaD protomer is similar to *Ec*MlaD_$P4_12_12$. However, as per the analysis, an additional α-helix (H1) is present between β6 and β7a strands in the former two structures (Figure C2A-C2B). A structure-based secondary structure estimation further reveals that *Ec*MlaD_MlaD1^{P41212} does not contain any helix, whereas *Ec*MlaD_ MlaD1^{I222} and *Ec*MlaD_ MlaD1^{P21221} are composed of ~2.9% helix (Table C2).

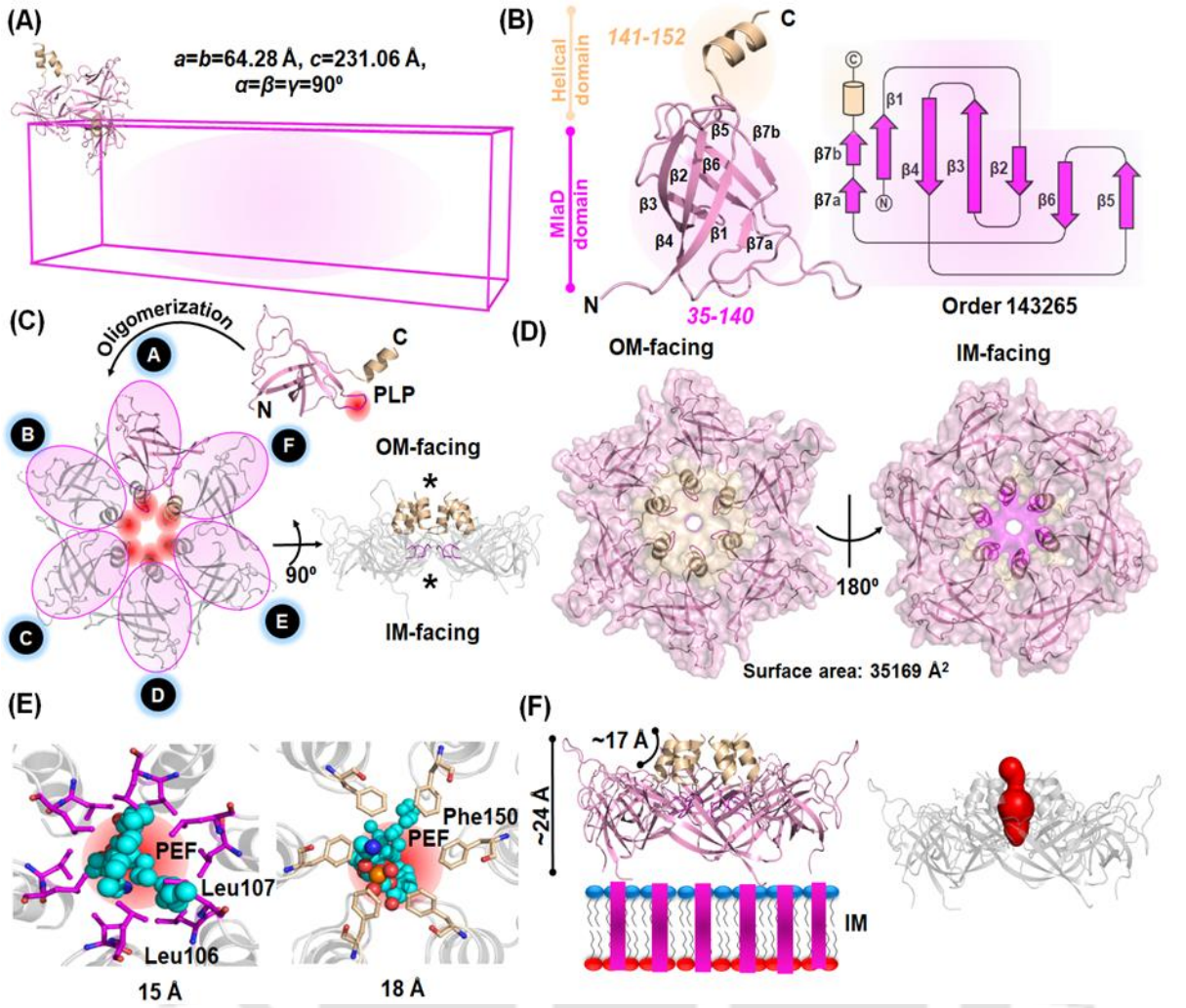


Figure 5.3. The overall structure of EcMlaD. *(A) Arrangement of three EcMlaD protomers in $P4_12_12$ space group. (B) The overall structure of EcMlaD protomer. (Left) Each protomer can be divided into two distinct parts, the N-terminal β-barrel region with seven β-strands (MlaD domain, residues 35-140, pink) and the C-terminal α-helical region (HD, residues 141-152, wheat). (Right) Topology diagram of EcMlaD_$P4_12_12$ protomer. The seven β-strands (pink arrows) are arranged in the order 143265 that give rise to the C-terminal HD (wheat cylinder). (C) Oligomerization of EcMlaD_$P4_12_12$ protomer. (Top) Orientation of EcMlaD_$P4_12_12$ protomer with respect to the PLP (highlighted in magenta). (Bottom left) Homo-hexameric arrangement of EcMlaD_$P4_12_12$. Six protomers of EcMlaD_$P4_12_12$ arrange in a circular fashion to give rise to a ring-like structure with a central channel (highlighted in red). (Bottom right) Lateral view of the EcMlaD_$P4_12_12$ hexamer. Two major constrictions are present at the HD (wheat) and the PLP (magenta) regions which are denoted by asterisks. (D) Overview of the OM- and IM-facing sides of EcMlaD_$P4_12_12$ hexamer with the*

constrictions at HD (wheat) and the PLP (magenta) regions. (E) Comparison of the two constrictions in the central channel. (Left) Constriction located at the PLP region formed by the residues Leu106 and Leu107, having a diameter of ~15 Å. (Right) Constriction located at the HD formed by the residue Phe150, having a diameter of ~18 Å. Both the constrictions are capable of accommodating the phosphatidylethanolamine (PEF, sphere model in cyan) molecule that was manually placed. (F) The proposed arrangement of EcMlaD hexamer in the bacterial cell. (Left) Lateral view of the EcMlaD_P4$_1$2$_1$2 hexamer resting on the IM. (Right) Dimensional mapping of the central channel of EcMlaD_P4$_1$2$_1$2 hexamer.

Structural analyses suggest that *Ec*MlaD_P4$_1$2$_1$2 favors a homo-hexamer with a six-fold symmetry, similar to the previously reported structure *Ec*MlaD_C2 (Figure 5.3C and Figure C2C), in contrast to the native-PAGE analysis. In the hexamer, each protomer of *Ec*MlaD_P4$_1$2$_1$2 would juxtapose in a circular fashion with the loop between β5 and β6 strands (L_{5-6}) lining the central channel of the homo-hexameric ring. This loop is referred to as Pore LooP (PLP) as it constitutes the inner wall of the channel (Figure 5.3C). On the other hand, the HD from each protomer protrudes from one side of the assembly and collectively forms a bundle-like arrangement giving rise to a channel that runs through the center of the ring (Figure 5.3C). Owing to the arrangement of the PLP and HD, the central channel of *Ec*MlaD_P4$_1$2$_1$2 would possess two major constrictions, one situated towards the IM-facing side and the other towards the OM-facing side (Figure 5.3C-5.3D). The overall surface area of the homo-hexamer of *Ec*MlaD_P4$_1$2$_1$2 is estimated to be 35169 Å^2 (Figure 1D). Thus, both constrictions are situated at the extreme ends of the channel. The first constriction is constituted by the residues Leu106 and Leu107 situated in the PLP with a diameter of 15 Å (Figure 5.3E). On the other hand, the second constriction is formed by the residue Phe150 situated in the HD with a diameter of 18 Å (Figure 5.3E). In order to find out whether the central channel with the two constrictions could be able to fit PLs, a molecule phosphatidylethanolamine (PEF) was manually placed in the channel. The modeling analysis clearly showed that the channel could easily accommodate the PL without any hindrance, thus, indicating its functional significance (Figure 5.3E). As MlaD remains IM-anchored, it can be proposed that *Ec*MlaD_P4$_1$2$_1$2 hexamer would evenly rest on the IM. In such a position, *Ec*MlaD_P4$_1$2$_1$2 traverses a height of ~24 Å,

whereas the HD covers a height of ~17 Å. The analysis further reveals that the central channel would give rise to a continuous translocation pathway with varying dimensions, thus, indicating an uninterrupted and dynamic passage of PL molecules (Figure 5.3F).

Interestingly, the interfacial analysis of *Ec*MlaD_ MlaD1^{I222} and *Ec*MlaD_ MlaD1^{P21221} reveals significant variations from that of *Ec*MlaD_P4$_1$2$_1$2. *Ec*MlaD_ MlaD1^{I222} structure favors a dodecameric assembly with an overall surface area of ~56158 Å^2. In contrast, *Ec*MlaD_ MlaD1^{P21221} prefers both hexameric as well as dodecameric assemblies with an overall surface area of ~32829 Å^2 and ~54360 Å^2, respectively (Figure C2A-C2B). An in-depth analysis of the dodecameric assembly reveals that the two layers of the MlaD hexamer remain stacked on top of each other in a manner that the C-terminal of each hexamer faces each other (Figure C3A-S3B). On the other hand, *Ec*MlaD_C2 prefers a similar hexameric assembly as *Ec*MlaD_P4$_1$2$_1$2 with a surface area of ~33792 Å^2 (Figure C2C). In *Ec*MlaD_ MlaD1^{I222}, the diameters of the two constrictions of the central channel are ~15 Å, while those in *Ec*MlaD_ MlaD1^{P21221} are ~14 Å (Figure C3C). Whereas those of the previously reported structure *Ec*MlaD_C2 are ~15 Å and ~18 Å (Figure C3C). Thus, the analysis reveals that irrespective of oligomeric states, the diameter of the central channel remains more or less comparable.

5.3.2. MlaD domains have structurally conserved cores

In addition to *Ec*MlaD, proteins PqiB (*Ec*PqiB) and YebT (*Ec*YebT) from *E. coli*, Mce from *Mycobacterium tuberculosis* (*Mt*Mce) and Trigalactosyldiacylglycerol 2 (*At*TGD2) from *Arabidopsis thaliana* possess the MlaD domain(s) (Ekiert et al., 2017; Dutta et al., 2021). To obtain structural insights, the MlaD domain(s) from all these proteins were compared. The proteins *Ec*PqiB (PDB id: 5UVN), *Ec*YebT (PDB id: 6V0C) and *Ec*MlaD_P4$_1$2$_1$2 (PDB id: 8HQA) contain 546, 877 and 183 amino acid residues and possess three (*Ec*PqiB_MlaD1-3), seven (*Ec*YebT_MlaD1-7) and one (*Ec*MlaD_MlaD1^{P41212}) MlaD domain(s), respectively (Figure 5.4A-5.4B). In *Ec*PqiB, three stacked MlaD domains, including the HD, span around ~162 Å in the periplasm, while in *Ec*YebT, it is ~210 Å with seven stacked MlaD domains without the HD. Although, like *Ec*MlaD, the proteins *Ec*PqiB and *Ec*YebT form hexamers, their domain architectures differ due to the presence of a different number of the MlaD domain(s)

(Dutta et al., 2021). However, there exists ambiguity regarding the MlaD domain boundaries. This is because the domain boundaries provided in the databases are not consistent with the available literature (Lu & Benning, 2009; Ekiert et al., 2017; Isom et al., 2020). Thus, to maintain uniformity, we have considered the signature β-barrel fold of the MlaD domain as the reference for defining the domain boundaries. The analysis suggests that, unlike *Ec*MlaD_MlaD1^{P41212}, *Ec*PqiB_MlaD1-3 and *Ec*YebT_MlaD1-7 possess seven continuous β-strands. We have previously reported a high sequence variability among the MlaD domains of *E. coli* proteins (Dutta et al., 2021). A structural comparison of *Ec*MlaD_MlaD1^{P41212} with *Ec*PqiB_MlaD1-3 and *Ec*YebT_MlaD1-7 reveals their structural conservation, except for the PLP region (Figure 5.4A-5.4B). The MlaD domains from other Gram-negative bacteria, such as *Acinetobacter baumannii* (*Ab*MlaD_MlaD1, PDB id: 6Z5U) and *Pseudomonas aeruginosa* (*Pa*MlaD_MlaD1, PDB id: 7CHA) have similar cores as *Ec*MlaD_MlaD1^{P41212}. However, *Ab*MlaD_MlaD1 possesses an insertion between the β4 and β5 strands (Figure C4A).

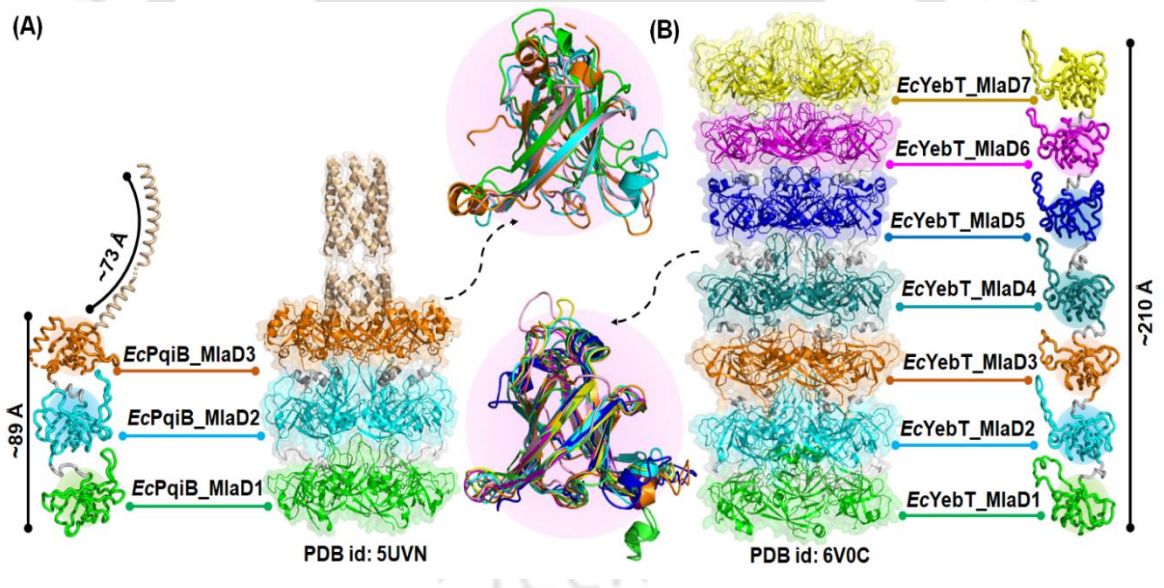


Figure 5.4. Structural comparison of the MlaD domains from E. coli. *(A) (Left, middle, side upper) Arrangement of the MlaD domains (EcPqiB_MlaD1-3) in EcPqiB protomer, hexamer and their superimposition with EcMlaD_MlaD1^{P41212} (pink), respectively. The domains EcPqiB_MlaD1-3 are in green, cyan and orange, respectively. (B) (Right, middle, side lower) Arrangement of the MlaD domains (EcYebT_MlaD1-7) in EcYebT protomer, hexamer and their superimposition with*

EcMlaD_MlaD1^{P41212} (pink), respectively. The domains EcYebT_MlaD1-7 are shown in green, cyan, orange, teal, blue, magenta and yellow, respectively.

In *M. tuberculosis*, proteins *Mt*Mce1A-4F possess a single copy of the MlaD domain (*Mt*Mce1A-4F_MlaD1). In addition to the MlaD domain, few Mce proteins contain a domain known as 'cholesterol uptake porter CUP1 of Mce4 (Mce4_CUP1)'. Based on the presence or absence of the Mce4_CUP1 domain, the Mce proteins were previously reported to belong to two different architectures, A (those having only the MlaD domain) and BI (those having both the MlaD and Mce4_CUP1 domains) (Dutta et al., 2021). The proteins *Mt*Mce1E, 3F, 4C, 4E and 4F belong to the archecture type A, whereas, *Mt*Mce1A-1D, 1F, 2A-2F, 3A-3E, 4A-4B and 4D belong to the architecture type BI (Figure C4B). For the purpose of analysis, the AlphaFold models of *Mt*Mce1A-4F were obtained and the MlaD domains from each of them were trimmed. Although the domains *Mt*Mce1A-4F_MlaD1 also depict low sequence similarities, they conserve their structures (Figure C4C-C4F and Table C3). Furthermore, *Mt*Mce1A-4F_MlaD1 possess seven β-strands similar to *Ec*MlaD_MlaD1^{P41212}; *Mt*Mce_1B-1F, 2B-2F, 3B, 3D-3F, 4A _MlaD1 contain a discontinous β7 strand (Table C2). The protein *At*TGD2 (381 residues) possesses a single copy of the MlaD domain (*At*TGD2_MlaD1). Like other proteins, *At*TGD2_MlaD1 also shares a low sequence similarity. However, it conserves its structure (Figure C4G and Table C2). In summary, the analysis suggests that the overall structural features of the MlaD domain(s) are conserved in prokaryotes and eukaryotes.

5.3.3. MlaD domains have local structural variations

Although the core structure of the MlaD domains is conserved, the loops interconnecting β-strands show high flexibility. For an in-depth insight, a structure-based domain profiling of the MlaD domains was performed. Akin to *Ec*MlaD_MlaD1^{P41212}, PLPs of *Ec*PqiB_MlaD3, *Mt*Mce1A-4F_MlaD1 and *At*TGD2_MlaD1 are devoid of any secondary structural elements (Figure 5.3D, Figure C5A-C5D and Table C2). On the other hand, the PLP of *Ec*PqiB_MlaD1 consists of a short helix (H_{L1}) while that of *Ec*PqiB_MlaD2 comprises two short β-strands (β_{L1} and β_{L2}) followed by a short helix (H_{L1}) (Figure 5.5A and Table C2). Similarly, PLPs of *Ec*YebT_MlaD1,3,5,6,7 are similar to *Ec*PqiB_MlaD1 while that of *Ec*YebT_MlaD2,4

are similar to the *Ec*PqiB_MlaD2 (Figure 5.5B and Table C2). In addition to PLP, the loops between β1-β2 (L_{1-2}), β2-β3 (L_{2-3}), β3-β4 (L_{3-4}), β4-β5 (L_{4-5}), β6-β7 ($L_{6-7/7a}$) and β7a-β7b (L_{7a-7b}) strands were investigated. A short helix (H1) is present in the loop L_{3-4} of *Ec*PqiB_MlaD3 and *Mt*Mce1B_MlaD1 and in the loop L_{4-5} of *Ec*PqiB_MlaD1-2, *Ec*YebT_MlaD1-5,7_MlaD1, *Mt*Mce(1-4)A_MlaD1. Notably, the loop L_{4-5} of *Ec*PqiB_MlaD3 contains two short helices (H2 and H3). Similarly, the discontinuous β7 (β7a and β7b) strand in *Mt*Mce1B-1E, 2B-2E, 3B, 3D-3F_MlaD1 contains a short helix (H1/H2) (Table C2). Further, the loop $L_{6-7/7a}$ in *Ec*MlaD_ $MlaD1^{I222}$, *Ec*MlaD_ $MlaD1^{P21221}$, *Ec*YebT_MlaD5, *Mt*Mce3C_MlaD1 and *At*TGD2_MlaD1 possess a short helix (H1/H2). Interestingly, β3 strand in *Pa*MlaD_MlaD1 is segmented (β3a and β3a) (Figure C6A). Notably, the MlaD domain of *Ab*MlaD_MlaD1 contains an insertion comprising four α-helices (H_{I1-I4}) and a β-strand (β_{I1}) in the loop L_{4-5} (Figure C6B). Previously, we reported the absence of cysteine residues in the MlaD domain (Dutta et al., 2021). However, this analysis revealed the presence of two cysteine residues in the loop L_{6-7} of *At*TGD_MlaD1 that can form an intra-chain disulfide bond (Figure C6C). The *Ec*MlaD orthologs in plants and cyanobacteria have been reported to play a critical role in the transport of phosphatidic acid (Lu & Benning, 2009). To affirm the confinement of cysteine residues in higher organisms, an evolutionary analysis of the MlaD sequences from prokaryotic and eukaryotic organisms was performed (Table C4). The analysis confirms the presence of cysteine residues in the higher orthologs in the loop L_{6-7}. Further, the MlaD domains can be grouped into three distinct clusters, I, II and III, belonging to Gram-negative bacteria, Actinomycetes and plants & cyanobacteria, respectively (Figure C6D).

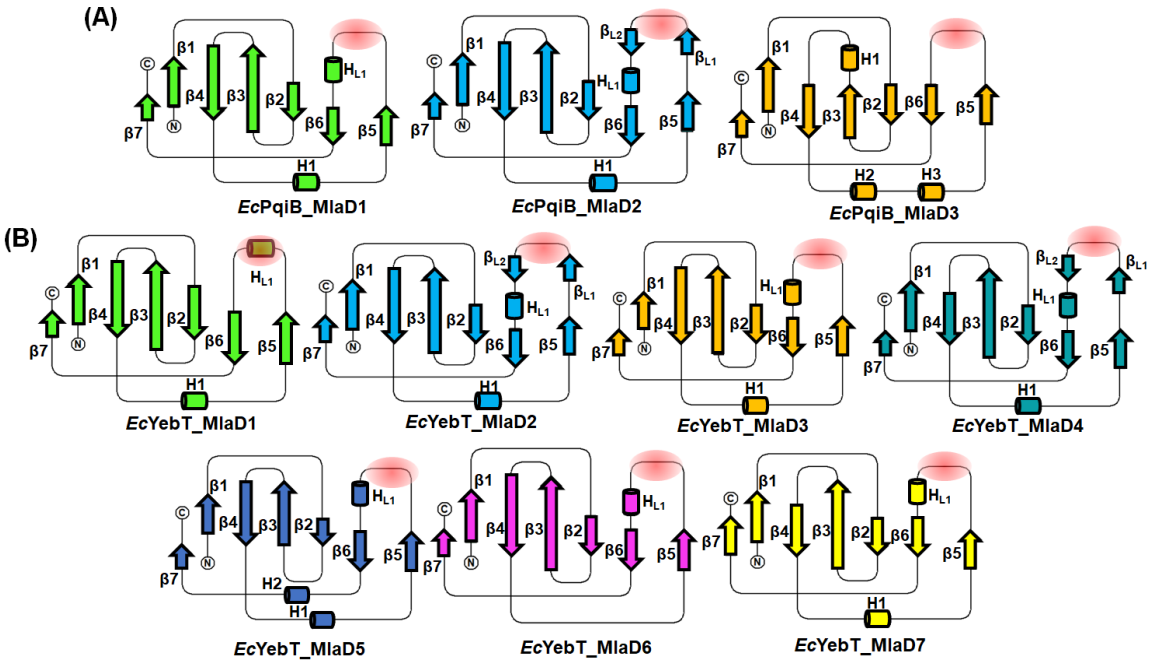


Figure 5.5. Comparison of topologies of the MlaD domains from E. coli. *(A) Topology diagrams of EcPqiB_MlaD1-3 have been shown in green, cyan and orange, respectively. (B) Topology diagrams of EcYebT_MlaD1-7 have been shown in green, cyan, orange, teal, blue, magenta and yellow, respectively. The α-helices and β-strands have been represented as cylinders and arrows, respectively. The PLP region is highlighted in red.*

5.3.4. The MlaD domains give rise to hydrophobic central channels of varying diameters

The MlaD domain-containing proteins *Ec*MlaD, *Ec*PqiB and *Ec*YebT oligomerize to form homo-hexameric rings, giving rise to central channels (Ekiert et al., 2017; Isom et al., 2020). In *Ec*MlaD_P4$_1$2$_1$2, the homo-hexamer of *Ec*MlaD_MlaD1^{P41212} gives rise to a ring having a diameter of ~92 Å (Figure 5.6A). Whereas, in *Ec*PqiB, three layers of homo-hexamers of *Ec*PqiB_MlaD1-3 have diameters of ~91, ~88 and ~92 Å, respectively (Figure 5.6B). Similarly, in *Ec*YebT, seven layers of homo-hexamers of *Ec*YebT_MlaD1-7 have diameters of ~91, ~90, ~96, ~92, ~85, ~93 and ~93 Å, respectively (Figure 5.6C). Further, channels formed by these homo-hexamers are extremely hydrophobic in nature (Figure C7A-C7C). To investigate whether the diameters of central channels of the hexameric layers formed by the MlaD domain remain the same, the distances between the central residue of the PLPs of two

opposite protomers in the MlaD domain were measured. The analysis reveals that the hexamers *Ec*MlaD_MlaD1^{P41212} and *Ec*PqiB_MlaD1-3 give rise to a central channel with a diameter of ~15, ~40, ~38 and ~20 Å, respectively. Similarly, central channels formed by the hexamers of *Ec*YebT_MlaD1-7 are of diameters ~21, ~22, ~24, ~19, ~22, ~10 and ~18 Å, respectively (Figure 5.6A-5.6C and Table C5). Furthermore, the diameters of central channels of the cryo-EM structures of *E. coli* and other Gram-negative bacteria were found to be similar (Figure C8A-C8C and Table C6). Once stacked over each other, these hexamers give rise to continuous channels with varying dimensions (Figure C8A-C8C).

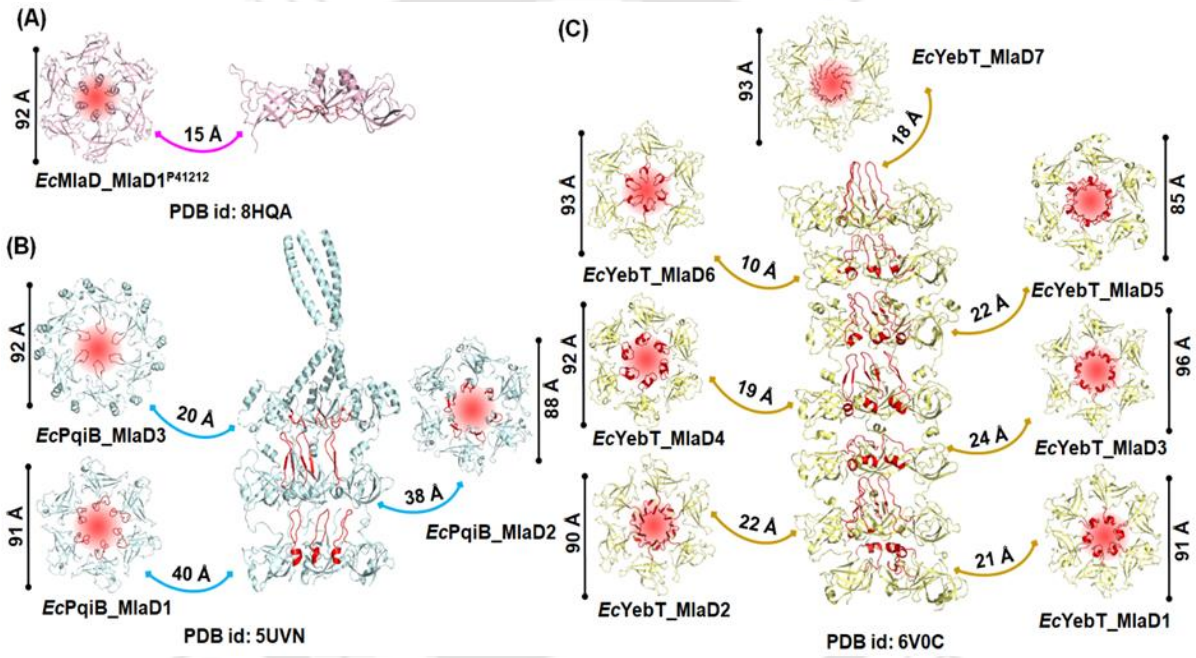


Figure 5.6. Comparison of the diameters of the MlaD-domain hexamers and their central channels from E. coli. *(A-C) The diameters of MlaD domains of EcMlaD (EcMlaD_MlaD1^{P41212}), EcPqiB (EcPqiB_MlaD1-3) and EcYebT (EcYebT_MlaD1-7), respectively. The values along the vertical lines denote the hexameric diameters, while the values along the curved lines on the sides denote the diameters of the central channels. The PLPs and the central channels are in red.*

5.3.5. *Ec*MlaD_MlaD1^{P41212} does not form domain-swapped dimers via the β5 strand

The crystal structure of the MlaD domain of *Mt*Mce4A (hereafter referred to as *Mt*Mce4A^{domain}) reveals that it forms a domain-swapped dimer through the β5 strand (PDB ids: 7AI2 and 7AI3). In both the crystal structures of *Mt*Mce4A^{domain}, four protomers (A-D and A'-D', respectively) form a tetramer. However, in solution, the *Mt*Mce4A^{domain} remains in a monomeric state (Asthana et al., 2021). To investigate the structural peculiarities, the protomers of the crystal structures were compared with *Ec*MlaD_MlaD1^{P41212}. It is observed that the typical β-barrel fold of the MlaD domains is lost in the domain-swapped dimer.

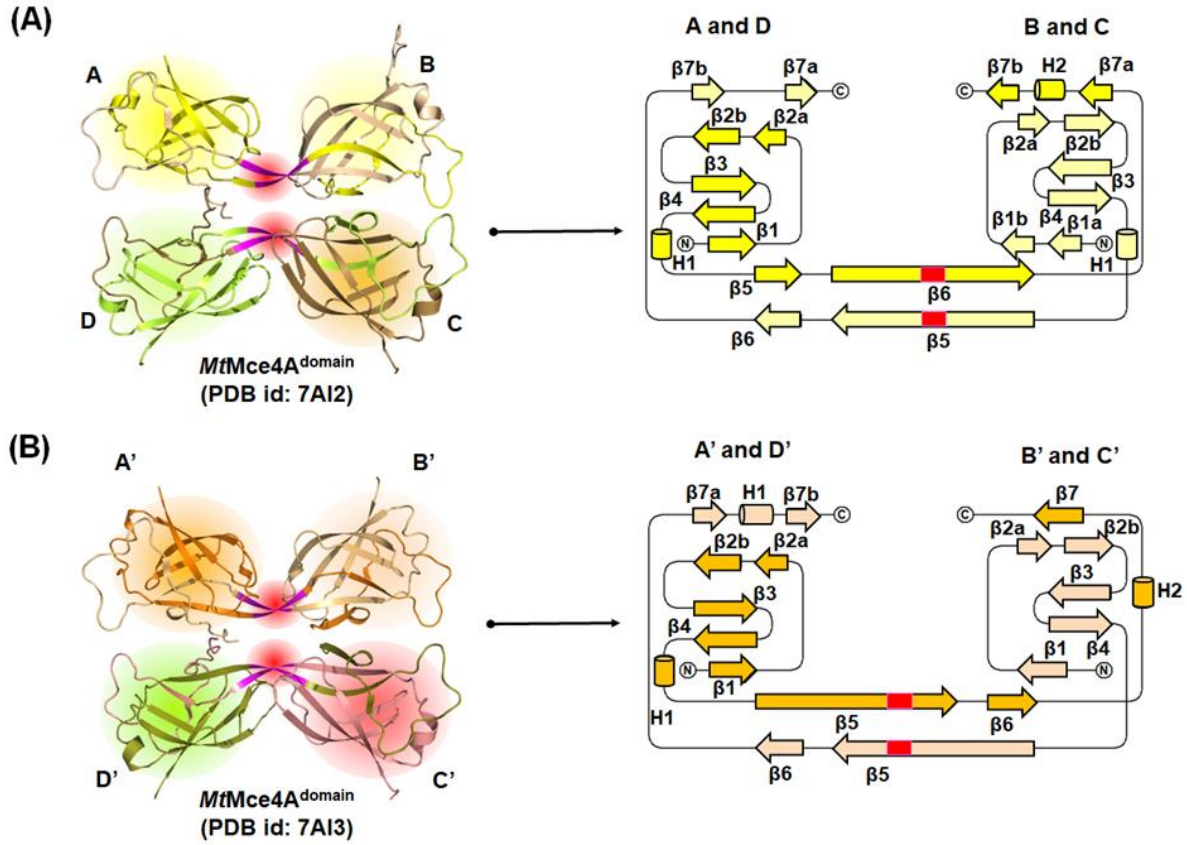


Figure 5.7. Overview of the structural features of MtMce4A. *(A-B) (Left) Crystal structures of MtMce4A^{domain} (PDB ids: 7AI2 and 7AI3) and their respective topology diagrams (right), respectively. The PLP regions are in magenta and highlighted in red in the crystal structures, whereas they are represented as red rectangles in the topology diagrams. The α-helices and β-strands are represented as cylinders and arrows, respectively.*

The domain profiling of the crystal structures of *Mt*Mce4A^{domain} (PDB id: 7AI2) revealed that the β2 and β7 strands fragmented into β2a, β2b, β7a and β7b. Besides, in two protomers (A and D), the β1 strand is also segmented into β1a and β1b. Furthermore, these two protomers possess a short helix H1 and H2 in the loops $L_{4\text{-}5}$ and $L_{7a\text{-}7b}$, respectively, whereas the remaining two protomers (B and C) lack the H2 helix (Figure 5.7A and Table C2). Similarly, in the other crystal structure (PBD id: 7AI3), all the protomers contain fragmented β2 strands, whereas two protomers (B' and C') have segmented β7 strands also. Furthermore, two protomers (A' and D') possess a short helix H1 and H2 in the loops $L_{4\text{-}5}$ and $L_{6\text{-}7}$, respectively, whereas the other two protomers (B' and C') possess a short helix H1 in the loop $L_{7a\text{-}7b}$. In both structures, the PLP region is part of the β5 strand except for one protomer (PDB id: 7AI2) (Figure 5.7B and Table C2). Overall, the results suggest that the β2 strand is segmented in all the structures of *Mt*Mce4A^{domain}. Also, β-strand content is reduced (up to 27-30%) as compared to *Ec*MlaD_MlaD1^{P41212} (Table C2). Considering the presence of a monomeric state in solution, it is likely that the individual molecules of *Mt*Mce4A^{domain} lose their β-barrel arrangement by extending the β5-β7 strands. Thus, a total of 19 different domain profiles of the MlaD domain could be identified (Figure C9). This clearly suggests that even though the core of the MlaD domain is conserved, local structural changes in the loop regions are prominent.

5.3.6. The *Ec*MlaD_MlaD1^{P41212} does not determine the oligomeric state of the protein

To investigate the role of the MlaD domain in oligomerization of the protein, the interfacial residues of the *Ec*MlaD_P4_12_12 hexamer were analyzed (Figure 5.8A and Table C7). For the convenience of the analysis, each dimeric interface has been divided into three regions, viz. head, body and base (Figure 5.8B-5.8D). The analysis reveals that a total of fifty-two residues are present at the *Ec*MlaD_P4_12_12 dimeric interface, of which ~73% are from the *Ec*MlaD_MlaD1^{P41212}. On the other hand, twelve residues from the HD participate in dimeric interface formation (Table C7). Among the interfacial residues, fourteen residues establish hydrogen bonds. Further, the residues Ile49, Gly50, Leu52, Val63, Val65, Leu73, Ile93, Ser104, Gly105 and Val116 are conserved with varying degrees (Figure C10). These results suggest that both the MlaD domain and HD participate in interface formation. However, in the absence of

HD, *Ec*MlaD adopts a dodecameric arrangement. This indicates that HD plays a critical role in determining the oligomeric state of the protein.

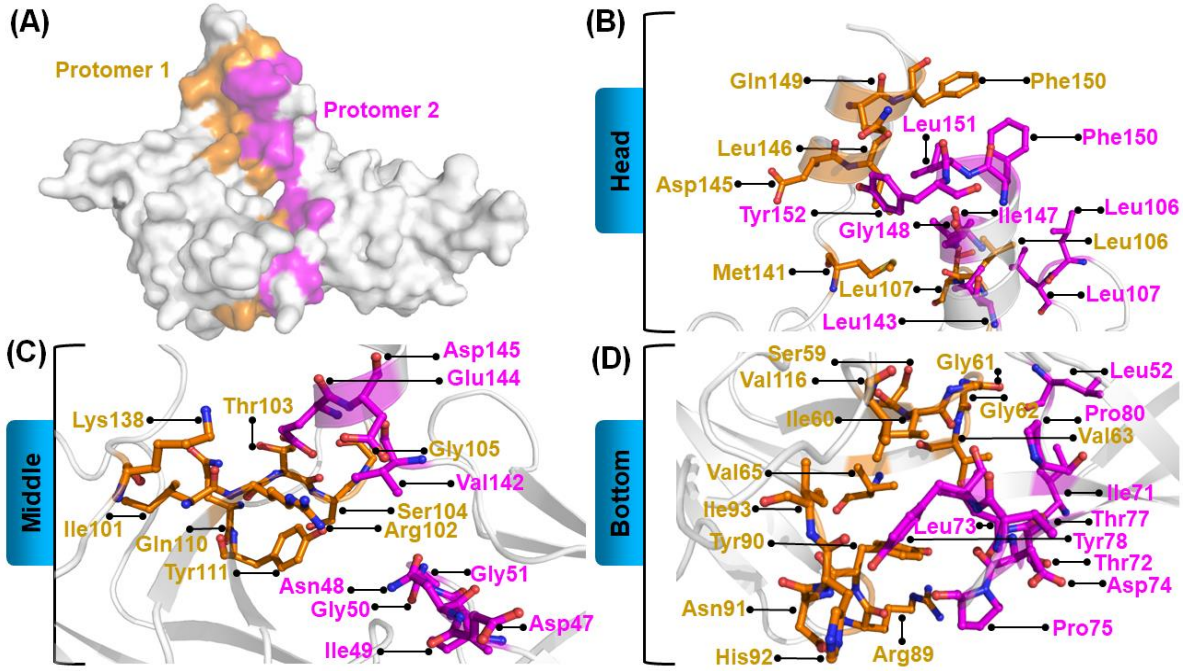


Figure 5.8. The residues present at the dimeric interface of EcMlaD*. (A) Surface representation of a dimeric interface where two protomers of protein EcMlaD are shown in orange and cyan. (B–D) The residues (ball-and-stick model) of each protomer present at the interface belonging to the head, body and base parts, respectively.*

5.3.7. *Ec*MlaD undergoes an "asymmetric protomer movement" as well as channel expansion during ligand transfer

The exact mechanism and the directionality of PLs transport followed by the Mla system remain unresolved to date. Although the Mla system was reported to be involved in the retrograde (OM to IM) transfer of PLs, structural similarities of *Ec*MlaE to ABC exporters and the conformations of bound PLs in the MlaFEDB complex and MlaC suggests its involvement in anterograde (IM to OM) transfer of PLs (Coudray et al., 2020; Dutta and Kanaujia, 2022). Furthermore, the PL transfer from *Ec*MlaD to *Ec*MlaC has recently been shown to be ATP-independent (Low et al., 2021). These reports suggest a significant conformational change in proteins MlaC and MlaD during PLs translocation. The *Ec*MlaC-*Ec*MlaD complex would provide significant insight into

the mechanism of PL transport. However, to date, structural details of the complex is not available. Hence, an attempt was made to crystallize the complex and determine the structure. For the crystallization of the *Ec*MlaC-*Ec*MlaD complex, equal volumes of 200 µM of *Ec*MlaC and *Ec*MlaD were incubated at room temperature for 1 hour. An initial screening for crystallization conditions was set up using the available screens at 4 and 20°C. The crystal hits obtained for the *Ec*MlaC-*Ec*MlaD complex produced diffractions at a resolution of ~2.9-2.4 Å. Unfortunately, structure determination revealed that only *Ec*MlaD was crystallized. As a consequence, in order to understand the mechanism, the PL-bound nucleotide-free cryo-EM structure of the MlaFEDB complex from *E. coli* (PDB id: 6XBD, *Ec*MlaFEDB) was utilized by first performing an orientation analysis. As anticipated, the *Ec*MlaE dimer (with the exception of the N-terminal region) and the TM helices of *Ec*MlaD hexamer (*Ec*MlaD_complex) are located in between the two leaflets of the IM. On the hand, *Ec*MlaF and *Ec*MlaB are located on the inner leaflet side (cytoplasm), while *Ec*MlaD_complex (protomers A'-F') is located on the outer leaflet side (periplasm). With respect to the outer leaflet of IM, the ring of *Ec*MlaD_complex makes an angle of ~16º. As a consequence, the protomers A'-C' are situated at an average distance of ~10 Å while the protomers D'-F' are situated at an average distance of ~17 Å from the outer leaflet (the average distance between Thr39 of each protomer and the outer leaflet was obtained, Figure 5.9A). Thus, one-half of the hexamer is close to the membrane, while the other half is away from the membrane. It hints at an asymmetric upward movement, which is in contrast to the proposition that the MlaD hexamer would evenly rest on the IM. Furthermore, the buried surface areas at the dimeric interface of the hexamers in *Ec*MlaD_$P4_12_12$ (uniform with an average of ~830 $Å^2$) and *Ec*MlaD_complex (vary with an average of ~838 $Å^2$) are different (Table C8).

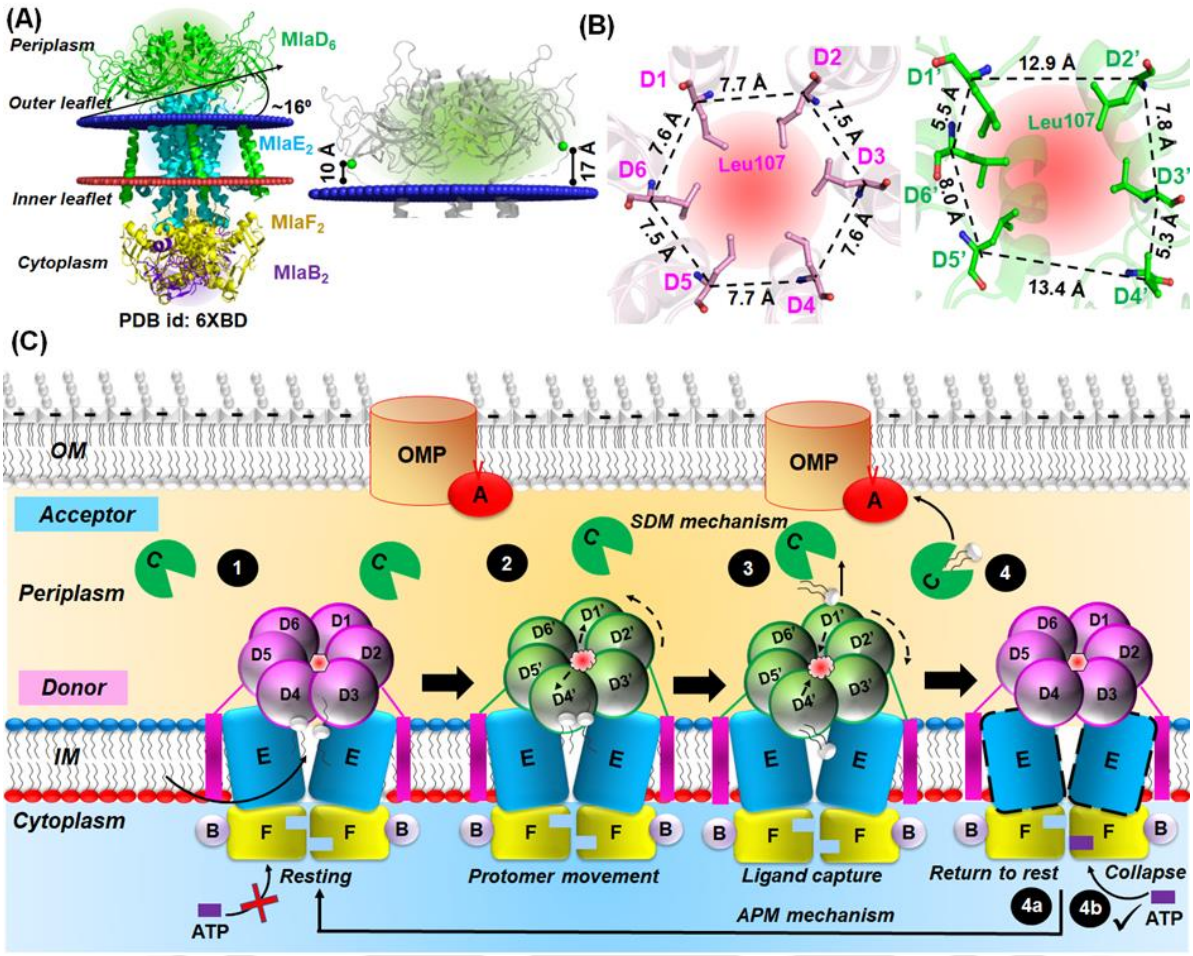


Figure 5.9. Mechanism of PL transport. *(A) Overview of the orientation of EcMlaFEDB complex (PDB id: 6XBD). (Left) Orientation of cryo-EM structure of MlaFEDB (MlaF, yellow; MlaE, cyan; MlaD, pink; MlaB, purple) in a membrane arrangement (outer leaflet, blue spheres; inner leaflet, red spheres) obtained from PPM 2.0 server (Lomize et al., 2012). With respect to the outer leaflet, the EcMlaD hexamer makes an angle of ~16º. (Right) Orientation of EcMlaD hexamer (EcMlaD_complex). Three protomers of EcMlaD_complex are at an average distance of ~17 Å and the remaining three are at an average distance of ~10 Å with respect to the outer leaflet. The distance was measured between Thr39 (green sphere) of EcMlaD_complex protomers and the outer leaflet. (B) Overview of the central channel in EcMlaD_P4$_1$2$_1$2 and EcMlaD_complex. (Left, right) Distance mapping between two adjacent Leu107 residues (ball-and-stick model) of the hexamers of EcMlaD_P4$_1$2$_1$2 (D1-D6) and EcMlaD_complex (D1'-D6'), respectively. The values along the dotted lines between the Cα atoms of two adjacent Leu107 residues denote the distances. (C) Proposed mechanism of PL transport. The mechanism comprises four steps, (1) Extraction of*

PLs from the IM into the EcMlaE binding pocket (outward-open) in the absence of ATP while the EcMlaD hexamer rests with its six-fold symmetry intact. (2) Asymmetric movement of the protomers followed by pore expansion resulting in the loss of six-fold symmetry. The change in channel geometry causes the re-orientation of PLs to attain the conformation required for export. (3) Ligand capture by periplasmic EcMlaC. (4) Returning of EcMlaD hexamer to the resting state by restoring its six-fold symmetry. (4a) The extraction and export PL cycle continues with EcMlaE and EcMlaD hexamer in their initial conformations. (4b) Otherwise, ATP binding leads to the collapse of the translocation pathway and halts PL export. SDM, segmented domain movement and APM, asymmetric protomer movement.

Upon superimposing the hexamers of *Ec*MlaD_P4$_1$2$_1$2 (A-F) with that of *Ec*MlaD_complex (A'-F'), the protomers A-F', B-E', C-D', D-E', E-F' and F-A' align to each other, referred to as D1-D1', D2-D2', D3-D3', D4-D4', D5-D5' and D6-D6', respectively (Figure C11A). The analysis reveals that the six-fold symmetry is lost in the *Ec*MlaD_complex. To investigate the impact on the central pore geometry due the loss of symmetry, the distances between the residue Leu107 of PLPs of two adjacent protomers were measured in both hexamers. The results reveal that PLPs of D1-D2 & D4-D5, D2-D3 & D5-D6 and D3-D4 & D6-D1 are situated at a distance of 7.7, 7.5 and 7.6 Å, respectively (Table C8). On the other hand, PLP distances of D1'-D2', D2'-D3', D3'-D4', D4'-D5', D5'-D6' and D6'-D1' were found to be 12.9, 7.8, 5.3, 13.4, 8.0 and 5.5 Å, respectively (Figure 5.9B and Table C8). This suggests that the loss of asymmetry in the *Ec*MlaD_complex has resulted in the complete rearrangement of the central channel. A closer inspection revealed that Leu107 makes an outward movement of 4.1 and 4.7 Å from D1 to D1' and from D4 to D4', respectively (Figure C11B). However, the spatial position of Leu107 in other protomers remains intact. This results in the movement of Leu107 in D1' closer to D6' (5.5 Å) and away from D2' (12.9 Å) (Figure C11B). Similarly, Leu107 in D4' moves closer to D3' (5.3 Å) and away from D5' (13.4 Å) (Figure C11B). Thus, in *Ec*MlaD_complex, the PLPs of two protomers (D1' and D4') from each half of the hexamer move outward (Figure C11B). This results in a side-wise expansion of the central channel near D1' and D4' in *Ec*MlaD_complex and changes the average PLP distance of 7.6 Å in *Ec*MlaD_P4$_1$2$_1$2 to 8.8 Å in *Ec*MlaD_complex (Table C8).

In order to understand the reason behind such a change in geometry, the central translocation pathway of *Ec*MlaFEDB was studied. The analysis revealed that at the points of side-wise expansion of *Ec*MlaD_complex, two PLs (PL1 and PL2) in two different conformations are present. In the case of PL1, the γ-chain (γ1) remains inserted into the *Ec*MlaE pocket while the β-chain (β1) projects into the *Ec*MlaD_complex pore. On the other hand, in the case of PL2, both the β- and γ- chains (β2, γ2) remain confined to the *Ec*MlaE pocket. Although the tails of these PLs are present in the central translocation pathway, their heads project into the lateral channels between *Ec*MlaE and *Ec*MlaD_complex (Figure C11C). Interestingly, the two conformations represent the two different states of PL export (Coudray et al., 2020). Additionally, *Ec*MlaE dimer in *Ec*MlaFEDB is in outward-open conformation as observed in other nucleotide-free PL un(bound) structures of MlaFEDB (Coudray et al., 2020; Ekiert et al., 2022). However, it has been previously reported that ATP binding leads to the collapse of the central translocation pathway restricting PL binding (Ekiert et al., 2022). Thus, it can be speculated that in the absence of ATP, the PLs from the IM would be extracted and accommodated in the binding pocket of *Ec*MlaE. The changes observed in the geometry of the central channel of *Ec*MlaD_complex, including the upward lift of the three protomers and loss of six-fold symmetry, would help in the re-orientation of accumulated PLs in attaining the conformation required for export. Based on these observations, it can be proposed that *Ec*MlaD_$P4_12_12$ (six-fold symmetry) and *Ec*MlaD_complex (loss of six-fold symmetry) represent the resting state and the ligand transporting state of *Ec*MlaD, respectively (Figure C11D). Thus, in the absence of ATP, the significant conformational changes in *Ec*MlaD and *Ec*MlaE would help in the transport of PLs from *Ec*MlaFEDB to *Ec*MlaC. This hints that the central translocation pathway is conformationally dynamic and adaptable, as reported previously (Mann et al., 2021). Also, the docking of *Ec*MlaC to the *Ec*MlaD (from the *Ec*MlaFEDB complex) might further induce changes that would enhance PL transfer. Such a conformation-driven ATP-independent mechanism of PL has been reported for MlaA (Abellón-Ruiz et al., 2017).

Based on these observations, we propose an “asymmetric protomer movement (APM)” mechanism of ligand transport. Wherein the MlaFEDB complex in the nucleotide-free state would extract the PLs from the IM and accumulate them in the binding pocket of the *Ec*MlaE dimer, which would be in the outward-open conformation. In such a state,

*Ec*MlaD would evenly rest on the top of *Ec*MlaE (step 1). This is followed by the upward movement of three protomers and outward shifts of two diagonally opposite PLPs of *Ec*MlaD. This results in the channel expansion and the loss of the six-fold symmetry of *Ec*MlaD, along with the re-orientation of PLs (step 2). The PLs exposed in the central channel of *Ec*MlaD are extracted by *Ec*MlaC present in the periplasm. Following the "segmented domain movement (SDM)" mechanism (Dutta & Kanaujia, 2022), *Ec*MlaC binds PL(s) (step 3). Eventually, *Ec*MlaC hands over the PL(s) to the *Ec*MlaA-OMP complex. Once PL(s) get(s) transferred, *Ec*MlaD can return to the initial stage restoring its six-fold symmetry and continuing the export of PL (Step 4a). Otherwise, the ATP binds to the *Ec*MlaFEDB complex, causing the collapse of the translocation pathway, thereby halting the export of PLs (Step 4b) (Figure 5.9C). Based on the requirement of the bacterial cell, the next cycle of export/import would be initiated.

5.3.8. The "asymmetric protomer movement" mechanism might not be demonstrated by all the MlaD domains-containing proteins

To investigate whether other MlaD domain-containing systems such as Pqi, Yeb, Mce and TGD can follow the proposed "asymmetric protomer movement" mechanism, a phylogeny analysis of these proteins was performed. The result reveals that *Ec*MlaD and *Mt*Mce1A-4F (cluster I) and *Ec*PqiB and *Ec*YebT (cluster II) clade separately (Figure 5.10A). This suggests that the Mce proteins might follow the "asymmetric protomer movement" mechanism of lipid transport, akin to the Mla system. As described, in the "asymmetric protomer movement" mechanism, the protomers D1'-D3' move upward while the PLPs of the protomers D1' and D4' shift outwards. Interestingly, the helices of the protomers D1' & D4' along with D2' & D5' of *Ec*MlaD interact with *Ec*MlaE (Figure C12A). It has previously been reported that Gly11 of motif LxxF/LG of *Ec*MlaD is involved in the interaction with the conserved Gly10 and Gly21 in the N-terminal interfacial helix (IF1) of *Ec*MlaE (Coudray et al., 2020). The MSA of the TM helix of *Ec*MlaD orthologs and IF1 of *Ec*MlaE orthologs from Gram-negative bacteria clearly shows the conservation of the glycine residues, suggesting the involvement of the "asymmetric protomer movement" mechanism (Figure C12A and Table C9). Thus, it can be anticipated that the Mce proteins and the associated TMDs would possess the conserved glycine residues. However, unlike Gram-negative bacteria, *Mt*Mce(1-4)E proteins are lipoproteins and do not contain the N-terminal IF1; rather, they contain N-terminal signal peptide that carries a positively-charged region

(n-region), hydrophobic region (h-region) and lipobox with invariant cysteine (c-region) (Figure 5.10B). Although *Mt*Mce(1-4)B-C proteins cluster with *Ec*MlaD, the residue Gly11 is conserved only in *Mt*Mce(1-4)C proteins (Figure 5.10C and Figure C12B-C12D). Similarly, *At*TGD1 also conserves the glycine residue, while *Ec*PqiB and *Ec*YebS proteins lack it (Figure C12E). Overall, the results suggest that since the proteins Mce, Pqi and Yeb lack the conserved glycine residue, they might not follow the "asymmetric protomer movement" mechanism of lipid transport.

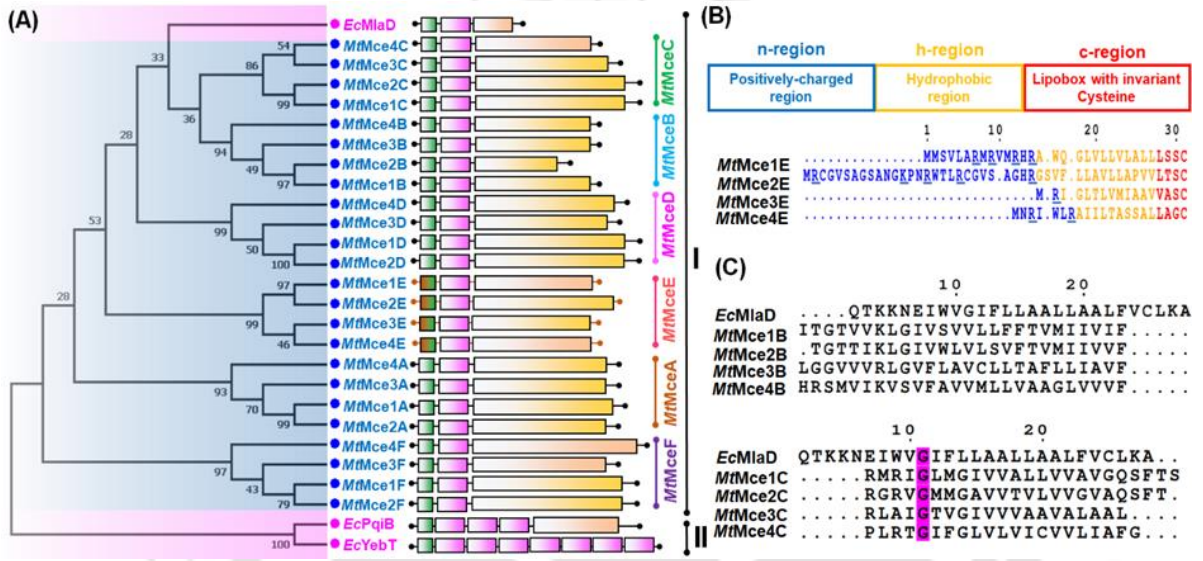


Figure 5.10. Evolutionary relationship of the MlaD domain-containing proteins. *(A) Phylogenetic tree displaying the evolutionary relationship of EcMlaD, MtMce1A-4F, EcPqiB and EcYebT. The branch tips of the proteins from E. coli and M. tuberculosis are highlighted in pink and blue, respectively. Numbers at each internal node of the phylogenetic tree signify the probability of its occurrence (in percentage) out of 1000 bootstrap replicates. The domain arrangements of each protein are provided on the side of the protein name. The domains TM helix (green), lipobox-containing region (brown-green), MlaD domain (pink), helical domain (wheat) and Mce4_CUP1 domain (yellow) are shown as rectangles. The two major clusters are labeled as I and II, while the subclusters of MtMce proteins are labeled as MtMceA (orange), MtMceB (cyan), MtMceC (green), MtMceD (magenta), MtMceE (red) and MtMceF (purple). (B) The N-terminal region of the MlaD domain contains three regions, viz. positively-charged n-region (blue), hydrophobic h-region (yellow) and lipobox with invariant cysteine c-region (red). The charged residues in the n-region are underlined. MSA of the N-terminal regions of MtMce(1-4)E. (C) MSA of TM helix of EcMlaD with MtMce(1-4)B*

(top) and MtMce(1-4)C (bottom). The conservation of Gly11 residue is highlighted in magenta.

5.3.9. The MlaD domain remains associated with diverse translocases

Since the residues, Gly10 and Gly21 in the N-terminal IF1 of *Ec*MlaE are critical for interaction with *Ec*MlaD (Coudray et al., 2020), their presence in other permeases such as *Ec*PqiA, *Ec*YebS, *Mt*YrbE1A-4B and *At*TGD1 was also investigated (Table C10). As per the Pfam database, the proteins *Ec*MlaE, *Mt*YrbE1A-4B and *At*TGD1 belong to the MlaE family (Pfam id: PF02405), whereas *Ec*PqiA and *Ec*YebS belong to the PqiA family (Pfam id: PF04403). Further, phylogenetic analysis reveals that *Ec*MlaE & *Mt*YrbE1A-4B and *Ec*PqiA & *Ec*YebS clusters in two separate groups, I and II (Figure C13A). It has previously been reported that the typical ABC and YrbE permeases locate their C-terminal helix on the cytoplasmic and periplasmic side of IM, respectively. On the other hand, the N-terminal helix in ABC permeases is located on the cytoplasmic side, while that of YrbE permeases can be equivocated (Casali and Riley, 2007).

The orientation analysis of *Ec*MlaFEDB (PDB id: 6XBD) exhibits that in *Ec*MlaE, the N-terminal is periplasmic while the C-terminal is cytoplasmic with the IF1 being almost parallel to the plane of IM (Figure 5.11A and Table C11). A topology analysis of *Ec*MlaE reveals that IF1 is parallel to the IM plane while IF2 remains at an angle (Figure 5.11B). Such an arrangement is critical for the interaction between TM helices of *Ec*MlaD and IF1 of *Ec*MlaE, thus, aiding in the “asymmetric protomer movement” mechanism. Notably, as per TMHMM predictions, *Mt*YrbE1A, *Mt*YrbE2B and *Mt*YrbE4A possess six transmembranes (TM) helices, whereas other members of the MlaE family contain five TM helices. Furthermore, both the N- and C-terminals of *Mt*YrbE1A are periplasmic, whereas those are cytoplasmic in *Mt*YrbE2B and *Mt*YrbE4A. The N- and C-terminals of *At*TGD1 are periplasmic and cytoplasmic, respectively. On the other hand, *Ec*PqiA and *Ec*YebS possess eight TM helices, with both terminals located on the cytoplasmic side. The theoretical models of *Mt*YrbE1A-4B exhibit five TM helices with N-terminal IF1 and IF2 having a similar orientation as in *Ec*MlaE (Figure C13B-C13I and Table C11). Although structures of *Mt*YrbE1A-4B and *Ec*MlaE are similar

(Figure 5.11C), the length and segmentation of their IF1 vary. Also, IF1 of *Mt*YrbE1A-4B contains the conserved Gly24 (Figure C13J).

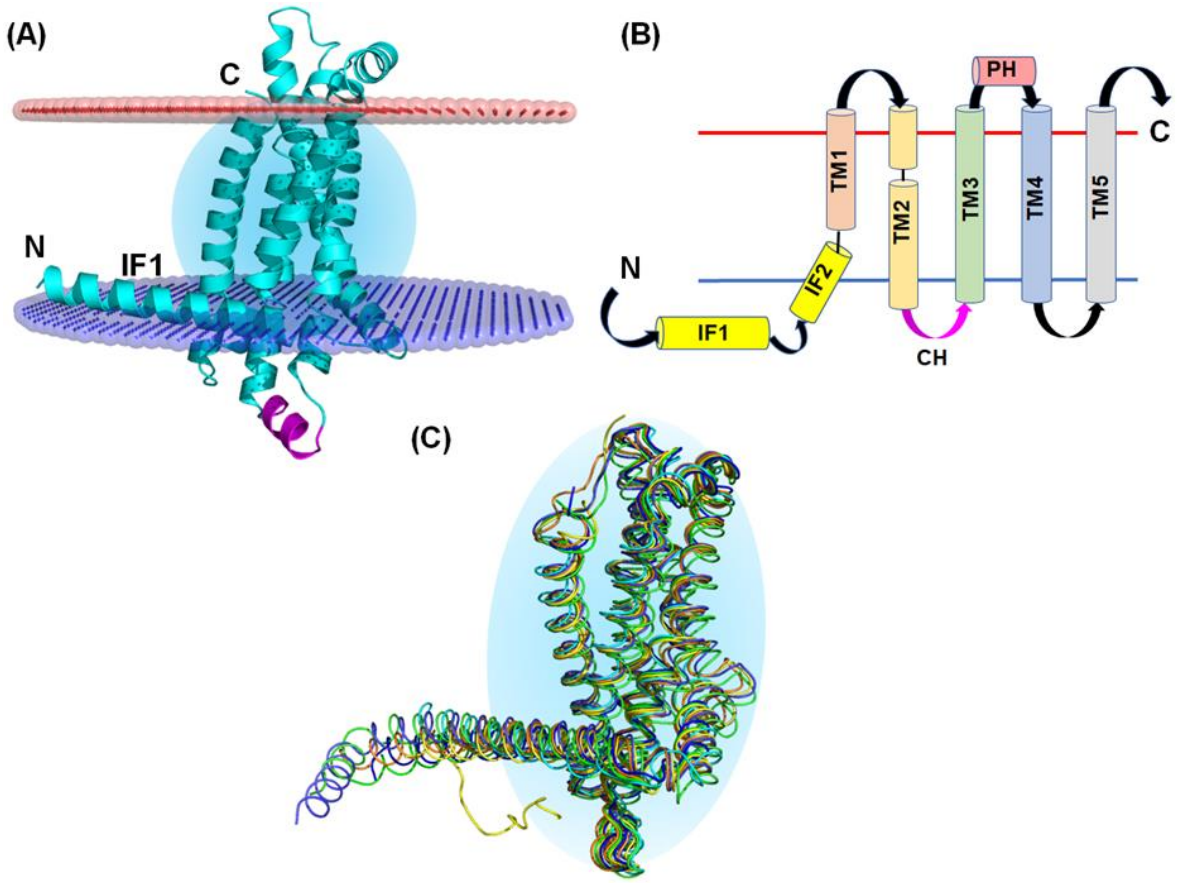


Figure 5.11. Structural analysis of EcMlaE. *(A-B) Orientation and topology analysis of EcMlaE (PDB id: 6XBD), respectively. The outer and inner leaflets of IM are represented by red and blue planes, respectively. IF: interfacial helix, TM transmembrane helix, PH: periplasmic helix and CH: coupling helix. (C) Structural comparison of EcMlaE (cyan) with its orthologs, MtYrbE1A-4B (green, splitpea, blue, purpleblue, yellow, yelloworange, orange and brightorange, respectively).*

Similar to *Ec*MlaE, the protein *At*TGD1 contains five TM helices, with the N- and C-terminals on the cytoplasmic and periplasm side, respectively. Further, its IF2 and TM1 are fused, as observed in some *Mt*YrbE proteins. Strikingly, the N-terminal region of IF1 is long and disordered (Figure C14A). In contrast, the proteins *Ec*PqiA and *Ec*YebS have eight TM helices with their N- and C- terminals on the cytoplasmic side and lack IF1 and IF2. Both these proteins possess an extended disordered N-terminal and multiple cytoplasmic (CS) and periplasmic (PS) segments (Figure C14B-C14C).

Pairwise sequence alignment of *Ec*MlaE and *At*TGD1 reveals the presence of an extension and an insertion in the N-terminal region of the latter (Figure C14D). In summary, the analysis suggests that the MlaD domain-containing proteins are not always typical MlaE permease dependent and thus may not follow the "asymmetric protomer movement" mechanism of lipid transport.

5.3.10. The MlaD protein shares structural similarity with EF/AMT-type beta(6)-barrel fold and constitutes a novel SBP cluster

A structural homology search suggests that the protein *Ec*MlaD is similar to the MlaD domain-containing proteins. In addition, *Ec*MlaD_MlaD1^{P41212} of *Ec*MlaD surprisingly shows structural similarity with the C-terminal region of the aminomethyltransferases from *Homo sapiens* (*Hs*Amt, PDB id: 1WSV), *Pelagibacter ubique* (*Pu*Amt, PDB id: 3TFH) and *Leptospirillum rubarum* (*Lr*Amt, PDB id: 3TTG) (Figure 5.12A and Table C12). As per the CATH database, *Hs*Amt comprises four domains, 1 (1a, 4-55; 1b, 147-240), 2 (56-146), 3 (241-286) and 4 (287-363), belonging the "probable tRNA modification gtpase trme; domain 1 (3.30.1360.120)", "aminomethyltransferase beta-barrel domains (3.30.70.1400)", "aminomethyltransferase fragment (4.10.1250.10)" and "aminomethyltransferase beta-barrel domains (2.40.30.110)" superfamilies, respectively (Figure 5.12B and Table C13). Interestingly, among these four domains, only domain 4 possesses a beta-barrel architecture (2.40), containing six anti-parallel beta strands (β1-β6) and is structurally similar to the *Ec*MlaD_MlaD1^{P41212}. On the other hand, *Pu*Amt and *Lr*Amt consist of only domain 1 (Table C13), which is structurally similar to all the four domains of *Hs*Amt (Figure 5.12B). Although the MlaD domain and domain 4 of *Hs*Amt, *Pu*Amt and *Lr*Amt are structurally similar, their topology varies. The MlaD domain has the β-strand order of 143265 (with an additional β7 strand) and domain 4 has 143256 (Figure 5.12C). Thus, domain 4 is similar to the MlaD domain (β1-β6, assigned as Core I; β7, assigned as Core II), wherein β5 and β6 strands flip the order in the former (Figure 5.12D). This indicates that the MlaD domain is structurally similar to the Aminomethyltransferase beta-barrel domain.

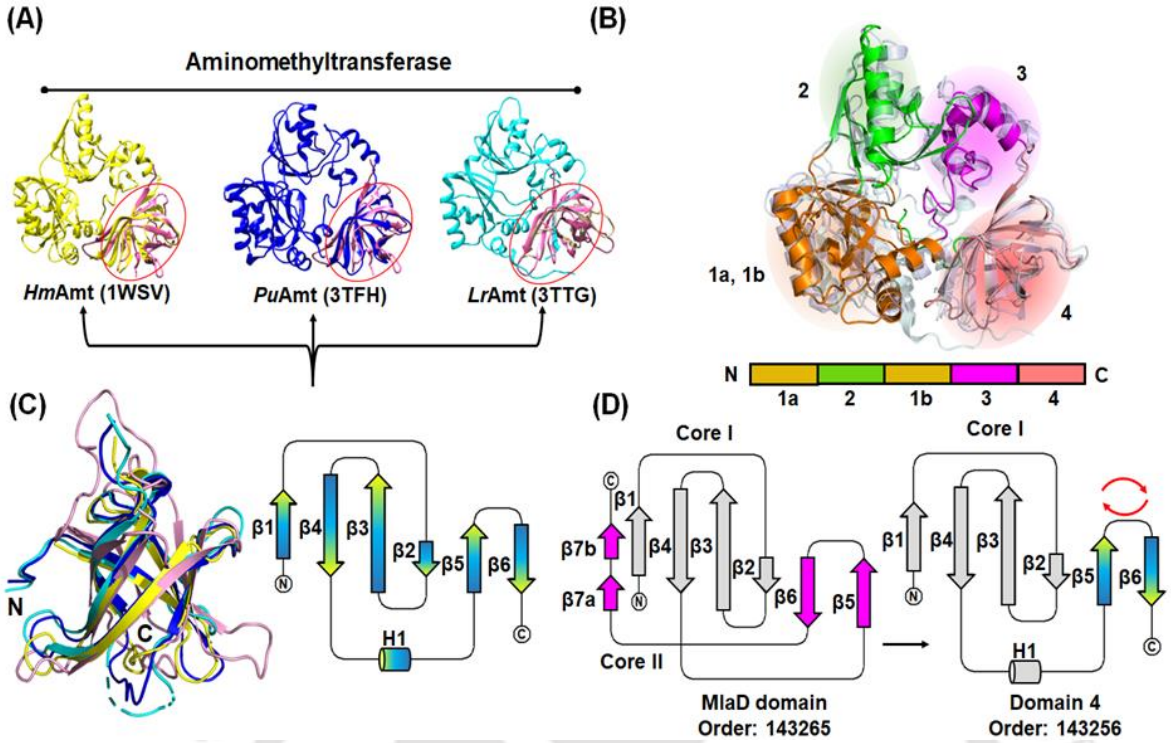


Figure 5.12. Ancestry analysis of the MlaD domain. *(A) Structural homologs of EcMlaD that do not contain the MlaD domain. EcMlaD_MlaD1^{P41212} (pink) shows similarity with the C-terminal region (encircled in red) of HmAmt (PDB id: 1WSV, yellow), PuAmt (PDB id: 3TFH, blue) and LrAmt (PDB id: 3TTG, cyan). (B) The overall structure and domain organization in different aminomethyltransferases, HmAmt (color spectrum), PuAmt (light blue) and LrAmt (pale cyan). (C) (Left) Structural superimposition of EcMlaD_MlaD1^{P41212} (pink) and domain 4 from HmAmt (yellow), PuAmt (blue) and LrAmt (cyan). (Right) A general topology diagram of the three aminomethlytransferases. The β-strands and helix are represented as arrows and cylinders (yellow-blue-cyan). (D) Topology comparison between EcMlaD_MlaD1^{P41212} and domain 4 of aminomethyltransferases. With respect to EcMlaD_MlaD1^{P41212}, the β5 and β6 are flipped in domain 4 denoted by red arrows. The region from β1 to β6 strand is designated as core I, whereas the region from β7a to β7b strands is designated as core II.*

In order to identify the origin of the fold of the MlaD domain, an ancestral analysis of aminomethyltransferase was performed using the SCOP database. However, there is no information available in the SCOP database for the proteins *Hs*Amt, *Pu*Amt and *Lr*Amt. Rather information about an ortholog aminomethyltransferase from

Thermotoga maritima (*Tm*Amt; PDB id: 1WOS) was available. As per CATH analysis, *Tm*Amt has similar domain architecture and is structurally similar to *Hs*Amt (Figure C15A). As per the SCOP database, the protein *Tm*Amt forms a domain (8056080), belonging to GcvT aminomethyltranferase-like family (4000062). Interestingly, the GcvT aminomethyltranferase-like family is involved in segment swapping. Thus, this family comprises members from aminomethyltransferase beta-barrel domain-like (3000024) and folate-binding domain-like (3000025) superfamilies. As anticipated, domain 4 belongs to the aminomethyltransferase beta-barrel domain-like superfamily, whereas the other three domains (1-3) belong to the folate-binding domain-like superfamily (3000025). The former belongs to folate-binding halfdomain-like (20000017) fold under the Alpha and beta proteins class (1000003), whereas the latter belongs to EF/AMT-type beta(6)-barrel (2000004) fold under All beta proteins class (1000001) (Figure C15B). As per the SCOP database, the All beta proteins class comprises 260 folds, each having its signature features. Among them, the EF/AMT-type beta(6)-barrel fold is marked by the presence of a barrel comprising six antiparallel strands that are in the order 143256. These observations indicate that core I of the MlaD domain (i.e. MlaD-type beta(6)-barrel) might have originated from EF/AMT-type beta(6)-barrel fold. Thus, it can be proposed that the EF/AMT-type beta(6)-barrel fold has originated from the All beta proteins class with the order 143256. However, in the course of evolution, the order of β5 and β6 strands flipped, giving rise to the core I of the MlaD domain. Further, this was followed by the elongation of the C-terminal (β7) of the domain leading to the incorporation of core II (Figure C15C).

Canonical SBPs, based on structural folds and substrate specificities, have been classified into eight clusters, A-H (Chandravanshi et al., 2021). The protein MlaD shows no structural similarity with the canonical SBPs. In our previous work, we classified SBPs into canonical and non-canonical based on the presence or absence of NTD-CTD arrangement and the mechanism of ligand binding (Dutta and Kanaujia, 2022). In the canonical SBPs, classes -I, II and III are included, while the non-canonical SBPs include MlaC, LptA and LolA. The SBP-I, II and III types again belong to the α/β class of proteins (SCOP id: 1000002). However, they possess the Rossmann(2x3)oid (SCOP id: 2000016, parallel β-strands are arranged in 21345 order), SBP2HA-like (SCOP id: 2000060, parallel β-strands are arranged in 21354 order, with anti-parallel β5 strand) and chelatase-like (SCOP id: 2001032, the four parallel β-strands are

arranged in 2134 order) fold, respectively (Figure C16A-C16B). Also, they belong to the superfamily Type 1 solute binding protein-like (SCOP id: 3000125), Type 2 solute binding protein-like (SCOP id: 3000083) and helical backbone metal receptor superfamily (SCOP id: 3001679), respectively (Figure C16A-C16C).

On the other hand, the non-canonical SBPs have diverse ancestries. MlaC belongs to the MlaC-like family (SCOP id: 4005548) of the NTF2-like superfamily (SCOP id: 3000472). It possesses the cystatin-like fold (SCOP id: 2000326) and belongs to the protein class α+β (SCOP id: 1000003) (Figure C17A). The functional analogs of MlaC, viz. LolA and LptA also have different ancestries and belong to the β-protein class (SCOP id: 1000001) (Figure C17B). However, the protein LolA belongs to the Outer-membrane lipoproteins carrier protein LolA family (SCOP id: 4003959) of the LolA-like superfamily (SCOP id: 3001964) and possesses LolA-like fold (SCOP id: 2001203). On the other hand, LptA belongs to the LptA-like family (SCOP id: 4004381) of the OstA-like beta-folder (SCOP id: 3002196) superfamily and has beta-folder fold (SCOP id: 2001334) (Figure C17C). The protein LptC is anchored to the IM by an N-terminal helix (Schultz et al., 2018). Thus, LptC seems to be a functional analog of MlaD. Surprisingly, LptC belongs to the same class, fold and superfamily to that of LptA. Based on these observations, we propose the inclusion of MlaD and LptC in the non-canonical SBPs having membrane anchors. Thus, the canonical class of SBP would comprise clusters A-H, including the SBP-I, II and III, while the non-canonical class is constituted by MlaC, MlaD, LolA, LptA and LptC (Figure C18).

5.4. Discussion

Identification of the Mla system has provided a new perspective on the translocation of PLs between the OM and IM. Unlike typical ABC transporters, the Mla system utilizes two different types of LTPs, ‘Box-like lipid shuttles’ (periplasmic MlaC) and ‘Tube-like lipid conduits’ (IM-associated MlaD), that work in coordination to perform the task of lipid shuttling (Wong et al., 2019). MlaD presents a unique case as it possesses the highly conserved MlaD domain and remains IM-associated. This study reports the crystal structures of the periplasmic component of *Ec*MlaD in three different space groups in the resolution range of 2.3-3.2 Å.

5.4.1. *Ec*MlaD oligomerizes to give rise to a dynamic central channel suitable for phospholipid transport

A single protomer of *Ec*MlaD comprises two distinct regions, the N-terminal MlaD domain and the C-terminal HD. The former consists of seven anti-parallel β-strands with interconnecting loops, while the latter comprises a short helix. However, the absence of electron densities of the entire HD and its poor pLDDT observed in the AphaFold model of *Ec*MlaD indicates that the HD would be extremely flexible or disordered. Six copies of *Ec*MlaD protomer form a homo-hexameric ring with a central channel. This channel has two major constrictions but possesses sufficient volume for the passage of PL molecules. Furthermore, pathway mapping reveals that the central channel is continuous but has varying dimensions and is extremely hydrophobic. The PL molecules are dynamic and tend to execute 'lipid gymnastics' (Neumann et al., 2017). Such an arrangement probably provides directionality during the passage of different PLs. Additionally, the absence of HD results in domain stacking, leading to a dodecameric assembly of *Ec*MlaD. Hence, this establishes that HD is critical for determining the oligomeric state of the protein. Furthermore, in *Ec*YebT, the stacked domains are connected by linker regions (Isom et al., 2020). This hints that HD in *Ec*MlaD would prevent the stacking of the MlaD domains, thus, giving rise to a hexameric assembly.

5.4.2. The MlaD domains have conserved cores but possess local structural differences

A structural comparison reveals that the MlaD domains from *E. coli*, *M. tuberculosis* and *A. thaliana* have conserved cores, which basically comprise seven anti-parallel β-strands. However, a topology study reveals that the MlaD domains demonstrate extreme variation in the loop regions. Furthermore, in some MlaD domains, the strands β1, 2, 3 and 7 are segmented. Also, the MlaD domains from plants and cyanobacteria have cysteine residues that have been previously reported to be less in the MlaD domain (Dutta et al., 2021). It is likely that the MlaD domains from higher organisms might have evolved additional structural features. The structure-based secondary structure estimation of *Ec*MlaD_MlaD1^{P41212} and the MlaD domains shows significant variations. Structure-based MlaD domain profiling led to the identification of 19 different profiles of the MlaD domains of varying lengths from *E. coli*, *M. tuberculosis*, *A. thaliana*, *A. baumannii* and *P. aeruginosa*. Thus, the MlaD domains present a case of

local structural differences, as observed across various classes (All-α, All-β, α/β and α+β) (Joseph et al., 2012). Such differences have been reported to contribute to diversity in thermal stability, interaction, functionality, substrate specificity, complex formation, etc. (Sandhya et al., 2009). This has been observed in the case of *Ec*YebT, wherein the MlaD domain rings show differential flexibilities (Isom et al., 2020). Thus, it can be speculated that the local structural differences give rise to the conformational dynamicity of the MlaD domain. Furthermore, owing to the presence of cysteine residues, it can be speculated that the MlaD domains from higher organisms might have developed separately and acquired additional features.

5.4.3. Phospholipid transport involves two complementary mechanisms

Unlike typical ABC transporters, the Mla system is dependent on two different SBPs, MlaC and MlaD. The canonical SBPs have been reported to follow one of the four mechanisms of ligand binding, (1) 'Venus Fly Trap' (Mao et al., 1982), (2) 'asymmetric domain movement' (Pandey et al., 2016), (3) 'one domain movement' (Chandravanshi et al., 2020a) and (4) 'subdomain movement' mechanism (Chandravanshi et al., 2020b) (Figure 5.13A). On the other hand, non-canonical SBP MlaC follows the "segmented domain movement (SDM)" mechanism of ligand binding. Unlike, MlaC and canonical SBPs, *Ec*MlaD possesses the conserved MlaD domain and oligomerizes to a homo-hexameric ring with a central channel. It has been reported that the MlaD domains, especially the PLPs, would be extremely flexible in the proteins *Ec*MlaD, *Ec*PqiB and *Ec*YebT (Isom et al., 2020; Dutta et al., 2021). This study shows that three protomers of the hexamer would demonstrate upward movements while the PLPs from two opposite protomers would demonstrate outward movement resulting in a change in the geometry of the central channel. This movement likely exposes the PLs to be transported in the central channel so that they could be captured by *Ec*MlaC. This has been referred to as the "asymmetric protomer movement (APM)" mechanism. Furthermore, *Ec*MlaC would bind to the exposed PL by following the SDM mechanism and take it to the *Ec*MlaA-Omp complex (Dutta & Kanaujia, 2022). Thus, both the SDM and APM mechanisms would be executed sequentially for the transport of PLs (Figure 5.13B). Also, this mechanism seems particularly favorable for the transport of flexible lipids exhibiting 'lipid gymnastics' (Neumann et al., 2017). Additionally, LPS and lipoprotein transport by Lpt and Lol systems, respectively, are dependent on non-

canonical SBPs and involve mechanisms different from canonical SBPs (Takeda et al., 2003; Chng et al., 2010; Mi et al., 2017) (Figure 5.13B).

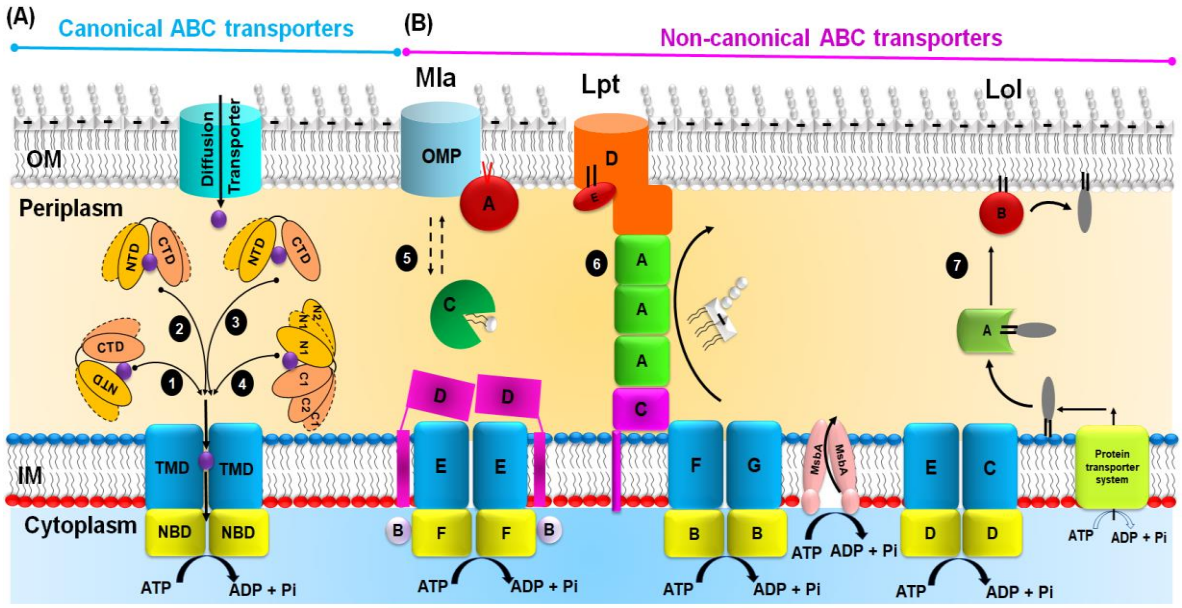


Figure 5.13. Overview of the two major types of ABC transporter mechanisms based on SBP types. *(A) Canonical ABC transporters. (1) The 'Venus Fly-Trap' (Mao et al., 1982). (2) The 'asymmetric domain movement' mechanism (Pandey et al., 2016). (3) The 'one domain movement' mechanism (Chandravanshi et al., 2020a). (4) The subdomain movement mechanism (Chandravanshi et al., 2020b). (B) Non-canonical ABC transporters. (5) The "segmented domain movement" mechanism (Dutta and Kanuajia, 2022). (6) The "asymmetric protomer movement" mechanism (this study). (7) The Lpt transport mechanism (Chng et al., 2010; Mi et al., 2017). (8) The Lipoprotein transport mechanism (Takeda et al., 2003).*

Interestingly, the execution of the APM is critically dependent on the interaction between TM of *Ec*MlaD and IF1 of *Ec*MlaE through glycine residues conserved in Gram-negative bacteria. This indicates that the APM mechanism would be observed in most of the Gram-negative bacteria. However, this study shows the absence of these residues or association of diverse TMDs in Pqi, Yeb, Mce and TGD systems. The Pqi and Yeb systems do not belong to the ABC transporters and also are independent of MlaC-like components (Dutta et al., 2021). On the other hand, the Mce and TGD systems might have different mechanisms of substrate transport in spite of the

presence of the MlaD domains. This might be because both the Mce and TGD systems are involved in the import of substrates and are independent of any free-floating SBP. This is further evidenced by the difference in the macromolecular arrangement of the two systems that involve additional and/or unique components. Thus, the SDM-APM complementary mechanism might be particularly favorable for the transport of PL between MlaC and MlaD. Furthermore, recent studies have reported that *Ec*MlaD orthologs are also involved in the transport of rubber degradation products in *Gordonia* polyisoprenivorans VH2, *Nocardia nova* SH22a, *Steroidobacter cummioxidans sp. nov*., strain 35Y and uptake of γ-hexachlorocyclohexane in *Sphingobium japonicum* UT26 (Lal et al., 2010; Nagata et al., 2011; Sharma et al., 2018). Owing to the diversity of the substrates, it can be speculated that *Ec*MlaD orthologs might participate in different mechanisms of substrate transport.

5.4.4. *Ec*MlaD constitutes a novel cluster of SBP

The peculiar mechanism of ligand transport can be attributed to the unique structure of *Ec*MlaD. Typical SBPs have originated from CheY-like proteins undergoing domain swapping and dimerization (Fukami-Kobayashi et al., 1999). The aberration of SBP structure has been previously reported for members belonging to the PAO family that are marked by the fusion of two SBP domains (van der Heide & Poolman, 2002). Even recently, MlaC, LolA and LptA have been recognized as non-canonical SBPs owing to their unique ancestries and mechanism of substrate binding (Dutta & Kanaujia, 2022). Surprisingly, the protein MlaD has not been considered under any such classification scheme of SBP. This study suggests that the MlaD domain is derived from the EF/AMT-type beta(6)-barrel fold present in domain 4 of aminomethyltransferase. Interestingly, the MlaD domain has previously been reported to be present with domains associated with methyltransferases, thus, giving rise to diverse domain architectures (Dutta et al., 2021). Owing to their unique ancestries as well as the mechanism, both MlaD and LptC have now been incorporated into the class of non-canonical SBPs. Furthermore, the transporters, e.g., Mla, Lpt and Lol, do not show a resemblance to the canonical ABC transporters. Thus, these transporters have been categorized as non-canonical ABC transporters. Furthermore, these three transporters are dependent on OM-associated lipoproteins, which ensure the proper localization of the substrates. Based on the number of SBPs or carrier-like proteins as well as the sequence of substrate transport, the mechanisms of transport of Mla, Lpt and Lol

resemble a relay arrangement in which one component hands over the substrate to multiple components ensuring proper localization.

5.5 Conclusion

This study provides significant insights into the structural and mechanistic aspects of *Ec*MlaD. To the best of our knowledge, it is the first comprehensive analysis of *Ec*MlaD. The work documents the structure-based domain profiling of the MlaD domains from Mla and its orthologous systems. Furthermore, based on an extensive-structural analysis, a novel ligand transport mechanism, viz. “asymmetric protomer movement (APM)”, has been proposed. The study presents mechanistic insights into the functioning of Mce proteins and firmly establishes the presence of two complementary mechanisms that are involved in ligand transport. Furthermore, the work also provides significant insights into the unique evolutionary ancestry of *Ec*MlaD. Altogether, the finding firmly establishes *Ec*MlaD to be a unique SBP.

Appendix C. Supplementary data

Supplementary figures, Figure C1-C18.

Supplementary tables, Table C1-C13.

Chapter 6

Molecular docking and molecular dynamics simulation studies of *Ec*MlaC protein

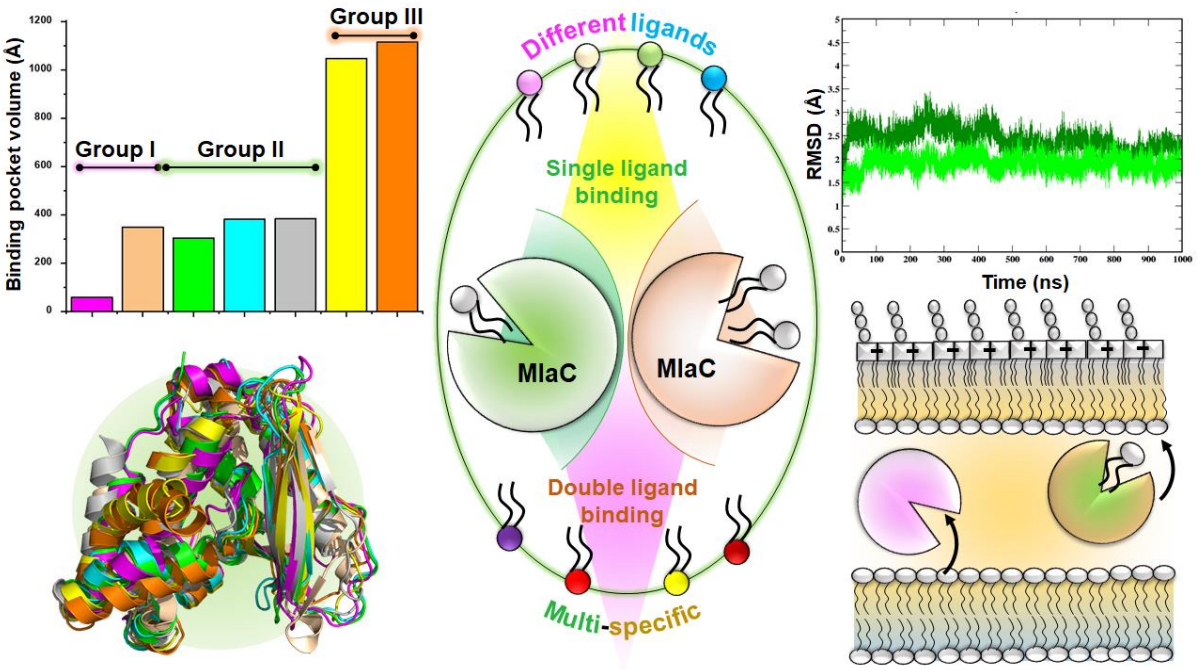
A part of this chapter is under preparation for publication:

1. **Dutta A**, Patel S and Kanaujia SP. Molecular dynamics simulation studies of MlaC. ***Under preparation*.**

Abstract

The outer membrane (OM) is a salient feature of Gram-negative bacteria, which serves as a formidable barrier against a plethora of detrimental compounds such as toxins, antibiotics, detergents, etc. This barrier function is contributed by the asymmetric nature of the OM owing to the presence of lipopolysaccharide (LPS) in the outer leaflet and phospholipid (PL) in the inner leaflet. The highly conserved intermembrane ATP-binding cassette (ABC) transporter, viz. maintenance of lipid asymmetry (Mla) system, is involved in the transfer of PL between the membranes in order to safeguard the PL asymmetry of the OM. The protein MlaC serves as the substrate-binding protein (SBP) component of the system that ferries PLs. Owing to its unusual structural features, MlaC has been recognized as a non-canonical SBP that follows a unique mechanism of ligand binding, viz. ‘’segmented domain movement (SDM)’’. However, knowledge about the ligand specificity as well as the atomic motions of MlaC, is still scarce. Hence, in this study, an extensive molecular docking and molecular dynamics simulation study was conducted on the MlaC protein from *Escherichia coli* (*Ec*MlaC) and its orthologues in order to gather insights into the mechanism of action of the protein. The study reveal that *Ec*MlaC orthologues can bind different PLs in varying conformations. The bound PLs have the liberty to change their orientation depending on the availability of space within the binding pocket. Furthermore, the simulation studies conducted for 1000 ns for the *Ec*MlaC orthologues reveals that they possess large conformational landscape owing to their dynamic structural features. The analyses also validate the previously reported SDM mechanism and further enrich the information due to the identification of new protein motions. Based on these findings, we have been able to put forward a working model that highlights the molecular details of PL transport by *Ec*MlaC.

Graphical abstract



6.1. Introduction

The cell envelope in Gram-negative bacteria consists of an inner membrane (IM) and an outer membrane (OM) separated by a spacious periplasmic space, which contains a thin peptidoglycan layer. The OM is a distinctive feature of Gram-negative bacteria as it acts as a protective shield against the assaults of detrimental compounds owing to its asymmetric nature. This is because the lipopolysaccharides (LPSs) and phospholipids (PLs) are located in the outer and the inner leaflet, respectively. Such an arrangement creates a strong barrier, which does not allow the passage of antibiotics, detergents, toxins, etc. (Delcour, 2009; Henderson et al., 2016). However, the PLs in the inner leaflet have the tendency to flip back and accumulate in the outer leaflet leading to the disruption of the asymmetric nature of OM and compromising the barrier function. In order to restore the ideal PL population in the OM, bacteria implement two major mechanisms, both involving OM β-barrel enzymes, viz. PldA and PagP. Both these enzymes re-install the OM asymmetry by modifying the unwanted PLs (Bishop, 2008). However, a third mechanism has been recently reported that utilizes a highly conserved intermembrane transporter, viz. maintenance of lipid

asymmetry (Mla) system. The system consists of an OM lipoprotein-Osmoporin complex (MlaA-OmpC/F), a periplasmic protein (MlaC) and an IM ATP-binding cassette (ABC) transporter complex (MlaFEDB) (Malinverni and Silhavy, 2009). As per the initial proposition, the Mla system would be involved in a retrograde movement wherein it would extract the mislocalized PLs from the OM and transfer them to the IM complex MlaFEDB with the aid of MlaC for probable internalization (Malinverni and Silhavy, 2009). However, recent reports have also highlighted that the Mla system participates in the anterograde and/or bidirectional movement of PLs between the membranes (Hughes et al., 2019; Coudray et al., 2020; Low et al., 2021; Dutta and Kanaujia, 2022; Ekiert et al., 2022). Irrespective of the directionality of movement, MlaC is the sole PL-transferring mobile component of the Mla system that serves as the substrate-binding protein (SBP) and connects the IM and OM.

The SBPs remain functionally associated with ABC importers and not with exporters. They impart specificity as well as directionality to the importers by capturing substrates and translocating them to the transmembrane domains (TMDs) for subsequent internalization (Maqbool et al., 2015). They share low sequence similarities among themselves but possess a highly conserved architecture of N- and C-terminal domains (NTD and CTD) connected by a flexible hinge region. In both the domains, α-helices flank the β-sheets although the topological arrangement of β-sheets in the core regions of NTD and CTD, along with the overall hinge region, can vary (Fukami-Kobayashi et al., 1999; Lee et al., 1999). Based on structural folds and substrate specificities, SBPs have been classified into eight clusters, A-H (Berntsson et al., 2010; Scheepers et al., 2016; Chandravanshi et al., 2021). Furthermore, in all the SBPs, the substrate-binding site is located between the NTD-CTD and the flexibility of the hinge region aids the movement of the two domains for the capture of ligands (Fukami-Kobayashi et al., 1999). Based on the movement of NTD and CTD for the entrapment of ligands, the ligand-binding mechanisms of SBPs are classified into four types. These are “Venus Fly Trap” (both the domains move equally), “asymmetric domain movement” (one domain moves more than the other), “one domain movement” (only one domain moves) and “subdomain movement” (different subdomains move differentially) mechanisms (Mao et al., 1982; Pandey et al., 2016; Chandravanshi et al., 2020a; Chandravanshi et al., 2020b).

However, structural studies have established that MlaC does not belong to either of the eight SBP clusters as it does not possess the typical NTD-CTD arrangement (Ekiert et al., 2017; Yero et al., 2021; Dutta and Kanaujia, 2022). Instead, it consists of two domains, viz. nuclear transport factor 2-like (NTF2-like, D1) and phospholipid-binding protein (PBP, D2). Each domain is further divided into two subdomains arranged in a discontinuous fashion (D1R1, D2R1, D1R2, D2R2). Unlike typical SBPs, the binding pocket in MlaC is in the shape of a decanter that is sandwiched between the subdomains D1R1-D2R1 and D1R2-D2R2 (Yero et al., 2021; Dutta and Kanaujia, 2022). Owing to these unique structural features, we categorized MlaC under the class of non-canonical SBP (Dutta and Kanaujia, 2022). Furthermore, we have recently reported the crystal structure of MlaC from *Escherichia coli* (*Ec*MlaC) in a quasi-open state and following an extensive structural analysis of *Ec*MlaC and its orthologs (in different liganded states), we were able to report for the first time the fifth mechanism of ligand-binding viz. “segmented domain movement’’ mechanism. This involves a sequence of structural changes in the subdomains leading to the opening of the binding pocket followed by ligand capture (Dutta and Kanaujia, 2022). Despite sufficient structural data, there is a dearth of studies dedicated to the identification of protein motions (at the global and local scales) in MlaC. MlaC has been recognized as a promising drug target and unraveling the dynamic properties of the protein would help design effective anti-microbial interventions. Hence, in this work, extensive molecular docking and dynamics simulations of *Ec*MlaC and its orthologues were performed in order to understand the binding and dynamic properties of the protein. Additionally, an attempt was made to identify the critical atomic motions of the protein and put forward an updated working model unveiling the molecular mechanism of ligand binding and transport.

6.2. Materials and methods

6.2.1. Analysis of structural data

The protein sequences and the atomic coordinates of structures analyzed in the study were retrieved from the UniProtKB database (The UniProt Consortium, 2023) and RCSB Protein Data Bank (PDB) (Berman et al., 2000), respectively. In order to determine the sequence similarities among the MlaC orthologs, pair-wise sequence

alignment was performed using the web tool BLAST with the default set of parameters (Altschul et al., 1990). To determine the subdomain boundaries, structure-based multiple sequence alignment was performed using the program PROMALS3D (Pie et al., 2008). The volumes of the ligand-binding pockets were calculated using the program CASTp with a default probe radius of 1.4 Å (Tian et al., 2018). Helix curvatures were analyzed using the tool Bendix embedded in the program Visual Molecular Dynamics (VMD) (Humphrey et al., 1996; Dahl et al., 2012). The hydrophobicity mapping of the proteins was performed by using the YRB color scheme in the program PyMOL (Hagemans et al., 2015). As per this scheme, the carbon atoms not bound to nitrogen and oxygen atoms are highlighted in yellow, nitrogen atoms in the side chains of lysine and arginine are in blue, oxygen atoms in the side chains of glutamate and aspartate are in red, and all remaining atoms are in white.

6.2.2. Molecular docking studies

The binding affinities of the different PLs to *Ec*MlaC and its selected orthologues were estimated by molecular docking analysis performed using the freely available program AutoDock v4.2 (Morris et al., 2009). The apo form of the proteins was obtained by manually removing the bound ligands from the binding pocket using the program Coot. A total of 104 PLs were selected for the docking studies. On the other hand, the three-dimensional atomic coordinates of the ligands were extracted from the available crystal/cryo-EM structures, which were retrieved from the RCSB Protein Data Bank (PDB). For the PLs whose structures are not available PDB, the atomic coordinates were obtained from PubChem in simulation description format (SDF), which were later converted to PDB format by using the software (O'Boyle et al., 2011; Kim et al., 2023). Wherever applicable, before each docking experiment, the hydrogen atoms were added to the protein and ligand using the module available in the program AutoDock. Subsequently, the partial charges were assigned to the protein atoms using the Gasteiger charge algorithm (Gasteiger and Marsili, 1980). To perform the blind docking of the ligands to the protein, a grid size of 126x126x126 with a grid spacing of 0.375 Å was generated by taking the center of mass of the targeted protein as the grid center. The rigid molecular docking was performed by keeping the protein atoms rigid and the ligand atoms flexible, along with the rotatable bonds. For each molecular docking experiment, a total of 2000 runs of the Lamarckian genetic algorithm (LGA) were set up. The docked ligand conformations were clustered with a root mean square

deviation (RMSD) cut-off of 2.0 Å. The molecular interactions of docked ligands and other structural figures were prepared using the program PyMOL.

6.2.3. Generation of the molecular systems

The crystal structures of MlaC orthologues in different liganded states were obtained from PDB. These included unliganded MlaC systems from *E. coli* (PDB id: 6GKI; *Ec*MlaC(Apo)) and *Neisseria gonorrhoeae* (PDB id: 8DTE; *Ng*MlaC(Apo)), single-liganded MlaC systems from *E. coli* (PDB ids: 7VR6, 5UWA; *Ec*MlaC(Holo1), *Ec*MlaC(Holo2), respectively) and *Ralstonia solanacearum* (PDB id: 2QGU; *Rs*MlaC(Holo)) and double-liganded MlaC systems from *Pseudomonas putida* (PDB id: 5UWB; *Pp*MlaC(Holo)) and *Pseudomonas aeruginosa* (PDB id: 6HSY; *Pa*MlaC(Holo)). Following the classification scheme we have previously proposed, these systems were categorized under three groups – I, II and III, respectively (Dutta and Kanuajia, 2022). One representative from each group was selected for analysis (*Ec*MlaC(Apo), Group I; *Ec*MlaC(Holo1). Group II; *Pa*MlaC(Holo)). As NgMlaC(Apo) presented a unique case of a transient state, it was also included in the study. By manually removing the endogenously bound PL(s) from the holo MlaC systems, the corresponding apo MlaC systems were generated.

6.2.4. Molecular dynamics simulations

Molecular dynamics (MD) simulations were performed using the package GROMACS 2021.1 (Abraham et al., 2015). The experimental setup can be divided into two different categories. These included simulations of the systems in the (A) ligand-free and (B) ligand-bound states. The simulations of both the systems in their ligand-free states were performed using the force field Amber ff99SB embedded in GROMACS (Ponder and Case, 2003; Hornak et al., 2006). The molecular topology and coordinate files were generated using the program *gmx pdb2gmx*. In the case of the simulations of ligand-bound states, the generation of topology and coordinate files of the ligands was done using the ACPYPE program (Sousa da Silva and Vranken, 2012). This was followed by the generation of a dodecahedron box around the systems using the program *gmx editconf*. The minimum distance between the solute and the edge of the box was kept at least 1.2 nm. The prepared systems were solvated with the TIP3P water model using the *gmx solvate* module (Price and Brooks, 2004). Chloride and sodium ions were utilized to neutralize the overall charge of the systems using the

module *gmx genion*. Energy minimization was performed using the steepest descent method with a maximum force cut-off of 1000 kJ/mol/nm to ensure that the system has no steric clashes or inappropriate geometry and that the structures are relaxed. The solvent and ions around the proteins were equilibrated in two phases. The long-range electrostatic interactions were evaluated using the particle mesh Ewald (PME) method (Dartan et al., 1993; Essmann et al., 1995). The short-range van der Waals interactions were computed using Verlet neighbor list calculation with a cut-off of 10 Å. The P-LINCS algorithm was used to constrain the bond lengths, and a time step of 2 fs was used to integrate the equations of motion (Hess, 1997). The equilibrations were first executed under an NVT (canonical or isothermal-isochoric) ensemble for 100 ps with a reference temperature of 300 K. The second equilibration was carried out under an NPT (isothermal-isobaric) ensemble for 100 ps with a reference pressure of 1 bar. A velocity-rescaling thermostat with a coupling constant of 0.1 ps was used to control the reference temperature at 300 K (Bussi et al., 2007). The reference pressure and coupling constant was set to 100 kPa (1 bar) and 2 ps, respectively, and were controlled using the Parrinello-Rahman barostat (Parrinello and Rahman, 1981). Production runs have been carried out, reaching 1 µs for each system, for a total of 6 µs of classical MD (i.e., ~1 µs x 6 systems). Coordinates of the systems were collected every 10 ps for a total of 100,000 frames for each run. All the simulations were performed on the Param-Isham high-performance computing facility at the Indian Institute of Technology Guwahati (IITG).

6.2.5. Analysis of simulation data

MD trajectories were analyzed using in-build GROMACS modules and Bio3D library in R-packages. Structural visualization of simulation trajectories was analyzed using VMD software and PyMol (Humphrey et al., 1996). The program XMGRACE v 5.1.25 was used to prepare the graphs (https://plasma-gate.weizmann.ac.il/Grace/). We measured generalized correlations that include mutual information of nonlinear contributions of the protein by applying Bio3D, which produces results in the form of a dynamic cross-correlation matrix (DCCM). Principal component analysis (PCA) and free-energy landscape (FEL) were performed to extract the functionally relevant motions from the simulated trajectories using GROMACS. The structural figures were prepared using the program PyMOL. The overall workflow of the simulation studies is provided in Figure 6.1.

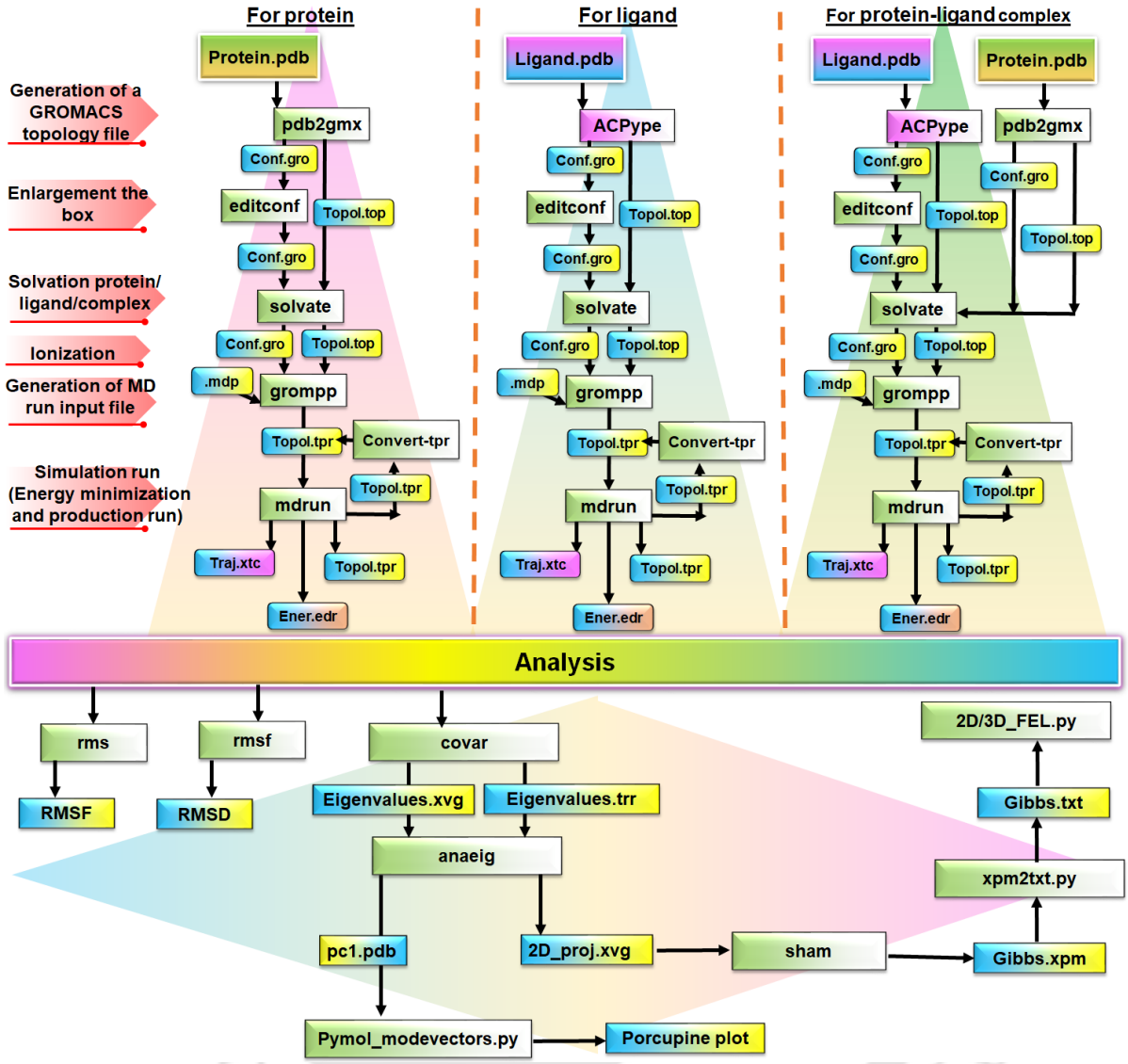


Figure 6.1. Overview of the workflow of molecular dynamics simulation studies.

6.3. Results

6.3.1. MlaC orthologues can exist in the open unliganded state

In order to understand the protein dynamics, it is critical to find out the different liganded states along with the plasticity of the binding pocket of *Ec*MlaC and its orthologues. We have previously reported that *Ec*MlaC orthologues possess hydrophobic binding pockets that remain closed in the unliganded state. This is in order to shield the pockets from the surrounding water molecules of the periplasm. Furthermore, as per our

proposed PL binding cycle, the open unliganded state would be transient in nature (Dutta and Kanuajia, 2022). Recently, the crystal structure of *Ec*MlaC orthologue from *Neisseria gonorrhoeae* in the unliganded state has been reported (PDB id: 8DTE; *Ng*MlaC(Apo)). A structural comparison of *Ec*MlaC(Holo1) with *Ng*MlaC(Apo) and other orthologs suggests overall similarity, with RMSD in the range of 0.7 to 3.9 Å for all Cα atoms (Table D1, D2; Figure D1A-D1D). However, the orthologues share diverse sequence similarities. Hydrophobicity mapping reveals that the binding pocket of *Ng*MlaC(Apo) would be extremely hydrophobic, similar to other MlaC orthologues (Figure 6.2A). A comparison of the binding pocket volumes reveals that the cavity of *Ng*MlaC(Apo) is ~6-fold larger (350 $Å^3$) than Group I member (closed; *Ec*MlaC(Apo), 59 $Å^3$), which is similar to that of Group II members (open single-liganded state; *Ec*MlaC(Holo1), 305 $Å^3$; *Ec*MlaC(Holo2), 382 $Å^3$; *Rs*MlaC(Holo), 384 $Å^3$) but less than Group III members (open double-liganded state; *Pp*MlaC(Holo), 1048 $Å^3$; *Pa*MlaC(Holo), 1115 $Å^3$) (Figure 6.2B).

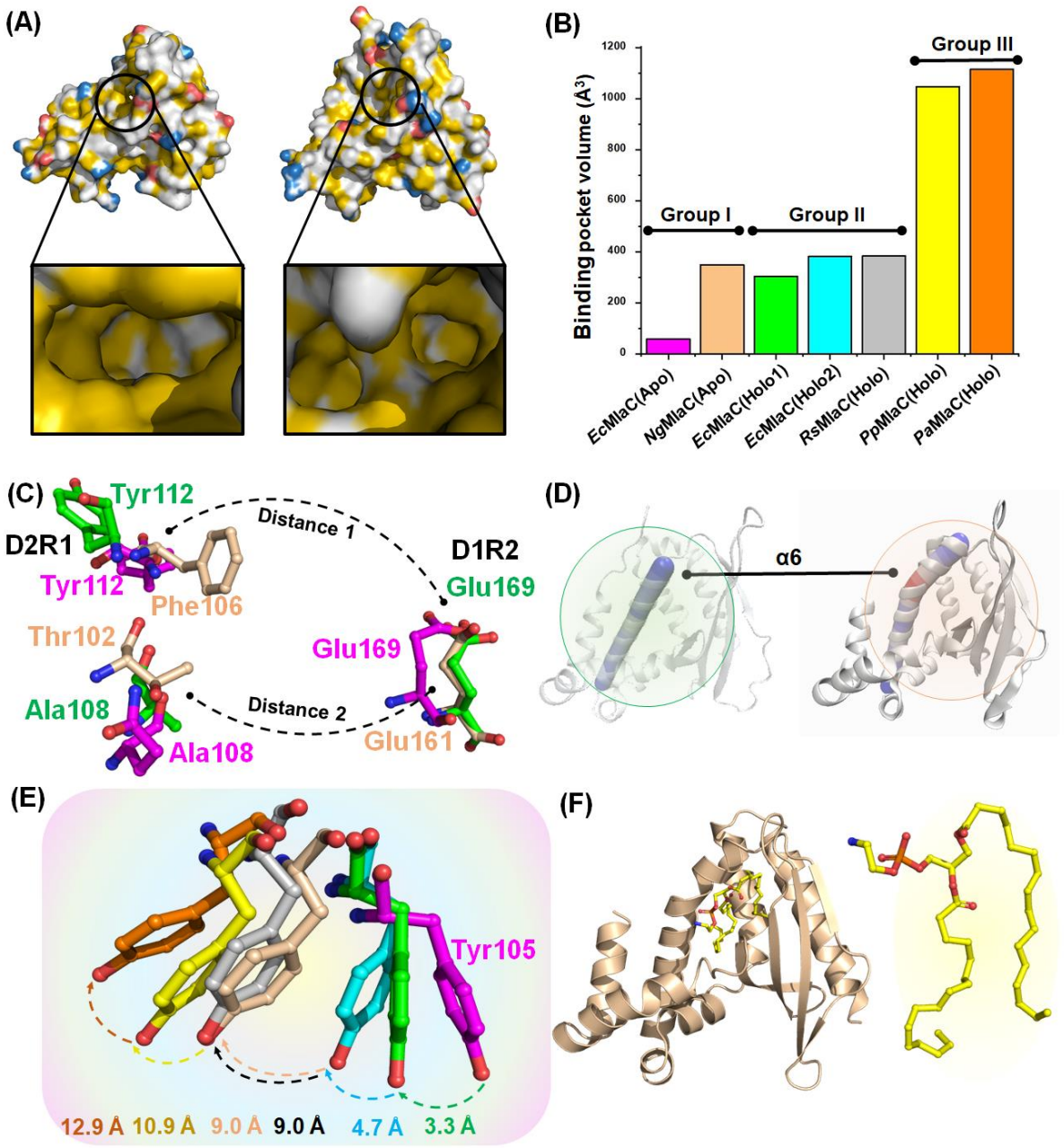


Figure 6.2. Structural analysis of EcMlaC and its orthologues. *(A) (Top, bottom) Hydrophobicity mapping of NgMlaC(Apo) and EcMlaC(Apo), respectively. (B) Volume analysis of EcMlaC and its orthologues. Histogram depicting the binding-pocket volumes of different EcMlaC_H3_holo1 and its orthologs. The apo, single-liganded and double-liganded MlaC species have been designated as Group I, II and III, respectively. (C) Movement of the residues from the D2R1 and D1R2 subdomains modulating the opening and closing of EcMlaC(Apo)(Magenta), EcMlaC(Holo1)(Green) and NgMlaC(Apo)(Wheat). (D) (Left, right) Analysis of helix curvatures of α6 helix of EcMlaC(Holo1) and NgMlaC(Apo). The curvature has been color-coded using the blue-white-red scheme. As per the scheme, blue, white, and red*

color signify straight curvature and angle, respectively. (E) Outward movement of Tyr105 residues in EcMlaC orthologues with respect to EcMlaC(Apo)(Magenta). The residues in the orthologues EcMlaC(Holo1)(Green), NgMlaC(Apo)(Wheat), RsMlaC(Holo)(Grey), PpMlaC(Holo)(Yellow) and PaMlaC(Orange), respectively are represented as ball-and-stick models. The distance moved by the residues with respect to EcMlaC(Apo) is mentioned below. (F) Molecular docking of NgMlaC(Apo) and PL. (Left) The docked structure of NgMlaC(Apo) and PL. (Right) Conformation of PL in the docked structure.

The analysis also reveals that *Ng*MlaC(Apo) possesses a similar segmented domain arrangement (D1R1, D2R1, D1R2 and D2R2) as observed in *Ec*MlaC and its orthologues (Figure D2A-D2B, Table D3). We have previously reported that the increase in distance between the residues Ala108 (D2R1) & Glu169 (D1R2) (Distance 1) and Tyr112 (D2R1) & Glu169 (D1R2) (Distance 2) in *Ec*MlaC results in the opening of the binding pocket (Dutta and Kanaujia, 2022). Hence, the distance between the corresponding residues Thr102 & Glu161 (Distance 1) and Phe106 & Glu161 (Distance 2) in *Ng*MlaC(Apo) were determined and compared with *Ec*MlaC(Apo) and *Ec*MlaC(Holo1). In the case of *Ec*MlaC(Apo), Distance 1 and 2 are 10.2 Å and 13.9 Å, respectively, whereas in the case of *Ec*MlaC(Holo1), they are 14.8 Å and 17.8 Å, respectively. In *Ng*MlaC(Apo), both distances are found to be 15.9 Å (Figure 6.2C). The increase in the subdomain distance was reported to be caused by ligand binding (Huang et al., 2016; Dutta and Kanaujia, 2022). However, this analysis shows that even in the absence of a ligand, an increase in the distance can take place. Additionally, the C-terminal region of the α6 helix in *Ng*MlaC(Apo) possesses a kink that is not observed in *Ec*MlaC(Holo1) (Figure 6.2D). Such a feature has been observed in the members of Group III (Dutta and Kanaujia, 2022). This causes an increase in Distances 1 and 2, leading to the outward movement of L6. Also, there is a distinct change in the curvature of the D1R2 subdomain (β1-β4 strands) and the inward movement of L5. Such movements have been previously reported to contribute to the increase in the binding pocket volume. Previously, we have reported the increase in the outward movement of the conserved Tyr105 residue in the MlaC orthologues is proportional to the opening of the binding pocket (Dutta and Kanuajia, 2022). The analysis shows that this residue in *Ng*MlaC(Apo) makes a movement of 9 Å, which is

comparable to that of Group II members. In fact, it is equal to *Ng*MlaC(Apo) and they both have comparable binding pocket volumes (Figure 6.2E). This again indicates that *Ng*MlaC(Apo) is in an open unliganded state. In order to further validate this, a molecular docking analysis was made wherein the PEF molecule (bound to *Ec*MlaC(Holo1)) was docked with *Ng*MlaC(Apo). As anticipated, the PEF was able to dock into the binding pocket of *Ng*MlaC(Apo), thus, confirming that the protein is indeed in an open unliganded state (Figure 6.2F).

6.3.2. *Ec*MlaC orthologues can bind different ligands in varying conformations

The binding pocket of MlaC is hydrophobic in nature and can accommodate the PL tails in different conformations. On the hand, the head regions of the PLs remain solvent (Dutta and Kanaujia, 2022). Owing to such an arrangement of PLs in the binding pocket, it has been speculated that MlaC would be able to bind a wide range of PLs (Ekiert et al., 2017). However, an in-depth study exploring the ligand binding capability is still lacking. Also, the presence of two PLs in case *Pa*/*Pp*MlaC(Holo) and one PL in case of *Ec*MlaC(Holo1)/*Ec*MlaC(Holo2)/*Rs*MlaC(Holo) raise questions regarding the ability of *Ec*MlaC and its orthologues to bind more than one ligand. Hence, to address these questions, an extensive molecular docking study was conducted where a total of 104 PLs were docked with selected *Ec*MlaC orthologues (Table D4). For the study, *Ec*MlaC(Apo), *Ec*MlaC(Holo1) and *Pa*MlaC(Holo) were taken as representative of Groups I, II and III, respectively and the analyses were performed. For the docking experiments, the ligands were removed from *Ec*MlaC(Holo1) and *Pa*MlaC(Holo) (referred to as *Ec*MlaC(Apo*) and *Pa*MlaC(Apo*), respectively). The ligand PEF was unable to bind to *Ec*MlaC(Apo). In fact, even smaller PLs, such as B3H, could not enter the binding pocket and tend to show a non-specific binding. This is observed for the rest of the ligands (Figure 6.3A). On the other hand, PEF was able to bind to *Ec*MlaC(Apo*) along with other ligands (Figure 6.3B). In the case of *Pa*MlaC(Apo*), the ligand GOT was able to bind in two different planes (Figure 6.3C-6.3D). Similar results were obtained for other ligands. Both *Ec*MlaC(Apo*) and *Pa*MlaC(Apo*) were able to bind most of the ligands, the latter being able to accommodate PLs in different orientations. This indicates that *Ec*MlaC and its orthologues are multi-specific in nature. In all the cases, the tails of the PLs remain embedded in the binding pocket while the head groups remain solvent exposed. The inability to bind PLs in a different orientation in the case of *Ec*MlaC(Apo*) can be

attributed to sits reduced binding pocket. *Ec*MlaC and *Pa*MlaC have been previously reported to bind to cardiolipin (CL) and PLs (two copies), respectively (Hughes et al., 2019; Yero et al., 2021). As the CL molecule is equivalent to two PLs, a docking experiment was set up for CL (PDB id: CDL) along with 5PL (with three acyl chains) with *Ec*MlaC(Apo*) and *Pa*MlaC(Apo*) (Figure D3A-D3B). Surprisingly, none of the proteins were unable to bind to CDL and 5PL. As per analysis, the binding is energetically unfavourable. However, it can be speculated restrictions in the torsional angle (in Autodock) might have also contributed to this.

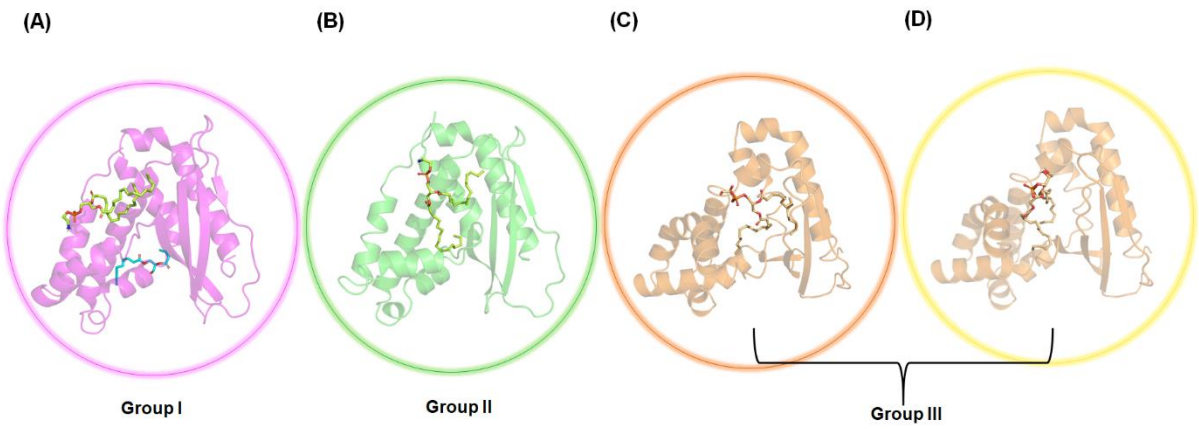


Figure 6.3. Molecular docking analyses. *(A-B) Docking of PLs to EcMlaC (Apo/Apo*), respectively. (C-D) Docking of PL to PaMlaC(Apo*) in two different planes. The ligands PEF (limon), B3H(cyan) and GOT (lightorange) are represented as ball-and-stick models.*

6.3.3. Ligand binding stabilizes the backbone of MlaC

In order to visualize the dynamic behaviour of *Ec*MlaC and its orthologues in a solvent environment, the selected MlaC systems were subjected to MD simulations (Table D5). Based on the convergence of RMSD plots, the equilibration of the MD trajectories was studied for each frame with respect to the initial energy- structure for a period of 1000 ns. Surprisingly, the six systems showed significant variations in regard to their time of convergence (Table D6). For *Ec*MlaC(Apo) and *Ng*MlaC(Apo), the convergence was attained at 300 and 400 ns, respectively. On the other hand, *Ec*MlaC(Apo*) and *Ec*MlaC(Holo), the time of convergence was observed to be 500 ns. However, for both *Pa*MlaC(Holo) and *Pa*MlaC(Apo*), no convergence was observed as a whole, although

in the concluding 50 ns, the systems seemed to have stabilized. The comparison of the two systems reveals that *Ec*MlaC(Apo) and *Ng*MlaC(Apo) have average RMSD values in the range of ~1.00-2.00 Å and ~2.50-3.50 Å. Thus, between the two systems, *Ng*MlaC(Apo) demonstrates higher fluctuations. On the hand, the RMSD distributions of *Ec*MlaC(Holo) and *Ec*MlaC(Apo*) indicate that the former is comparatively stable with an RMDS range of ~1.50-2.25 Å while the latter has a range of ~2.50-3.00 Å. A similar pattern was observed in the case of *Pa*MlaC(Holo) and *Pa*MlaC(Apo*) but with larger fluctuations. *Pa*MlaC(Holo) has an RMSD range of ~1.50-3.00, while *Pa*MlaC(Apo*) possesses a range of ~4.00-5.00 Å. As a matter of fact, *Pa*MlaC(Apo*) shows a broad distribution pattern with multiple peaks, suggesting structural deviations during the simulation. Thus, in both the cases, *Ec*MlaC(Holo/Apo*) and *Pa*MlaC(Holo/Apo*), the unliganded systems demonstrate higher fluctuations as compared to the liganded systems (Figure 6.4A-6.4C). This suggests that ligand binding stabilizes the backbone of MlaC by restricting the motions. As a result, the open unliganded states are relatively less stable. A corollary of this analysis is that the greater the opening of the unliganded state, the higher the fluctuations and the lesser the stability.

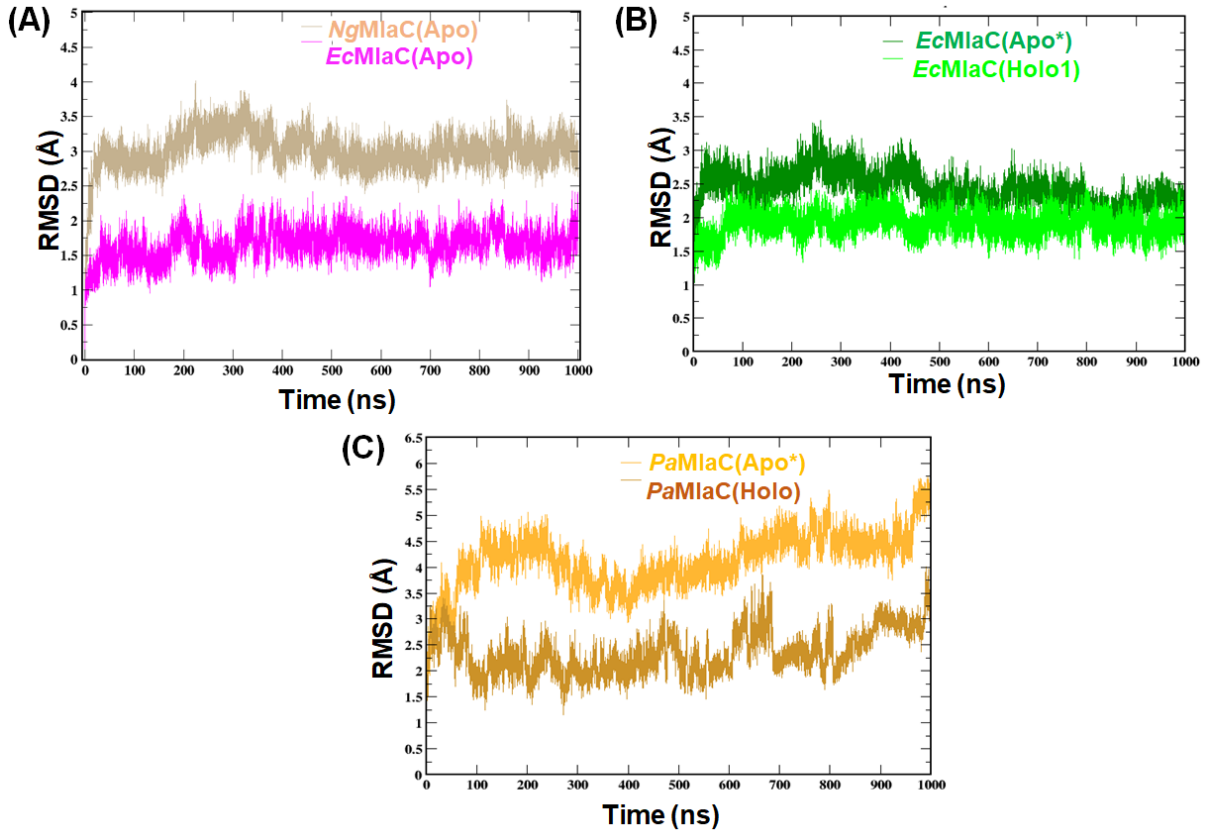


Figure 6.4. RMSD analysis of MlaC systems. *(A-C). RMSD plots of Ec/NgMlaC(Apo), EcMlaC(Apo*/Holo1) and PaMlaC(Apo*/Holo), respectively.*

6.3.4. The bound ligands demonstrate dynamic movements in MlaC

Typical SBPs accommodate ligands in the cleft between NTD and CTD (Fukami-Kobayashi et al., 1999). The orientations and conformations of the ligands are extremely specific and are guided by the structural features of the ligands as well as the interaction with the SBP (Chandravanshi et al., 2020). Interestingly, the orientation, as well as the conformation of the PLs present in the *Ec*MlaC and the different liganded states of its orthologs, vary (Figure D4A-D4E). In order to get further insight into the PL movement, the simulation of *Ec/Pa*MlaC(Holo1/Holo) was monitored every 100 ns till the completion of the simulation. In the case of *Ec*MlaC(Holo1), at 100 ns, the PEF molecule makes a rightward shift. This causes a significant change in the conformation of the head as well as the tail regions. With the progression of the simulation, the conformational changes continued with the most prominent movements observed in the head region. It seems that the change in the conformation of the PL molecule inside the cavity is based on its length and complexity, in addition to cavity plasticity (Figure D5A). A similar observation was made in the case of *Pa*MlaC(Holo), where the two PLs demonstrated a change in the conformation. However, the change is most observed in the head regions (Figure D5B). The RMSD analysis further supports this, which clearly shows that the distribution is more fluctuating in the tail regions of PEF of *Ec*MlaC(Holo1). In the case of *Pa*MlaC(Holo), the RMSD fluctuations are comparatively less in the tails and most prominent in the head regions (Figure 6.5A-6.5C). It can be speculated that high RMSD in case PEF is due to its unusual conformation and its tendency to attain a better state within the cavity of MlaC. Thus, the analysis clearly highlights that bound PEF would exhibit dynamic movements within the cavity of MlaC.

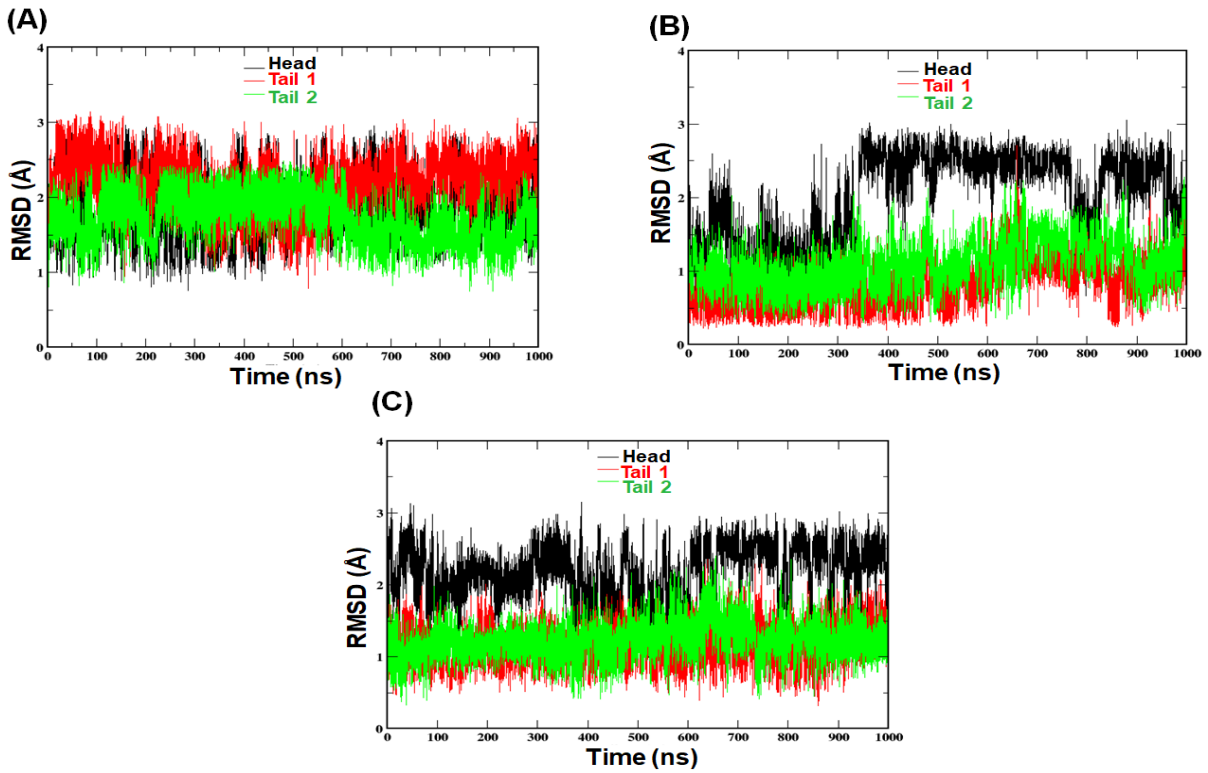


Figure 6.5. RMSD analysis of PLs bound to MlaC systems. *(A-C) RMSD plots of PEF (bound to EcMlaC(Holo1)), GOT (bound to PaMlaC(Holo)) and H3T (bound to PaMlaC(Holo)).*

6.3.5. Opening and closing of the phospholipid-binding site is modulated by D2R1 and D1R2 subdomains aided by angular motions

We have previously reported that the increase in distance between the residues Ala108 (D2R1) & Glu169 (D1R2) (Distance 1) and Tyr112 (D2R1) & Glu169 (D1R2) (Distance 2) in *Ec*MlaC results in the opening of the binding pocket. This change in the distance can be used as a marker to determine the state of the binding pocket (Dutta and Kanaujia, 2022). Hence, the corresponding residues in the other systems were determined and during the simulation run, the variations in the selected distances (Distance 1 and 2) were monitored (Table D7). The analysis clearly shows that upon ligand binding, the distances undergo significant fluctuations (Figure 6.6A-6.6E, Figure D6A).

On the other hand, Gly182 residue has been reported to cause an angular shift in H7, leading to the closing and opening of the binding pocket (Hughes et al., 2019). Additionally, we have already reported that H7 would impart conformational dynamicity

to MlaC (Dutta and Kanaujia, 2022). Hence, the angle formed by the residues Tyr112-Gly182-Glu169 was used as a marker and the change in the angular motion was monitored during the simulation run. Analysis of the crystal structures reveals that *Ec*MlaC(Apo), *Ng*MlaC(Apo), *Ec*MlaC(Holo) and *Pa*MlaC(Holo) possess angles of 23.9°, 29.5°, 34.6° and 30.4°, respectively. As observed in the case of Distances 1 and 2, the analysis shows that the ligand binding induces higher fluctuations in angular motions (Figure D6B, Figure D7A-D7B). Altogether, the study clearly shows that ligand binding induces increments of Distance 1 and 2 along with the increase in angular motions leading to the opening of the binding pocket.

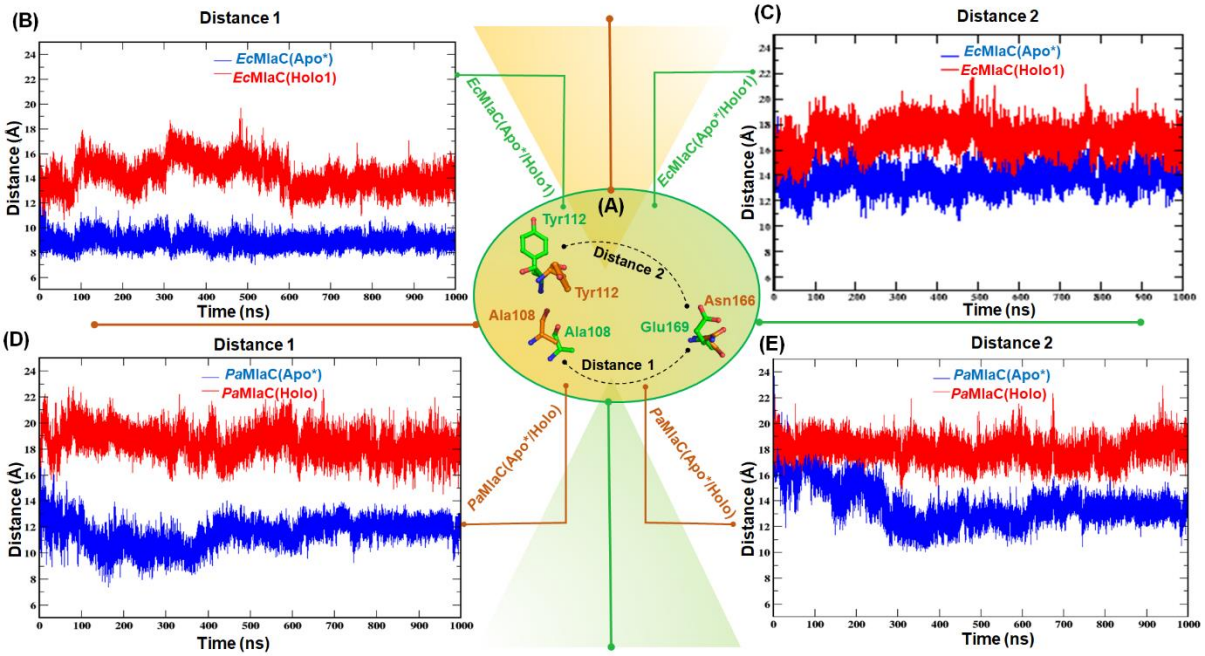


Figure 6.6. Analysis of the change in distance between residues around the binding pocket. *(A) Overview of Distance 1 and Distance 2 in EcMlaC(Holo1)(Green) and PaMlaC(Holo)(Orange). The residues of interest are represented as ball-and-stick models and curved dotted lines represent the change in distance. (B-C) Fluctuations of Distance 1 and 2 in EcMlaC(Holo1), respectively. (D-E) Fluctuations of Distance 1 and 2 in PaMlaC(Holo), respectively.*

6.3.6. The subdomain D2R1 is the most plastic among the other regions

Previously, the loops between β2 and β3, as well as β4 and β5, were reported to be flexible (Huang et al., 2016). However, the role of the subdomain still remained

spurious. In order to get insight into local motions, the RMSF analysis of the selected system was performed. Surprisingly, contrary to the previous report, all the subdomains showed fluctuations. However, the comparison of *Ec*MlaC(Apo) and *Ng*MlaC(Apo) reveals that RMSF fluctuations are more in *Ng*MlaC(Apo). Also, in both cases, sharp peaks can be observed at the D2R1 subdomain. A similar pattern of Cα fluctuations is found in the other systems, with the major peaks confined to the D2R1 subdomain (Figure 6.7A-6.7E). This implies that the plasticity of this region allows MlaC to alter its shape in order to accommodate the PL.

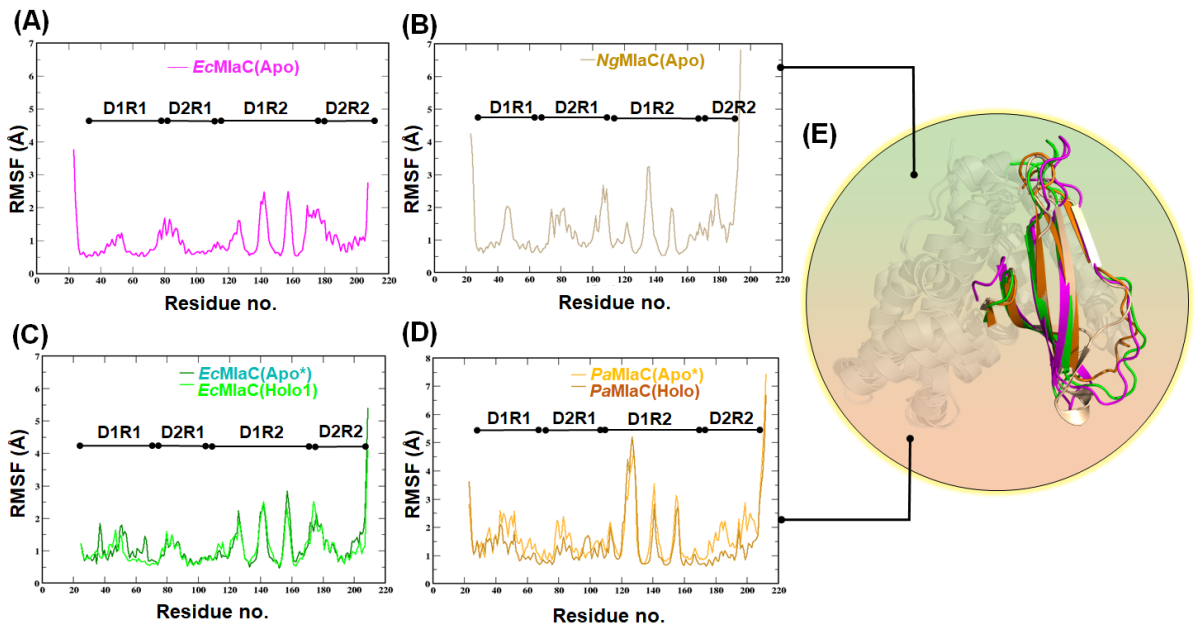


Figure 6.7. RMSF analysis of MlaC systems. *(A-D) RMSF plots of EcMlaC(Apo), NgMlaC(Apo), EcMlaC(Apo*/Holo1) and PaMlaC(Apo*/Holo), respectively. (E) Structural superimposition of EcMlaC(Apo)(Magneta)), NgMlaC(Apo)(Wheat), EcMlaC(Holo1)(Green) and PaMlaC(Holo)(Orange). The D1R2 subdomain is highlighted in colours, while the rest of the subdomains are in grey.*

6.3.7. Unliganded MlaC systems exhibit more motions

The RMSF analysis clearly shows that all the sub-domains of *Ec*MlaC and its orthologues are extremely dynamic. In order to get better insight, a holistic approach of DCCM plots was utilized to identify different types of intrinsic motion of the proteins that occur in different states. This would help us understand the conformational

changes associated with the dynamic motions of the segmented domains. A previous report has highlighted that α6 and loops would contribute to the change in the conformations of the binding pocket via correlated motions (Huang et al., 2016). However, the DCCM plots of the selected systems show that a number of correlated and anti-correlated motions contribute to the dynamics of the systems. The comparison of the apo and holo forms reveals that larger correlated and anti-correlated motions are associated with the former. It might be because upon binding with the ligand, the motions of the protein got restricted, due to which the motions got reduced. In the case of *Ec*MlaC(Apo) and *Ng*MlaC(Apo), two important correlated and anti-correlated motions are observed. However, the matrix is much more dynamic in the case of *Ng*MlaC(Apo) (Figure D8A-D8B).

In the case of *Ec*MlaC(Apo*), two correlated and three anti-correlated motions are observed. In the case of *Ec*MlaC(Holo1), three correlated motions along with one anti-correlated motion are present. The analysis shows that one D1R1-D1R2 motions are completely absent in *Ec*MlaC(Holo1) (Figure 6.8A-6.8B). A similar trend of reduced motions was seen in the case of *Pa*MlaC(Holo). In fact, in the case of *Pa*MlaC(Holo), only the D2R2-D2R2 motion was observed, which can be seen in all the systems. On the other hand, in the case of *Pa*MlaC(Apo*), two correlated and anti-correlated motions are observed (Figure D8C-D8D). The D2R2-D2R2 correlated motion corresponds to a twisted movement of the α7 and α8 helix. As the α7 helix has been reported to play a significant role in the opening and closing of the protein, it can be speculated that this twisting motion would lead to the conformational dynamicity of the MlaC system aiding in ligand binding. Furthermore, the analyses firmly establish that ligand binding restrains the conformational landscape of MlaC by reducing the intrinsic motions. This is in accordance with the results obtained through the RMSD study.

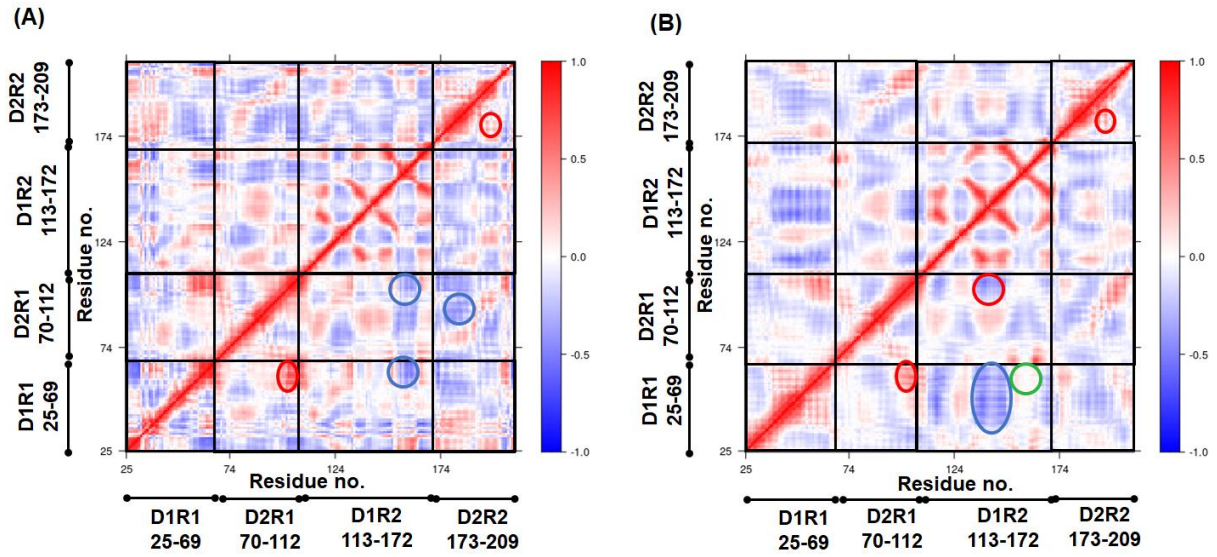


Figure 6.8. Correlation maps of EcMlaC systems. *(A-B) Correlation map of EcMlaC(Apo*) and EcMlaC(Holo1), respectively. The change in the colour bar from blue to red signifies the transition from anti-correlated to correlated motions. The regions encircled with blue and red circles represent anti-correlated and correlated motions of interest. The regions encircled in green represent the absence of any motion in the corresponding regions. The maps are divided into quadrants that correspond to the four subdomains (D1R1, D2R1, D1R2, and D2R2) as per the position of the residues. The boundaries of the subdomains are provided along the axes.*

6.3.8. Unliganded MlaC systems have larger conformational spaces

To study the important conformational changes, PCA was applied to the simulated systems. The first two PCs (PC1 and PC2) of all systems were considered in order to extract the maximum variance in all-atom displacements during the simulations. The global energy minima conformations are designated in blue. The size and shape of the minimal energy area signify the stability of systems. Since the thermodynamically most stable structure resides in a minimum free-energy (ΔG) surface, more concentrated blue areas indicate more stability of the corresponding structures during MD simulation (Pontiggia et al., 2015). As per the analysis, *Ec*MlaC(Apo) spans a small area of conformational space having three minima (I-III, high-density regions or low energy funnels) at different time frames (~66 ns, ~575 ns and ~750 ns). Among them, the deepest minima are obtained at ~750 ns (III) (Figure D9A). In order to acquire in-depth

insight, the conformations extracted from all the minima were superimposed to observe the structural variations. After extracting the structures, the path of the trajectory connecting the minima and the sub-conformational spaces of the systems from the free energy landscape along PC1 and PC2 was inspected. The minima are numbered as per the appearance of the basins with respect to time during simulation and porcupine plots were made that reflect the atomic motions in the same or opposite directions. As per the study, the subdomains D2R1 (α6), D1R2 (β1-β4) and D2R2 (α7-α8) exhibit conformational changes. Furthermore, α7 and α8 helices exhibit a twister motion, as observed in the DCCM analysis. These observations are further evidenced by the porcupine plots that clearly highlight the observed motions (Figure D9A). In contrast, *Ng*MlaC(Apo) confines a larger area of conformational space having five minima (I-V; ~53 ns, ~226 ns, ~560 ns, ~731 ns and ~838 ns, respectively) and the deepest minima are observed at ~225 ns. Structural alignment of the extracted conformations reveals that all the sub-domains except for a small region exhibit variation (α4-α5). These movements were again confirmed by the porcupine plots (Figure D9B).

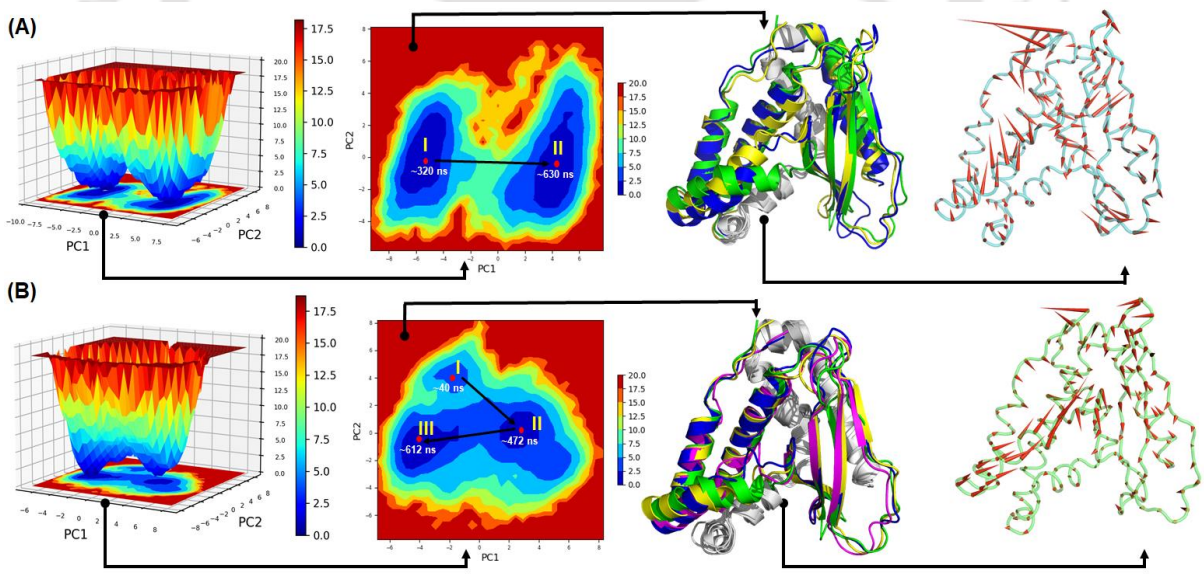


Figure 6.9. Free energy landscape (FEL) analysis and porcupine plots of EcMlaC systems. *(A-B) FEL plots and structural analysis of EcMlaC(Apo*) and EcMlaC(Holo), respectively. (First) 3D FEL between PC1 (X-axis) and PC2 (Y-axis). (Second) 2D FEL between PC1 (X-axis) and PC2 (Y-axis). The low-energy minima states of the systems and the paths connecting these minima are marked in red dots (the time frames are*

mentioned below) in the 2D FEL plots. The colour bar in the first and second panels represents the Gibbs free energy (kcal mol^{-1}) ranging from the lowest energy (blue) to the highest (red) conformational states. (Third) Structural superimposition of minima extracted from FEL plots. For EcMlaC(Apo), conformations from two minima (I, II; yellow and blue, respectively; green, reference structure) and for EcMlaC(Holo), conformations from three minima (I-III; yellow, blue and magenta, respectively; green, reference) were superimposed. The regions showing movements are coloured, while the rest of the regions are grey. (Fourth) Porcupine plots of the systems. The red arrows represent the direction of motion.*

On the other hand, *Ec*MlaC(Apo*) confines a large area of conformational space, having two distinct large minima at ~320 ns (I) and ~629 ns (II). A comparison of the extracted structures shows that all the sub-domains, with the exemption of a few regions (α2, α3, α4 and α5), exhibit movements. Whereas, *Ec*MlaC(Holo1) possesses a smaller area of conformational space having three minima (I-III; ~40 ns, ~472 ns and ~612 ns, respectively), but the deepest minima were observed at ~612 ns (III). Structural alignment of the conformations obtained from minima reveals that compared to *Ec*MlaC(Apo*), the movements have decreased in *Ec*MlaC(Holo1) with a lower number of regions (α6, β1-β4, α7 and α8) participating in the movement (Figure 6.9A-6.9B).

On the hand, *Pp*MlaC(Apo*) also confines are large conformational spaces, having seven minima (I-VII; ~70 ns, ~150 ns, ~350 ns, ~565 ns, ~470 ns, ~565 ns, ~908 ns and ~988 ns, respectively) and the deepest was found at ~908 ns (VI). A structural comparison reveals that all the sub-domains exhibit movements, thus, contributing to such a large conformational space (Figure D10A). On the contrary, *Pa*MlaC(Holo) confines relatively small conformational space having five minima (I-V; ~61 ns, ~202 ns, ~622 ns, ~700 ns and ~ 984 ns, respectively) and the deepest minima were observed at ~202 ns (II) (Figure D10B). The analysis again shows that *Pa*MlaC(Holo) possesses reduced movements contributed by fewer regions (α6, β1-β4, α7 and α8). Interestingly, throughout the analysis, the twisted motion between α7 and α8 is observed as seen in the DCCM study (Figure D11A-D11F). Also, the unliganded

systems confine larger conformational spaces as compared to the liganded systems owing to conformational plasticity.

6.4. Discussion

MlaC serves as the SBP of the Mla system and is involved in the transport of PLs between the IM and the OM. Unlike typical SBPs that possess NTD-CTD arrangement, MlaC possesses a unique segmented domain arrangement, which consists of two domains – NTF2-like (D1) and PBP (D2). These two domains can be further divided into subdomains that are placed in a discontinuous fashion (D1R1, D2R1, D1R2 and D2R2. Through extensive analysis of available crystal structures, we have previously reported the novel "segmented domain movement (SDM)" mechanisms, which MlaC implements to capture ligands. However, there exists a dearth of knowledge regarding the ability of MlaC to bind different ligands. Also, the characterization of the atomic details of different global and local motions associated with MlaC has not been done to date. Thus, in this study, extensive molecular docking and MD simulation studies were conducted in order to get detailed insights into the binding affinity of MlaC along with the conformational landscape of the protein. For this, the crystal structure of *Ec*MlaC, which we had previously determined, was taken as a point of reference and analyses were performed.

6.4.1. MlaC can exist in different transient states

Based on the binding pocket volume and the presence of ligand(s), we have categorized *Ec*MlaC and its orthologues into three groups – I (unliganded, closed), II (single liganded, semi-open) and III (double liganded, open). After thorough analysis, we also proposed that ligand binding cause the opening of the binding pocket, which is contrary to typical SBPs. Thus, *Ec*MlaC and its orthologues would follow a reverse mechanism of binding pocket opening and would remain close in the unliganded state (Dutta and Kanaujia, 2022). A similar process has previously been observed in the case of the periplasmic molecular chaperone LolA, which belongs to the LolABCDE ABC transporter involved in ferrying lipoproteins from IM to OM (Oguchi et al., 2008). The reverse mechanism can be attributed to its extremely hydrophobic binding site

required for PLs accommodation. This is because, in the apo state, the binding site needs to be closed and shielded from the water molecules that have a substantial presence in the periplasm (Stock et al., 1977). It is achieved by dehydrating the binding site through intra-molecular contacts among the hydrophobic residues (Huang et al., 2016). MlaC orthologs have the tendency to co-purify with PLs, indicating their strong affinity for the ligand. The unliganded open state of *Ec*MlaC(Apo) was only attained by delipidation and refolding experiments (Hughes et al., 2019). Thus, the *Ng*MlaC(Apo) presents a unique case as it is in an unliganded open state. We have further proved that it is in the open state by performing a study docking. The authors have not mentioned how they were able to obtain this unliganded open state which biologically should be unstable and transient. Under such circumstances, it can only be speculated that *Ng*MlaC(Apo) represents a pre- or post-ligand binding state. Nevertheless, this highlights that MlaC orthologues have a large conformational space.

6.4.2. MlaC is multi-specific in nature and can bind to different ligands

The available crystal structures of MlaC orthologues suggest that it can bind different ligands in varying numbers (Ekiert et al., 2017; Yero et al., 2021; Dutta and Kanuajia, 2022). Such features have not been observed in the case of canonical SBPs. Owing to its ability to transport PL, MlaC has also been categorized as an LTP (Wong et al., 2019). Interestingly, LPTs are known to simultaneously accommodate more than one ligand (Madni et al., 2022). Furthermore, we have also previously reported that during double ligand binding, the binding plane of PLs would undergo a change in orientation (Dutta and Kanaujia, 2022). Our molecular docking studies conducted on *Ec*MlaC and its orthologues reveal that the proteins (Group II and III) can accommodate different PLs. However, the PLs are unable to enter the binding pocket of the member of Group I as it remains closed. Although this seems reasonable for large ligands like PL, which demand an open binding pocket, the inability of the smaller ligand to dock is surprising. In fact, smaller ligands tend to show a non-specific binding instead of entering the pocket. Based on this, it can be speculated that the size, as well as the structural features of PL, might have a filtering effect. Due to this, smaller ligands are unable to access the binding pocket even if they do not require a completely open binding pocket. The results also show that in the docked structures, the tails remain deeply buried while the head groups remain solvent exposed. However, the conformation of docked ligands varies extensively among the docked structures, indicating that *Ec*MlaC and its

orthologues can accommodate ligands in dynamic conformations. The analysis also reveals that the *Ec*MlaC (Group I) can dock ligands in only one plane. However, *Pa*MlaC (Group III) can bind ligands in two different planes – one resembling that of *Ec*MlaC and the other similar to that observed in the crystal structure of *Pa*MlaC. This clearly indicates that the opening of the binding pocket allows the ligand to change the orientation of the binding plane of the ligand. This is in accordance with our previously reported hypothesis (Dutta and Kanaujia, 2022).

6.4.3. MlaC is an extremely dynamic conformational landscape

The ability of PLs to be present in different conformations can be attributed to 'lipid gymnastics'. Owing to this, the PLs in non-membrane arrangement display conformational dynamicity (Neumann et al., 2017). As a result, in the binding pocket, the PLs can attain different conformations and orientations both in crystal and docked structures. This is further evidenced by the different conformations PLs attain as the simulation progresses. Such observations have been previously made for other LPTs (Madni et al., 2022). In order to accommodate such a dynamic molecule, the protein also needs to be extremely plastic. The simulation studies clearly show that *Ec*MlaC and its orthologues are dynamic in nature. Nearly all the subdomains show movements. In fact, the motion between D2R1 and D1R2 subdomains aided by angular motions modulate the opening and closing of the binding pocket. The subdomain motions are further highlighted by RMSF analysis which shows the D1R2 subdomain to be the most dynamic. The RMSD, DCCM and FEL analyses further helped identify global and local motions and one the most prominent being the motion between α7 and α8. The study clearly establishes that the unliganded (apo) state is more dynamic than the liganded (holo) state. Owing to the absence of any ligand in the binding pocket, the former is able to attain various conformations and is more flexible. Overall, the analysis not only validates the SDM mechanism but has also helped enrich the information related to protein motions.

6.4.4. MlaC exhibits a series of dynamic movements via the subdomains resulting in a unique ligand binding mechanism

Canonical SBPs generally follow one of the four main mechanisms of ligand binding – "Venus fly-trap'', "Asymmetric domain movement'', "One domain movement'' and "Subdomain movement'' mechanisms (Chandravanshi et al., 2020). These movements

are mainly executed by the motions of the NTD and CTD that leads to the entrapment of ligands in the cleft situated between the two domains (Fukami-Kobayashi et al., 1999). However, MlaC defies such structural as well as mechanistic similarities. MlaC has been categorized as an LTP. The ligand binding and transport process of LTPs is known to have multiple steps that are sequentially followed (Wong et al., 2017, 2019). Based on the analysis of the available crystal structures, we have reported the ligand binding mechanism of *Ec*MlaC and its orthologues as an SDM mechanism (Dutta and Kanaujia, 2022). To further substantiate this finding, we performed extensive structural analysis via MD simulations to understand the intricate molecular details of ligand binding as well as the transport mechanism followed by *Ec*MlaC and its orthologues. This has aided in putting forward a working model for the process of the ligand transport mechanism of MlaC. As per this model, the MlaC (apo) would approach the IM complex and eventually dock onto MlaD. At the time of docking, the complex would be in a nucleotide-free state (resting state) and the translocation pathway would have the required PL molecules that are meant to be transported. The docking of MlaC would induce MlaD hexamer to demonstrate the asymmetric upward lifting of three protomers, followed by the change in the geometry of the central channel. This would create the necessary force for the transfer of PL from MlaD to MlaC. The entrapment of PL by MlaC would require a series of structural changes which would include the outward movement of the D2R1 and D1R2 subdomains, which would be aided by the increased angular motions between the two subdomains. This would also involve correlated motion between α7 and α8, which would be followed by the entrapment of PL(s) and the release of MlaC (holo). The IM complex would then return to the resting state or bind to the ATP molecule to start the import cycle, as observed in the APM mechanism. MlaC (Holo) would approach the OM and dock with the MlaA-OMP complex. MlaC(Holo) would hand over the PL to the complex, thus, attaining the original state, MlaC (Apo) (Figure 6.10A-6.10F).

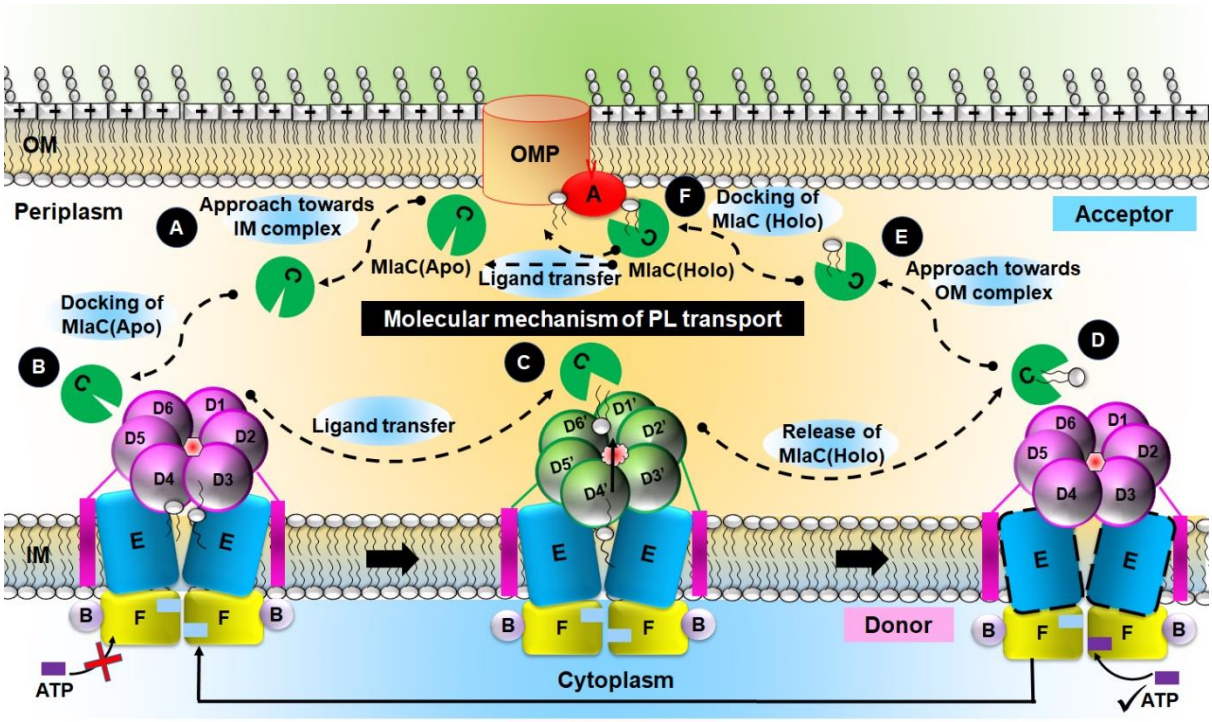


Figure 6.10. Proposed working model for PL transport by MlaC. *(A) MlaC(Apo) would approach the IM complex. (B) Docking of MlaC(Apo) upon MlaD of IM complex (nucleotide-free state). (C) Transfer of ligand from IM complex to MlaC. The MlaC and IM complex undergo conformational changes. (D) Release of MlaC(Holo). The IM complex returns to its resting state/binds to ATP. (E) The approach of MlaC(Holo) towards OM complex. (F) Docking of MlaC(Holo) upon MlaA-OMP complex. This is followed by the transfer of PL to the OM complex and MlaC(Holo) returns to MlaC(Apo).*

6.5. Conclusion

The study provides a comprehensive characterization of the ligand binding and transport mechanism *Ec*MlaC and its orthologues. The extensive analysis suggests that *Ec*MlaC orthologues would be multi-specific in nature and can bind to the protein in different conformations as well as orientations depending upon the opening of the binding pocket. This work not only validates the previous SDM mechanism but also enriches it by identifying new protein motions that are critical for ligand binding. To sum up, the study provided significant insight into the mechanism of action of *Ec*MlaC orthologues and firmly established them to be unique SBPs.

Appendix D. Supplementary data

Supplementary figures, Figure D1-D11.

Supplementary tables, Table D1-D6.

Summary

This study reports the computational as well as structural characterization of Mla proteins from the Mla system of *E. coli* and presents for the first time two distinct molecular mechanisms associated with the transport of PLs. The overall findings of the work are listed below.

- *In silico* analysis suggests that the MlaD domain possesses diverse domain architectures. Here we report a classification scheme wherein a total of 20 different (sub)groups (A–E) of MlaD domain architectures have been placed. We have also performed the profiling of the MlaD domain across these architectures. The results suggest that despite such architectural variations, the MlaD domain has conserved amino acid profiles marked by an abundance of glycine and hydrophobic residues and the lack of cysteine residues.
- A noticeable feature of the MlaD domain is the conserved N-terminal region which contains a conserved glycine residue that constitutes a consensus motif referred to as the MlaD domain motif across different architectures.
- The study further reveals the evolutionary relatedness of the MlaD domain of *Ec*MlaD to the MlaD domains of *M. tuberculosis* and *A. thaliana*. The analysis also hints at a synergetic association between Mla and Lpt systems.
- Both *Ec*MlaE and *Ec*MlaF possess unique evolutionary lineages, which are evident via their sequence and structural features. *Ec*MlaF-*Ec*MlaB association results in additional interfaces that contribute to the functioning of the Mla system. On the other hand, *Ec*MlaA possesses a ring-like structure, which resembles a doughnut and gives rise to a flexible yet convoluted central channel that aids in the transport of PLs. Association of *Ec*MlaA with *Ec*OmpC/F aids in the proper localization of the protein, while CTH might be critical for the maintenance of proper orientation.
- The work reports the crystal structure of *Ec*MlaC in a quasi-open state and in a complex with a PL at a resolution of 2.5 Å. A thorough structural analysis reveals that *Ec*MlaC possesses a unique segmented domain arrangement in which two domains, viz. nuclear transport factor 2-like (NTF2-like, D1) and phospholipid-binding protein (PBP, D2), remain in a discontinuous fashion (D1R1, D2R1, D1R2 and D2R2).
- The study also led to the identification of three different states of *Ec*MlaC and its orthologues (Group I-III). This, in turn, suggested that *Ec*MlaC would follow a reverse mechanism of binding-site opening when the protein would open upon ligand binding. This mechanism is contrary to that observed in the case of typical

SBPs. After extensive analysis, we propose for the first time a novel mechanism of ligand binding, viz. “segmented domain movement (SDM)’’, which involves a series of inherent structural changes in the subdomains of *Ec*MlaC in order to bind PL.

- The study also establishes *Ec*MlaC and its orthologues as non-canonical SBPs. This led us to introduce a new scheme of SBP classification comprising two classes – canonical SBPs (possessing NTD-CTD arrangement) and non-canonical SBPs (not possessing NTD-CTD arrangement).
- The work also reports the crystal structures of the periplasmic domain of *Ec*MlaD in three different space groups at a resolution range of 2.3-3.2 Å. *Ec*MlaD protomers consist of seven β-strands adopting a β-barrel fold (MlaD domain) and C-terminal α-helical domain (HD). Furthermore, *Ec*MlaD protomers form a homo-hexameric ring with a hydrophobic central channel and possess a six-fold symmetry. The channel serves as the translocation pathway for PLs and has non-uniform dimensions.
- In-depth structural analysis reveals that *Ec*MlaD would demonstrate an asymmetric upward movement of the protomers followed by a change in the geometry of the central channel. These conformational changes help in the transport of PL from *Ec*MlaD to *Ec*MlaC without the hydrolysis of the ATP molecule. Based on this, we have proposed another novel mechanism of PL transport. The ”asymmetric protomer movement (APM)’’ mechanism. This has, in turn, helped us put forward a working model for the ATP-independent export process of PL.
- The work also shows that *Ec*MlaD is structurally homologous to EF/AMT-type beta (6)-barrel and has a unique ancestry, unlike canonical SBPs. This has led us to further update the SBP classification scheme.
- *Ec*MlaC and its orthologues are multi-specific and can bind different PLs in varying conformations. Also, the PLs can change their orientations within the binding pocket in the presence of sufficient space.
- *Ec*MlaC orthologues have large and diverse conformational landscapes. In the unliganded states, *Ec*MlaC exhibits higher atomic motions. This is because ligand binding reduces the plasticity of the protein. Also, a new movement has been identified that involves a twisted motion α7 and α8. Based on the observations, a detailed model highlighting the molecular mechanism of PL transport by *Ec*MlaC has been put forward.

Altogether, the outcomes of this study provide significant insight into all the components of the Mla system. The characterizations performed through this work highlight the presence of structural and functional diversity among the various components of ABC transporters. The in-depth analysis of *Ec*MlaC and *Ec*MlaD firmly establishes the versatility of SBPs leading to the proposition of two novel mechanisms of ligand transport. To sum up, the work sheds a sufficient amount of light on the machinations of the Mla system.

Appendix A
Supplementary data to Chapter 3

Supplementary Tables, A1

Table A1.1. List of the MlaD domain architectures found in different proteins obtained from Pfam and InterPro databases. The MlaD domains are in default black text, while the conserved domains, families and sites are highlighted in red, green and purple, respectively.

Group	Architecture no.	Domain architecture	No. of copies of the MlaD domain(s)
A	1-8	MlaD x (1-8)	36
B-I	9	MlaD; Mce4_CUP1	1
	10	MlaD; RmuC	1
	11	MlaD; ABC_trans_aux	1
	12	MlaD; AP2	1
	13	MlaD; OmpA-like domain	1
	14	MlaD; TonB-dependent receptor-like, β-barrel	1
	15	MlaD; LXG	1
	16	MlaD; MCD_N	1
	17	MlaD; MobC	1
	18	MlaD; YkyA	1
	19	MlaD; FlgN	1
	20	MlaD; Cas_DxTHG	1
	21	MlaD; ABC transporter-like	1
	22	MlaD; Phytochrome, central region	1
	23	MlaD; Target SNARE coiled-coil homology domain	1
	24	MlaD; Forkhead-associated (FHA) domain	1
	25	MlaD; Tetratricopeptide repeat-containing domain	1
	26	MlaD; Chromosome partition protein MukF, middle domain	1
	27	MlaD; Bacteriophage lambda, GpH, tail tape measure, N-terminal	1
	28	MlaD; EF-hand domain	1
	29	MlaD; HAMP domain	1
	30	MlaD; Toluene tolerance Ttg2/ phospholipid-binding protein MlaC	1
	31	MlaD; ABC transporter permease MalE	1
	32	MlaD; Protein-export membrane protein SecD/ SecF, archaeal and bacterial	1
	33	MlaD; DUF2397	1
	34	MlaD; Protein of unknown function DUF2913	1

	35	MlaD; Protein of unknown function DUF2884	1
	36	MlaD; TRAF3-interacting protein 1, C-terminal domain	1
	37	MlaD; Gram-negative bacterial TonB protein C-terminal	1
B-II	38	Mce4_CUP1; MlaD	1
	39	NDUFA12; MlaD	1
	40	Membrane lipoprotein, lipid attachment site; MlaD	1
	41	ABC_tran; MlaD	1
	42	TRAP C4-dicarboxylate transport system permease DctM subunit; MlaD	1
	43	PDDEXK-like family of unknown function; MlaD	1
	44	PRC-barrel domain; MlaD	1
B-III	45	NDUFA12; MlaD; DUF2155	1
	46	Aminomethyltransferase, folate-binding domain; MlaD; Glycine cleavage T-protein, C-terminal barrel domain	1
	47	Peptidase M56; MlaD; Prepilin-type processing-associated H-X9-DG domain	1
B-IV	48	MlaD; Mce4_CUP1 x 2	1
	49	MlaD; DhaL domain; Mce4_CUP1	1
	50	MlaD; PAS domain; Mce4_CUP1	1
	51	MlaD; Septation ring formation regulator EzrA; Mce4_CUP1	1
	52	MlaD; Mce4_CUP1; MCD_N	1
	53	MlaD; DUF148; DUF1216	1
	54	MlaD; DNA recombination and repair protein RecA-like, ATP-binding domain; AAA+ ATPase domain	1
B-V	55	OTCace_N; OTCace; MlaD	1
B-VI	56	OTCace_ N ;OTCace ; MlaD; AP2/ERF domain	1
B-VII	57	MlaD; Pyruvate flavodoxin/ferredoxin oxidoreductase, N-terminal; Thiamine pyrophosphate enzyme, C-terminal TPP-binding; 4Fe-4S ferredoxin-type, iron-sulphur binding domain x 2	1
B-VIII	58	MlaE; MlaD; MlaC; STAS_2; BolA	1

C-I	59	MlaD; Mce4_CUP1; MlaD	2
C-II	60	MlaD x 2; Mce4_CUP1	2
C-III	61	PqiA; MlaD x 2	2
C-IV	62	MlaD; Mce4_CUP1; MlaD; Mce4_CUP1	2
C-V	63	PqiA; MlaD x 2; Methyltranf_PUA	2
C-VI	64	PqiA x 3; MlaD x 2; ABC_trans_aux; FdhD-NarQ	2
D-I	65	MlaD x 3; ABC_trans_aux	3
	66	MlaD x 3; Signal transduction histidine kinase, dimerisation/phosphoacceptor domain	3
D-II	67	PqiA; MlaD x 3	3
D-III	68	PqiA x 2; MlaD x 3	3
D-IV	69	PqiA x 2; MlaD x 3; ABC_trans_aux	3
E	70	PqiA x 2; MlaD x 5; Methyltr_RsmF_N; Methyltr_RsmB-F; Methyltranf_PUA	5

Note: 'x' followed by a number signifies the number of copies of a domain present adjacent to each other in the architecture.

Table A1.2. Pairwise sequence identity/similarity (query coverage) in percentage between the MlaD domains of proteins MlaD, PqiB and YebT from *E. coli*. Those alignments having 'not significant' score are mentioned as 'NS'.

	***Ec*MlA_ MlaD1**	***Ec*PqA_ MlaD1**	***Ec*PqA_ MlaD2**	***Ec*PqA_ MlaD3**	***Ec*YeA_ MlaD1**	***Ec*YeA_ MlaD2**	***Ec*YeA_ MlaD3**	***Ec*YeA_ MlaD4**	***Ec*YeA_ MlaD5**	***Ec*YeA_ MlaD6**	***Ec*YeA_ MlaD7**
***Ec*MlA_ MlaD1**	100	41/63 (22)	26/36 (60)	32/64 (23)	55/65 (30)	25/47 (67)	38/57 (62)	NS	NS	50/60 (6)	30/48 (65)
***Ec*PqA_ MlaD1**	-	100	NS	75/100 (3)	39/63 (100)	24/37 (67)	NS	16/46 (47)	28/39 (47)	32/49 (79)	28/48 (70)
***Ec*PqA_ MlaD2**	-	-	100	55/63 (20)	26/46 (69)	29/48 (95)	NS	21/47 (74)	30/45 (79)	NS	37/53 (98)
***Ec*PqA_ MlaD3**	-	-	-	100	31/51 (46)	67/78 (29)	NS	NS	NS	NS	32/48 (82)
***Ec*YeA_ MlaD1**	-	-	-	-	100	25/53 (57)	32/54 (100)	NS	31/47 (50)	30/51 (100)	29/59 (72)
***Ec*YeA_ MlaD2**	-	-	-	-	-	100	22/50 (90)	32/50 (95)	29/48 (72)	NS	35/53 (92)
***Ec*YeA_ MlaD3**	-	-	-	-	-	-	100	28/50 (74)	47/82 (14)	32/50 (69)	32/50 (75)
***Ec*YeA_ MlaD4**	-	-	-	-	-	-	-	100	26/62 (58)	23/48 (68)	29/52 (88)
***Ec*YeA_ MlaD5**	-	-	-	-	-	-	-	-	100	NS	33/54 (75)
***Ec*YeA_ MlaD6**	-	-	-	-	-	-	-	-	-	100	27/47 (72)
***Ec*Ye7_ MlaD7**	-	-	-	-	-	-	-	-	-	-	100

Table A1.3. List of TMD sequences used for the construction of phylogenetic tree.

S. No.	Protein name	UniProt id.	Organism name	Referred to as
1.	Intermembrane phospholipid transport system permease protein MlaE	P64606	*Escherichia coli K-12*	*Ec*MlaE
2.	Intermembrane transport protein PqiA	P0AFL9		*Ec*PqiA
3.	Intermembrane transport protein YebS	P0AD03		*Ec*YebS
4.	Conserved integral membrane protein YrbE1A	O07412	*Mycobacterium tuberculosis* H37Rv	*Mt* YrbE1A
5.	Conserved integral membrane protein YrbE1B	L0T2Q9		*Mt* YrbE1B
6.	Conserved hypothetical integral membrane protein YrbE2A	I6Y870		*Mt* YrbE2A
7.	Conserved hypothetical integral membrane protein YrbE2B	O07790		*Mt* YrbE2B
8.	Conserved hypothetical integral membrane protein YrbE3A	O53965		*Mt* YrbE3A
9.	Conserved hypothetical integral membrane protein YrbE3B	O53966		*Mt* YrbE3B
10.	Conserved integral membrane protein YrbE4A. Possible ABC transporter	O53546		*Mt* YrbE4A
11.	Conserved integral membrane protein YrbE4B. Possible ABC transporter	I6Y3P5		*Mt* YrbE4B
12.	Protein TRIGALACTOSYLDIACYLGLYCEROL 1, chloroplastic	Q8L4R0	*Arabidopsis thaliana*	*At*TGD1

Table A1.4. List of amino acid residues forming the PLP regions of the MlaD domain of the proteins MlaD, PqiB and YebT from *E. coli* and the respective diameters of the central pore openings.

S. No.	Domain name	Residues forming the PLPs (number of residues)	Diameter of central pore opening (Å)*
1	*Ec*MlA_MlaD1	G**L**LG (4)	16.3
2	*Ec*PqA_MlaD1	PQIGR**E**GISGLGTLLS (16)	39.5
3	*Ec*PqA_MlaD2	GIAVDLTS**A**GMRVEMG (16)	37.5
4	*Ec*PqA_MlaD3	TGN**L**VTGAL (9)	19.9
5	*Ec*YeA_MlaD1	KAS**L**AGVSGLDALVGG (16)	21.2
6	*Ec*YeA_MlaD2	VSGVDANVSI**S**GAKVKLESLA (21)	20.3
7	*Ec*YeA_MlaD3	NPKLS**L**SDANLSALLTG (17)	22.1
8	*Ec*YeA_MlaD4	SRVDVKVGLD (10)	**
9	*Ec*YeA_MlaD5	GGAKVQLNG**S**GLTVQASPLSRALKG (25)	28.4
10	*Ec*YeA_MlaD6	TPQIS**A**AGVEHLDTILQPY (19)	9.6
11	*Ec*YeA_MlaD7	GYSLDFGL**T**GGVVKTGTF (18)	20.6

*Diameter at the opening of the pore. The distance was measured between the C_α atoms of the amino acids (highlighted in bold), each belonging to an opposite protomer. **The diameter of the pore could not be measured due to the poor quality of the electron density for the PLP region.

Table A1.5. The amino acid residues participating in the dimeric interface of the homo-hexamer of the protein MlaD from *E. coli*.

Interface	**Residues (position in the MlaD domain) present at the interface of**	
	Protomer 1	**Protomer 2**
Protomer 1 - Protomer 2	Ile60(23), Gly61(24), Gly62(25), Val63(26), Val65(28), Arg89(52), Tyr90(53), Asn91(54), His92(55), Ile93(56), Ile101(64), Arg102(65), Thr103(66), Ser104(67), Gly105(68), Leu106(69), Leu107(70), Tyr111(74), Val116(79), Lys138, Met141, Leu146, Gln149, Phe150	Asp47(10), Asn48(11), Ile49(12), Gly50(13), Gly51(14), Ile71(34), Thr72(35), Leu73(36), Pro75(38), Lys76(39), Thr77(40), Tyr78(41), Leu79(42), Pro80(43), Leu106(69), Leu107(70), Gln110(73), Val142, Leu143, Glu144, Asp145, Ile147, Gly148, Phe150, Leu151, Tyr152
Note: The positions of the residues located within the MlaD domain are provided in parathesis. The protomers of the homo-hexamer of the protein MlaD are named A, B, C, D, E and F. Residue Gly51(14) is not present at the A-B and D-E interfaces. Residue Pro75(38) is not present at the A-B, C-D, D-E and F-A interfaces. Residue Lys76(39) is not present at the A-B and D-E interfaces. Residue Thr77(40) is not present at the B-C, C-D, D-E, E-F and F-A interfaces. Residue Leu79(42) is not present at the B-C, C-D, E-F and F-A interfaces. Residue Gly148 is not present at the C-D and F-A interfaces.		

Table A1.6. List of proteins interacting with MlaC and MlaD from *E. coli*. Only the top ten interacting proteins sorted by the best confidence score in decreasing order are listed.

Reference protein name	Details of the interacting protein	
	Gene	***Protein name**
MlaC	*mlaD*	Intermembrane phospholipid transport system binding protein MlaD
	mlaB	Intermembrane phospholipid transport system binding protein MlaB
	mlaA	Intermembrane phospholipid transport system lipoprotein MlaA
	mlaE	Intermembrane phospholipid transport system permease protein MlaE
	mlaF	Intermembrane phospholipid transport system ATP-binding protein MlaF
	ompF	Outer membrane porin F
	lptC	Lipopolysaccharide export system protein LptC
	ibaG	Acid stress protein IbaG
	macB	Macrolide export ATP-binding/permease protein MacB
	argA	Amino-acid acetyltransferase
MlaD	*mlaC*	Intermembrane phospholipid transport system binding protein MlaC
	mlaE	Intermembrane phospholipid transport system permease protein MlaE
	mlaF	Intermembrane phospholipid transport system ATP-binding protein MlaF
	mlaB	Intermembrane phospholipid transport system binding protein MlaB
	ibaG	Acid stress protein IbaG
	mlaA	Intermembrane phospholipid transport system lipoprotein MlaA
	yebT	Intermembrane transport protein YebT
	pqiB	Intermembrane transport protein PqiB
	murA	UDP-N-acetylglucosamine 1-carboxyvinyltransferase
	lptC	Lipopolysaccharide export system protein LptC

*Protein names are extracted from the UniProtKB database.

Table A1.7. List of TMDs analyzed in the study.

Sl. No.	Protein name	Source organism	UniProtKb accession no.	Length	Annotation
1	Intermembrane phospholipid transport system permease protein MlaE	*Escherichia coli* (strain K-12)	P64606	260	*Ec*MlaE
2	ABC transporter permease (YrbE1A)	*Mycobacterium smegmatis* (strain mc^{2}155)	A0QNR0	266	*Ms*YrbE1A
3	ABC-transporter integral membrane protein (YrbE1B)		A0QNR1	289	*Ms*YrbE1B
4	ABC transporter permease (YrbE3A)		A0QPB9	256	*Ms*YrbE3A
5	ABC transporter permease (YrbE3B)		A0QPC0	272	*Ms*YrbE3B
6	ABC transporter permease (YrbE4A)		A0R4P0	254	*Ms*YrbE4A
7	TrnB2 protein (YrbE4B)		A0R4N9	287	*Ms*YrbE4B
8	ABC transporter permease (YrbE5A)		A0QW85	270	*Ms*YrbE5A
9	ABC transporter permease (YrbE5B)		A0QW86	281	*Ms*YrbE5B
10	ABC transporter permease (YrbE5bA)		A0R1L4	257	*Ms*YrbE5bA
11	ABC transporter permease (YrbE5bB)		A0R1L3	282	*Ms*YrbE5bB
12	ABC-transporter integral membrane protein (YrbE7A)		A0QRJ7	289	*Ms*YrbE7A
13	ABC-transporter integral membrane protein (YrbE7B)		A0QRJ8	285	*Ms*YrbE7B
14	Conserved integral membrane protein YrbE1A	*Mycobacterium tuberculosis* (strain H37Rv)	O07412	265	*Mt*YrbE1A
15	Conserved integral membrane protein YrbE1B		L0T2Q9	289	*Mt*YrbE1B
16	Conserved hypothetical integral membrane protein YrbE2A		I6Y870	265	*Mt*YrbE2A
17	Conserved hypothetical integral membrane protein YrbE2B		O07790	295	*Mt*YrbE2B
18	Conserved hypothetical integral membrane protein YrbE3A		O53965	265	*Mt*YrbE3A
19	Conserved hypothetical integral membrane protein YrbE3B		O53966	271	*Mt*YrbE3B

20	Conserved integral membrane protein YrbE4A. Possible ABC transporter		O53546	254	*Mt*YrbE4A
21	Conserved integral membrane protein YrbE4B. Possible ABC transporter		I6Y3P5	280	*Mt*YrbE4B
22	Protein trigalactosyldiacylglycerol 1, chloroplastic	*Arabidopsis thaliana*	Q8L4R0	350	*At*TGD1
23	Intermembrane transport protein PqiA	*Escherichia coli* (strain K-12)	P0AFL9	417	*Ec*PqiA
24	Intermembrane transport protein YebS		P0AD03	427	*Ec*YebS
25	Lipopolysaccharide export system permease protein LptF		P0AF98	366	*Ec*LptF
26	Lipopolysaccharide export system permease protein LptG		P0ADC6	360	*Ec*LptG
27	Maltose/maltodextrin transport system permease protein MalF		P02916	514	*Ec*MalF
28	Maltose/maltodextrin transport system permease protein MalG		P68183	296	*Ec*MalG
29	Intermembrane phospholipid transport system permease protein MlaE	*Shigella flexneri*	P64609	260	*Sf*MlaE
30	Intermembrane phospholipid transport system permease protein MlaE	*Klebsiella pneumoniae*	W9BBH6	260	*Kp*MlaE
31	Intermembrane phospholipid transport system permease protein MlaE	*Acinetobacter baumannii*	V5V9F4	258	*Ab*MlaE
32	Intermembrane phospholipid transport system permease protein MlaE	*Yersinia pestis*	A0A0H2W819	260	*Yp*MlaE
33	Intermembrane phospholipid transport system permease protein MlaE	*Vibrio azureus*	U3CAX8	263	*Va*MlaE
34	Intermembrane phospholipid transport system permease protein MlaE	*Pseudomonas synxantha*	A0A5D3GB67	265	*Ps*MlaE

Table A1.8. List of NBDs analyzed in the study.

Sl. No.	Protein name	Source organism	UniProtKb accession no.	Length	Annotation
MkI NBDs (Gram-negative bacteria)					
1	Intermembrane phospholipid transport system ATP-binding protein MlaF	*Escherichia coli* (strain K12)	P63386	269	*Ec*MlaF
2	Putative ATP-binding component of a transport system	*Shigella flexneri*	A0A2S4N0X4	269	*Sf*_A0A2S4N0X4
3	ABC transporter ATP-binding protein	*Salmonella typhimurium*	A0A0D6IAP2	270	*St*_A0A0D6IAP2
4	Putative ABC transporter ATP-binding protein YrbF	*Klebsiella pneumoniae subsp. pneumoniae* (strain HS11286)	A0A0H3GY27	282	*Kp*_A0A0H3GY27
5	Phospholipid ABC transporter ATP-binding protein MlaF	*Yersinia pestis*	A0A3G5L1P8	272	*Yp*_A0A3G5L1P8
6	ABC transporter, ATP-binding protein	*Vibrio cholerae serotype O1 (strain M66-2)*	C3LRI4	267	*Vc*_C3LRI4
7	Intermembrane phospholipid transport system ATP-binding protein	*Haemophilus influenzae (strain KW20)*	P45031	264	*Hi*_P45031
8	ABC transporter ATP-binding protein	*Acinetobacter baumannii*	A0A086HZU3	272	*Ab*_A0A086HZU3
9	Phospholipid ABC transporter ATP-binding protein MlaF	*Legionella pneumophila*	A0A128YXY1	265	*Lp*_A0A128YXY1
10	Probable ATP-binding component of ABC transporter	*Pseudomonas aeruginosa (strain ATCC 15692)*	Q9HVW1	269	*Pa*_Q9HVW1
11	ABC transporter ATP-binding protein	*Francisella tularensis subsp. holarctica*	A0A0B3VVK2	267	*Ft*_A0A0B3VVK2

12	ABC transporter ATP-binding protein	*Neisseria gonorrhoeae*	A0A1D3HJR8	266	*Ng*_A0A1D3HJR8
Mkl NBDs (Actinomycetales)					
13	Probable ribonucleotide transport ATP-binding protein mkl	*Mycobacterium tuberculosis* (strain H37Rv)	P9WQL5	359	*Mt*_P9WQL5
14	ABC transporter, ATP-binding protein	*Mycolicibacterium smegmatis*	A0QS64	360	*Ms*_A0QS64
15	Probable ribonucleotide transport ATP-binding protein mkl	*Mycobacterium leprae* (strain TN)	P30769	347	*Ml*_P30769
16	Probable ribonucleotide transport ATP-binding protein mkl	*Mycobacterium bovis* (strain ATCC BAA-935)	P63358	359	*Mb*_P63358
17	ABC transporter domain-containing protein	*Mycolicibacterium paratuberculosis* (strain ATCC BAA-968)	Q73SE5	336	*Mp*_Q73SE5
18	Putative ABC transporter ATP-binding protein	*Streptomyces avermitilis* (strain ATCC 31267)	Q82B03	326	*Sa*_Q82B03
19	Putative ABC-transporter ATP-binding protein	*Streptomyces coelicolor* (strain ATCC BAA-471)	Q9KXZ3	343	*Sc*_Q9KXZ3
20	Probable bifunctional ABC transport system	*Rhodococcus jostii (strain RHA1)*	Q0SFA1	363	*Rj*_Q0SFA1
21	ABC superfamily, ATP-binding component		Q0SD37	346	*Rj*_Q0SD37
22	Putative ABC transporter ATP-binding protein	*Nocardia farcinica* (strain IFM 10152)	Q5YPC9	376	*Nf*_Q5YPC9
23	Putative ABC transporter ATP-binding protein	*Amycolatopsis mediterranei*	Q7BUF5	390	*Am*_Q7BUF5
24	Putative ABC transporter, ATP-binding protein	*Nocardia nova SH22a*	W5TKV8	333	*Nn*_W5TKV8
TGD3 NBDs (Plants)					

25	Protein trigalactosyldiacylglycerol 3, chloroplastic	*Arabidopsis thaliana*	Q9AT00	345	*At*_Q9AT00
26	Protein trigalactosyldiacylglycerol 3, chloroplastic	*Nicotiana attenuata*	A0A314KP13	372	*Na*_A0A314KP13
27	Protein TRIGALACTOSYLDIACYLGLYCEROL 3, chloroplastic	*Anthurium amnicola*	A0A1D1YCC8	348	*Aa*_A0A1D1YCC8
28	Protein TRIGALACTOSYLDIACYLGLYCEROL 3, chloroplastic	*Zea mays*	A0A3L6E1Y4	379	*Zm*_A0A3L6E1Y4
29	TGD3	*Urochloa ruziziensis*	A0A8A2IEG4	353	*Ur*_A0A8A2IEG4
Non-Mkl NBDs (*Escherichia coli*)					
30	Maltose/maltodextrin import ATP-binding protein MalK	*Escherichia coli* (strain K12)	P68187	371	*Ec*_P68187
31	Spermidine/putrescine import ATP-binding protein PotA		P69874	378	*Ec*_P69874
32	Taurine import ATP-binding protein TauB		Q47538	255	*Ec*_Q47538
33	Lipopolysaccharide export system ATP-binding protein LptB		P0A9V1	241	*Ec*_P0A9V1
34	High-affinity branched-chain amino acid transport ATP-binding protein LivG		P0A9S7	255	*Ec*_P0A9S7
35	Zinc import ATP-binding protein ZnuC		P0A9X1	251	*Ec*_P0A9X1
36	Iron(3+)-hydroxamate import ATP-binding protein FhuC		P07821	265	*Ec*_P07821
37	Phosphonates import ATP-binding protein PhnC		P16677	262	*Ec*_P16677

38	Cell division ATP-binding protein FtsE		P0A9R7	222	*Ec*_P0A9R7
39	Lipoprotein-releasing system ATP-binding protein LolD		P75957	233	*Ec*_P75957
40	Uncharacterized ABC transporter ATP-binding protein YbbA		P0A9T8	228	*Ec*_P0A9T8

Table A1.9. Analysis of the interfacial residues in the MlaF-MlaB complex.

Sl. No.	PDB	Interfaces	
1	6XBD	**I1**	**I2**
		MlaF (Chain I, F1)-MlaB (Chain K, B1)	MlaF (Chain J, F2)-MlaB (Chain L, B2)
		F1 & F2: Leu110, Thr114, Gln115, Leu116, Pro117, Pro119, Leu120, Ser123, Thr124, Met127, Lys128, Glu130, Ala13, Ile159, Ala160, Leu16, Glu162, Glu189, Leu190, Ala193, Leu194 **B1 & B2**: Gln22, Asp23, Leu25, Leu26, Trp29, Arg49, Asp5, Thr52, Gly53, Leu55, Ala56, Leu57, Leu59, His60, Lys80, Thr83, Leu84, Leu87, Tyr88, Asn89, Leu90	
		Average interface area (Å^2)	775
		I1'	**I2'**
		MlaF (Chain I, F1)-MlaB (Chain L, B2)	MlaF (Chain J, F2)-MlaB (Chain K, B1)
		F1 & F2: Arg255, Tyr26, His262, Leu265, Leu266, Pro267 **B1 & B2**: Leu59, His60, Asp63, Leu87, Tyr88, Asn89, Leu90, Pro91, Asp93, Val94	
		Average interface area (Å^2)	307
2	7D06	**I1**	**I2**
		MlaF (Chain B, F1)-MlaB (Chain C, B1)	MlaF (Chain E, F2)-MlaB (Chain F, B2)
		F1 & F2: His118, Thr119, Lys120, Leu121, Ser122, Asn124, Leu125, Glu128, Leu129, Leu132, Lys133, Glu135, Ser136, Val137, Gly138, Arg140, Leu166, Asp167, Ser194, Leu195, Glu197, Ala198, Leu199, Asp200 **B1 & B2**: Phe17, Ala20, Glu21, Tyr24, Asn50, Thr51, Leu52, Leu54, Ala55, Val56, Val58, Gln59, Arg62, Lys78, Lys81, Ile82, Ala85, Cys86, His87, Leu88	
		Average interface area (Å^2)	790
		I1'	**I2'**
		MlaF (Chain B, F1)-MlaB (Chain F, B2)	MlaF (Chain E, F2)-MlaB (Chain C, B1)

		F1 & F2: Tyr259, Gln260, Ala265, Tyr266, Leu267, Asp268, Asn269. Glu270, Val271 **B1 & B2**: Val58, Gln59, Leu6, Arg62, Lys67, Cys86, His87, Leu88, Asp9, Leu92	
		Average interface area (Å^2)	344
3	7CH9	**I1**	**I2**
		MlaF (Chain I, F1)-MlaB (Chain K, B1)	MlaF (Chain J, F2)-MlaB (Chain L, B2)
		F1 & F2: Asp82, Lys85, Leu110, His113, Thr114, Gln115, Leu116, Pro117, Glu119, Met120, Arg122, Asp123, Ile124, Met127, Lys128, Gln130, Ala131, Ile159, Leu161, Asp162, Gln164, Leu186, Leu189, Leu190, Asp192, Ala193, Leu194, Gly195 **B1 & B2**: Tyr22, Ser23, Gly25, Pro26, Arg29, Glu30, Arg53, Ser55, Ser56, Val57, Ile59, Ser60, Leu61, Leu63, Ala64, Arg67, Glu87, Ile88, Ala89, Lys90, Val91, Ser92, Ser93, Leu94	
		Average interface area (Å^2)	829
		I1'	**I2'**
		MlaF (Chain I, F1)-MlaB (Chain L, B2)	MlaF (Chain J, F2)-MlaB (Chain K, B1)
		F1 & F2: His255, Tyr261, Arg262, Asp264, Leu265, Leu266, Gly267 **B1 & B2**: Leu63, Ile66, Arg67, Arg70, Ser92, Ser93, Leu94, Glu96, Ile97	
		Average interface area (Å^2)	281

Table A1.10. List of STAS domain-containing proteins analyzed in the study.

.

Sl. No.	Protein name	Source organism	UniProtKb accession no.	Length	Annotation
STAS domain-containing proteins					
1	Anti-sigma F factor antagonist	*Bacillus subtilis* (strain 168)	P10727	117	*Bs*_P10727
2	Anti-sigma F factor antagonist	Bacillus stearothermophilus	O32726	116	*Bs*_O32726
3	C4-dicarboxylic acid transporter DauA	*Escherichia coli* (strain K12)	P0AFR2	111* (449-559)	*Ec*_P0AFR2
4	Putative anti-sigma factor antagonist TM_1442	*Thermotoga maritima* (strain ATCC 43589	Q9X1F5	110	*Tm*_Q9X1F5
5	Probable sulfate transporter Rv1739c	*Mycobacterium tuberculosis* (strain H37Rv)	P9WGF7	116* (442-557)	*Mt*_P9WGF7
6	Anti-sigma F factor antagonist	*Bacillus sphaericus*	O32723	117	*Bs*_O32723
7	Sulfate transporter 1.2	*Arabidopsis thaliana*	Q9MAX3	124* (522-645)	*At*_Q9MAX3
8	Chloride anion exchanger	*Homo sapiens*	P40879	196* (525-720)	*Hs*_P40879
9	Prestin	*Homo sapiens*	P58743	189* (525-713)	*Hs*_P58743
MlaB orthologs from Gram-negative bacteria					
1	Intermembrane phospholipid transport system binding protein MlaB	*Escherichia coli* (strain K12)	P64602	97	*Ec*MlaB
2	Intermembrane phospholipid transport system binding protein MlaB	*Shigella flexneri*	P64603	97	*Sf*_P64603

3	Lipid asymmetry maintenance protein MlaB	*Salmonella typhimurium* (strain SL1344)	A0A719DKL9	98	*St*_A0A719DKL9
4	Lipid asymmetry maintenance protein MlaB	*Klebsiella pneumoniae subsp. pneumoniae*	A0A4S7M9J7	95	*Kp*_A0A4S7M9J7
5	Lipid asymmetry maintenance protein MlaB	*Yersinia pestis*	A0A2S9PIQ5	100	*Yp*_A0A2S9PIQ5
6	Putative anti-sigma B factor antagonist	*Vibrio cholerae* serotype O1 (strain M66-2)	C3LRI0	109	*Vc*_C3LRI0
7	Intermembrane phospholipid transport system binding protein MlaB	*Haemophilus influenzae (strain KW20)*	Q57407	105	*Hi*_Q57407
8	Anti-sigma factor antagonist	*Acinetobacter baumannii*	V5V9K5	95	*Ab*_V5V9K5
9	STAS domain-containing protein	*Pseudomonas aeruginosa* (strain ATCC 15692)	Q9HVW5	102	*Pa*_Q9HVW5
10	STAS domain-containing protein	*Legionella pneumophila subsp. pneumophila* (strain Philadelphia 1 / ATCC 33152 / DSM 7513)	Q5ZX87	93	*Lp*_Q5ZX87
11	STAS domain protein	*Francisella tularensis subsp. holarctica*	A0A0B6EDN6	97	*Ft*_A0A0B6EDN6
12	STAS domain-containing protein	*Neisseria gonorrhoeae* 35/02	A0A7U8QES7	92	*Ng*_A0A7U8QES7

Appendix A

Supplementary data to Chapter 3

Table A1. 11. Available crystal strcutres of MlaA and OmpC/F complexes.

Sl. No.	PDB id	PDB description	Protein (Annotation)	Organism
1	5NUQ	Structural basis for maintenance of bacterial outer membrane lipid asymmetry	MlaA (*Sm*MlaA)	*Serratia marcescens*
			OmpF (*Ec*OmpF)	*Escherichia coli*
2	5NUP		MlaA (*Kp*MlaA)	*Klebsiella pneumoniae*
			OmpC (*Kp*MlaA)	
3	5NUO		MlaA (*Kp*MlaA)	*Klebsiella pneumoniae*
			OmpF (*Ec*OmpF)	*Escherichia coli*
4	5NUR		MlaA (*Kp*MlaA)	*Klebsiella pneumoniae*
			OmpF (*Ec*OmpF)	*Escherichia coli*

Supplementary Tables, A2

Table A2.1. Overview of the MlaD domain architectures.

Architecture group	Architecture no.	Protein name	Organism	Organism type	UniProt accession no.	No. of MlaD domains	Domain annotation	Length of the domain	Length of the protein
A	1	Intermembrane phospholipid transport system binding protein MlaD	*Escherichia coli* (strain K12)	Gram-negative bacteria	P64604	1	*Ec*MlA_MlaD1	80	183
	1	Mce-family protein Mce1B	*Mycobacterium tuberculosis* H37Rv	Actinobacteria	O07414	1	*Mt*1BA_MlaD1	79	346
	1	Mce-family protein Mce1C	*M. tuberculosis* H37Rv	Actinobacteria	O07415	1	*Mt*1CA_MlaD1	76	515
	1	Mce-family protein Mce1E	*M. tuberculosis* H37Rv	Actinobacteria	O07417	1	*Mt*1EA_MlaD1	77	390
	1	Mce-family protein Mce1F	*M. tuberculosis* H37Rv	Actinobacteria	L0T2W6	1	*Mt*1FA_MlaD1	77	515
	1	Mce-family protein Mce2B	*M. tuberculosis* H37Rv	Actinobacteria	O07788	1	*Mt*2BA_MlaD1	79	275
	1	Mce-family protein Mce2C	*M. tuberculosis* H37Rv	Actinobacteria	O07787	1	*Mt*2CA_MlaD1	76	481
	1	Mce-family protein Mce2D	*M. tuberculosis* H37Rv	Actinobacteria	I6WYT7	1	*Mt*2DA_MlaD1	77	508

	1	Mce-family protein Mce2E	*M. tuberculosis* H37Rv	Actinobacteria	I6Y461	1	*Mt*2EA_MlaD1	77	402
	1	Mce-family protein Mce3B	*M. tuberculosis* H37Rv	Actinobacteria	O53968	1	*Mt*3BA_MlaD1	78	342
	1	Mce-family protein Mce3C	*M. tuberculosis* H37Rv	Actinobacteria	O53969	1	*Mt*3CA_MlaD1	77	410
	1	Mce-family protein Mce3D	*M. tuberculosis* H37Rv	Actinobacteria	O53970	1	*Mt*3DA_MlaD1	77	423
	1	Mce-family protein Mce3E	*M. tuberculosis* H37Rv	Actinobacteria	O53971	1	*Mt*3EA_MlaD1	76	377
	1	Mce-family protein Mce3F	*M. tuberculosis* H37Rv	Actinobacteria	O53972	1	*Mt*3FA_MlaD1	77	437
	1	Mce-family protein Mce4B	*M. tuberculosis* H37Rv	Actinobacteria	I6X7G8	1	*Mt*4BA_MlaD1	78	350
	1	Mce-family protein Mce4C	*M. tuberculosis* H37Rv	Actinobacteria	I6YGB1	1	*Mt*4CA_MlaD1	77	357
	1	Mce-family protein Mce4E	*M. tuberculosis* H37Rv	Actinobacteria	I6Y3P1	1	*Mt*4EA_MlaD1	82	384
	1	Mce-family protein Mce4F	*M. tuberculosis* H37Rv	Actinobacteria	I6YC95	1	*Mt*4FA_MlaD1	77	564
	1	Protein Trigalactosyldiacylglycerol 2, chloroplastic	*Arabidopsis thaliana*	Plant	Q9LTR2	1	*At*TgA_MlaD1	78	381
	1	Mce related family protein	*M. ulcerans str. Harvey*	Actinobacteria	X8F9T8	1	*Mu*McA_MlaD1	78	151
	1	MlaD domain-containing protein	*Euryarchaeota archaeon*	Archaea	A0A2E5YUC0	1	*Ea*MlA_MlaD1	74	316

	2	ABC-type transport system involved in resistance to organic solvents, periplasmic component	*Rhodanobacter denitrificans*	Gram-negative bacteria	I4WVY1	2	*Rd*AbA_MlaD1	80	283
							*Rd*AbA_MlaD2	80	283
	3	Intermembrane transport protein PqiB	*E. coli* (strain K12)	Gram-negative bacteria	P43671	3	*Ec*PqA_MlaD1	93	546
							*Ec*PqA_MlaD2	94	546
							*Ec*PqA_MlaD3	106	546
	3	Uncharacterized protein HI_1672	*Haemophilus influenzae* (strain KW20)	Gram-negative bacteria	P44288	3	*Hi*UnA_MlaD1	92	881
							*Hi*UnA_MlaD2	86	881
							*Hi*UnA_MlaD3	63	881
	4	Paraquat-inducible protein B	*Photobacterium iliopiscarium*	Gram-negative bacteria	A0A0D8PPY1	4	*Pi*PiA_MlaD1	93	875
							*Pi*PiA_MlaD2	94	875
							*Pi*PiA_MlaD3	73	875
							*Pi*PiA_MlaD4	91	875
	5	Uncharacterized protein	*Mannheimia haemolytica serotype* A1/A6 str. PKL10	Gram-negative bacteria	A0A011PA02	5	*Mh*UnA_MlaD1	92	886
							*Mh*UnA_MlaD2	56	886
							*Mh*UnA_MlaD3	61	886

							*Mh*UnA_MlaD4	88	886
							*Mh*UnA_MlaD5	59	886
	6	Multivalent adhesion molecule 7	*Shewanella oneidensis (*strain MR-1*)*	Gram-negative bacteria	Q8EDY5	6	*So*MaA_MlaD1	93	878
							*So*MaA_MlaD2	90	878
							*So*MaA_MlaD3	90	878
							*So*MaA_MlaD4	98	878
							*So*MaA_MlaD5	93	878
							*So*MaA_MlaD6	99	878
	7	Intermembrane transport protein YebT	*E. coli (*strain K12*)*	Gram-negative bacteria	P76272	7	*Ec*YeA_MlaD1	93	877
							*Ec*YeA_MlaD2	88	877
							*Ec*YeA_MlaD3	83	877
							*Ec*YeA_MlaD4	74	877
							*Ec*YeA_MlaD5	111	877
							*Ec*YeA_MlaD6	91	877
							*Ec*YeA_MlaD7	85	877
	7	Mce-related protein			F9RRM4	7	*Vs*McA_MlaD1	91	877

			Vibrio scophthalmi LMG 19158	Gram-negative bacteria			*Vs*McA_MlaD2	60	877
							*Vs*McA_MlaD3	90	877
							*Vs*McA_MlaD4	64	877
							*Vs*McA_MlaD5	53	877
							*Vs*McA_MlaD6	78	877
							*Vs*McA_MlaD7	70	877
	8	Uncharacterized protein	*Motiliproteus* sp. MSK22-1	Gram-negative bacteria	A0A1R1LYR3	8	*Ms*UnA_MlaD1	92	1015
							*Ms*UnA_MlaD2	85	1015
							*Ms*UnA_MlaD3	93	1015
							*Ms*UnA_MlaD4	86	1015
							*Ms*UnA_MlaD5	93	1015
							*Ms*UnA_MlaD6	81	1015
							*Ms*UnA_MlaD7	92	1015
							*Ms*UnA_MlaD8	86	1015
B-I	9	Mce-family protein Mce1A	*M. tuberculosis* H37Rv	Actinobacteria	Q79FZ9	1	*Mt*1ABI_MlaD1	84	454

	9	Mce-family protein Mce1D	*M. tuberculosis* H37Rv	Actinobacteria	O07416	1	*Mt*1DBI_MlaD1	77	530
	9	Mce-family protein Mce2A	*M. tuberculosis* H37Rv	Actinobacteria	Q79FY7	1	*Mt*2ABI_MlaD1	84	404
	9	Mce-family protein Mce2F	*M. tuberculosis* H37Rv	Actinobacteria	O07784	1	*Mt*2FBI_MlaD1	77	516
	9	Mce-family protein Mce3A	*M. tuberculosis* H37Rv	Actinobacteria	L7N698	1	*Mt*3ABI_MlaD1	83	425
	9	Mce-family protein Mce4A	*M. tuberculosis* H37Rv	Actinobacteria	I6YC99	1	*Mt*4ABI_MlaD1	80	400
	9	Mce-family protein Mce4D	*M. tuberculosis* H37Rv	Actinobacteria	I6XHD6	1	*Mt*4DBI_MlaD1	77	451
	9	ABC transporter substrate-binding protein	*Streptomyces* sp. Tu 6176	Actinobacteria	A0A022ML59	1	*Ss*AbBI_MlaD1	77	417
	10	Mammalian cell entry related domain protein	*Caldithrix abyssi* DSM 13497	Gram-negative bacteria	H1XUM8	1	*Ca*McBI_MlaD1	77	294
	11	Putative ABC transporter substrate-binding protein	*Oligotropha carboxidovorans* (strain KCTC 32145)	Gram-negative bacteria	B6JE48	1	*Oc*AbBI_MlaD1	80	371
	12	AP2/ERF domain-containing protein	*Brassica oleracea var.oleracea*	Plant	A0A0D3CI28	1	*Bo*ApBI_MlaD1	78	652
	13	ABC-type transport system involved in resistance to organic solvents, periplasmic component	*Singulisphaera acidiphila* (strain DSM 18658)	Gram-negative bacteria	L0DA47	1	*Sa*AbBI_MlaD1	79	371

	14	Virulence factor Mce	*Megasphaera hexanoica*	Gram-negative bacteria	A0A344MGI1	1	*Mh*McBI_MlaD1	76	414
	15	Uncharacterized protein	*Candidatus Cloacimonetes bacterium* 4572_65	Unclassified	A0A1W9VNF8	1	*Cc*UnBI_MlaD1	78	301
	16	Uncharacterized protein	*Aeromicrobium* sp. PE09-221	Actinobacteria	A0A200H5R3	1	*As*UnBI_MlaD1	79	434
	17	Phospholipid/cholesterol/gamma-HCH transport system substrate-binding protein	*Halomonas ilicicola* DSM 19980	Gram-negative bacteria	A0A1M5AVQ8	1	*Hi*PlBI_MlaD1	81	295
	18	MlaD domain-containing protein	*Candidatus Melainabacteria bacterium* GWA2_34_9	Cyanobacteria	A0A1F6PFB4	1	*Cm*MlBI_MlaD1	78	300
	19	Phospholipid/cholesterol/gamma-HCH transport system substrate-binding protein	*Haloechinothrix alba*	Actinobacteria	A0A239A9M2	1	*Ha*PlBI_MlaD1	78	450
	20	Phospholipid/cholesterol/gamma-HCH transport system substrate-binding protein	*Fibrobacter* sp. UWB8	Gram-negative bacteria	A0A1M5EVA1	1	*Fs*PlBI_MlaD1	80	308
	21	ATP-binding cassette domain-containing protein	*Lactobacillus rhamnosus*	Gram-positive bacteria	A0A370MIM1	1	*Lr*PlBI_MlaD1	72	124

	22	Uncharacterized protein	*Starkeya novella*	Gram-negative bacteria	A0A2W5MK11	1	*Sn*UnBI_MlaD1	75	526
	23	Phospholipid/cholesterol/gamma-HCH transport system substrate-binding protein	*Parapedobacter indicus*	Gram-negative bacteria	A0A1I3SUZ7	1	*Pi*PlBI_MlaD1	71	314
	24	Phospholipid/cholesterol/gamma-HCH transport system substrate-binding protein	*Gordonia westfalica*	Actinobacteria	A0A1H2JIS7	1	*Gw*PlBI_MlaD1	79	339
	25	Paraquat-inducible protein B	*Pseudomonas alcaligenes*	Gram-negative bacteria	A0A1N6TND3	1	*Pa*PiBI_MlaD1	95	316
	26	Phospholipid/cholesterol/gamma-HCH transport system substrate-binding protein	*Duganella* sp. CF402	Gram-negative bacteria	A0A1H8DZL8	1	*Ds*PlBI_MlaD1	76	312
	27	Uncharacterized protein	*Planctomycetaceae bacterium*	Gram-negative bacteria	A0A3C0WEC4	1	*Pb*UnBI_MlaD1	76	401
	28	EF-hand domain-containing protein	*Rhodopirellula sp.*	Gram-negative bacteria	A0A2D9Z6K7	1	*Rs*EfBI_MlaD1	77	443
	29	ABC-type transport system involved in resistance to organic solvents, periplasmic component	*Oscillatoria acuminata* PCC 6304	Cyanobacteria	K9TI08	1	*Oa*AbBI_MlaD1	76	567

	30	Outer membrane lipid asymmetry maintenance protein MlaD	Gammaproteobacteria bacterium ESL0073	Gram-negative bacteria	A0A2Z3HXY3	1	*Gb*MlBI_MlaD1	77	405
	31	Putative Mce family protein	*M. abscessus*	Actinobacteria	A0A0U1A5V8	1	*Ma*McBI_MlaD1	72	209
	32	Protein translocase subunit SecD	*M. bohemicum* DSM 44277	Actinobacteria	A0A0U0W890	1	*Mb*SdBI_MlaD1	74	889
	33	Mammalian cell entry-related domain protein	Uncultured bacterium	Unclassified	A0A2I6QL88	1	*Ub*McBI_MlaD1	77	386
	34	MCE family protein	Betaproteobacteria bacterium HGW-Betaproteobacteria-16	Gram-negative bacteria	A0A2N2THZ3	1	*Bb*McBI_MlaD1	97	343
	35	MCE family protein	*Sulfitobacter* sp. SK012	Gram-negative bacteria	A0A345QPF4	1	*Sc*McBI_MlaD1	78	571
	36	MCE family protein	*Anaerolineaceae* bacterium	Gram-negative bacteria	A0A3B9PYL4	1	*Ab*McBI_MlaD1	63	277
	37	Outer membrane lipid asymmetry maintenance protein MlaD	*Comamonadaceae bacterium* OH3737_COT-264	Gram-negative bacteria	A0A3P1W4X3	1	*Cb*MlBI_MlaD1	78	285
B-II	38	Virulence factor Mce family protein	*Nocardioides* sp. CF8	Actinobacteria	R7XWT8	1	*Ns*McBII_MlaD1	77	659
	39	MlaD domain-containing protein	*Glossina austeni*	Arthropoda	A0A1A9VJV1	1	*Ga*MlBII_MlaD1	79	246

	40	ABC transporter substrate-binding protein	*Actinophytocola xinjiangensis*	Actinobacteria	A0A1Q8C9G6	1	*Ax*AbBII_MlaD1	77	390
	41	Putative ABC Transporter ATP-Binding Protein	*Liberibacter crescens (*strain BT-1*)*	Gram-negative bacteria	L0EV30	1	*Lc*AbBII_MlaD1	74	726
	42	Uncharacterized protein	Marine sediment metagenome	Unclassified	A0A0F9HI77	1	*Ms*UnBII_MlaD1	79	509
	43	MlaD domain-containing protein	*Panicum miliaceum*	Plant	A0A3L6PQ90	1	*Pm*MlBII_MlaD1	76	593
	44	MCE family protein	*Flavivirga* sp. RZ03	Gram-negative bacteria	A0A4S1DUZ0	1	*Fs*McBII_MlaD1	66	310
B-III	45	ADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 12	*Trichoplax adhaerens*	Placozoa	B3SFA0	1	*Ta*AdBIII_MlaD1	79	367
	46	Uncharacterized protein	SAR202 cluster bacterium Io17 ChloroG7	Unclassified *Chloroflexi*	A0A2N0LT42	1	*Sc*UnBIII_MlaD1	41	367
	47	Regulatory protein BlaR1	*Candidatus Hydrogenedentesbacterium* ADurb.Bin101	Unclassified	A0A1V6F3K8	1	*Ch*BlBIII_MlaD1	67	2690
B-IV	48	Virulence factor Mce family protein	*Saccharomonospora viridis (*strain DSM 43017*)*	Gram-negative bacteria	C7MTI2	1	*Sv*McBIV_MlaD1	77	394
	49	Phospholipid/cholesterol/gamma-HCH transport system substrate-binding protein	*Nocardioides sp.* CF204	Actinobacteria	A0A2T5W7V8	1	*Ns*PlBIV_MlaD1	76	361

	50	ABC transporter substrate-binding protein	*Saccharopolyspora rectivirgula*	Actinobacteria	A0A073BBV5	1	*Sr*AbBIV_MlaD1	76	448
	51	Phospholipid/cholesterol/gamma-HCH transport system substrate-binding protein	*Saccharopolyspora spinose*	Actinobacteria	A0A2N3Y3K5	1	*Ss*PlBIV_MlaD1	75	341
	52	MCE family protein	*Marmoricola solisilvae*	Actinobacteria	A0A3N0DYD7	1	*Ms*McBIV_MlaD1	77	384
	53	Uncharacterized protein	*Candidatus Cloacimonetes bacterium*	Unclassified	A0A2G6BTQ6	1	*Cc*UnBIV_MlaD1	75	300
	54	DNA repair protein RadA	*Candidatus Rokubacteria bacterium*	Unclassified	A0A2V6Q134	1	*Cr*RaBIV_MlaD1	79	901
B-V	55	Uncharacterized protein	*B. rapa subsp. pekinensis*	Plant	M4F3Z1	1	*Br*UnBV_MlaD1	78	777
B-VI	56	AP2/ERF domain-containing protein	*Brassica campestris*	Plant	A0A3P5ZD14	1	*Bc*ApBVI_MlaD1	76	1011
B-VII	57	Uncharacterized protein	*Phycisphaerales* bacterium	Gram-negative bacteria	A0A353T9A2	1	*Pb*UnBVII_MlaD1	82	874
B-VIII	58	Permease and Tol Tol Ttg2 and MCE and BolA and ST AS 2 domain containing protein	*Trichuris trichiura*	Nematoda	A0A077ZFB4	1	*Tt*PrBVIII_MlaD1	80	709
C-I	59	Mce family protein Mce4D	*M. bovisvariant bovis* BCG	Actinobacteria	A0A109SAP2	2	*Mb*4DCI_MlaD1	75	833
							*Mb*4DCI_MlaD2	80	833

C-II	60	Virulence factor Mce family protein	*M. ahvazicum*	Actinobacteria	A0A2K4YJ81	2	*Ma*McCII_MlaD1	74	826
							*Ma*McCII_MlaD2	75	826
C-III	61	Paraquat-inducible protein A	*Isosphaera pallida (*strain DSM 9630*)*	Gram-negative bacteria	E8R5N3	2	*Ip*PiCIII_MlaD1	92	892
							*Ip*PiCIII_MlaD2	84	892
C-IV	62	Uncharacterized protein	*M. aviumsubsp. avium* 2285	Actinobacteria	X8B5I0	2	*Ma*UnCIV_MlaD1	73	846
							*Ma*UnCIV_MlaD2	74	846
C-V	63	Uncharacterized protein	*Acyrthosiphon pisum*	Arthropoda	X1WKZ8	2	*Ap*UnCV_MlaD1	92	972
							*Ap*UnCV_MlaD2	78	972
C-VI	64	Paraquat-inducible protein b	*Lasius niger*	Arthropoda	A0A0J7NPH8	2	*Ln*PiCVI_MlaD1	93	1328
							*Ln*PiCVI_MlaD2	108	1328
D-I	65	Paraquat-inducible protein B	*Klebsiella pneumoniae* ISC21	Gram-negative bacteria	W1GT48	3	*Kp*PiDI_MlaD1	93	733
							*Kp*PiDI_MlaD2	93	733
							*Kp*PiDI_MlaD3	106	733
	66	Paraquat-inducible protein B	*Pseudomonas* sp. LAMO17WK12:I10	Gram-negative bacteria	A0A285J231	3	*Ps*PiDI_MlaD1	87	549
							*Ps*PiDI_MlaD2	61	549
							*Ps*PiDI_MlaD3	112	549

D-II	67	Paraquat-inducible protein B	*Candidatus Paraburkholderia kirkii*	Gram-negative bacteria	A0A0L0LZ56	3	*Cp*PiDII_MlaD1	93	763
							*Cp*PiDII_MlaD2	92	763
							*Cp*PiDII_MlaD3	108	763
D-III	68	Paraquat-inducible protein B	*Klebsiella aerogenes*	Gram-negative bacteria	A0A3S4JAN9	3	*Ka*PiDIII_MlaD1	92	964
							*Ka*PiDIII_MlaD2	61	964
							*Ka*PiDIII_MlaD3	105	964
D-IV	69	MCE and PqiA and DUF330 domain containing protein	*Trichuris trichiura*	Nematoda	A0A077ZKM6	3	*Tt*McDIV_MlaD1	93	1044
							*Tt*McDIV_MlaD2	93	1044
							*Tt*McDIV_MlaD3	106	1044
E	70	Uncharacterized protein YebT	*Beauveria bassiana* D1-5	Fungi	A0A0A2VIN0	5	*Bb*YeE_MlaD1	93	1597
							*Bb*YeE_MlaD2	92	1597
							*Bb*YeE_MlaD3	80	1597
							*Bb*YeE_MlaD4	92	1597
							*Bb*YeE_MlaD5	85	1597

Table A2.2. Overview of the APOs, SDs and medians of amino acids of MlaD domains.

Organisms		Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val
E. coli	APO	5.9	5.5	3.5	5.7	0.0	3.4	4.5	10.7	1.2	6.1	11.5	4.9	1.7	3.0	4.4	7.1	7.1	0.6	2.9	10.4
	SD	2.3	1.5	1.6	1.7	0.0	1.8	1.8	2.1	1.0	1.8	3.1	1.9	1.5	1.9	1.4	2.8	2.7	0.6	1.5	2.6
	Median	4.7																			
M. tuberculosis	APO	9.6	6.1	4.0	5.3	0.0	3.0	3.6	9.3	1.1	6.1	7.7	4.4	1.5	2.3	4.7	6.8	8.3	0.4	3.7	12.1
	SD	2.6	2.8	2.0	1.5	0.0	1.6	1.7	3.3	1.1	2.2	2.4	2.0	1.4	1.4	2.3	2.4	2.4	0.8	1.6	2.8
	Median	4.6																			
A. thaliana	APO	3.8	7.7	5.1	3.8	1.3	2.6	7.7	6.4	1.3	14.1	6.4	3.8	3.8	1.3	5.1	6.4	7.7	0	1.3	10.3
	SD	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Median	4.5																			
Overall	**APO**	**6.9**	**5.5**	**3.9**	**5.7**	**0.1**	**3.6**	**4.6**	**9.6**	**1.3**	**7.0**	**9.4**	**4.9**	**1.4**	**2.9**	**4.3**	**7.1**	**6.8**	**0.5**	**3.3**	**11.3**
	SD	**3.1**	**2.4**	**2.2**	**2.0**	**0.4**	**1.8**	**2.1**	**2.7**	**1.2**	**3.0**	**2.9**	**2.5**	**1.3**	**1.7**	**1.9**	**2.6**	**2.6**	**0.7**	**1.8**	**3.2**
	Median	**4.7**																			

Supplementary Figures, A1

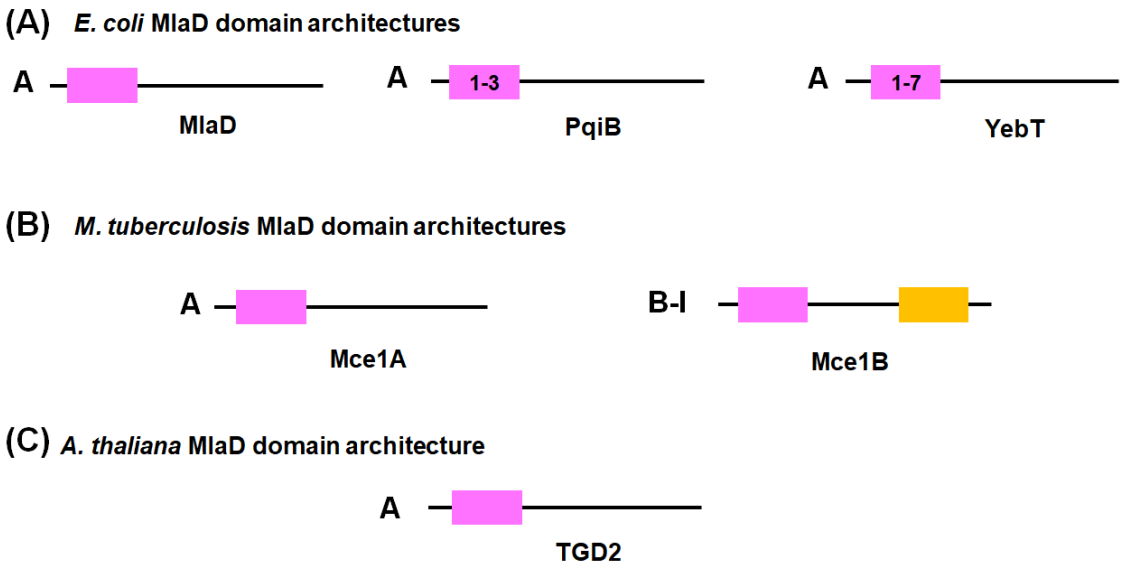


Figure A1.1. Schematic representation of the MlaD domain architectures. *The MlaD domain architecture(s) of (A) E. coli, (B) M. tuberculosis and (C) A. thaliana. The MlaD domain is represented as a pink rectangle, while the Mce4_CUP1 domain as an orange rectangle.*

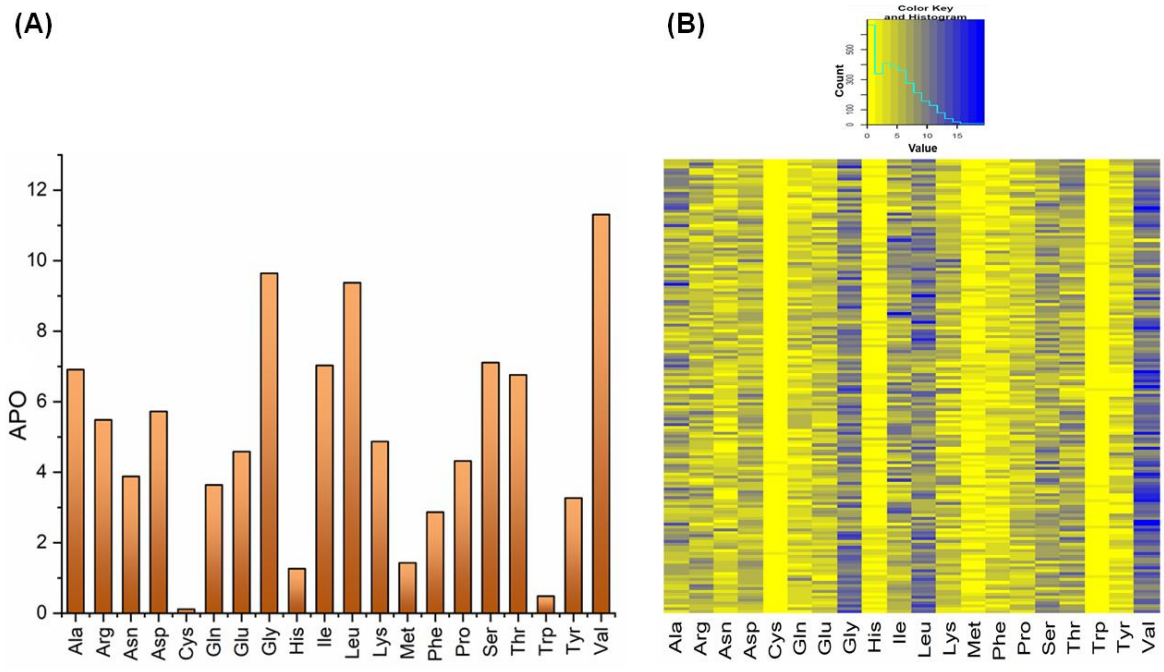


Figure A1.2. The distribution of amino acids in the MlaD domains of different domain architectures. *(A) Graph depicting the average percentage of occurrences (APOs) of amino acids present in the MlaD domains belonging to different architecture types. The amino acids and their APOs are represented along the X- and Y-axis, respectively. (B) A heat map showing the distribution of the 20 amino acids in MlaD domains across different architectures. The amino acids are represented in columns and the MlaD domains are in rows. The X-axis in the color key represents the number of cells and the Y-axis represents the values in each cell. The increasing gradient of blue signifies an increase in the PO values. The histogram within the color key indicates the number of cells in the heat map carrying a percentage occurrence (PO). The domain profiles are arranged in the heatmap as per the order of the domains in Table A2.1.*

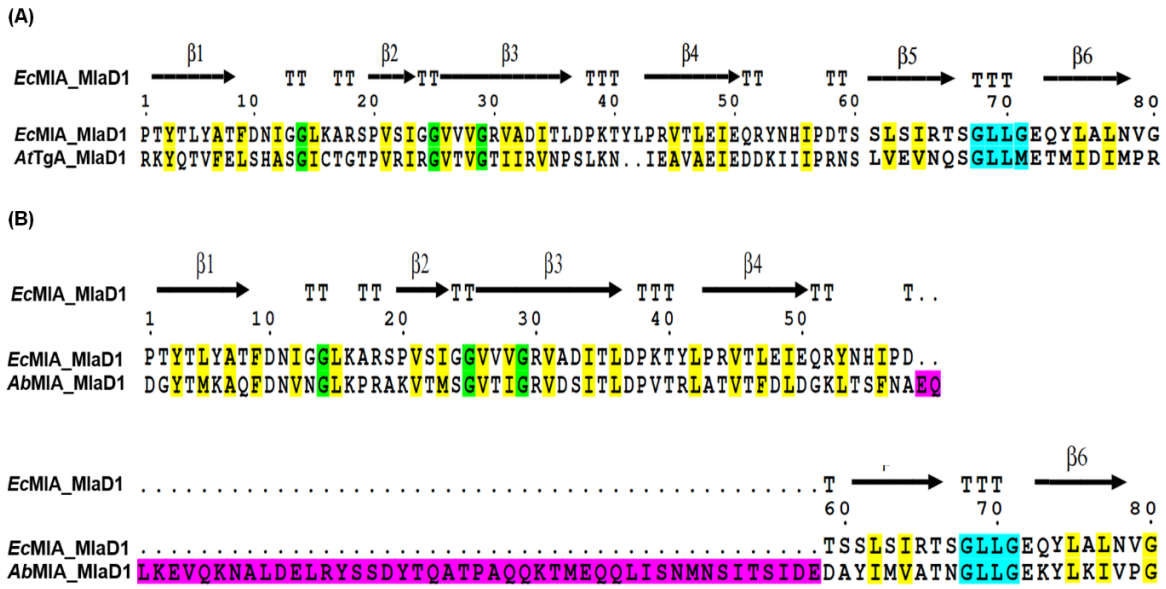


Figure A1.3. Pairwise sequence alignment of the MlaD domains. *(A-B) Sequence alignment of EcMlA_MlaD1 with AtTgA_MlaD1 and AbMlA_MlaD1, respectively. The conserved glycine and hydrophobic residues are highlighted in green and yellow, respectively. The conserved PLP region is highlighted in cyan. The insertion in the AbMlA_MlaD1 sequence is highlighted in pink.*

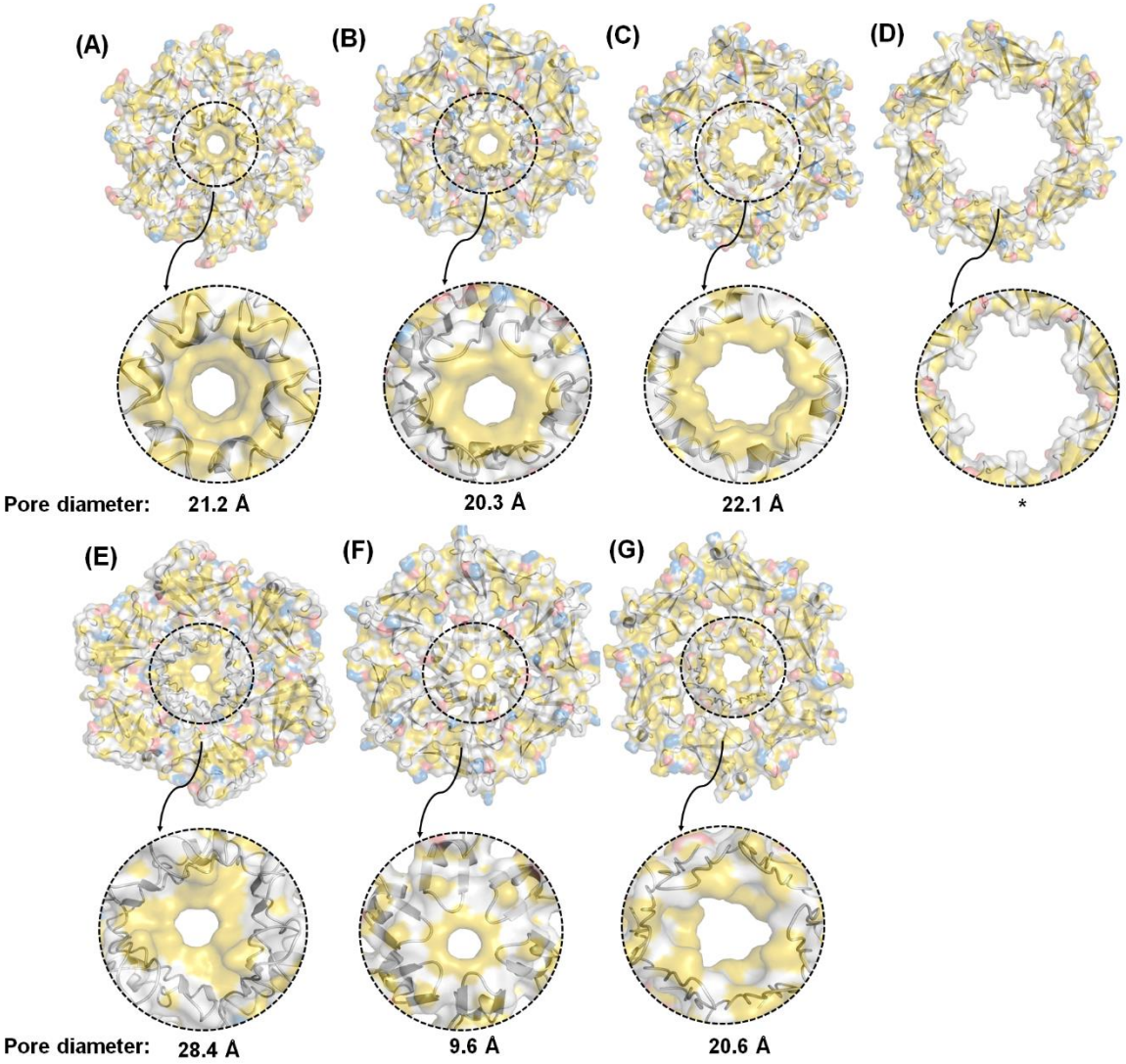


Figure A1.4. Overview of the hexameric assemblies of the MlaD domains and their respective pores in the protein YebT from E. coli. *(A-G) The surface YRB representation of EcYeA_MlaD1-7 hexamers is shown at the top, while the magnified view of their respective central pores lined by the PLP regions along the pore diameters is shown at the bottom. The pore diameter for EcYeA_MlaD4 (marked with an asterisk) could not be measured due to the low quality of the electron density for the PLP region.*

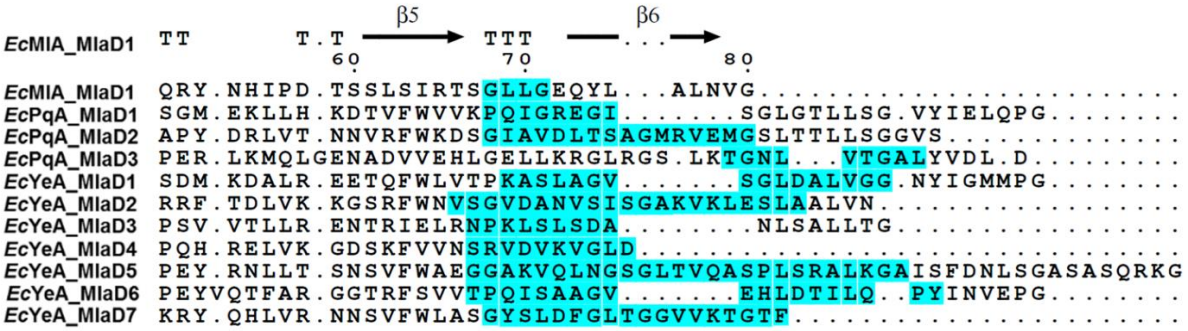


Figure A1.5. Multiple sequence alignment of the MlaD domains of the proteins MlaD, PqiB and YebT from E. coli. *The PLP region of each of the domains of the proteins MlaD, PqiB and YebT of E. coli is highlighted in cyan. For the clarity of the figure, only the C-terminal region of the MlaD domains is shown here.*

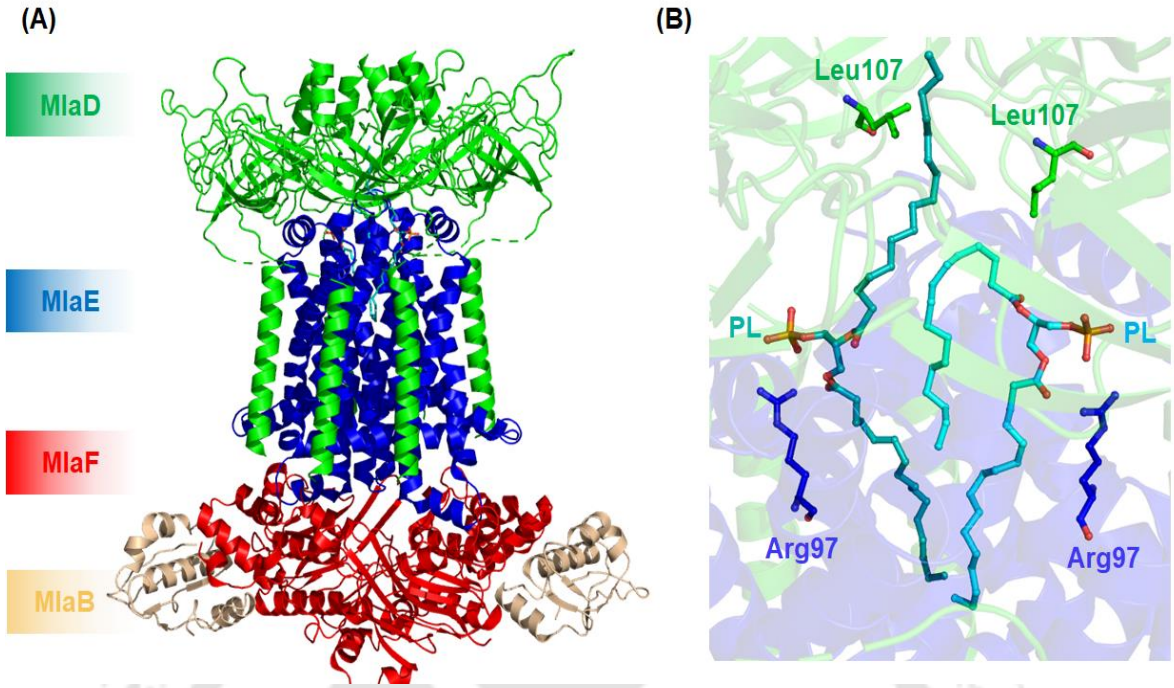


Figure A1.6. Cryo-EM structure of the MlaFEDB complex and the interactions of the bound PLs. *(A) Overview of the $MlaF_2E_2D_6B_2$ complex (PDB id: 6XBD). The proteins MlaF, MlaE, MlaD and MlaB are coloured in red, blue, green and wheat, respectively. (B) Interaction of the two PLs bound in the binding pocket formed by MlaD and MlaE. The two PLs are coloured teal and cyan. The residue Arg97 from each MlaE subunit forms a salt bridge with the phosphate head group. The protein MlaD makes asymmetric interactions via Leu107 located in the PLP region with one fatty acid tail of PL.*

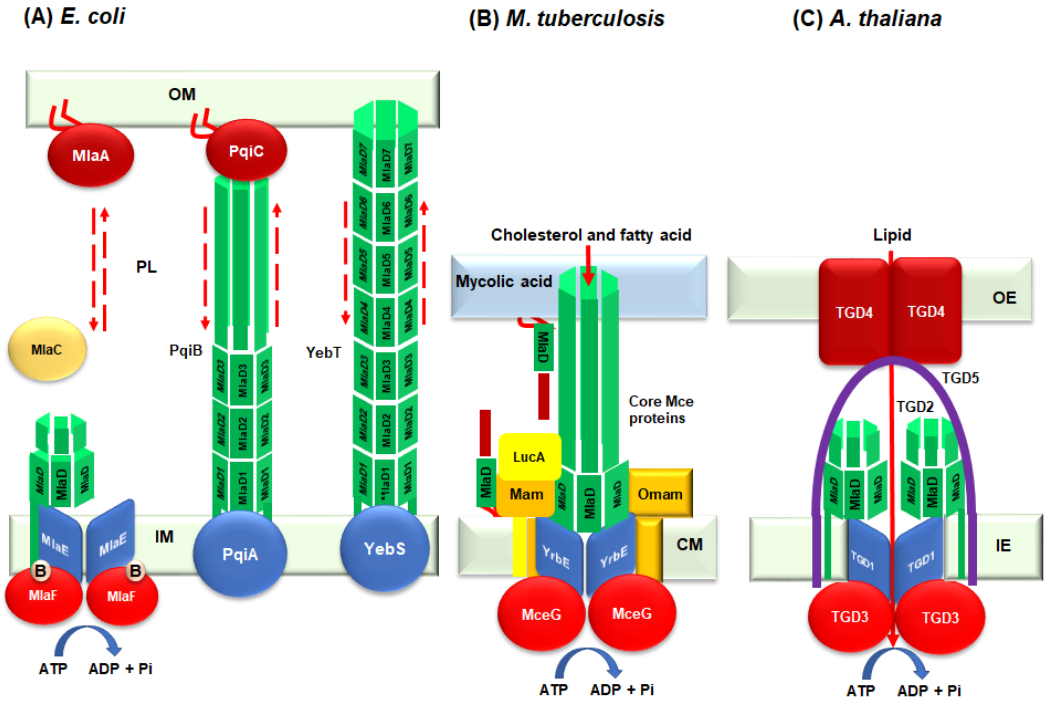


Figure A1.7. Organization of the MlaD domain-containing systems in E. coli, M. tuberculosis and A. thaliana. *(A) The Mla, Pqi and Yeb systems of E. coli. (Left) The Mla system comprises an IM-associated ABC transporter complex MlaBDEF, a free-floating periplasmic protein MlaC and OM-associated lipoprotein MlaA. (Center) The Pqi system consists of the IM protein PqiA, a periplasm-spanning protein PqiB and an OM-associated lipoprotein PqiC. (Right) The Yeb system consists of the IM protein YebS and the periplasm-spanning protein YebT. The proposed directionality of the PLs movement is represented by dotted red arrows. (B) The Mce system of M. tuberculosis comprises the core MCE proteins, the permease YrbE and the associated proteins Mam, Omam, LucA and MceG. The internalization of the hydrophobic molecules through the Mce protein complex is represented by a solid red arrow. (C) The TGD system of A. thaliana, consisting of an inner envelope associated ABC transporter complex TGD1-3, an outer envelope (OE) protein TGD4 and an inner envelope (IE) associated protein TGD5 present in the intermembrane space. The lipid molecules are internalized into the chloroplast via the protein TGD4 (represented by a solid red arrow). (OM, outer membrane; IM, inner membrane; CM, cytoplasmic membrane; IE, inner envelope; OE, outer envelope).*

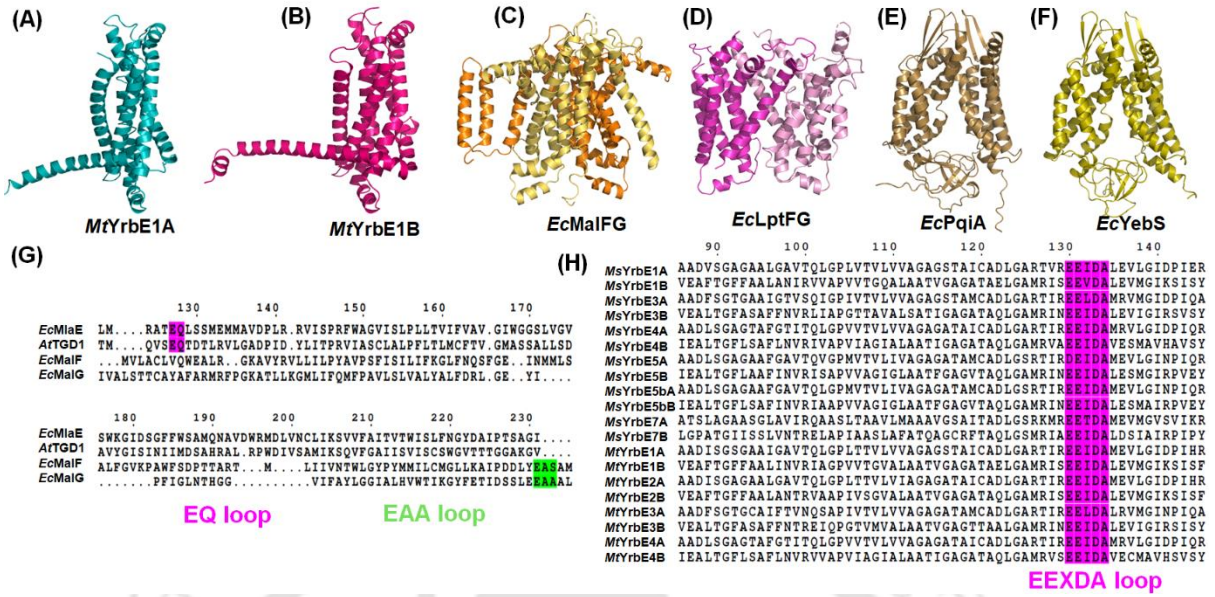


Figure A1.8. Structure and sequence analysis of EcMlaE orthogs and other TMDs. *(A-F). The overall structure of MtYrbE1A (modeled, teal), MtYrbE1B (modeled, hotpink), EcMalFG (Pdb id: 2R6G, orange and yelloworange, respectively), EcLptFG (PDB id: 6MHZ, pink and lightpink, respectively), EcPqiA (modeled, sand) and EcYebS (modeled, olive), respectively. (G) MSA of EcMlaE, AtTGD1, EcMalF and MalG. The EQ and EAA loops are highlighted in magenta and green, respectively. (H) MSA of YrbE proteins from M. smegmatis (MsYrbE1A-7B) and M. tuberculosis (MtYrbE1A-4B). The EEXDA loop is highlighted in magenta.*

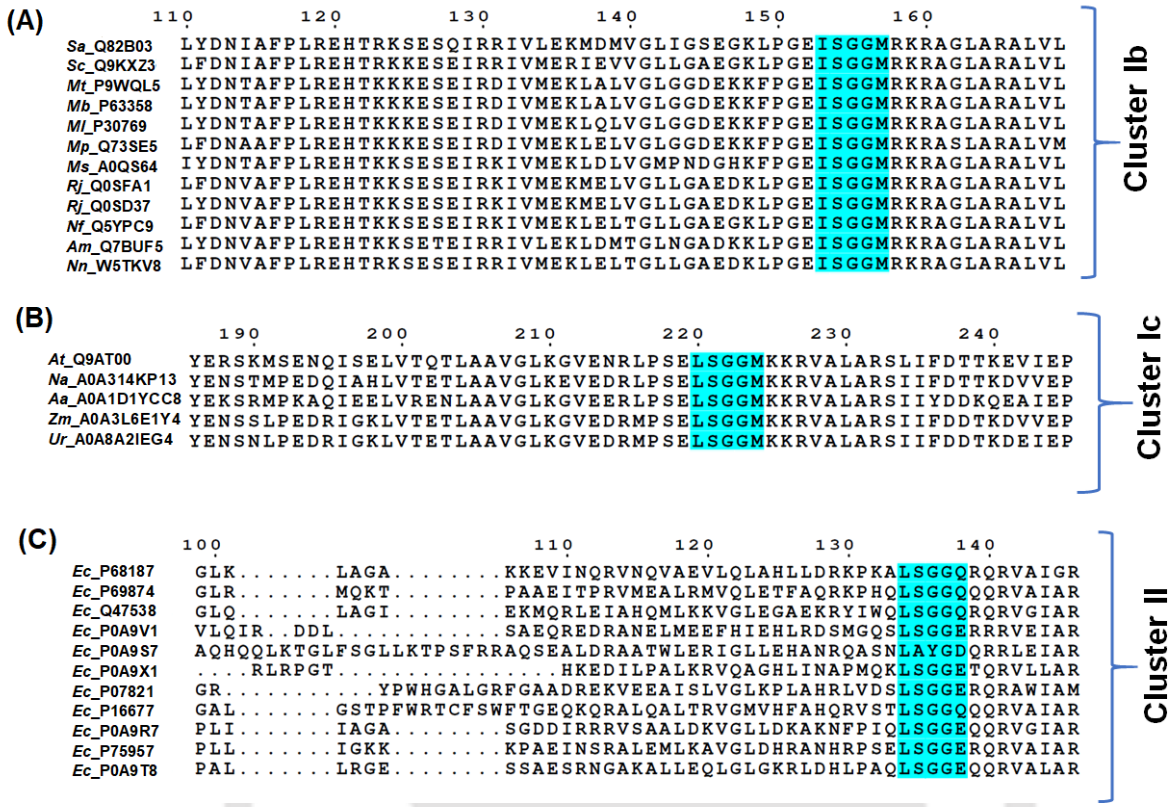


Figure A1.9. Sequence of EcMlaF orthologs. *(A-B) MSA of EcMlaF orthologues from Actinomycetales (Cluster Ib) and plants (Cluster Ic), respectively. (C) MSA of non-Mkl NBDs (Cluster II). The conservation of ISGGM, LSGGM and LSGGQ motifs are highlighted in cyan*

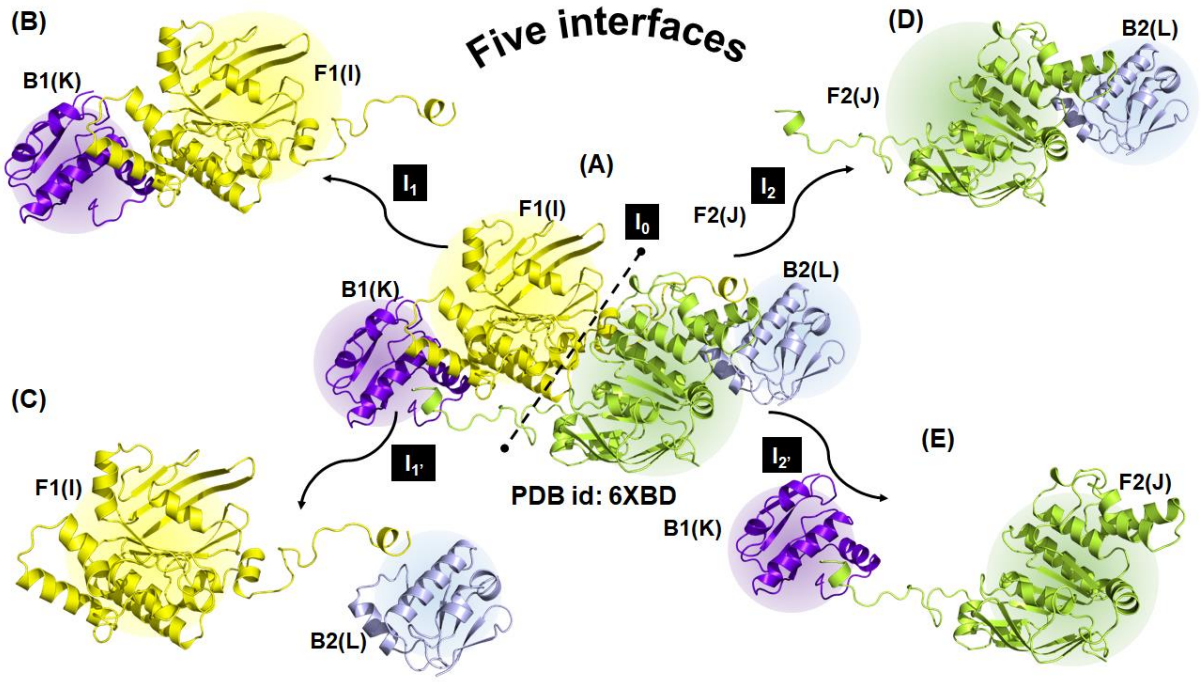


Figure A1.10. Overview of the different interfaces in the EcMlaF-EcMlaB complex. *(A) Interface analysis of EcMlaF-EcMlaB complex (PDB id: 6XBD). The complex can be divided into two halves (shown with a dotted line), each containing one unit of EcMlaF (F1, yellow, chain I; F2, limon, chain J) and EcMlaB (B1, purpleblue, chain K; B2, lightblue, chain L). The interfaces of F1-F2, F1-B1, F1-B2, F2-B2 and F2-B1 are designated as I_0, I_1, $I_{1'}$, I_2 and $I_{2'}$, respectively. (B-E) Interfaces between F1-F2, F1-B1, F1-B2, F2-B2 and F2-B1, respectively.*

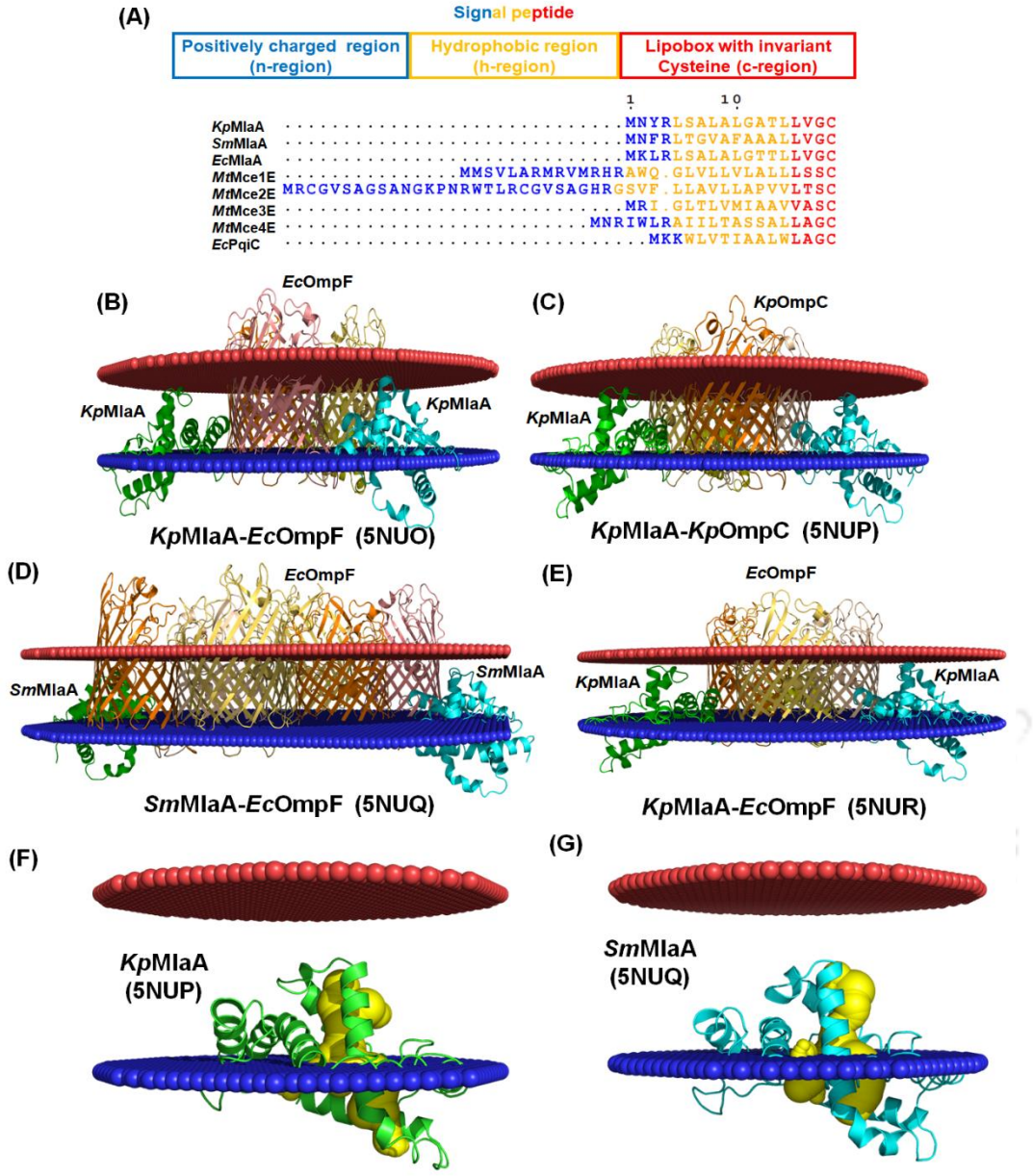


Figure A1.11. Sequence and structure analyses of MlaA. *(A) MSA of Kp/Sm/EcMlaA, MtMce(1-4)E and EcPqiC. The n-, h- and c-regions of the signal peptides are highlighted in blue, yellow and red, respectively. (B-E) Orientation analysis of available Kp/SmMlaA-Ec/SmOmpC/F complexes (PDB ids: 5NUO, 5NUP, 5NUQ and 5NUR, respectively). (F-G) Orientation analysis of KpMlaA (PDB id: 5NUP) and SmMlaA (PDB id: 5NUQ), respectively. The central channels are represented as yellow surfaces. The outer and inner leaflets of OM are represented by red and blue planes, respectively (Figure B-G).*

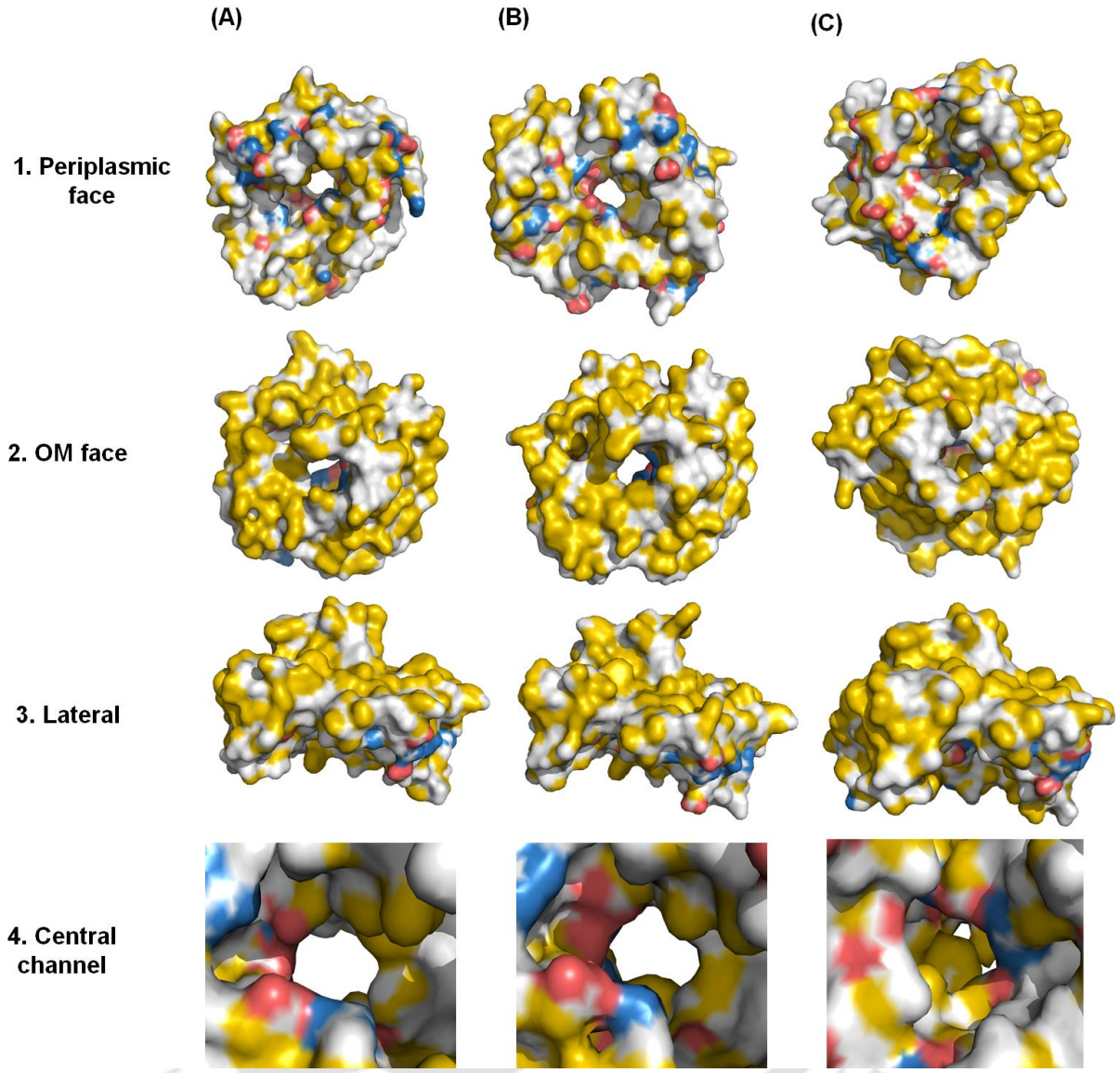


Figure A1.12. (A-C) Hydrophobicity mapping of KpMlaA, SmMlaA and EcMlaA$_m$, respectively. *(Row 1-4) Periplasmic face, OM-facing face, lateral view and magnified view of the central channel, respectively.*

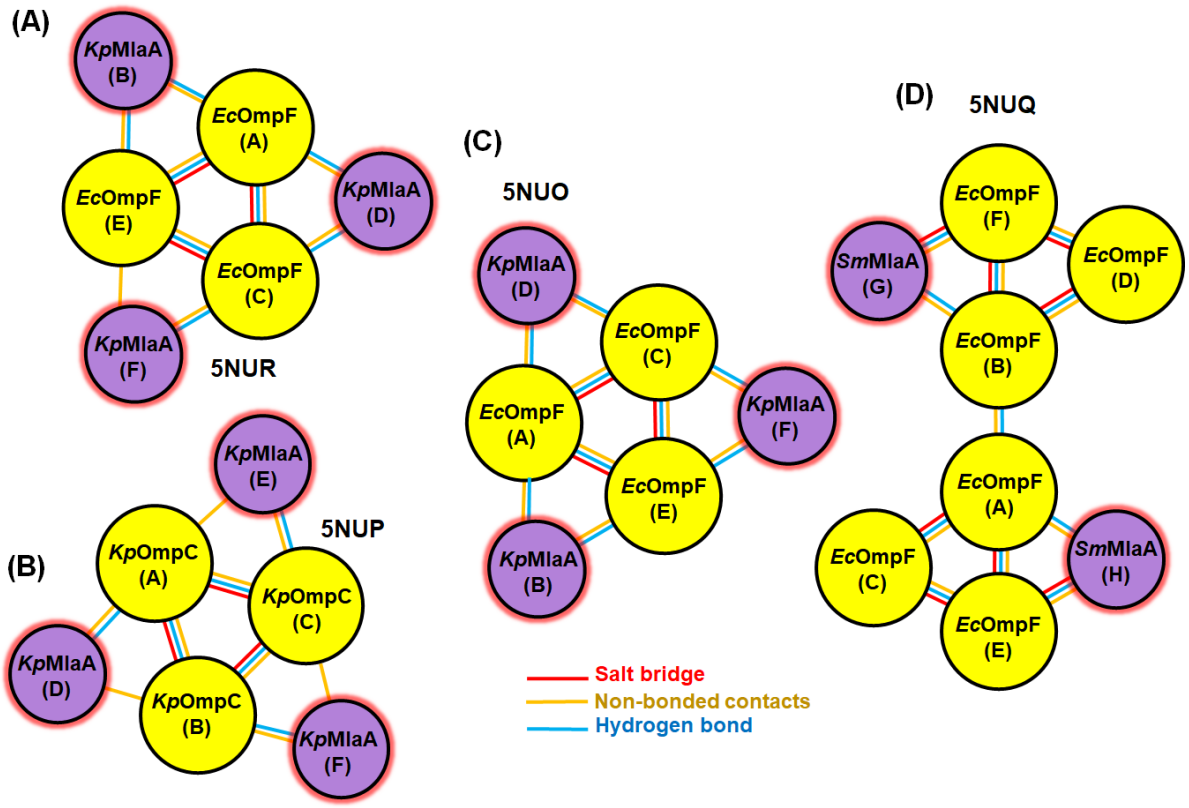
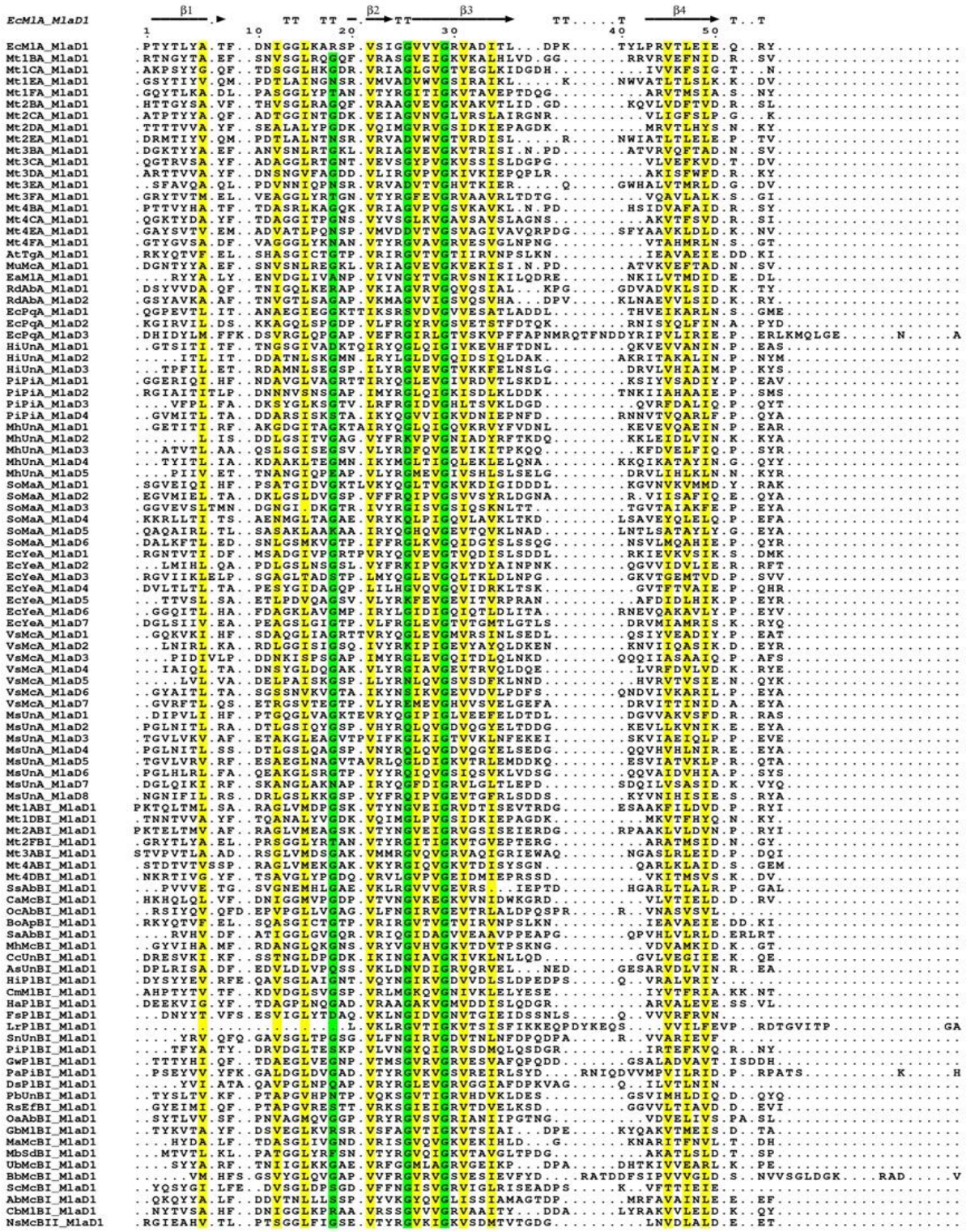
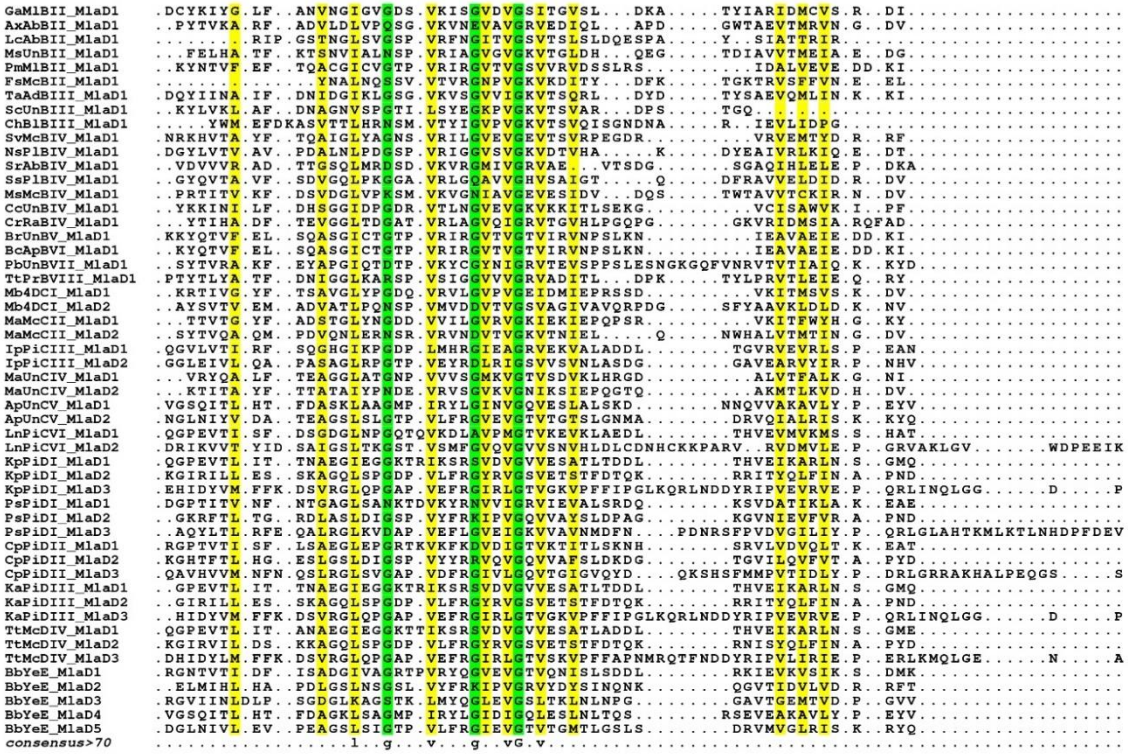
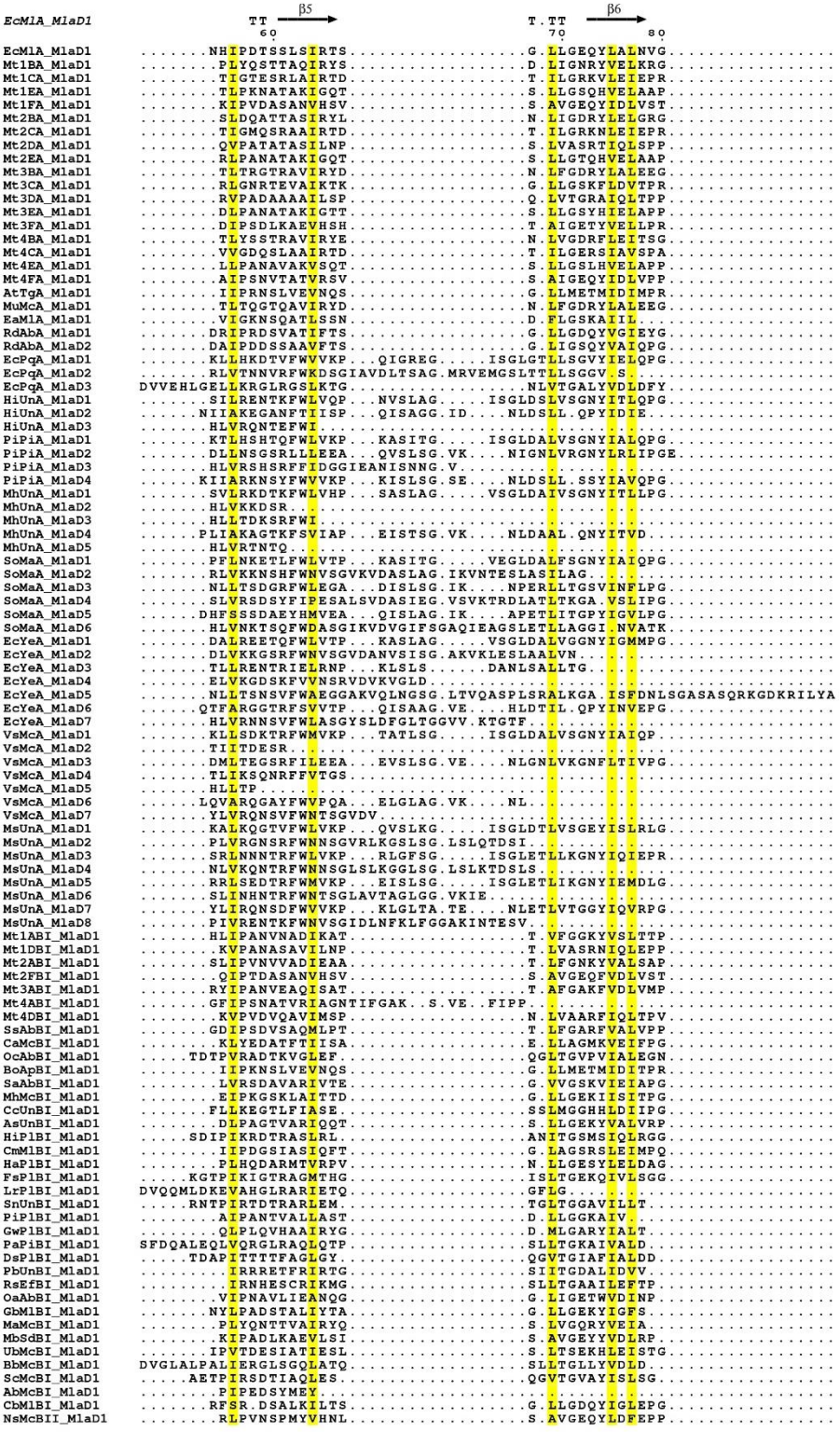


Figure A1.13. Schematic representation of the Kp/SmMlaA-Ec/KpOmpC/F complexes. *(A-D) Overview of the interactions observed in the available crystal structures of Kp/SmMlaA-Ec/KpOmpC/F complexes (PDB ids: 5NUR, 5NUP, 5NUO and 5NUQ, respectively). Kp/SmMlaA are highlighted in red.*

Supplementary Figure, A2







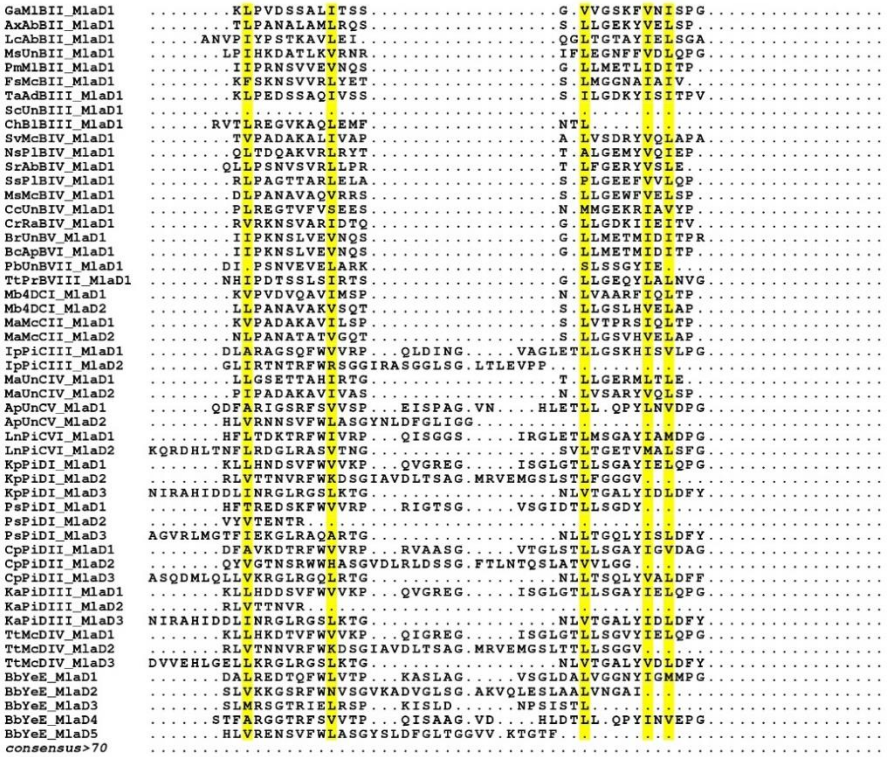


Figure A2.1. *MSA of the MlaD domains from different architectures. The similar hydrophobic and the conserved residues are highlighted in yellow and green, respectively.*

Appendix B

Supplementary data to Chapter 4

Supplementary Tables

Table B1. X-ray crystallographic data collection and refinement statistics for *Ec*MlaC. The values in parenthesis represent the statistics for the last resolution shell.

Parameters	*Ec*MlaC
Wavelength (Å)	1.5418
Temperature (K)	100
Space group	*H*3
Unit-cell parameters (Å, °)	*a*=*b*=114.94, *c*=46.10, *α*=*β*=90, *γ*=120
Resolution (Å)	57.47-2.50 (2.60-2.50)
No. of observed reflections	35454 (3998)
No. of unique reflections	7857 (898)
Mn(I) CC(1/2)	0.996 (0.865)
Completeness (%)	100 (100)
V_M (Å^3 Da^{-1})	2.58
Solvent content (%)	52.41
Mosaicity (°)	0.55
Mean I/σ(I)	13.5 (3.8)
R_{merge}† (%)	8.4 (36.2)
R_{pim} (%)	6.8 (29.1)
R_{meas} (%)	10.9 (46.7)
Multiplicity	4.5 (4.5)
R_{work}/R_{free} (%)	16.91/21.70
Protein model	
No. of subunits in the asymmetric unit	1
Protein atoms	1507
Water molecules	58
Phosphatidylethanolamine (PEF)	1
Other molecules	1
Deviation from ideal geometry	
Bond length (Å)	0.014
Bond angles (°)	1.984
Average *B*-factor (Å^2)	
Protein atoms	23.48
Water molecules	31.78
Phosphatidylethanolamine (PEF)	58.18
Ramachandran plot (%)	
Favoured	96.69
Allowed	2.76
Remaining	0.55
PDB id	7VR6

† $R_{merge} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I(hkl)$ is the intensity of reflection *hkl,* $\sum hkl$ is the sum overall reflections and $\sum i$ is the sum over *i* measurements of reflection *hkl.*

Appendix B

Supplementary data to Chapter 4

Table B2. The list of MlaC orthologs analyzed in this study.

S. No.	PDB id	Structure title	Organism	UniProt id	Ligand id	# of PLs bound	Ligand naming	Protein naming
1.	6GKI	Structure of *E. coli* MlaC in variously loaded states	*Escherichia coli*	P0ADV7	-	-	-	*Ec*MlaC_P3_121_apo
2.	5UWA	Structure of *E. coli* phospholipid-binding protein MlaC	*Escherichia coli*	P0ADV7	8ND	1	8ND	*Ec*MlaC_P$2_12_12_1$_holo1
3.	2QGU	Three-dimensional structure of the phospholipid-binding protein from *Ralstonia solanacearum* Q8XV73_RALSQ in complex with a phospholipid at the resolution 1.53 Å	*Ralstonia solanacearum*	Q8XV73	PEF		PEF'	*Rs*MlaC_holo1
4.	7VR6	Crystal structure of MlaC from *Escherichia coli* in quasi-open state	*Escherichia coli*	P0ADV7	PEF		PEF	*Ec*MlaC_H3_holo1
5.	5UWB	Re-refined 4FCZ: lipid-bound crystal structure of toluene-tolerance protein from *Pseudomonas putida*	*Pseudomonas putida*	Q88P91	PEF	2	PEF_1, PEF_2	*Pp*MlaC_holo2
6.	6HSY	Two-phospholipid-bound crystal structure of the substrate-binding protein Ttg2D from *Pseudomonas aeruginosa*	*Pseudomonas aeruginosa*	Q9HVW4	GOT, H3T		GOT, H3T	*Pa*MlaC_holo2

Appendix B

Supplementary data to Chapter 4

Table B3. List of protein-ligand interactions. The polar and non-polar contacts are considered within the distance cut-off of 3.5 and 4.0 Å, respectively.

S. No.	Group	PDB id	Subdomain	Protein residue	Protein atom	Ligand id	Ligand atom(s)
1.	II	7VR6	D1R1	Met31	C^{ε}	PEF	¶C45'
2.				Phe39	$C^{\varepsilon 1}$		¶C40'
3.					C^{ζ}		¶C39', ¶C40'
4.				Val68	$C^{\gamma 2}$		†C21', †C22', †C23'
5.			D2R1	Tyr100	$C^{\varepsilon 1}$		*N'
6.					C^{ζ}		*N'
7.					O^{η}		*C5', *N'
8.				Leu101	$C^{\delta 2}$		†C14'
9.				Ala104	C		*O4'
10.					O		*O4'
11.					C^{β}		*O4'
12.				Tyr105	N		*O4'
13.					C^{γ}		*O4'
14.					$C^{\delta 2}$		*O4'
15.					$C^{\varepsilon 1}$		†C12', †C13', †C21'
16.					C^{ζ}		†C21'
17.					O^{η}		†C21'
18.				Ala108	O		¶C36'
19.					C^{β}		*C3'
20.				Leu109	$C^{\delta 2}$		¶C40'
21.				Met111	Cβ		¶C36'
22.					C^{ε}		¶C31', ¶C32'
23.			D1R2	Gln115	$O^{\varepsilon 1}$		¶C37'
24.				Phe150	$C^{\delta 1}$		†C23'
25.					$C^{\varepsilon 1}$		†C22', †C23', †C24'

26.				Trp152	$C^{\zeta 2}$		†C25'
27.				Ala163	C^{β}		†C23', †C25'
28.				Asp165	C^{α}		†C18'
29.				Met166	C^{ε}		*O5'
30.			D2R2	Met173	C^{β}		†C16'
31.					C^{ε}		†C11'
32.				Lys177	N^{ζ}		*C4', *N'
33.		5UWA	D1R1	Ala34	C^{β}	8ND	¶C42'
34.				Phe39	$C^{\varepsilon 1}$		¶C36', ¶C37'
35.					C^{ζ}		¶C36'
36.			D2R1	Tyr105	$C^{\delta 1}$		†C14', †C16'
37.					$C^{\varepsilon 1}$		†C16'
38.				Ala108	O		¶C33'
39.				Leu109	$C^{\delta 2}$		¶C36'
40.				Met111	O		¶C31'
41.					C^{β}		¶C31'
42.			D1R2	Gln115	$N^{\varepsilon 2}$		*O5', *C30', *C31'
43.				Val146	$C^{\gamma 1}$		*O2'
44.				Phe150	$C^{\varepsilon 2}$		¶C43'
45.					C^{ζ}		¶C43'
46.				Tyr164	O		†C23'
47.				Glu169	$O^{\varepsilon 2}$		*C1'
48.			D2R2	Met173	C^{β}		†C25'
49.				Ile174	N		†C25'
50.					$C^{\gamma 1}$		†C25'
51.		2QGU	D1R1	Val39	$C^{\gamma 2}$	PEF'	¶C38'
52.				Ile68	$C^{\delta 1}$		¶C39'
53.					$C^{\gamma 2}$		¶C42'
54.			D2R1	Phe74	$C^{\delta 1}$		†C24'
55.				Leu105	C^{β}		†C22'

56.				Tyr109	C^{β}		†C18'
57.					C^{γ}		†C18'
58.					$C^{\delta 2}$		†C18'
59.				Ala112	C^{β}		*O4'
60.			D1R2	Gln119	C^{γ}		*O5'
61.					C^{δ}		*O5'
62.					$N^{\varepsilon 2}$		*O5'
63.				Asn143	C^{β}		*O5'
64.					C^{γ}		*C3', *O5'
65.					$O^{\delta 1}$		*C3', *O5'
66.					$N^{\delta 2}$		*C1', *C3', *O1P'
67.				Leu150	$C^{\delta 1}$		¶C34'
68.				Tyr152	O^{η}		¶C38'
69.				Val163	$C^{\gamma 1}$		¶C42', ¶C44'
70.				Ile168	$C^{\gamma 2}$		†C11', †C12'
71.	III	5UWB	D1R1	Leu59	O	PEF_1	¶C44'
72.					$C^{\delta 1}$		¶C39'
73.				Asn60	C^{α}		¶C44'
74.					C^{γ}		¶C44'
75.					$O^{\delta 1}$		¶C44'
76.				Leu63	C		¶C45'
77.					C^{β}		¶C45'
78.				Gly64	N		¶C45'
79.					C^{α}		¶C45'
80.			D2R1	Ile72	$C^{\gamma 2}$		†C24'
81.					$C^{\delta 1}$		†C24'
82.				Phe97	$C^{\delta 2}$		†C25'
83.				Lys98	C^{β}		¶C42'
84.					C^{γ}		¶C42', ¶C43'
85.					C^{δ}		¶C42', ¶C43'

86.				Leu101	$C^{\delta1}$		†C18', †C20', †C21', †C22'
87.				Tyr105	O		¶C35'
88.					C^{β}		¶C35'
89.					$C^{\varepsilon1}$		†C13'
90.					C^{ζ}		†C13'
91.				Ala108	C^{β}		*C30', ¶C31', *O5'
92.				Leu109	C^{γ}		¶C34'
93.					$C^{\delta1}$		¶C34', ¶C35', ¶C36'
94.			D1R2	Asn165	$N^{\delta2}$		*O4'
95.			D1R1	Thr34	$O^{\gamma1}$		†C21'
96.				Leu37	$C^{\delta2}$		†C19'
97.			D2R1	Ile72	$C^{\delta1}$		¶C44'
98.				Tyr112	$C^{\delta1}$		†C13'
99.			D1R2	Asn114	$N^{\delta2}$		*O4P'
100.				Gly138	C		*O4P'
101.					C^{α}		*O3P', *O4P'
102.				Asn139	C^{α}		*O4P'
103.					C^{β}		*O4P'
104.					$O^{\delta1}$		*O4'
105.					N	PEF_2	*O2P', *O4P', *P'
106.					$N^{\delta2}$		*O4P'
107.				Tyr144	C^{β}		*O5'
108.					C^{γ}		*O5'
109.					$C^{\delta2}$		*O5'
110.				Tyr148	$C^{\varepsilon1}$		†C22', ¶C39'
111.					O^{η}		†C20', †C22', ¶C38', ¶C39'
112.				Val159	$C^{\gamma1}$		¶C43', ¶C44'
113.					$C^{\gamma2}$		†C25'
114.				Val164	$C^{\gamma2}$		¶C34'
115.				Asn165	$N^{\delta2}$		*O5'

116.		6HSY	D1R1	Val67	$C^{\gamma1}$	GOT	¶C17'
117.			D2R1	Ile72	$C^{\gamma2}$		†C49', †C50'
118.					$C^{\delta1}$		¶C18'
119.				Leu101	$C^{\delta2}$		†C41'
120.			D1R2	Tyr145	$C^{\delta1}$		*O22', *O26'
121.					$C^{\varepsilon1}$		*C27', *O22', *O26'
122.					C^{ζ}		*O26'
123.					O^{η}		*C27', *O26'
124.				Tyr149	$C^{\varepsilon2}$		¶C12'
125.					O^{η}		¶C10', ¶C12'
126.				Asn162	$O^{\delta1}$		†C50'
127.				Asn166	$N^{\delta2}$		*O22'
128.			D2R2	Ile170	$C^{\gamma2}$		†C48'
129.				Gly171	N		†C50'
130.					C^{α}		†C50'
131.				Lys208	C^{δ}		*O34'
132.					C^{ε}		*C33', *O25', *O32', *O34'
133.				Glu211	C^{δ}		*O24'
134.					$O^{\varepsilon1}$		*C21', *O22', *O24', *P23'
135.					$O^{\varepsilon2}$		*O22', *O24'
136.			D1R1	Leu37	$C^{\delta2}$	H3T	†C44'
137.				Asp60	$O^{\delta1}$		¶C17'
138.				Gly64	C^{α}		¶C19', ¶C20'
139.				Val66	$C^{\gamma2}$		†C50'
140.			D2R1	Lys98	C^{ε}		¶C20'
141.					N^{ζ}		¶C20'
142.				Met102	C^{γ}		¶C15'
143.				Tyr105	O		¶C10', ¶C11'
144.					$C^{\delta1}$		¶C6'
145.				Ala108	C^{β}		¶C6'

146.				Leu109	$C^{\delta 2}$		†C42', †C43'
147.				Tyr112	$C^{\delta 2}$		*O34'
148.			D1R2	Ile117	$C^{\delta 1}$		†C40', †C41'
149.				Ile137	$C^{\delta 1}$		†C40'
150.				Tyr149	$C^{\delta 2}$		†C49'
151.					$C^{\varepsilon 2}$		†C47', †C48'
152.					O^{η}		†C45', †C46'
153.				Met151	C^{ε}		†C49', †C50'
154.				Val160	$C^{\gamma 2}$		†C49', †C50'
155.			D2R2	Lys204	C^{ε}		*C27', *O3'
156.					N^{ζ}		*C27', *C28', *O29'
157.				Lys208	C^{ε}		*O25'
158.					N^{ζ}		*O25', *C27', *C30'

¶γ tail, †β tail and *Head region.

Table B4. Top 20 structural homologs of the protein *Ec*MlaC obtained using the web server Dali (Holm, 2020).

S. No.	Description	PDB id	Z-score	SI (%)	RMSD (Å)
1.	Probable phospholipid-binding protein MlaC	5UWA	29.8	100	1.0
2.	Toluene tolerance protein ttg2d	6HSY	18.5	18	2.6
3.	Putative ntf2-like transpeptidase	3K7C	7.8	16	2.9
4.	Iota-carbonic anhydrase	7C5X	7.0	5	3.3
5.	Ketosteroid isomerase-like protein	3D9R	6.9	12	2.9
6.	Putative orphan protein	3BB9	6.5	5	2.7
7.	Uncharacterized protein	4R1K	6.4	6	3.5
8.	Rd1ntf2_04	6W3G	6.4	10	2.7
9.	Putative cytoplasmic protein	4HZ9	6.3	9	3.4
10.	Uncharacterized protein with a ntf2-like fold	3GZR	6.2	12	3.0
11.	Uncharacterized protein	3B7C	6.1	8	3.4
12.	Hypothetical protein pa1314	1TP6	6.1	12	3.6
13.	Ntf2 fold protein loop-helix-loop design nt-9	6W90	6.0	7	3.9
14.	Uncharacterized ntf2-like protein	3F7S	6.0	10	2.7
15.	Ntf2-like protein of unknown function	3LYG	6.0	11	2.5
16.	Duf4440 domain-containing protein	6BJT	5.9	5	3.2
17.	RD1NTF2_05_I64F_A80G_T94P_D101K_L106W	6W3F	5.9	13	3.0
18.	Uncharacterized conserved protein	3ROB	5.8	10	3.3
19.	Ntf2-like protein of unknown function	3DUK	5.8	10	3.3
20.	Putative uncharacterized protein	4EC6	5.8	8	3.4

Table B5. Subdomain boundaries in the MlaC structures.

S. No.	Subdomain (superfamily)	PDB id	Subdomain boundary
1.	D1R1 (NTF2-like)	7VR6	25-69
		5UWA	24-69
		6GKI	23-69
		2QGU	29-73
		5UWB	23-68
		6HSY	23-68
2.	D2R1 (PBP)	7VR6	70-112
		5UWA	70-112
		6GKI	70-112
		2QGU	74-116
		5UWB	69-112
		6HSY	69-112
3.	D1R2 (NTF2-like)	7VR6	113-172
		5UWA	113-172
		6GKI	113-172
		2QGU	117-172
		5UWB	113-168
		6HSY	113-169
4.	D2R2 (PBP)	7VR6	173-209
		5UWA	173-208
		6GKI	173-207
		2QGU	173-207
		5UWB	169-209
		6HSY	170-212

Appendix B

Supplementary data to Chapter 4

Table B6. Summary of the sequence and structure comparison of MlaC orthologs.

Pairwise sequence comparison (identity/similarity (query coverage) in percentage)						
UniProt (PDB) id	**P0ADV7 (7VR6)**	**P0ADV7 (6GKI)**	**P0ADV7 (5UWA)**	**Q8XV73 (2QGU)**	**Q88P91 (5UWB)**	**Q9HVW4 (6HSY)**
P0ADV7 (7VR6)	100/100 (100)	100/100 (100)	100/100 (100)	29/51 (63)	22/46 (78)	22/42 (81)
P0ADV7 (6GKI)	-	100/100 (100)	100/100 (100)	29/51 (63)	22/46 (78)	22/42 (81)
P0ADV7 (5UWA)	-	-	100/100 (100)	29/51 (63)	22/46 (78)	22/42 (81)
Q8XV73 (2GQU)	-	-	-	100/100 (100)	24/50 (73)	26/48 (88)
Q88P91 (5UWB)	-	-	-	-	100/100 (100)	66/80 (95)
Q9HVW4 (6HSY)	-	-	-	-	-	100/100 (100)
Pairwise structural comparison (RMSD in Å)						
PDB id	**7VR6**	**6GKI**	**5UWA**	**2QGU**	**5UWB**	**6HSY**
7VR6	0	1. 4	0.7	1.7	3.2	3.7
6GKI	-	0	1.8	2.1	3.6	3.9
5UWA	-	-	0	1.9	3.6	3.9
2GQU	-	-	-	0	1.4	1.2
5UWB	-	-	-	-	0	1.0
6HSY	-	-	-	-	-	0

Appendix B

Supplementary data to Chapter 4

Table B7. List of MlaC protein sequences selected for analysis from Group III.

S. No.	UniProt id	Protein Name	Organism name
1.	Q6F7P1	Putative toluene tolerance protein (Ttg2D)	*Acinetobacter baylyi* strain ATCC 33305
2.	F6CSF5	Toluene tolerance family protein	*Marinomonas posidonica* strain CECT 7376
3.	A0A063Y388	Putative ABC transporter, auxiliary component YrbC	*Nitrincola lacisaponensis*
4.	S6GHK2	Uncharacterized protein	*Osedax symbiont* Rs1
5.	A0A7U8C6N6	Uncharacterized protein	*Neptuniibacter caesariensis*
6.	E1V9I2	Phospholipid-binding protein MlaC	*Halomonas elongata* strain ATCC 33173
7.	Q0VS42	Toluene-tolerance protein	*Alcanivorax borkumensis* strain ATCC 70065
8.	A4BC84	ABC-type transport system involved in resistance to organic solvents, auxiliary component	*Reinekea blandensis* MED297
9.	A0A0P8AWH8	ABC-type uptake system auxiliary component MlaC	*Marinobacter sp.* HL-58
10.	Q2SBJ0	ABC-type transport system involved in resistance to organic solvents, auxiliary component	*Hahella chejuensis* strain KCTC 2396
11.	A0A0F7M049	ABC-type transport system involved in resistance to organic solvents, auxiliary component	*Spongiibacter sp.* IMCC21906
12.	B8KR79	Uncharacterized protein	*Luminiphilus syltensis* NOR5-1B
13.	Q1YU09	Toluene tolerance protein	*Gamma proteobacterium* HTCC2207
14.	A0A143HQP8	Toluene tolerance protein	*Microbulbifer thermotolerans*
15.	B3PBY7	Toluene tolerance, Ttg2 superfamily	*Cellvibrio japonicus* strain Ueda107

Appendix B

Supplementary data to Chapter 4

Table B8. Details of the molecular docking results of various ligands targeted with *Ec*MlaC_H3_holo1. The polar and non-polar contacts are considered within the distance cut-off of 3.5 and 4.0 Å, respectively.

S. No.	Ligand	EFBE (kcal mol^{-1})	Subdomain	Protein residue	Protein atom	Ligand atom
1.	PEF	-4.0	D1R1	Phe39	$C^{\varepsilon 2}$	†C25’
2.					C^{ζ}	†C25’
3.				Val68	$C^{\gamma 1}$	¶C45’
4.					$C^{\gamma 2}$	¶C45’
5.				Leu101	$C^{\delta 2}$	¶C36’
6.				Ala104	C^{α}	*O5’
7.					C	*O5’
8.					O	*O5’, *C30’
9.					C^{β}	*O5’, ¶C31’
10.				Tyr105	$C^{\delta 1}$	†C15’
11.					$C^{\varepsilon 1}$	†C15’
12.					$C^{\varepsilon 2}$	¶C35’, ¶C36’, ¶C45’
13.					C^{ζ}	¶C35’
14.					O^{η}	¶C35’, ¶C36’, ¶C41’, ¶C44’
15.				Gln107	$N^{\varepsilon 2}$	*N’, *O1P’
16.					O	*P’
17.					C^{γ}	*P’, *O1P’, *O2P’
18.					C^{α}	*O1P’
19.					C	*O1P’
20.					O	*O1P’
21.					C^{β}	*O1P’
22.					C^{δ}	*O1P’
23.				Ala108	N	*O1P’
24.					C^{β}	†C13’, †C14’, †C15’
25.					C	†C21’

26.					O	†C21', †C22', †C23'
27.				Met111	C^{γ}	*O4P', *C4'
28.					S^{δ}	*O4P', *C4'
29.					C^{ε}	*O2', *C10', †C11'
30.					C^{β}	†C22'
31.				Tyr112	N	†C23'
32.					C^{α}	†C23'
33.					$C^{\delta1}$	†C25'
34.			D1R2	Gln115	$O^{\varepsilon1}$	†C25'
35.				Leu148	$C^{\delta1}$	†C16', †C17', †C18', †C19'
36.					$C^{\delta2}$	†C17', †C18'
37.				Tyr164	C	¶C42', ¶C43'
38.					O	¶C42'
39.				Asp165	C^{α}	¶C40', ¶C41, ¶C42'
40.					$O^{\delta1}$	¶C40'
41.					N	¶C42', ¶C43'
42.					C	¶C41'
43.				Met166	C^{ε}	†C14', †C17'
44.					S^{δ}	†C17'
45.					N	¶C41'
46.			D2R2	Met173	C^{ε}	¶C32', ¶C33', ¶C34'
47.					C^{γ}	¶C33', ¶C34', ¶C35'
48.					S^{δ}	¶C33'
49.					C^{β}	¶C35', ¶C37', ¶C40'
50.				Ile174	$C^{\gamma1}$	¶C39', ¶C40'
51.				Glu208	$O^{\varepsilon2}$	*C4', *C5'
52.	8ND	-6.91	D1R1	Phe39	$C^{\varepsilon1}$	*O5', ¶C32', ¶C34'
53.					$C^{\varepsilon2}$	*O5'
54.					C^{ζ}	*O5', *C30', ¶C32'
55.				Leu42	$C^{\delta1}$	*C3', *O3'
56.				Leu64	C^{β}	¶C39'

57.					C^{γ}	¶C39’
58.					$C^{\delta1}$	¶C36’, ¶C37’, ¶C38’, ¶C39’, ¶C40’
59.					$C^{\delta2}$	¶C35’, ¶C36’, ¶C37’, ¶C38’, ¶C39’
60.				Val68	$C^{\gamma2}$	¶C41’
61.			D2R1	Tyr100	O^{η}	†C21’
62.				Leu101	$C^{\delta2}$	†C22’, †C23’
63.				Ala104	C	†C18’
64.					O	†C18’
65.					C^{β}	†C18’, †C19’
66.				Tyr105	N	C18’
67.					C^{γ}	†C16’
68.					$C^{\delta1}$	†C13’, †C16’, ¶C40’
69.					$C^{\varepsilon1}$	†C13’, †C15’, †C16’, ¶C40’
70.					$C^{\varepsilon2}$	†C22’
71.					C^{ζ}	†C16’, ¶C40’, ¶C41’, ¶C42’
72.					O^{η}	¶C42’, ¶C43’, ¶C44’
73.				Ala108	O	*C2’, *O2’, *C3’, *O4’, *C10’
74.					C^{β}	†C12’, †C13’, †C14’, †C15’, †C16’
75.				Leu109	C^{α}	*C2’, *C3’
76.					O	*C3’
77.					$C^{\delta2}$	*C3’, *O3’
78.				Met111	C	*C1’
79.					O	*C1’
80.					Cε	*O4’
81.				Tyr112	N	*C1’, *C2’
82.					C^{α}	*C1’
83.			D1R2	Gln115	C^{δ}	*O3P’, *C1’
84.					$O^{\varepsilon1}$	*O3P’, *C1’
85.					$N^{\varepsilon2}$	*O3P’, *C1’
86.				Ile137	$C^{\gamma1}$	¶C31’
87.					$C^{\gamma2}$	*O3P’

88.					$C^{\delta1}$	¶C31', ¶C32'
89.				Phe150	$C^{\delta1}$	¶C42'
90.					$C^{\varepsilon1}$	¶C38', ¶C39', ¶C40', ¶C42'
91.					C^{ζ}	¶C37', ¶C38', ¶C39'
92.				Tyr164	C	¶C45'
93.					O	¶C45'
94.				Asp165	N	¶C45'
95.					C^{α}	¶C45'
96.					C	¶C45'
97.				Met166	C^{ε}	†C15'
98.			D2R2	Met173	C^{β}	†C24'
99.					C^{γ}	†C20'
100.					C^{ε}	†C20'
101.				Ile174	$C^{\gamma1}$	†C25'
102.			D1R1	Phe39	$C^{\varepsilon1}$	†C43', †C44'
103.					C^{ζ}	†C43', †C44', †C45', †C46', †C47'
104.					$C^{\varepsilon2}$	†C47'
105.				Leu42	$C^{\delta1}$	†C47', †C48'
106.			D2R1	Tyr100	O^{η}	¶C09', ¶C10'
107.				Leu101	$C^{\delta2}$	¶C15'
108.				Ala104	C^{β}	¶C09', ¶C10'
109.				Tyr105	$C^{\varepsilon2}$	¶C12', ¶C13'
110.	GOT	-3.13			O^{η}	¶C13', ¶C18', ¶C19'
111.					$C^{\delta1}$	†C39'
112.					$C^{\varepsilon1}$	†C39'
113.				Ala108	C^{β}	†C36', †C37', †C38'
114.					O	†C49', †C50'
115.				Leu109	$C^{\delta2}$	†C40', †C41', †C42', †C48'
116.					C^{α}	†C48', †C49'
117.					C	†C49'
118.					O	†C49'

119.				Met111	S^{δ}	*C27', *O31'
120.					C^{ε}	*C27', *O29'
121.					C	†C49'
122.				Tyr112	$C^{\delta1}$	†C47'
123.					C^{β}	†C48', †C49'
124.					N	†C49'
125.					C^{α}	†C49'
126.			D1R2	Gln115	$O^{\varepsilon1}$	†C47'
127.				Ile137	$C^{\delta1}$	†C43', †C44', †C45'
128.					$C^{\gamma1}$	†C45'
129.					$C^{\gamma2}$	†C45', †C46'
130.				Val146	$C^{\gamma1}$	*O34'
131.				Tyr164	C	¶C20'
132.					O	¶C20'
133.				Asp165	$O^{\delta1}$	¶C17'
134.					C^{α}	¶C19'
135.					C	¶C19'
136.				Met166	N	¶C19'
137.				Ala168	C^{β}	*O32', *C01'
138.			D2R2	Met173	S^{δ}	*O03', *C04', *O05', *C06'
139.					C^{ε}	*C04',*O05', *C06', *C07'
140.					C^{β}	¶C13', ¶C14'
141.					C^{γ}	¶C13'
142.				Ile174	$C^{\gamma1}$	¶C16', ¶C17'
143.				Lys177	C^{ε}	¶C10'
144.	H3T	-2.78	D1R1	Leu42	$C^{\delta1}$	†C50'
145.				Val68	$C^{\gamma1}$	¶C20'
146.			D2R1	Ala73	C^{β}	¶C15'
147.				Tyr100	$C^{\varepsilon1}$	*O25'
148.					$C^{\varepsilon2}$	¶C11'
149.					C^{ζ}	*O25'

150.					O^{η}	*O03', *C04', *O05, ¶C11', *O25'
151.				Leu101	$C^{\delta1}$	¶C20'
152.					$C^{\delta2}$	¶C12', ¶C13', ¶C20'
153.				Ala104	C^{α}	*O03', C06'
154.					C	¶C06'
155.					O	¶C06'
156.					C^{β}	*O03', *C04', ¶C06'
157.				Tyr105	$C^{\varepsilon2}$	¶C12', ¶C19', ¶C20'
158.					O^{η}	¶C18', ¶C19'
159.				Ala108	O	†C48'
160.					C^{β}	†C39', †C40'
161.				Met111	C	†C48'
162.					O	†C48'
163.					C^{β}	†C46', †C48'
164.					C^{ε}	†C36', †C37', †C38', †C39'
165.				Tyr112	N	†C48'
166.					C^{α}	†C48', †C50'
167.					C^{β}	†C50'
168.					C^{γ}	†C50'
169.					$C^{\delta1}$	†C50'
170.			D1R2	Gln115	C^{δ}	†C47'
171.					$O^{\varepsilon1}$	†C48'
172.					$N^{\varepsilon2}$	†C47'
173.				Ile137	$C^{\gamma1}$	†C43'
174.				Leu148	$C^{\delta1}$	†C42', †C43'
175.				Tyr164	C	¶C16', ¶C17'
176.					O	¶C16', ¶C17'
177.				Asp165	N	¶C16', ¶C17'
178.					C^{α}	¶C16', ¶C17'
179.					C	¶C17'
180.					$O^{\delta1}$	¶C16'

181.				Met166	N	¶C17’
182.			D2R2	Met173	C^{β}	¶C09’
183.					C^{γ}	¶C08’, ¶C09’, ¶C10’
184.					S^{δ}	¶C08’, ¶C09’
185.					C^{ε}	¶C08’, ¶C09’, ¶C10’
186.				Ile174	$C^{\gamma 1}$	¶C15’, ¶C16’
187.					$C^{\delta 1}$	¶C16’
188.				Lys177	C^{ε}	¶C10’
189.					N^{ζ}	*P23’, *O05’, *O24’, *O25’
190.				Glu180	C^{δ}	*O24’
191.					$O^{\varepsilon 2}$	*P23’, *O24’
192.				Ile200	$C^{\gamma 1}$	*C27’, *C30’
193.					$C^{\delta 1}$	*O26’, *C27’
194.				Gln203	C^{δ}	*C30’, *O31’
195.					$O^{\varepsilon 1}$	*C30’, *O31’
196.					$N^{\varepsilon 2}$	*C30’, *O31’

†β tail, ¶γ tail and *Head region. Abbreviations: EFBE, estimated free energy of binding.

Supplementary Figures

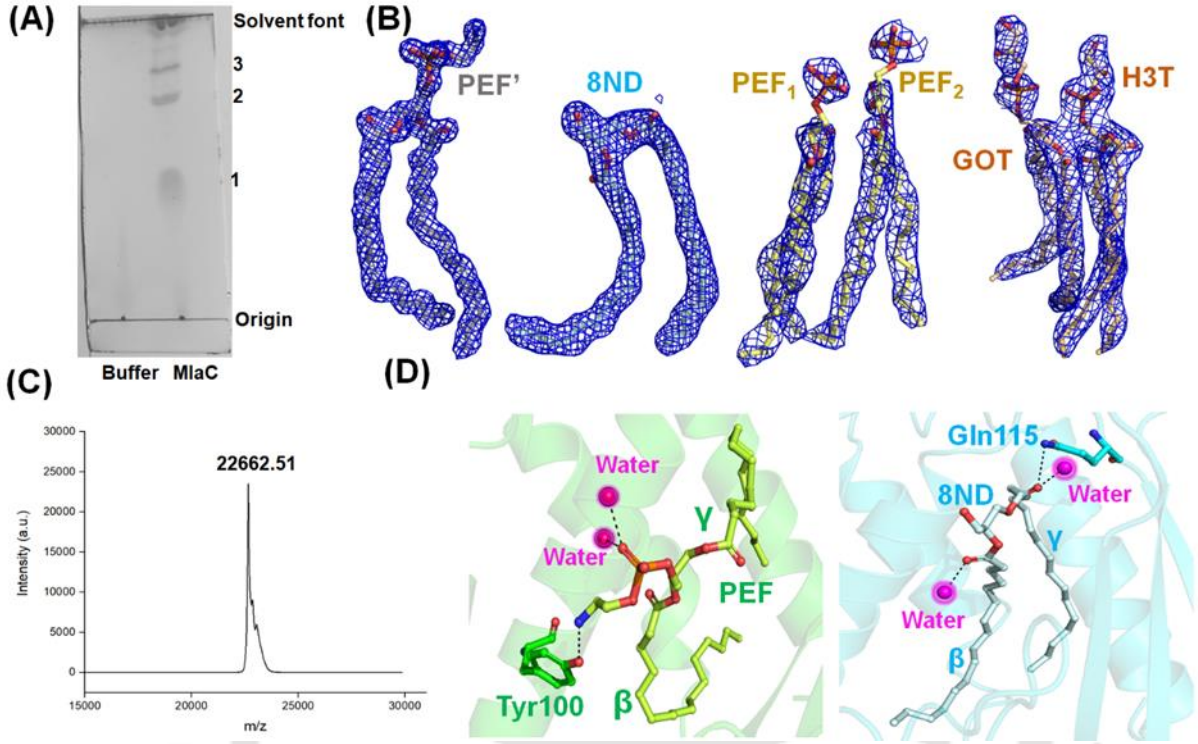


Figure B1. Analysis of the endogenously bound PLs. *(A) TLC of PLs bound to EcMlaC. Three distinct bands (1, 2, 3) corresponding to three different PLs are visible on the plate. (B) The composite omit maps of PEF' (first), 8ND (second), PEF_1-PEF_2 (third) and GOT-H3T (fourth) bound to MlaC orthologs (PDB ids: 2QGU, 5UWA, 5UWB and 6HSY, respectively). The maps are contoured at 1.0 σ and displayed in a blue mesh. (C) Mass spectrometric analysis of EcMlaC using MALDI-TOF. The obtained molecular weight of the protein in Daltons (unit) is labeled. (D) Polar interactions involving the head group of PEF (left) and 8ND (right). The head group of PEF establishes hydrogen bonds (dotted lines) with the residue Tyr100 and water molecules (magenta sphere). The head group of 8ND establishes hydrogen bonds (dotted line) with the residue Gln115 and water molecules (magenta sphere).*

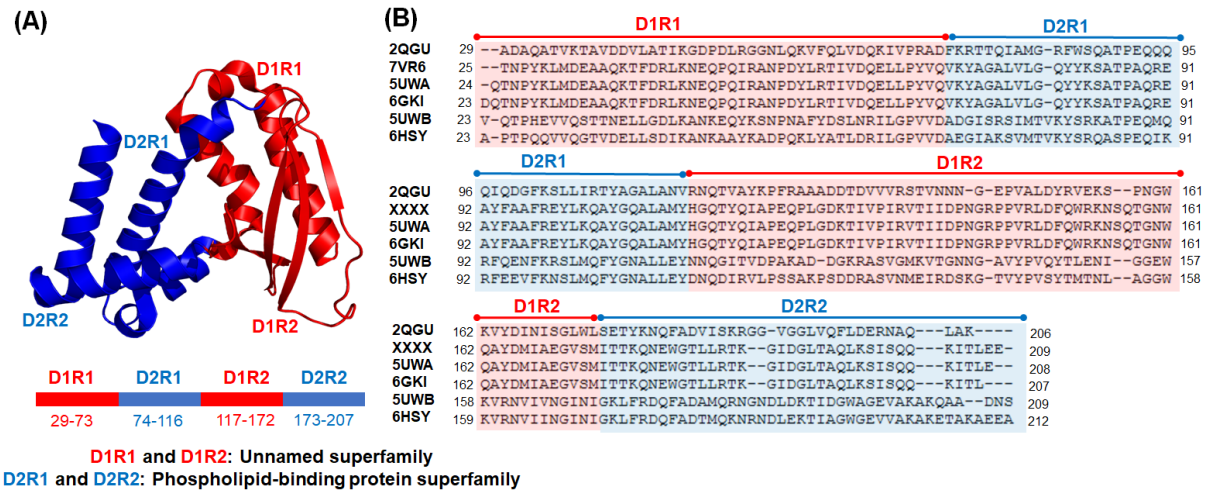


Figure B2. Domain arrangement in EcMlaC and its orthologs. *(A) Domain prediction provided by the CATH database for MlaC from Ralstonia solanacearum (PDB id: 2QGU). (Top) Three-dimensional structure of MlaC, highlighting the domain arrangement. The protein consists of two main domains belonging to the unnamed (D1) and phospholipid-binding protein (D2) superfamily, respectively. Each of the domains is divided into two subdomains- D1R1 & D1R2 (red) and D2R1 & D2R2 (blue), arranged in a discontinuous fashion. (Bottom) Schematic representation of the segmented domain arrangement in MlaC. D1R1 and D1R2 subdomains, along with their boundaries, are highlighted in red, while D2R1 and D2R2 subdomains are in blue. (B) Structure-based multiple sequence alignment of MlaC orthologs. All the six crystal structures (PDB ids: 2QGU, 7VR6, 5UWA, 6GKI, 5UWB, and 6HSY) of MlaC orthologs were considered in order to obtain the respective subdomain boundaries. The D1R1 and D1R2 subdomains are highlighted in red, while the D2R1 and D2R2 subdomains are in blue.*

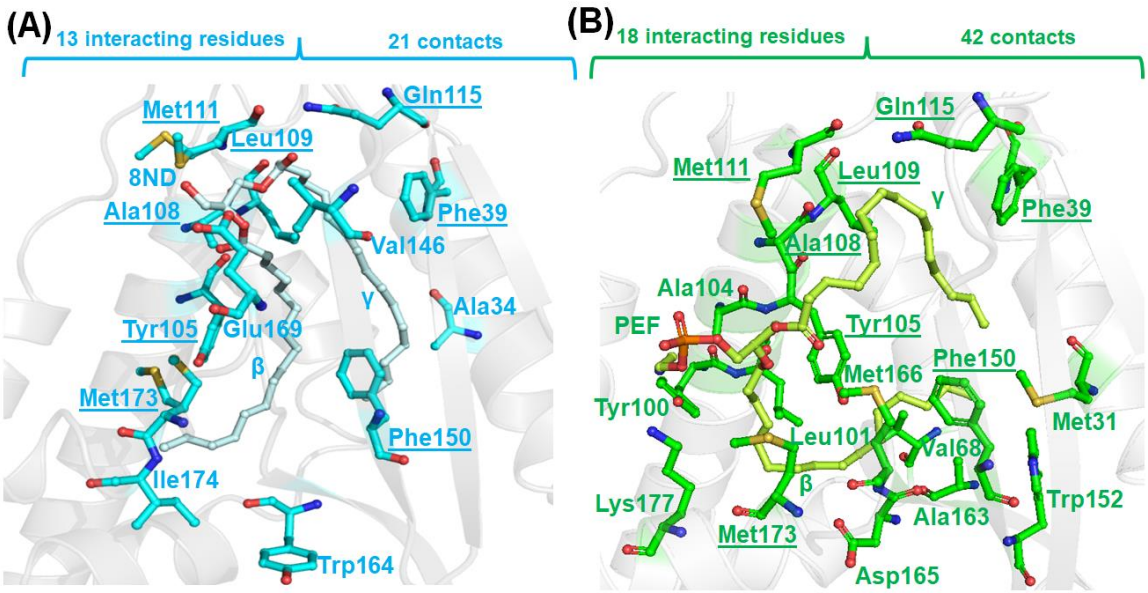


Figure B3. Comparison of the residues involved in non-polar contacts in EcMlaC_P2₁2₁2₁_holo1 and EcMlaC_H3_holo1. *(A) Overview of the contacts between EcMlaC_P2₁2₁2₁_holo1 and 8ND. The endogenously bound 8ND makes 21 non-polar contacts involving thirteen amino acid residues. (B) Overview of the non-polar contacts between EcMlaC_H3_holo1 and PEF. The endogenously bound PEF makes 42 non-polar contacts with EcMlaC_H3_holo1 involving eighteen amino acid residues. The residues commonly involved in interacting with the PEF and 8ND molecules have been underlined.*

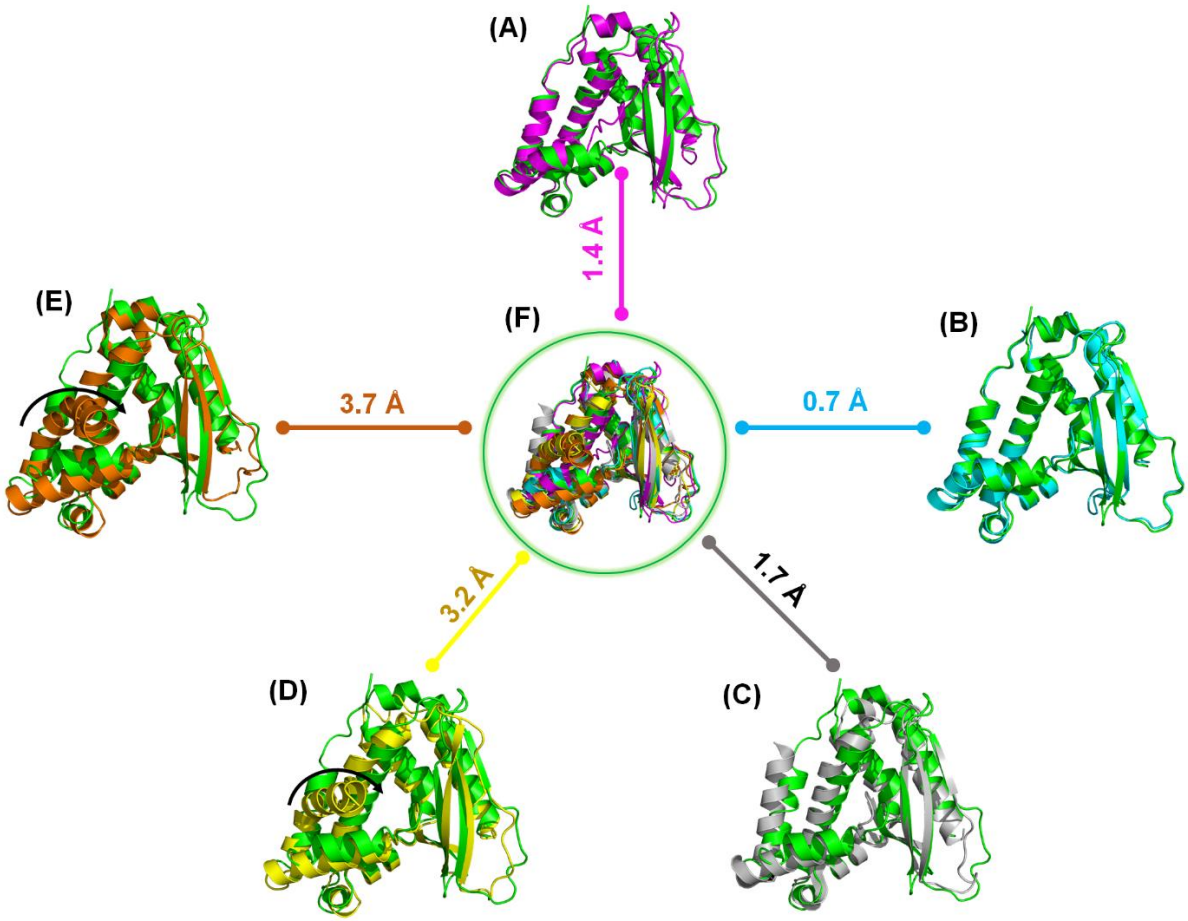


Figure B4. Structural comparison of EcMlaC_H3_holo1 and its orthologs. *Structural superimposition of EcMlaC_H3_holo1 (green) with (A) EcMlaC_P3$_1$21_apo (magenta), (B) EcMlaC_P2$_1$2$_1$2$_1$_holo1 (cyan), (C) RsMlaC_holo1 (grey), (D) PpMlaC_holo2 (yellow), and (E) PaMlaC_holo2 (orange). The bending in the α8 helix in PpMlaC_holo2 and PaMlaC_holo2 is denoted by a curved black arrow. The respective RMSD is mentioned along these lines. (F) Structural superimposition of all the MlaC orthologs.*

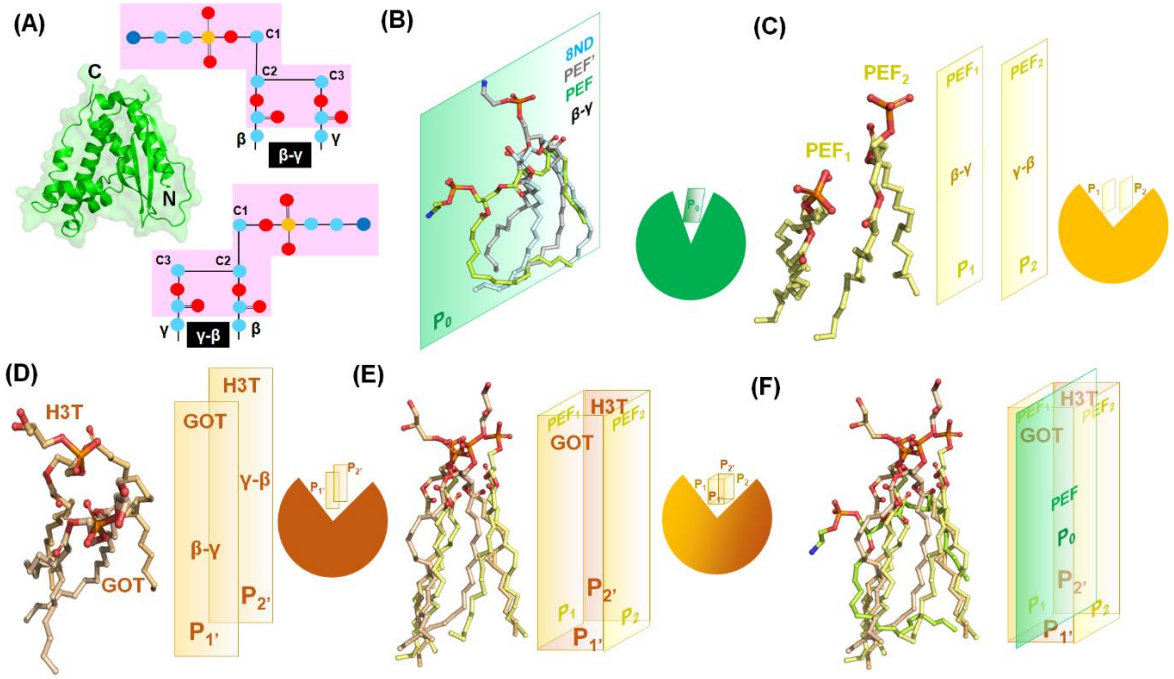


Figure B5. Analysis of ligand conformation and orientation. *(A) Conformation of ligands in MlaC. (Middle) The crystal structure of the protein is fixed in C-terminal to N-terminal orientation. (Top) The ligand is in β-γ conformation when the β-tail comes first, followed by the γ-tail. (Bottom) The ligand is in γ-β conformation when the γ-tail comes first, followed by the β-tail. The C1, C2 and C3 atoms of the glycerol moiety are marked and the head group of the PL is highlighted in pink. The C, N, O and P atoms are represented as cyan, blue, red and yellow circles, respectively. (B) Orientation of ligand-binding planes in Group II members. In single-liganded states, the bound PLs are present in β-γ conformation in one plane, P_0 (green parallelogram). (Right bottom) A schematic diagram showing the orientation of P_0 in MlaC (green partial-circle). (C-D) Orientation of ligand-binding planes in Group III members. In double-liganded states of MlaC, the bound PLs are present in β-γ and γ-β conformations in two non-overlapping binding planes P_1 and P_2 (yellow rectangles) or $P_{1'}$ and $P_{2'}$ (orange rectangles), respectively. (Right bottom) A schematic diagram showing the orientation of the binding planes in MlaC (yellow/orange partial-circle). (E) Superimposition of the ligands and their bindings planes in Group III members. (Right bottom) A schematic diagram showing the binding planes of ligands of Group III members in MlaC (orange-yellow partial-circle). (F) Superimposition of the ligands and their binding planes in Group II and III members.*

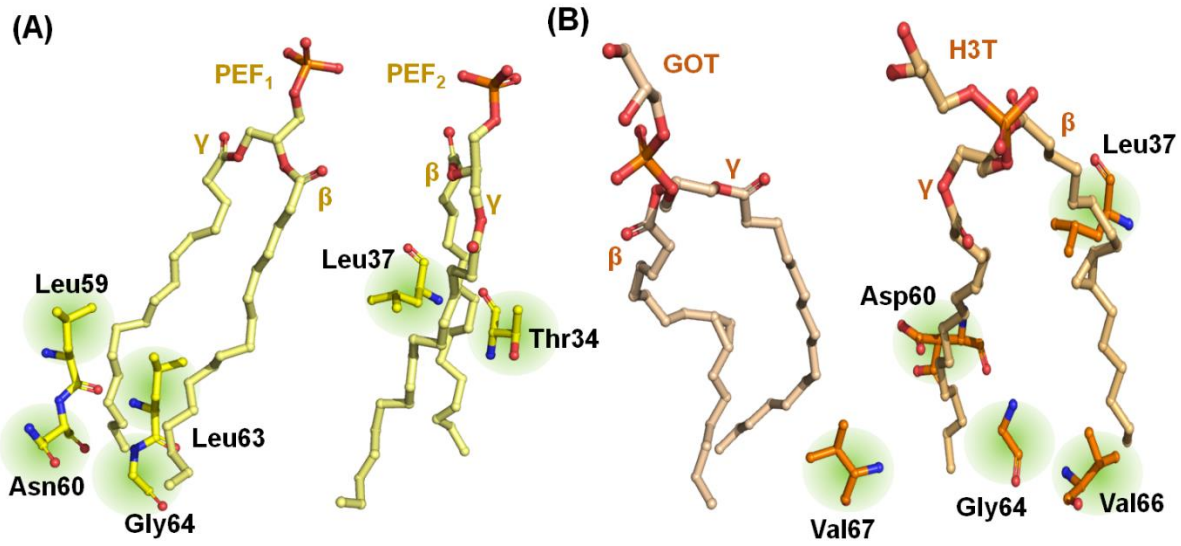


Figure B6. Structural details of the ligand interaction with the D1R1 subdomain of Group III. *(A) Interaction of PEF_1 (left) and PEF_2 (right) with the D1R1 subdomain in PpMlaC_holo2. (B) Interaction of GOT (left) and H3T (right) with the D1R1 subdomain in PaMlaC_holo2. The interacting residues are represented as ball-and-stick models. Further details of the interactions are provided in Table B3.*

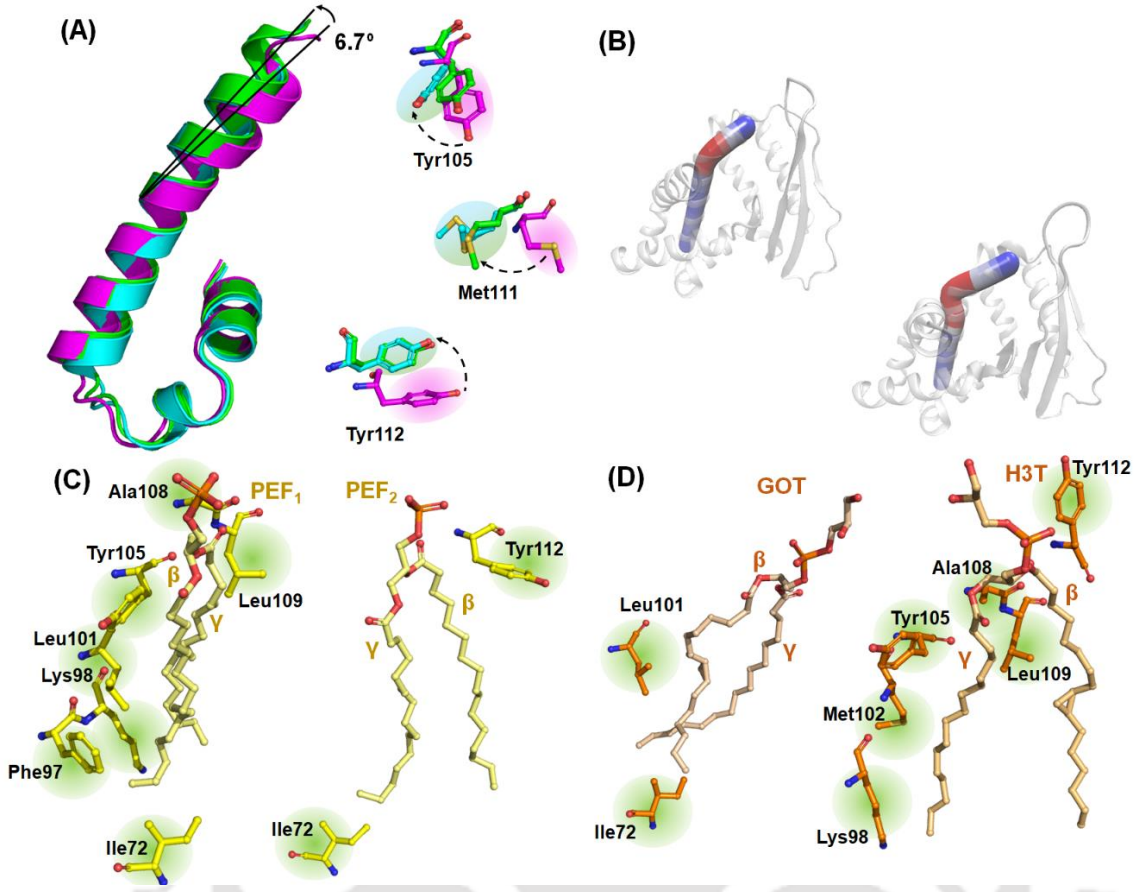


Figure B7. Structural details of the ligand interaction with the D2R1 subdomain of Group III. *(A) (Left) Structural superimposition of the D2R1 subdomain of EcMlaC_P3₁21_apo (magenta), EcMlaC_P2₁2₁2₁_holo1 (cyan), and EcMlaC_H3_holo1 (green) displaying the ligand-binding shifts of the α6 helix. (Right) The residues showing significant movements. (B) The curvatures of the α6 helices of PpMlaC_holo2 (top left) and PaMlaC_holo2 (bottom right). The curvatures have been color-coded using the blue-white-red scheme. As per the scheme, blue, white, and red colors signify straight, curvature, and angled helix, respectively. (C) Interaction of PEF_1 (left) and PEF_2 (right) with the D2R1 subdomain in PpMlaC_holo2. (D) Interaction of GOT (left) and H3T (right) with the D2R1 subdomain in PaMlaC_holo2. The interacting residues are represented as ball-and-stick models. Further details of the interactions are provided in Table B3.*

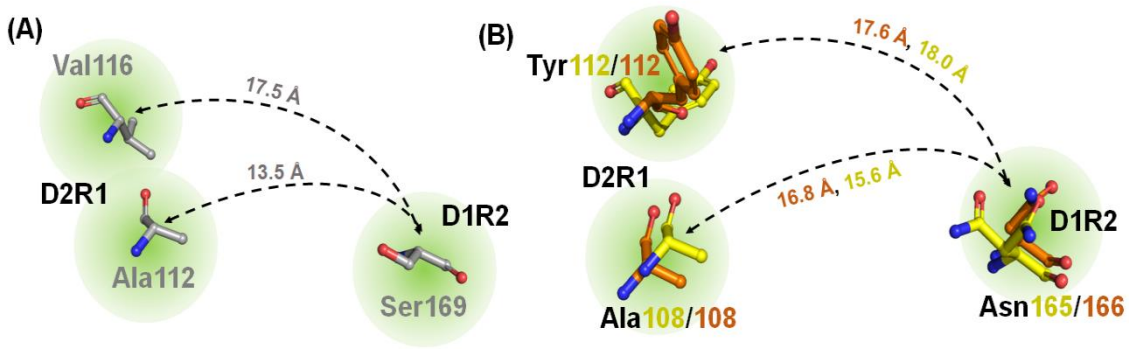


Figure B8. Movement of residues regulating the opening and closing of MlaC. *(A-B) Movement of the residues from the D1R2 and D2R1 subdomains. The distances between the residues Ala112 & Ser169 and Val116 & Ser169 in RsMlaC_holo1 (grey); Ala108 & Asn165 and Tyr112 & Asn165 in PpMlaC_holo2 (yellow); Ala108 & Asn166 and Tyr112 & Asn166 in PaMlaC_holo2 (orange) are mentioned. The movements are denoted by double-headed dotted arrows.*

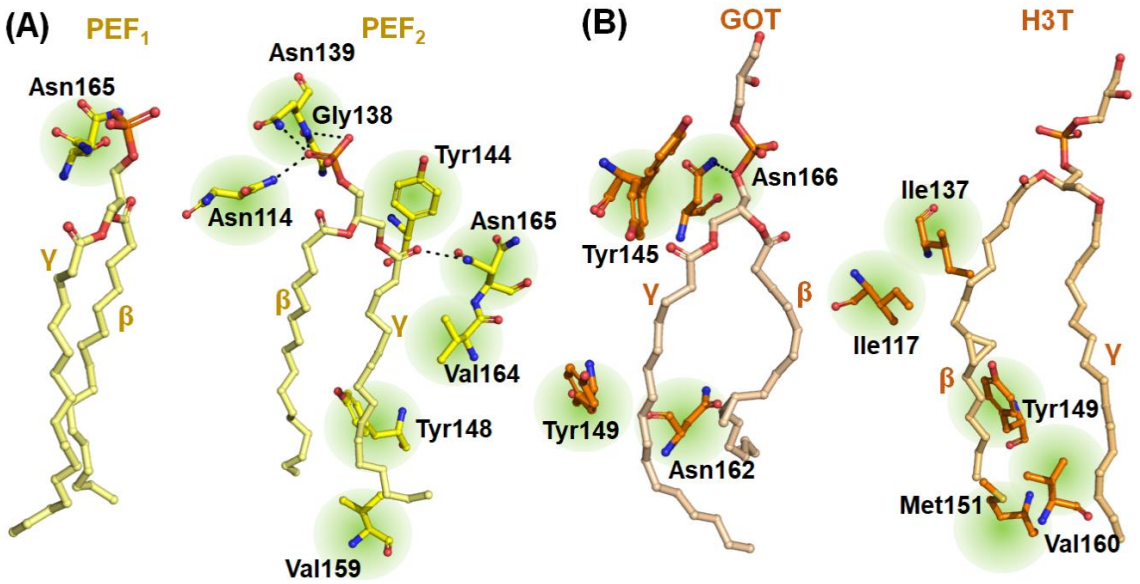


Figure B9. Structural details of the ligand interaction with the D1R2 subdomain of Group III members. *(A) Interaction of PEF_1 (left) and PEF_2 (right) with the D1R2 subdomain in PpMlaC_holo2. (B) Interaction of GOT (left) and H3T (right) with the D1R2 subdomain in PaMlaC_holo2. The interacting residues are represented as ball-and-stick models. The hydrogen bonds are denoted by dotted lines. Further details of the interactions are provided in Table B3.*

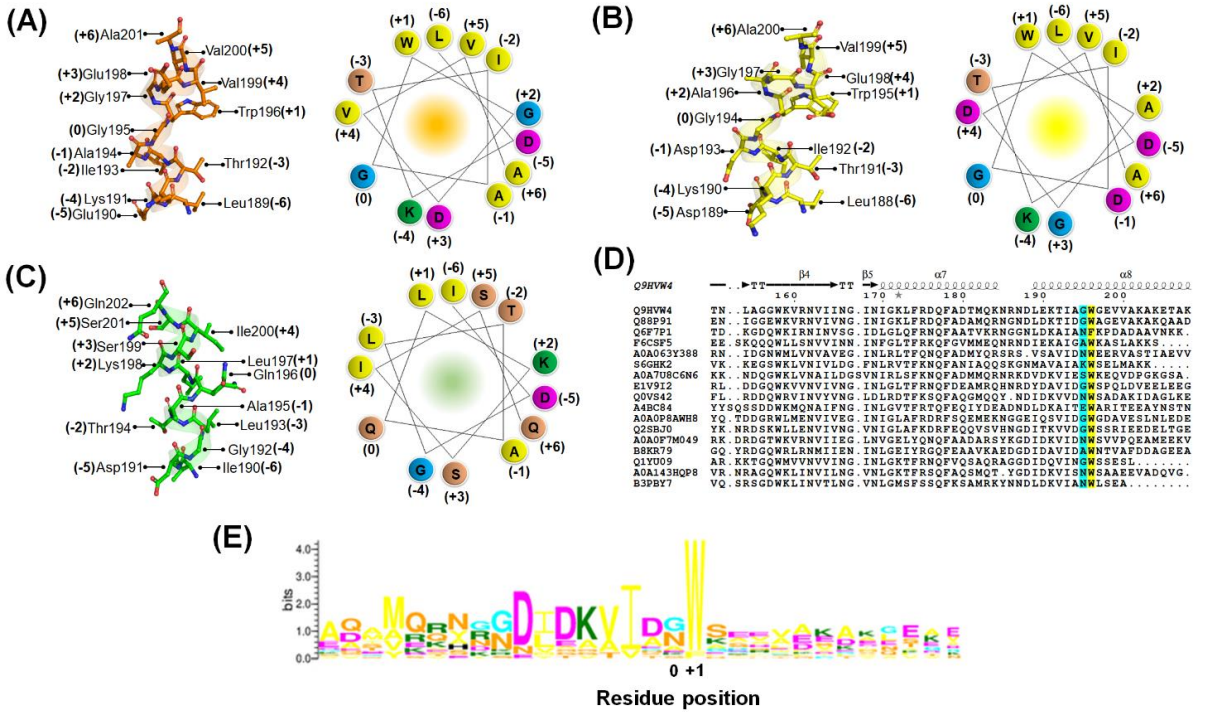


Figure B10. Kink analysis of MlaC orthologs. *(A-C) Ball-and-stick (left) and helical wheel (right) representations of residues around the kink forming residues of α8 helix in PaMlaC_holo2, PpMlaC_holo2, and EcMlaC_H3_holo1, respectively. A total of thirteen residues around a kink (position -6 to +6) were considered for the analysis. The kink residue is numbered as zero, while the residues around it are numbered relatively. The hydrophobic, polar uncharged, polar positively charged, polar negatively charged, and glycyl residues are colored in yellow, orange, green, magenta, and cyan, respectively. The positions of the residues around the kink are mentioned in brackets. (D) The extent of conservation of Gly195 and Trp196 is highlighted in cyan and yellow, respectively, in Group III members. (E) Sequence logo representation of the conservation of the kink forming residues of α8 helix in Group III. The figure suggests the residue Gly195 is not conserved at the kink (0) position, while Trp196 is well conserved at +1 position. However, the residue Trp196 is not exclusive to Group III and is replaced by phenylalanine. The amino acids are color-coded as per the coloring scheme followed in the helical wheel representation. Details of the sequences used for the analysis are provided in Table B8.*

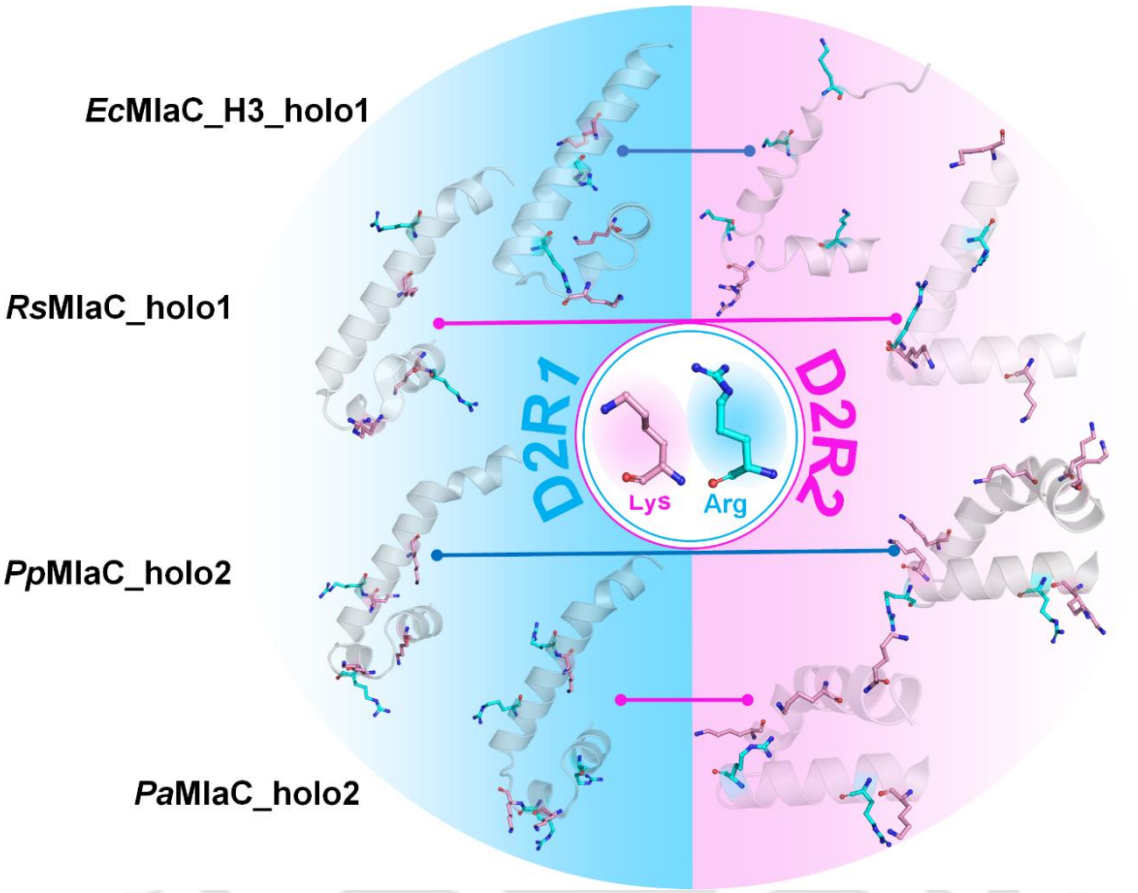


Figure B11. Overview of the distribution of arginine and lysine residues in the D2R1 and D2R2 subdomains of EcMlaC orthologs. *Distribution of arginine and lysine residues in the D2R1 (cyan) and D2R2 (pink) subdomains. The lysine and arginine residues are represented as ball-and-stick models in pink and cyan, respectively. The remaining parts of the subdomains are represented as a cartoon in grey.*

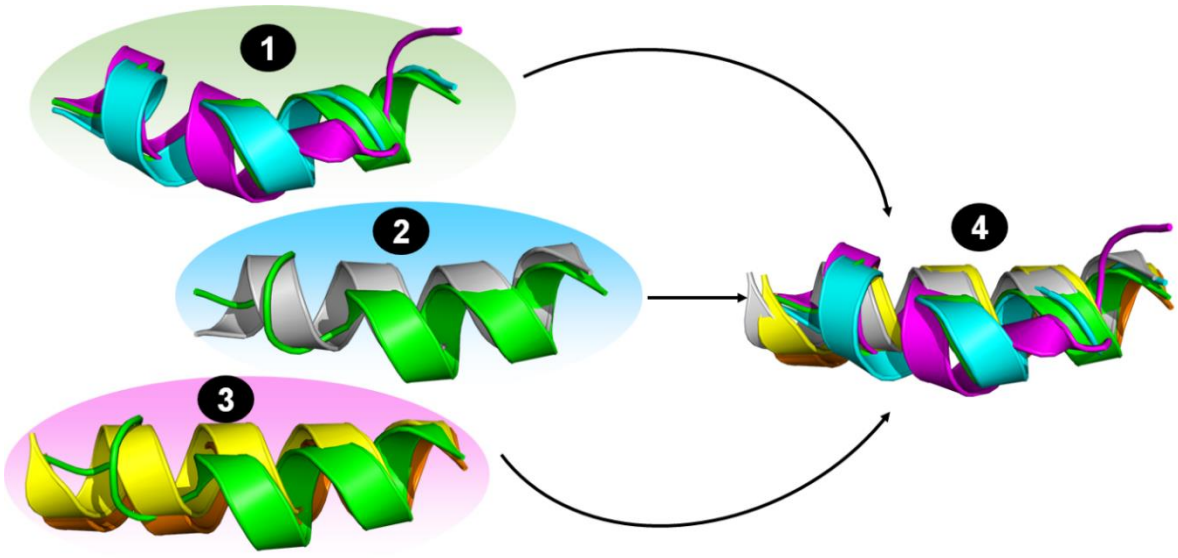


Figure B12. Structural details of the ligand interaction with the D2R2 subdomain of Group III. *(A) Structural superimposition of the α7 helix of EcMlaC_H3_holo1 (green) with (1) EcMlaC_P3$_1$21_apo (magenta) and EcMlaC_P2$_1$2$_1$2$_1$_holo1 (cyan), (2) RsMlaC_holo1 (grey), (3) PpMlaC_holo2 (yellow) and PaMlaC_holo2 (orange), and (4) all MlaC orthologs.*

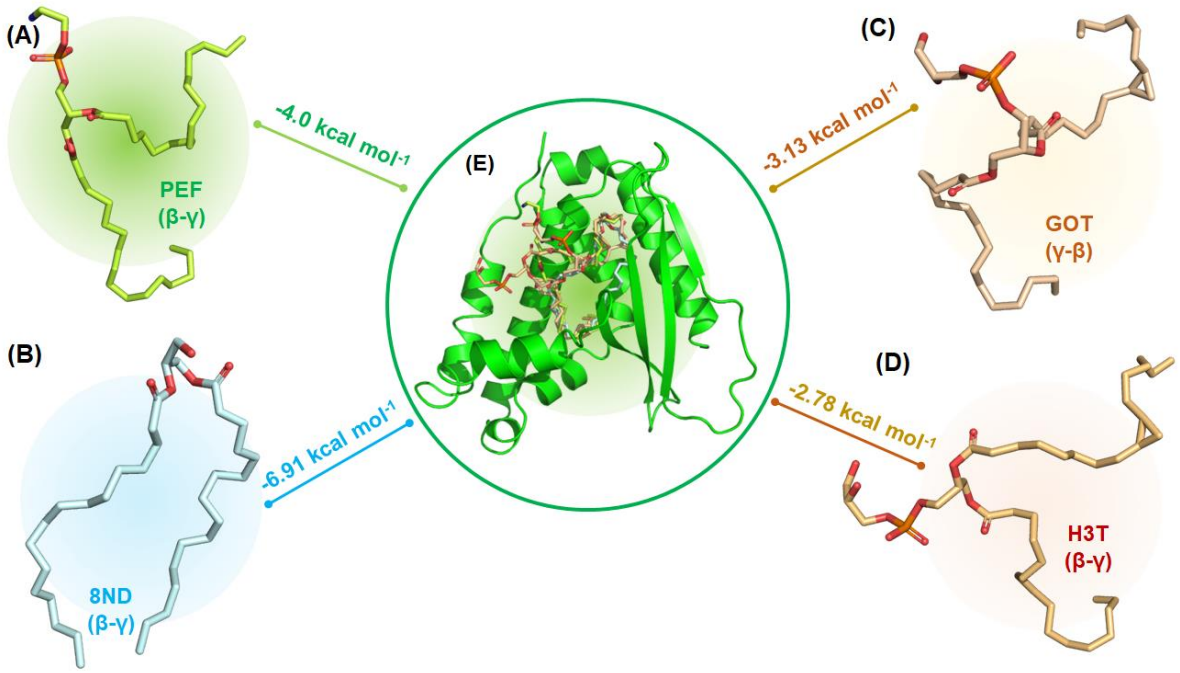


Figure B13. Overview of the PL confirmations and orientations docked to EcMlaC_H3_holo1. *(A-D) Conformation of the docked PEF (limon), 8ND (palecyan), GOT (wheat) and H3T (light orange) molecules, respectively. (E) Superimposition of the docked PLs. The binding energies obtained after docking are provided along these lines.*

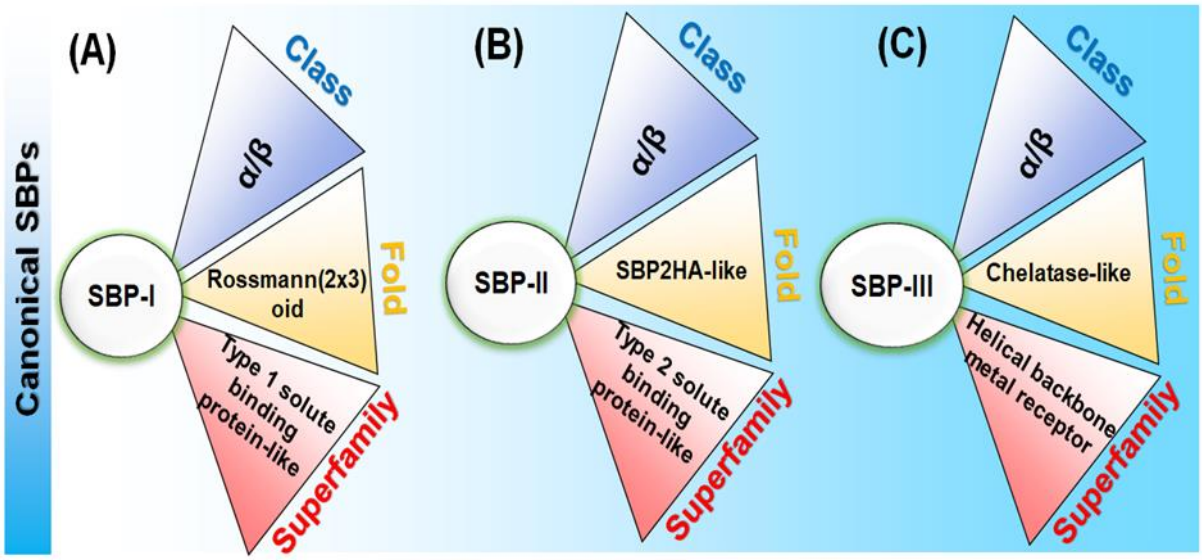


Figure B14. An overview of the ancestries of canonical SBPs. *(A-C) Ancestries of SBP-I, -II, and -III. Class, fold, and superfamily are highlighted in blue, yellow, and red, respectively.*

Appendix C

Supplementary data to Chapter 5

Supplementary tables

Table C1. X-ray crystallographic data collection and refinement statistics for *Ec*MlaD. Values in parenthesis are for the highest resolution shell.

Parameters	***Ec*MlaD_P4$_1$2$_1$2**	***Ec*MlaD_MlaD1^{I222}**	***Ec*MlaD_MlaD1^{P21221}**
Wavelength (Å)	1.5418	1.5418	1.5418
Temperature (K)	100	100	100
Space group	*P*4$_1$2$_1$2	*I*222	*P*2$_1$22$_1$
Unit-cell parameters (Å, °)	*a*=*b*=64.28, *c*=231.06, *α*=*β*=*γ*=90	*a*=60.28, *b*=107.29, *c*=117.92, *α*=*β*=*γ*=90	*a*=60.42, *b*=106.95, *c*=117.12, *α*=*β*=*γ*=90
Resolution (Å)	77.02-3.20 (3.28-3.20)	79.36-2.30 (2.36-2.30)	78.98-2.70 (2.77-2.70)
No. of observed reflections	86434 (15458)	137255 (13294)	190212 (24884)
No. of unique reflections	8680 (1516)	17387 (1678)	21557 (2805)
Mn(I) CC(1/2)	0.994 (0.934)	0.993 (0.936)	0.996 (0.943)
Completeness (%)	99.8 (100.0)	99.9 (100.0)	100.0 (100.0)
V_M (Å^3 Da^{-1})	2.14	2.46	2.44
Solvent content (%)	42.64	49.95	49.57
Mean I/σ(I)	9.6 (3.5)	13.6 (3.3)	12.6 (3.9)
R_{merge}† (%)	16.1 (53.1)	7.6 (54.9)	14.1 (57.5)
R_{pim} (%)	7.1 (24.2)	4.3 (30.7)	7.2 (30.0)
R_{meas} (%)	17.6 (58.6)	8.8 (63.1)	15.8 (65.0)
Multiplicity	10.0 (10.2)	7.9 (7.9)	8.8 (8.9)
R_{work}/R_{free} (%)	20.64/27.54	22.50/27.25	20.41/26.11
Protein model			
No. of subunits in ASU	3	3	6
Protein atoms	2651	2400	4852
Water molecules	22	81	173
Other molecules	None	CO2 (1), EDO (1)	Mg^{2+} (1), CO2 (1)
Deviation from ideal geometry			
Bond length (Å)	0.010	0.013	0.012
Bond angles (°)	1.812	1.867	1.827

Average *B*-factor (Å²)			
Protein atoms	45.77	35.33	27.59
Water molecules	38.63	52.33	29.30
Ramachandran plot (%)			
Favoured	95.18	97.73	98.53
Allowed	4.22	2.27	1.47
Remaining	0.60	0	0
PDB id	8HQA	8HPZ	8HQ9

† $R_{merge} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I(hkl)$ is the intensity of reflection *hkl,* $\sum hkl$ is the sum overall reflections and $\sum i$ is the sum over *i* measurements of reflection *hkl.*

Table C2. Structure-based domain profiling of domains. The TM helix boundary of the cryo-EM structures and AlphaFold models were analyzed using the DSSP server (Kabsch & Sander, 1983). The PLP region is marked in bold. The α-helix(s) and β-strand(s) within the PLP region are designated as $H_{L(\text{helix number})}$ and $\beta_{L(\text{strand number})}$, respectively.

S. No	MlaD domains (PDB id)	UniProt id	TM helix boundary (length)	Domain boundary (length as per PDB)	Structural component	PDBsum		DSSP	Secondary structural content (% of helix, sheet, others)
						Hera	PROMOTIF		
1.	*Ec*MlaD_MlaD1^{P41212} (8HQA)	P64604	2-28 (27)	35-140 (106)	β1	38-46	39-45	39-45	0, 45.3, 54.7
					β2	56-60	57-60	57-60	
					β3	63-74	63-73	63-73	
					β4	79-88	80-87	80-87	
					β5	97-104	98-103	98-103	
					PLP	**105-108**	**105-108**	**105-108**	
					β6	109-115	110-115	110-115	
					β7a	132-134	132-134	132-134	
					β7b	136-138	136-138	136-138	
2.	*Ec*MlaD_MlaD1^{I222} (8HPZ)			37-141 (105)	β1	38-46	39-45	39-45	2.9, 44.8, 52.4
					β2	56-60	57-60	57-60	
					β3	63-74	63-73	63-73	
					β4	79-88	80-87	80-87	
					β5	97-104	98-103	98-103	
					PLP	**105-108**	**105-108**	**105-108**	
					β6	109-115	110-115	110-115	
					H1	120-124	121-123	121-123	
					β7a	132-134	132-133	132-133	
					β7b	136-138	136-138	136-138	
3.	*Ec*MlaD_MlaD1^{P21221} (8HQ9)			37-140 (104)	β1	38-46	39-45	39-45	2.9, 43.3, 53.8
					β2	56-60	57-60	57-60	
					β3	63-74	63-73	63-73	
					β4	79-88	80-87	80-87	

					β5	97-104	98-103	98-103	
					PLP	**105-108**	**105-108**	**105-108**	
					β6	109-115	110-115	110-115	
					H1	120-124	121-123	121-123	
					β7a			132-133	
					β7b	136-138	136-138	136-138	
4.	*Ec*PqiB_MlaD1 (5UVN)	P43671	21-40 (20)	42-147 (106)	β1	43-49	44-49	44-49	9.5, 36.8, 53.8
					β2	60-64	61-64	61-64	
					β3	67-78	67-77	67-77	
					β4	83-91	84-90	84-90	
					H1	92-98	92-97	92-97	
					β5	103-104	103-106	103-106	
					PLP	**109-124**	**109-124**	**109-124**	
					H_{L1}	**120-123**	**121-124**	**121-124**	
					β6	127-131	128-131	128-131	
					β7	143-147	144-146	144-146	
5.	*Ec*PqiB_MlaD2 (5UVN)			156-273 (118)	β1	158-163	158-163	158-163	7.6, 39.8, 52.5
					β2	174-179	175-178	175-178	
					β3	181-192	181-191	181-191	
					β4	197-205	198-204	198-204	
					H1	208-212	209-211	209-211	
					β5	217-221	217-221	217-221	
					PLP	**223-238**	**223-238**	**223-238**	
					$β_{L1}$	**225-227**	**224-227**	**225-227**	
					$β_{L2}$	**234-236**	**234-236**	**234-236**	
					H_{L1}	**238-245**	**239-244**	**239-244**	
					β6	247-251	247-251	247-251	
					β7	266--270	267-269	267-269	20.4, 33.3, 46.3
6.	*Ec*PqiB_MlaD3 (5UVN)			284-412 (128)	β1	286-292	286-292	286-292	
					β2	303-307	304-307	304-307	
					β3	310-316	310-315	310-315	
					H1	322-333	323-332	323-332	
					β4	335-342	335-341	335-341	
					H2	343-346	343-345	343-345	

					H3	355-365	356-364	356-364	
					β5	369-375	370-374	370-374	
					PLP	**375-382**	**375-382**	**375-382**	
					β6	381-387	382-386	382-386	
					β7	405-407	406-407	406-408	
7.	*Ec*YebT_MlaD1 (6V0C)	P76272	25-43 (19)	46-149 (104)	β1	49-53	49-53	49-53	10.6, 40.4, 49.0
					β2	64-68	65-68	65-68	
					β3	71-82	71-82	71-82	
					β4	87-94	87-94	87-94	
					H1	95-102	96-101	96-101	
					β5	107-111	107-111	107-111	
					PLP	**114-129**	**114-129**	**114-129**	
					H_{L1}	**116-122**	**117-121**	**117-121**	
					β6	131-135	131-135	131-135	
					β7	143-147	144-146	144-146	
8.	*Ec*YebT_MlaD2 (6V0C)			160-272 (113)	β1	161-167	161-167	161-167	6.2, 35.4, 58.4
					β2	177-181	178-181	178-181	
					β3	184-193	184-193	184-193	
					β4	201-208	201-207	201-207	
					H1	209-212	209-211	209-211	
					β5	220-224	221-223	221-223	
					PLP	**224-244**	**224-244**	**224-244**	
					β_{L1}		**228-229**	**228-229**	
					β_{L2}		**238-239**	**238-239**	
					H_{L1}	**243-248**	**244-247**	**244-247**	
					β6	249-254	250-253	250-253	
					β7	266-271	267-271	267-271	
9.	*Ec*YebT_MlaD3 (6V0C)			279-382 (104)	β1	281-286	281-286	281-286	8.7, 41.3, 50.0
					β2	299-302	299-301	299-302	
					β3	305-316	305-316	305-316	
					β4	320-327	320-327	320-327	
					H1	332-334	332-334	332-334	
					β5	340-344	340-344	340-344	
					PLP	**345-361**	**345-361**	**345-361**	

					H_{L1}	**355-360**	**355-360**	**355-360**	
					β6	363-367	363-367	363-367	
					β7	376-378	376-378	376-378	
10.	*Ec*YebT_MlaD4 (6V0C)			391-499 (109)	β1	393-399	393-399	393-399	9.2, 44.0, 46.8
					β2	409-413	410-413	410-413	
					β3	416-427	416-426	416-426	
					β4	430-438	431-437	431-437	
					H1	441-446	442-445	442-445	
					β5	448-454	449-453	449-453	
					PLP	**455-479**	**455-479**	**455-479**	
					$β_{L1}$	**458-460**	**458-460**	**458-460**	
					$β_{L2}$	**467-469**	**467-469**	**467-469**	
					H_{L1}	**472-479**	**473-478**	**473-478**	
					β6	480-486	481-485	481-485	11.7, 28.8, 59.5
					β7	493-497	494-496	494-496	
11.	*Ec*YebT_MlaD5 (6V0C)			515-625 (111)	β1	517-521	517-521	517-521	
					β2	532-536	533-535	533-535	
					β3	542-547	542-547	542-547	
					β4	553-558	553-558	553-558	
					H1	563-567	564-566	564-566	
					β5	570-576	571-575	571-575	
					PLP	**577-601**	**577-601**	**577-601**	
					H_{L1}	**594-600**	**595-599**	**595-599**	
					β6	602-608	603-607	603-607	
					H2	611-617	612-616	612-616	
					β7	622-624	623-624	623-624	
12.	*Ec*YebT_MlaD6 (6V0C)			634-737 (104)	β1	635-641	635-641	635-641	2.9, 41.3, 55.8
					β2	651-655	652-655	652-655	
					β3	658-669	658-669	658-669	
					β4	674-682	674-681	674-681	
					β5	695-699	695-699	695-699	
					PLP	**700-718**	**700-718**	**700-718**	
					H_{L1}	**711-715**	**712-714**	**712-714**	
					β6	718-722	718-722	718-722	

					β7	732-734	732-733	732-733	
13.	*Ec*YebT_MlaD7 (6V0C)			746-862 (117)	β1	748-754	749-754	749-754	6, 36.8, 57.3
					β2	764-769	765-768	765-768	
					β3	771-782	771-781	771-781	
					β4	787-795	788-794	788-794	
					H1	798-802	799-801	799-801	
					β5	807-811	807-811	807-811	
					PLP	**813-830**	**813-830**	**813-830**	
					H_{L1}	**830-835**	**831-834**	**831-834**	
					β6	837-841	837-841	837-841	
					β7	856-861	857-861	857-861	
14.	*Pa*MlaD_MlaD1 (7CHA)	Q9HVW3	3-20 (18)	32-137 (106)	β1	38-46	39-46	39-46	0, 40.6, 59.4
					β2	57-60	58-60	58-60	
					β3a	63-66	63-66	63-66	
					β3b	67-74	68-73	68-73	
					β4	79-88	80-87	80-87	
					β5	98-104	98-103	98-103	
					PLP	**105-108**	**105-108**	**105-108**	
					β6	109-115	110-115	110-115	
					β7	131-133	132-133	132-133	
15.	*Ab*MlaD_MlaD1 (6Z5U)	V5V921	3-32 (30)	31-183 (153)	β1	38-46	39-46	39-46	22.2, 32.0, 45.8
					β2	56-60	57-60	57-60	
					β3	63-74	63-73	63-74	
					β4	79-88	80-87	79-87	
					H_{I1}	89-91	88-90	88-90	
					H_{I2}	95-109	96-108	96-108	
					H_{I3}	112-117	113-116	113-116	
					H_{I4}	119-135	120-134	120-134	
					β_{I1}	136-142	137-141	137-141	
					β5	147-151	147-150	147-150	
					PLP	**152-155**	**152-155**	**152-155**	
					β6	156-161	157-160	157-160	
					β7a	168-172	169-171	169-171	
					β7b	174-176	175-176	175-176	

16.	*Mt*Mce1A_MlaD1 (model)	Q79FZ9	15-37 (23)	44-149 (106)	β1	45-51	45-51	45-51	
					β2	62-67	63-66	63-66	
					β3	69-82	69-82	70-82	3.8, 50.9, 45.3
					β4	85-94	85-94	85-94	
					H1	95-100	96-99	96-99	
					β5	107-113	107-112	107-112	
					PLP	**114-116**	**113-116**	**113-116**	
					β6	118-124	119-124	119-124	
					β7	138-146	139-145	139-145	
17.	*Mt*Mce1B_MlaD1 (model)	O07414	3-28 (26)	35-139 (105)	β1	35-43	36-43	36-43	4.8, 46.7, 48.6
					β2	53-58	54-57	54-57	
					β3	60-71	60-71	60-71	
					H1	72-75	72-74	72-74	
					β4	76-85	76-84	76-84	
					β5	94-100	94-99	94-99	
					PLP	**101-104**	**101-104**	**101-104**	
					β6	105-111	106-111	106-111	
					β7a	126-128	127-128	127-128	
					H2	129-131	130-132	130-132	
					β7b	132-134	133-134	133-134	
18.	*Mt*Mce1C_MlaD1 (model)	O07415	9-32 (24)	38-136 (99)	β1	40-46	41-46	41-46	2.0, 45.5, 52.5
					β2	56-61	57-60	57-60	
					β3	63-74	63-74	63-74	
					β4	77-84	77-83	77-83	
					β5	94-100	94-99	94-99	
					PLP	**101-104**	**101-104**	**101-104**	
					β6	105-111	106-111	106-111	
					β7a	122-124	123-124	123-124	
					H1	125-127	126-128	126-128	
					β7b	128-130	129-130	129-130	
19.	*Mt*Mce1D_MlaD1 (model)	O07416	17-44 (28)	45-145 (101)	β1	45-53	46-53	46-53	2.0, 47.5, 50.5
					β2	63-68	64-66	64-67	
					β3	70-81	70-81	70-81	
					β4	84-93	84-92	84-92	

					β5	102-106	102-106	102-106	
					PLP	**109-112**	**109-112**	**109-112**	
					β6	115-119	115-119	115-119	
					β7a	132-134	132-134	132-134	
					H1	135-137	136-138	136-138	
					β7b	138-140	139-140	139-140	
20.	*Mt*Mce1E_MlaD1 (model)	O07417	-	50-153 (104)	β1	52-58	52-58	52-58	1.9, 46.2, 51.9
					β2	68-72	69-72	69-72	
					β3	75-86	75-86	75-86	
					β4	89-97	89-96	89-96	
					β5	107-113	107-112	107-112	
					PLP	**114-117**	**114-117**	**114-117**	
					β6	118-124	119-124	119-124	
					β7a	138-140	139-140	139-140	
					H1	141-143	142-144	142-144	
					β7b	144-148	145-147	145-147	
21.	*Mt*Mce1F_MlaD1 (model)	L0T2W6	4-34 (31)	38-136 (99)	β1	38-46	39-46	39-46	0, 49.5, 50.5
					β2	56-61	57-60	57-60	
					β3	63-74	63-74	63-74	
					β4	77-86	77-85	78-85	
					β5	95-101	95-100	95-100	
					PLP	**102-105**	**102-105**	**102-105**	
					β6	106-112	107-112	107-112	
					β7a	125-127	126-127	126-127	
					β7b	129-131	130-131	130-131	
22.	*Mt*Mce2A_MlaD1 (model)	Q79FY7	7-36 (30)	42-147 (106)	β1	44-50	44-50	44-50	3.8, 50.9, 45.3
					β2	61-66	62-65	62-65	
					β3	68-81	68-81	68-81	
					β4	84-93	84-93	84-93	
					H1	94-99	95-98	95-98	
					β5	106-112	106-111	106-111	
					PLP	**113-116**	**113-116**	**113-116**	
					β6	117-123	118-123	118-123	
					β7	137-145	138-144	138-144	

23.	*Mt*Mce2B_MlaD1 (model)	O07788	4-28 (25)	34-136 (103)	β1	35-43	36-43	36-43	1.9, 47.6, 50.5
					β2	53-58	54-57	54-57	
					β3	60-71	60-71	60-71	
					β4	76-85	76-84	76-84	
					β5	94-100	94-99	94-99	
					PLP	**101-104**	**101-104**	**101-104**	
					β6	105-111	106-111	106-111	
					β7a	123-125	124-125	124-125	
					H1	126-128	127-129	127-129	
					β7b	129-131	130-131	130-131	
24.	*Mt*Mce2C_MlaD1 (model)	O07787	9-31 (23)	38-134 (97)	β1	40-46	41-46	41-46	2.1, 46.4, 51.5
					β2	56-60	57-60	57-60	
					β3	63-74	63-74	63-74	
					β4	77-84	77-83	77-83	
					β5	94-100	94-99	94-99	
					PLP	**101-104**	**101-104**	**101-104**	
					β6	105-111	106-111	106-111	
					β7a	122-124	123-124	123-124	
					H1	125-127	126-128	126-128	
					β7b	128-130	129-130	129-130	
25.	*Mt*Mce2D_MlaD1 (model)	I6WYT7	17-44 (28)	46-142 (97)	β1	47-53	47-53	46-53	2.1, 44.3, 53.6
					β2	63-68	64-67	64-67	
					β3	70-81	70-80	70-80	
					β4	84-92	85-91	85-92	
					β5	102-106	102-106	102-106	
					PLP	**109-112**	**109-112**	**109-112**	
					β6	115-119	115-119	115-119	
					β7a	132-134	133-134	133-134	
					H1	135-137	136-138	136-138	
					β7b	138-140	139-140	139-140	
26.	*Mt*Mce2E_MlaD1 (model)	I6Y461	-	65-164 (100)	β1	67-73	67-73	67-73	2, 48, 50
					β2	83-88	84-87	84-87	
					β3	90-101	90-101	90-101	
					β4	104-112	104-111	104-111	

					β5	122-128	122-127	122-127	
					PLP	**129-132**	**129-132**	**129-132**	
					β6	133-139	134-139	134-139	
					β7a	153-155	154-155	154-155	
					H1	156-158	157-159	157-159	
					β7b	159-163	160-162	160-162	
27.	*Mt*Mce2F_MlaD1 (model)	O07784	4-34 (31)	38-135 (98)	β1	38-46	39-46	39-46	
					β2	56-61	57-60	57-60	
					β3	63-74	63-74	63-74	0, 50, 50
					β4	77-86	77-85	77-85	
					β5	95-101	95-100	95-100	
					PLP	**102-105**	**102-105**	**102-105**	
					β6	106-112	107-112	107-112	
					β7a	125-127	126-127	126-127	
					β7b	129-131	130-131	130-131	
28.	*Mt*Mce3A_MlaD1 (model)	L7N698	12-35 (24)	41-143 (103)	β1	42-48	42-48	42-48	2.9, 46.6, 50.5
					β2	59-64	60-63	60-63	
					β3	66-77	66-76	66-76	
					β4	83-91	84-90	84-90	
					H1	94-98	95-97	95-97	
					β5	103-109	103-108	103-108	
					PLP	**110-113**	**110-113**	**110-113**	
					β6	114-120	115-120	115-120	
					β7	134-142	135-141	135-141	
29.	*Mt*Mce3B_MlaD1 (model)	O53968	5-30 (26)	37-136 (100)	β1	39-45	39-45	39-45	2.0, 45, 53
					β2	55-60	56-59	56-59	
					β3	62-73	62-72	62-72	
					β4	77-84	78-84	78-84	
					β5	95-101	95-100	95-100	
					PLP	**102-105**	**102-105**	**102-105**	
					β6	106-112	107-112	107-112	
					β7a	125-127	126-127	126-127	
					H1	128-130	129-131	129-131	
					β7b	131-133	132-133	132-133	

30.	*Mt*Mce3C_MlaD1 (model)	O53969	9-28 (20)	38-133 (96)	β1	38-46	39-46	39-46	
					β2	56-61	57-60	57-60	2.1, 51, 46.9
					β3	63-74	63-74	63-74	
					β4	77-86	77-85	77-85	
					β5	95-101	95-101	95-101	
					PLP	**102-105**	**102-105**	**102-105**	
					β6	106-112	106-112	106-112	
					H1	121-126	125-127	125-127	
					β7	127-129	128-129	128-129	
31.	*Mt*Mce3D_MlaD1 (model)	O53970	2-32 (31)	34-133 (100)	β1	35-41	35-41	34-41	2, 46, 52
					β2	51-56	52-55	52-55	
					β3	58-69	58-69	58-69	
					β4	72-80	72-79	72-80	
					β5	90-94	90-94	90-94	
					PLP	**97-100**	**97-100**	**97-100**	
					β6	103-107	103-107	103-107	
					β7a	120-122	120-122	120-122	
					H1	123-125	124-126	124-126	
					β7b	126-128	127-128	127-128	
32.	*Mt*Mce3E_MlaD1 (model)	O53971	-	37-136 (100)	β1	38-44	38-44	38-44	2, 49, 49
					β2	54-59	55-58	55-58	
					β3	61-72	61-72	61-72	
					β4	75-83	75-82	75-82	
					β5	93-99	93-98	93-98	
					PLP	**100-103**	**100-103**	**100-103**	
					β6	104-110	105-110	105-110	
					β7a	125-127	125-127	125-127	
					H1	128-130	129-131	129-131	
					β7b	131-135	132-134	132-134	
33.	*Mt*Mce3F_MlaD1 (model)	O53972	6-38 (33)	41-139 (99)	β1	41-49	42-49	42-49	2, 51.5, 46.5
					β2	59-64	60-63	60-63	
					β3	66-77	66-77	66-77	
					β4	80-89	80-88	80-88	
					β5	96-104	97-103	97-103	

					PLP	**105-108**	**105-108**	**105-108**	
					β6	109-117	110-116	110-117	
					β7a	128-130	129-130	129-130	
					H1	131-133	132-134	132-134	
					β7b	134-136	135-136	135-136	
34.	*Mt*Mce4A_MlaD1 (model)	I6YC99	4-33 (30)	40-140 (101)	β1	41-47	41-47	40-48	3, 47.5, 49.5
					β2	58-63	59-62	59-62	
					β3	65-76	65-76	65-76	
					β4	79-87	79-86	79-87	
					H1	90-94	91-93	91-93	
					β5	99-105	99-104	99-104	
					PLP	**106-109**	**106-109**	**106-109**	
					β6	110-116	111-116	111-116	
					β7a	130-134	131-133	131-133	
					β7b	136-138	136-137	136-137	
35.	*Mt*Mce4B_MlaD1 (model)	I6X7G8	9-34 (26)	40-134 (95)	β1	41-49	42-49	42-49	0, 49.5, 50.5
					β2	59-64	60-63	60-63	
					β3	66-77	66-76	66-76	
					β4	81-90	82-89	82-89	
					β5	97-105	98-104	98-104	
					PLP	**106-109**	**106-109**	**106-109**	
					β6	110-118	111-117	111-117	
					β7	129-131	130-131	130-131	
36.	*Mt*Mce4C_MlaD1 (model)	I6YGB1	14-34 (21)	42-130 (89)	β1	43-51	44-51	44-51	0, 52.8, 47.2
					β2	61-66	62-65	62-65	
					β3	68-79	68-79	68-79	
					β4	82-91	82-90	82-90	
					β5	100-106	100-105	100-105	
					PLP	**107-110**	**107-110**	**107-110**	
					β6	111-117	112-117	112-117	
					β7	125-127	126-127	126-127	
37.	*Mt*Mce4D_MlaD1 (model)	I6XHD6	10-36 (27)	38-130 (93)	β1	38-46	39-46	39-46	0, 47.3, 52.7
					β2	56-61	57-60	57-60	
					β3	63-74	63-73	63-73	

					β4	77-86	78-85	78-85	
					β5	95-99	95-99	95-99	
					PLP	**102-105**	**102-105**	**102-105**	
					β6	108-112	108-112	108-112	
					β7	125-127	125-127	125-127	
38.	*Mt*Mce4E_MlaD1 (model)	I6Y3P1	-	39-139 (101)	β1	42-48	42-48	42-48	0, 47.5, 52.5
					β2	58-62	59-62	59-62	
					β3	65-78	65-77	65-77	
					β4	82-92	83-91	83-91	
					β5	102-108	102-107	102-107	
					PLP	**109-112**	**109-112**	**109-112**	
					β6	113-119	114-119	114-119	
					β7	134-136	134-136	134-136	
39.	*Mt*Mce4F_MlaD1 (model)	I6YC95	4-34 (31)	38-131 (94)	β1	38-46	39-46	39-46	
					β2	56-61	57-60	57-60	0, 51.1, 48.9
					β3	63-74	63-74	63-74	
					β4	77-86	77-85	77-85	
					β5	95-101	95-100	95-100	
					PLP	**102-105**	**102-105**	**102-105**	
					β6	106-112	107-112	107-112	
					β7	126-128	126-128	126-128	
40.	*At*TGD2_MlaD1 (model)	Q9LTR2	91-125 (35)	128-239 (112)	β1	129-135	129-135	129-135	3.6, 44.6, 51.8
					β2	145-150	146-149	146-149	
					β3	152-163	152-162	152-162	
					β4	166-174	167-173	167-173	
					β5	185-191	185-191	185-191	
					PLP	**192-195**	**192-195**	**192-195**	
					β6	196-202	196-202	197-202	
					H1	218-223	219-222	219-222	
					β7	230-237	230-236	230-236	
41.	*Mt*Mce4A^{domain} (Protomers A and D) (7AI2)	I6YC99	4-33 (30)	39-145 (107)	β1	39-47	40-47	40-47	6.5, 30.8, 62.6
					β2a		52-54	52-54	
					β2b	58-62	59-62	59-62	

					β3	65-76	65-76	65-76	
					β4	79-88	79-87	79-87	
					H1	89-94	90-93	90-93	
					β5		99-101	99-101	
					PLP	**106-109**	**106-109**	**106-109**	
					β6		104-116	104-116	
					β7a		131-132	131-132	
					H2	133-137	134-136	134-136	
					β7b		137-139	137-139	
42.	*Mt*Mce4A^{domain} (Protomers B and C) (7AI2)	I6YC99		32-144 (113)	β1a	39-41	40-41	40-41	4.4, 27.4, 68.1
					β1b	43-47	44-47	44-47	
					β2a		52-54	52-54	
					β2b	58-62	59-62	59-62	
					β3	65-76	65-76	65-76	
					β4	79-88	79-87	79-87	
					H1	89-94	89-93	89-93	
					β5		99-110	99-110	
					PLP	**106-109**	**106-109**	**106-109**	
					β6		114-116	114-116	
					β7a		131-132	131-132	
					β7b		137-138	137-138	
43.	*Mt*Mce4A^{domain} (Protomers A' and D') (7AI3)	I6YC99		39-145 (107)	β1	39-47	40-46	40-46	4.7, 28, 67.3
					β2a		52-54	52-54	
					β2b	58-62	59-62	59-62	
					β3	65-74	65-74	65-74	
					β4	80-88	81-87	81-87	
					H1	89-91	88-90	88-90	
					β5		99-116	99-116	
					PLP	**106-109**	**106-109**	**106-109**	
					β6		131-132	131-132	
					H2	133-137	134-136	134-136	
					β7		137-141	137-141	
44.		I6YC99		32-144 (113)	β1	39-47	40-46	40-46	2.7, 28.3, 69
					β2a		53-54	53-54	

	*Mt*Mce4A^{domain} (Protomers B' and C') (7AI3)				β2b	58-62	59-62	59-62	
					β3	65-76	65-76	65-76	
					β4	79-88	79-87	79-87	
					β5		99-110	99-110	
					PLP	**106-109**	**106-109**	**106-109**	
					β6		113-116	113-116	
					β7a		131-132	131-132	
					H1	133-137	134-136	134-136	
					β7b		137-141	137-141	
45.	*Hm*Amt (1WSV)	P48728	-	287-363 (77)	β1	289-297	290-297	290-297	2.6, 55.8, 41.6
					β2	310-396	307-309	307-309	
					β3	314-325	315-325	315-325	
					β4	330-339	330-338	330-338	
					H1	340-342	339-341	339-341	
					β5	346-352	347-351	347-351	
					β6	355-362	356-362	356-362	
46.	*Pu*Amt (3TFH)	Q4FP21	-	286-369 (84)	β1	289-298	290-297	290-297	2.4, 54.8, 42.9
					β2	308-312	308-311	308-311	
					β3	316-327	317-327	317-327	
					β4	332-341	332-340	332-340	
					H1	342-344	341-343	341-343	
					β5	348-354	349-354	349-354	
					β6	357-364	357-364	357-364	
47.	*Lr*Amt (3TTG)	-	-	251-334 (81)	β1	255-263	256-263	256-262	2.5, 49.4, 48.1
					β2	275-279	275-278	275-278	
					β3	284-295	285-295	285-295	
					β4	300-309	300-308	300-308	
					H1	310-312	309-311	309-311	
					β5	316-320	317-319	317-319	
					β6	324-329	325-329	325-329	
48.	*Tm*Amt (1WOS)	Q9WY54	-	276-361 (86)	β1	278-287	279-286	279-286	0, 54.7, 45.3
					β2	296-300	297-300	297-300	
					β3	303-314	303-314	303-314	
					β4	319-328	319-327	319-327	

					β5	334-340	335-340	335-340	
					β6	344-351	344-351	344-351	

Table C3. Pairwise sequence identity/similarity (query coverage) in percentage between the MlaD domains of proteins *Ec*MlaD and *Mt*Mce1A-4F. NS: not significant. The pairwise structural comparison (RMSD in Å) between the corresponding MlaD domain is also provided.

***Ec*MlaD_MlaD1 and *Mt*Mce1A-1F_MlaD1**							
	*Ec*MlaD_MlaD1	*Mt*Mce1A_MlaD1	*Mt*Mce1B_MlaD1	*Mt*Mce1C_MlaD1	*Mt*Mce1D_MlaD1	*Mt*Mce1E_MlaD1	*Mt*Mce1F_MlaD1
*Ec*MlaD_MlaD1	100/100 (100), 0	NS, 1.4	31/69 (48), 1.1	30/46 (73), 1.3	NS, 1.3	28/49 (92), 1.3	33/52 (72), 1.3
*Mt*Mce1A_MlaD1	-	100/100 (100), 0	30/50 (59), 1.3	NS, 1.3	NS, 1.1	NS, 1.2	30/50 (74), 1.3
*Mt*Mce1B_MlaD1	-	-	100/100 (100), 0	31/58 (89), 0.8	NS, 1.1	28/51 (74), 0.9	29/45 (68), 1.1
*Mt*Mce1C_MlaD1	-	-	-	100/100 (100), 0	29/52 (89), 1.1	NS, 1.0	NS, 1.0
*Mt*Mce1D_MlaD1	-	-	-	-	100/100 (100), 0	29/51 (75), 1.1	28/47 (71), 1.0
*Mt*Mce1E_MlaD1	-	-	-	-	-	100/100 (100), 0	23/49 (90), 1.1
*Mt*Mce1F_MlaD1	-	-	-	-	-	-	100/100 (100), 0
***Ec*MlaD_MlaD1 and *Mt*Mce2A-2F_MlaD1**							
	*Ec*MlaD_MlaD1	*Mt*Mce2A_MlaD1	*Mt*Mce2B_MlaD1	*Mt*Mce2C_MlaD1	*Mt*Mce2D_MlaD1	*Mt*Mce2E_MlaD1	*Mt*Mce2F_MlaD1
*Ec*MlaD_MlaD1	100/100 (100), 0	NS, 1.4	31/50 (76), 1.1	31/42 (80), 1.2	31/44 (91), 1.4	32/47 (88), 1.1	32/53 (73), 1.3
*Mt*Mce2A_MlaD1	-	100/100 (100), 0	30/48 (60), 1.3	NS, 1.3	NS, 1.1	34/48 (79), 1.2	27/48 (78), 1.3
*Mt*Mce2B_MlaD1	-	-	100/100 (100), 0	NS, 0.9	NS, 1.0	NS, 1.0	NS, 1.0
*Mt*Mce2C_MlaD1	-	-	-	100/100 (100), 0	24/48 (100), 1.1	31/52 (71), 1.2	25/43 (96), 1.0
*Mt*Mce2D_MlaD1	-	-	-	-	100/100 (100), 0	28/50 (74), 1.1	30/47 (88), 1.1

*Mt*Mce2E_MlaD1	-	-	-	-	-	100/100 (100), 0	26/47 (90), 1.3
*Mt*Mce2F_MlaD1	-	-	-	-	-	-	100/100 (100), 0
***Ec*MlaD_MlaD1 and *Mt*Mce3A-3F_MlaD1**							
	*Ec*MlaD_MlaD1	*Mt*Mce3A_MlaD1	*Mt*Mce3B_MlaD1	*Mt*Mce3C_MlaD1	*Mt*Mce3D_MlaD1	*Mt*Mce3E_MlaD1	*Mt*Mce3F_MlaD1
*Ec*MlaD_MlaD1	100/100 (100), 0	NS, 1.5	35/51 (88), 1.1	32/56 (68), 1.2	24/49 (92), 1.3	30/53 (53), 1.2	38/47 (34), 1.3
*Mt*Mce3A_MlaD1	-	100/100 (100), 0	25/46 (61), 1.6	29/43 (75), 1.5	44/77 (18), 1.3	NS, 1.4	31/54 (86), 1.3
*Mt*Mce3B_MlaD1	-	-	100/100 (100), 0	36/48 (88), 0.9	NS, 1.3	31/42 (90), 1.2	NS, 1.3
*Mt*Mce3C_MlaD1	-	-	-	100/100 (100), 0	30/44 (98), 1.2	NS, 1.1	27/54 (73), 1.3
*Mt*Mce3D_MlaD1	-	-	-	-	100/100 (100), 0	31/46 (88), 1.2	NS, 1.2
*Mt*Mce3E_MlaD1	-	-	-	-	-	100/100 (100), 0	22/43 (97), 1.4
*Mt*Mce3F_MlaD1	-	-	-	-	-	-	100/100 (100), 0
***Ec*MlaD_MlaD1 and *Mt*Mce4A-4F_MlaD1**							
	*Ec*MlaD_MlaD1	*Mt*Mce4A_MlaD1	*Mt*Mce4B_MlaD1	*Mt*Mce4C_MlaD1	*Mt*Mce4D_MlaD1	*Mt*Mce4E_MlaD1	*Mt*Mce4F_MlaD1
*Ec*MlaD_MlaD1	100/100 (100), 0	25/48 (49), 1.6	34/51 (82), 1.2	31/55 (80), 1.2	NS, 1.2	NS, 1.0	33/49 (96), 1.3
*Mt*Mce4A_MlaD1	-	100/100 (100), 0	NS, 1.7	31/54 (66), 1.5	NS, 1.1	30/54 (41), 1.5	32/51 (79), 1.0
*Mt*Mce4B_MlaD1	-	-	100/100 (100), 0	35/59 (92), 0.8	33/55 (72), 1.2	NS, 1.1	NS, 1.3
*Mt*Mce4C_MlaD1	-	-	-	100/100 (100), 0	31/51 (83), 1.2	NS, 0.9	NS, 1.2
*Mt*Mce4D_MlaD1	-	-	-	-	100/100 (100), 0	NS, 1.1	NS, 1.1

*Mt*Mce4E_MlaD1	-	-	-	-	-	100/100 (100), 0	28/46 (96), 1.1
*Mt*Mce4F_MlaD1	-	-	-	-	-	-	100/100 (100), 0

Table C4. List of protein sequences analyzed in order to study the conservation of cysteine residue.

S. No.	UniProt id	Protein name	Organism name	Category	Annotation
1.	Q9LTR2	Protein TRIGALACTOSYLDIACYLGLYCEROL 2, chloroplastic	*Arabidopsis thaliana*	Plant	*At*Q9LTR2
2.	A5AJN1	MlaD domain-containing protein	*Vitis vinifera*		*Vv*A5AJN1
3.	A2WZM3	MlaD domain-containing protein	*Oryza sativa*		*Os*A2WZM3
4.	A9TJV6	Predicted protein	*Physcomitrium patens*		*Pp*A9TJV6
5.	Q3M9P9	MlaD domain-containing protein	*Trichormus variabilis*	Cyanobacteria	*Tv*Q3M9P9
6.	B7JVN3	Mammalian cell entry related domain protein	*Rippkaea orientalis*		*Ro*B7JVN3
7.	A8YH52	Similar to tr|Q8Z0B7|Q8Z0B7	*Microcystis aeruginosa*		*Ma*A8YH52
8.	B0C226	MlaD domain-containing protein	*Acaryochloris marina*		*Am*B0C226
9.	Q9KP58	MlaD domain-containing protein	*Vibrio cholerae*	Gram-negative bacteria	*Vc*Q9KP58
10.	P45029	Intermembrane phospholipid transport system binding protein MlaD	*Haemophilus influenzae*		*Hi*P45029
11.	P64604	Intermembrane phospholipid transport system binding protein MlaD	*Escherichia coli*		*Ec*P64604
12.	Q9HVW3	MlaD domain-containing protein	*Pseudomonas aeruginosa*		*Pa*Q9HVW3
13.	Q82B06	Putative mce (Mammalian cell entry)-related protein	*Streptomyces avermitilis*	Actinomycetes	*Sa*Q82B06
14.	I6YC99	Mce-family protein Mce4A	*Mycobacterium tuberculosis*		*Mt*I6YC99
15.	Q5Z0T1	Putative Mce family protein	*Nocardia farcinica*		*Nf*Q5Z0T1
16.	C1BB82	Putative Mce family protein	*Rhodococcus opacus*		*Ro*C1BB82

Table C5. The residues forming the PLP regions of the MlaD domain of the proteins *Ec*MlaD, *Ec*PqiB and *Ec*YebT and the respective diameters of the central channels.

S. No.	Domain name	Residues forming the PLPs (Number of residues)	Approx. diameter of the central channel (Å)*
1.	*Ec*MlaD_MlaD1^{P41212}	GL**L**G (4)	15.0
2.	*Ec*PqiB_MlaD1	PQIGR**E**GISGLGTLLS (16)	40.0
3.	*Ec*PqiB_MlaD2	GIAVDLTS**A**GMRVEMG (16)	38.0
4.	*Ec*PqiB_MlaD3	TGN**L**VTGAL (9)	20.0
5.	*Ec*YebT_MlaD1	KAS**L**AGVSGLDALVGG (16)	21.0
6.	*Ec*YebT_MlaD2	VSGVDANVSI**S**GAKVKLESLA (21)	22.0
7.	*Ec*YebT_MlaD3	NPKLS**L**SDANLSALLTG (17)	24.0
8.	*Ec*YebT_MlaD4	SRVDVKVG**L**DGVEFLGASASEWING (25)	19.0
9.	*Ec*YebT_MlaD5	GGAKVQLNG**S**GLTVQASPLSRALKG (25)	22.0
10.	*Ec*YebT_MlaD6	TPQIS**A**AGVEHLDTILQPY (19)	10.0
11.	*Ec*YebT_MlaD7	GYSLDFGL**T**GGVVKTGTF (18)	18.0

*Diameter was determined by measuring the distance between the Cα atoms of the amino acids (bold), each belonging to an opposite protomer.

Table C6. List of cryo-EM structures analyzed for the determination of central channel diameters.

S. No.	PDB name	PDB id	Organism	Approx. diameter of the central channel (Å)
1.	Cryo-EM structure of MlaFEDB in nanodiscs with phospholipid substrates	6XBD	*E. coli*	16.0
2.	Cryo-EM structure of MlaFEDB in complex with AMP-PNP	6ZY9		17.0
3.	Cryo-EM structure of apo MlaFEDB	6ZY2		15.0
4.	Cryo-EM structure of MlaFEDB in complex with ADP	6ZY4		16.0
5.	Cryo-EM structure of MlaFEDB in complex with phospholipid	6ZY3		16.0
6.	The overall structure of nucleotide-free MlaFEDB complex	7CGE		17.0
7.	The overall structure of the MlaFEDB complex in ATP-bound EQtall conformation (Mutation of E170Q on MlaF)	7CGN		19.0
8.	The overall structure of the MlaFEDB complex in ATP-bound EQclose conformation (Mutation of E170Q on MlaF)	7CH0		18.0
9.	Cryo-EM structure of the *A. baumannii* MLA complex at 8.7 A resolution	6IC4	*A. baumannii*	20.0
10.	Cryo-EM structure of the *A. baumannii* MlaBDEF complex bound to APPNHP	6Z5U		16.0
11.	Acinetobacter MlaFEDB complex in ATP-bound Vtrans1 conformation	7D08		17.0
12.	Acinetobacter MlaFEDB complex in ADP-vanadate trapped Vclose conformation	7D0A		16.0
13.	Cryo-EM structure of the nucleotide-free Acinetobacter MlaFEDB complex	7D06		17.0

14.	Acinetobacter MlaFEDB complex in ATP-bound Vtrans2 conformation	7D09		17.0
15.	Cryo-EM structure of *P.aeruginosa* MlaFEBD with AMPPNP	7CHA	*P. aeruginosa*	15.0
16.	Cryo-EM structure of *P. aeruginosa* MlaFEBD with ADP-V	7CH8		15.0
17.	Cryo-EM structure of *P. aeruginosa* MlaFEBD	7CH9		16.0

*Diameter was determined by measuring the distance between the Cα atoms of the amino acids, each belonging to an opposite protomer.

Table C7. The amino acid residues participating in the dimeric interface of the homo-hexamer of the protein *Ec*MlaD.

Interface	Residues present at the interface of	
	Protomer 1	**Protomer 2**
Protomer 1 - Protomer 2	Ser59, Ile60, Gly61, Gly62, Val63, Val65, Arg89, Tyr90, Asn91, His92, Ile93, Ile101, Arg102, Thr103, Ser104, Gly105, Leu106, Leu107, Gln110, Tyr111, Val116, Lys138, Met141, Asp145, Leu146, Gln149, Phe150	Asp47, Asn48, Ile49, Gly50, Gly51, Leu52, Ile71, Thr72, Leu73, Asp74, Pro75, Thr77, Tyr78, Pro80, Leu106, Leu107, Val142, Leu143, Glu144, Asp145, Ile147, Gly148, Phe150, Leu151, Tyr152
	Residues involved in hydrogen bond	
	Gly61[O]-Ile49[N], Gly61[O]-Gly50[N], Asn91[N]-Tyr78[OH], His92[N]-Tyr78[OH], Arg102[NH2]-Asp145[OD1], Thr103[N]-Glu144[OE2], Gln149[OE1]-Tyr152[OH]	

Table C8. Analysis of interfacial areas of hexamers and PLP distances of different protomers.

MlaD hexamer	*Ec*MlaD_P4$_1$2$_1$2		*Ec*MlaD_complex	
Interfacial areas				
S. No.	**Interface No.†**	**Buried area (Å^2)**	**Interface No.†**	**Buried area (Å^2)**
1.	I1	849	I1’	968
2.	I2	848	I2’	905
3.	I3	825	I3’	885
4.	I4	825	I4’	860
5.	I5	818	I5’	722
6.	I6	813	I6’	690
Average (Å^2)		830		838
PLP distances of different protomers				
S. No.	**Protomers¶**	**Distance (Å)**	**Protomers¶**	**Distance (Å)**
1.	D1-D2	7.7	D1’-D2’	12.9
2.	D2-D3	7.5	D2’-D3’	7.8
3.	D3-D4	7.6	D3’-D4’	5.3
4.	D4-D5	7.7	D4’-D5’	13.4
5.	D5-D6	7.5	D5’-D6’	8.0
6.	D6-D1	7.6	D6’-D1’	5.5
Average (Å)		7.6		8.8

†The interfaces of *Ec*MlaD_P4$_1$2$_1$2 and *Ec*MlaD_complex are designated as I1-6 and I’1-6, respectively.

¶The PLP distances between two protomers in *Ec*MlaD_P4$_1$2$_1$2 and *Ec*MlaD_complex are designated as D(1-6)-D(1-6) and D’(1-6)-D’(1-6), respectively.

Table C9. List of *Ec*MlaD and *Ec*MlaE orthologs from Gram-negative bacteria analyzed in the study.

S. No.	Organism	Protein name	UniProt id	Protein name	UniProt id	Annotation
1.	*Escherichia coli*	Intermembrane phospholipid transport system binding protein MlaD	P64604	Intermembrane phospholipid transport system permease protein MlaE	P64606	*Ec*MlaD, *Ec*MlaE
2.	*Acinetobacter baumannii*	MCE family protein	V5V921	Intermembrane phospholipid transport system permease protein MlaE	V5V9F4	*Ab*MlaD, *Ab*MlaE
3.	*Pseudomonas aeruginosa*	MlaD domain-containing protein	Q9HVW3	Intermembrane phospholipid transport system permease protein MlaE	Q9HVW2	*Pa*MlaD, *Pa*MlaE
4.	*Haemophilus influenzae*	Intermembrane phospholipid transport system binding protein MlaD	P45029	Intermembrane phospholipid transport system permease protein MlaE	P45030	*Hi*MlaD, *Hi*MlaE
5.	*Nitrospira moscoviensis*	Putative ABC transporter, periplasmic binding component, ATP-dependent toluene efflux transporter	A0A0K2G8H8	Toluene transporter subunit: membrane component of ABC superfamily	A0A0K2G968	*Nm*MlaD, *Nm*MlaE
6.	*Vibrio cholerae*	MlaD domain-containing protein	Q9KP58	Intermembrane phospholipid transport system permease protein MlaE	Q9KP57	*Vc*MlaD, *Vc*MlaE
7.	*Francisella tularensis*	Outer membrane lipid asymmetry maintenance protein MlaD	A0A0B3WHJ2	Intermembrane phospholipid transport system permease protein MlaE	A0A0B3VBI4	*Ft*MlaD, *Ft*MlaE

8.	*Legionella pneumophila*	Outer membrane lipid asymmetry maintenance protein MlaD	A0A140IYF9	Intermembrane phospholipid transport system permease protein MlaE	A0A140J2J2	*Lp*MlaD, *Lp*MlaE
9.	*Klebsiella quasipneumoniae*	ABC transporter periplasmic substrate-binding protein	W9B331	Intermembrane phospholipid transport system permease protein MlaE	A0A1C1EY82	*Kq*MlaD, *Kq*MlaE
10.	*Arabidopsis thaliana*	Protein TRIGALACTOSYLDIACYLGLYCEROL 2, chloroplastic	Q9LTR2	Protein TRIGALACTOSYLDIACYLGLYCEROL 1, chloroplastic	Q8L4R0	*At*TGD2, *At*TGD1

Table C10. List of TMDs associated with MlaD domain-containing proteins.

S. No.	Protein name	Organism name	UniProt id	Protein family (Pfam id)	Notation
1.	Intermembrane phospholipid transport system permease protein MlaE	*E. coli*	P64606	MlaE (PF02405)	*Ec*MlaE
2.	Conserved integral membrane protein YrbE1A	*M. tuberculosis*	O07412	MlaE (PF02405)	*Mt*YrbE1A
3.	Conserved integral membrane protein YrbE1B		L0T2Q9		*Mt*YrbE1B
4.	Conserved hypothetical integral membrane protein YrbE2A		I6Y870		*Mt*YrbE2A
5.	Conserved hypothetical integral membrane protein YrbE2B		O07790		*Mt*YrbE2B
6.	Conserved hypothetical integral membrane protein YrbE3A		O53965		*Mt*YrbE3A
7.	Conserved hypothetical integral membrane protein YrbE3B		O53966		*Mt*YrbE3B
8.	Conserved integral membrane protein YrbE4A. Possible ABC transporter		O53546		*Mt*YrbE4A
9.	Conserved integral membrane protein YrbE4B. Possible ABC transporter		I6Y3P5		*Mt*YrbE4B
10.	Protein trigalactosyldiacylglycerol 1, chloroplastic	*A. thaliana*	Q8L4R0		*At*TGD1
11.	Intermembrane transport protein PqiA	*E. coli*	P0AFL9	PqiA (PF04403)	*Ec*PqiA
12.	Intermembrane transport protein YebS		P0AD03		*Ec*YebS

Table C11. Orientation analysis of TMDs associated with the MlaD domain-containing proteins.

<table>
<tr><th rowspan="2">S. No.</th><th rowspan="2">Protein name</th><th colspan="3">TMHMM</th><th colspan="3">Cryo-EM structures / AlphaFold models</th></tr>
<tr><th>No. of TM helices</th><th>Side of N-terminal</th><th>Side of C-terminal</th><th>No. of TM helices</th><th>Side of N-terminal</th><th>Side of C-terminal</th></tr>
<tr><td>1.</td><td>EcMlaE</td><td>5</td><td>Cytoplasmic</td><td rowspan="3">Periplasmic</td><td rowspan="10">5</td><td rowspan="12">Cytoplasmic</td><td rowspan="10">Periplasmic</td></tr>
<tr><td>2.</td><td>MtYrbE1A</td><td>6</td><td>Periplasmic</td></tr>
<tr><td>3.</td><td>MtYrbE1B</td><td rowspan="2">5</td><td rowspan="7">Cytoplasmic</td></tr>
<tr><td>4.</td><td>MtYrbE2A</td><td rowspan="2">Periplasmic</td></tr>
<tr><td>5.</td><td>MtYrbE2B</td><td>6</td><td>Cytoplasmic</td></tr>
<tr><td>6.</td><td>MtYrbE3A</td><td rowspan="2">5</td><td rowspan="2">Periplasmic</td></tr>
<tr><td>7.</td><td>MtYrbE3B</td></tr>
<tr><td>8.</td><td>MtYrbE4A</td><td>6</td><td>Cytoplasmic</td></tr>
<tr><td>9.</td><td>MtYrbE4B</td><td rowspan="2">5</td><td>Periplasmic</td></tr>
<tr><td>10.</td><td>AtTGD1†</td><td>Periplasmic</td><td rowspan="3">Cytoplasmic</td></tr>
<tr><td>11.</td><td>EcPqiB</td><td rowspan="2">8</td><td rowspan="2">Cytoplasmic</td><td rowspan="2">8</td><td rowspan="2">Cytoplasmic</td></tr>
<tr><td>12.</td><td>EcYebS</td></tr>
</table>

†For *At*TGD1, periplasmic and cytoplasmic refer to intermembrane space and stroma, respectively.

Table C12. Structural homologs, other than the MlaD domain-containing proteins, of the protein *Ec*MlaD obtained using the web server Dali (Holm, 2022).

S. No.	Dali result (UniProtKb name; Accession No.)	Organism name	PDB id	Z-score	RMSD	SI (%)
1.	Aminomethyltransferase (Aminomethyltransferase, mitochondrial; P48728)	*Homo sapiens*	1WSV	4.8	3.7	15
2.	Pseudorabies virus protease (Capsid scaffolding protein; Q83417)	Suid alphaherpesvirus 1	4CX8	4.5	3.0	16
3.	Probable trna pseudouridine synthase (Probable tRNA pseudouridine synthase B; Q7LWY0)	*Pyrococcus furiosus*	2EY4	4.2	2.4	17
4.	Gcvt-like aminomethyltransferase (Dimethylsulfoniopropionate demethylase DmdA; Q4FP21)	*Pelagibacter ubique*	3TFH	4.2	3.9	13
5.	Telomerase cajal body protein 1 (Telomerase Cajal body protein 1; Q9BUR4)	*Homo sapiens*	7BGB	4.2	3.5	9
6.	Ribonuclease mrp rna subunit nme1 (Ribonucleases P/MRP protein subunit POP1; P41812)	*Saccharomyces cerevisiae*	7C79	3.9	3.0	11
7.	Prohead core protein protease (Prohead core protein protease; P06807)	*Escherichia* virus T4	5JBL	3.8	3.4	7
8.	Capsid protein p40 (Capsid scaffolding protein; P03234)	Human gammaherpesvirus 4	1O6E	3.5	3.1	12
9.	Regulator of nonsense transcripts 1 (Regulator of nonsense transcripts 1; Q92900)	*Homo sapiens*	2WJY	3.5	4.5	9
10.	Atpase h+-transporting v1 subunit (H(+)-transporting two-sector ATPase; D4A133)	*Rattus norvegicus*	6VQ6	3.4	5.3	9
11.	Cell shape protein mrec (Cell shape-determining protein MreC; P16926)	*Escherichia coli*	7EFT	3.4	2.6	15
12.	Putative aminomethyltransferase (NA; UPI00022CC7F7)	*Leptospirillum rubarum*	3TTG	3.4	3.0	7

RMSD: root-mean-square deviation; SI: sequence identity; NA: Not applicable.

Table C13. Classification of aminomethyltransferases as provided by the CATH database.

PDB id	Domain	Range	Level	CATH code	Description
1WSV	1a,b	4-55, 147-240	C	3	Alpha Beta
			A	3.30	2-Layer Sandwich
			T	3.30.1360	Gyrase A; domain 2
			H	3.30.1360.120	Probable tRNA modification gtpase trme; domain 1
	2	56-146	C	3	Alpha Beta
			A	3.30	2-Layer Sandwich
			T	3.30.70	Alpha-Beta Plaits
			H	3.30.70.1400	Aminomethyltransferase beta-barrel domains
	3	241-286	C	4	Few Secondary Structures
			A	4.10	Irregular
			T	4.10.1250	Aminomethyltransferase fragment
			H	4.10.1250.10	Aminomethyltransferase fragment
	4	287-363	C	2	Mainly Beta
			A	2.40	Beta Barrel
			T	2.40.30	Elongation Factor Tu (Ef-tu); domain 3
			H	2.40.30.110	Aminomethyltransferase beta-barrel domains
3TFH	1	1-369	C	3	Alpha Beta
			A	3.30	2-Layer Sandwich
			T	3.30.1360	Gyrase A; domain 2
			H	3.30.1360.120	Probable tRNA modification gtpase trme; domain 1
3TTG	1	-5-334	C	3	Alpha Beta
			A	3.30	2-Layer Sandwich
			T	3.30.1360	Gyrase A; domain 2
			H	3.30.1360.120	Probable tRNA modification gtpase trme; domain 1
1WOS	1a, b	1-49, 137-234	C	3	Alpha Beta
			A	3.30	2-Layer Sandwich
			T	3.30.1360	Gyrase A; domain 2
			H	3.30.1360.120	Probable tRNA modification gtpase trme; domain 1
	2	50-136	C	3	Alpha Beta
			A	3.30	2-Layer Sandwich

			T	3.30.70	Alpha-Beta Plaits
			H	3.30.70.1400	Aminomethyltransferase beta-barrel domains
	3	235-275	C	4	Few Secondary Structures
			A	4.10	Irregular
			T	4.10.1250	Aminomethyltransferase fragment
			H	4.10.1250.10	Aminomethyltransferase fragment
	4	276-361	C	2	Mainly Beta
			A	2.40	Beta Barrel
			T	2.40.30	Elongation Factor Tu (Ef-tu); domain 3
			H	2.40.30.110	Aminomethyltransferase beta-barrel domains

Supplementary figures

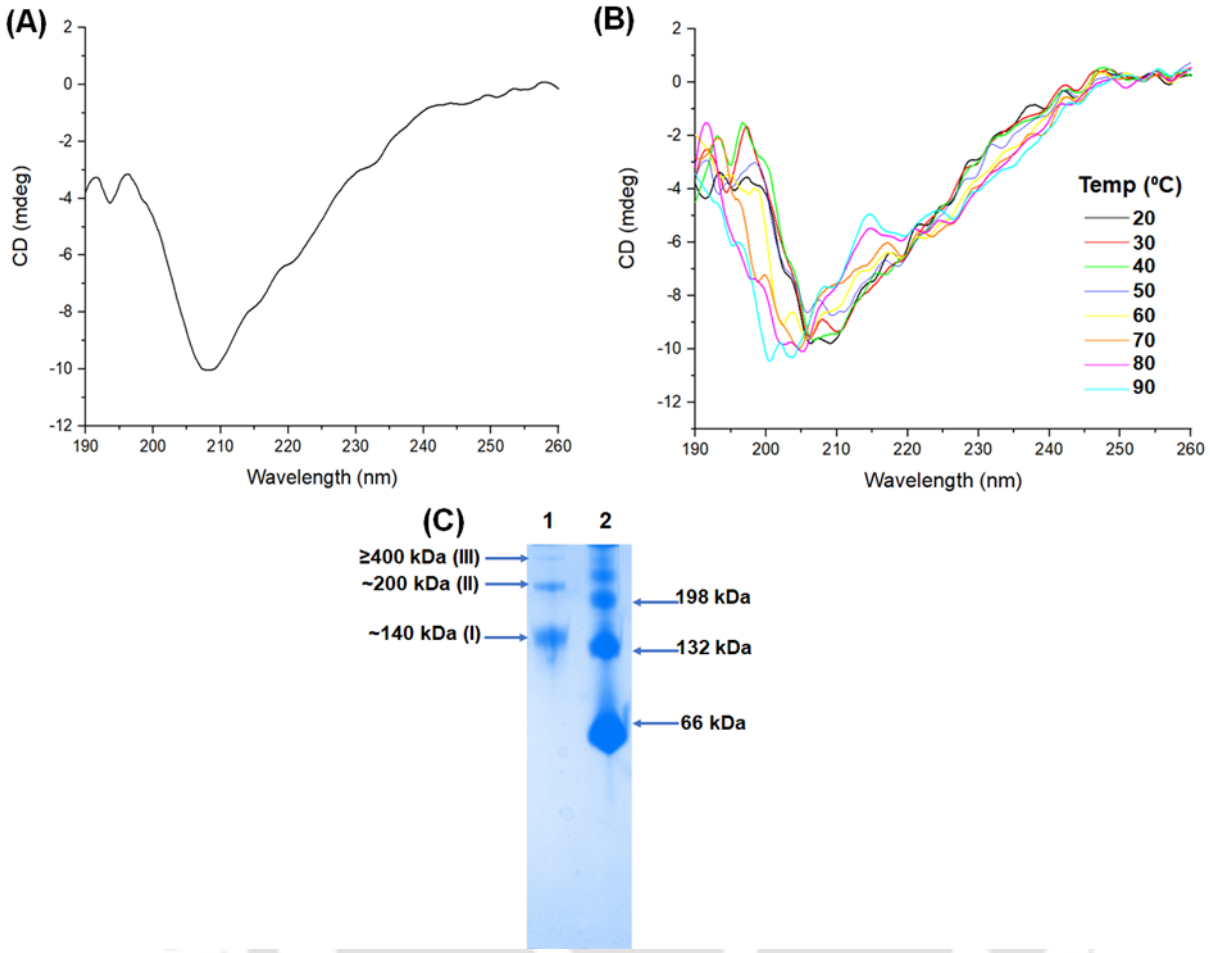


Figure C1. Biophysical characterization of EcMlaD. *(A) The circular dichroism (CD) spectrum of EcMlaD measured in the wavelength range of 190-260 nm. (B) Thermal melting plots depicting the change in secondary structure elements of EcMlaD at different temperatures in the range of 20-90 ºC. (C) Oligomeric forms of EcMlaD resolved on the native-polyacrylamide gel. Bands of different sizes of EcMlaD (I, II and III) and Bovine Serum Albumin (BSA) are shown in lanes 1 and 2, respectively.*

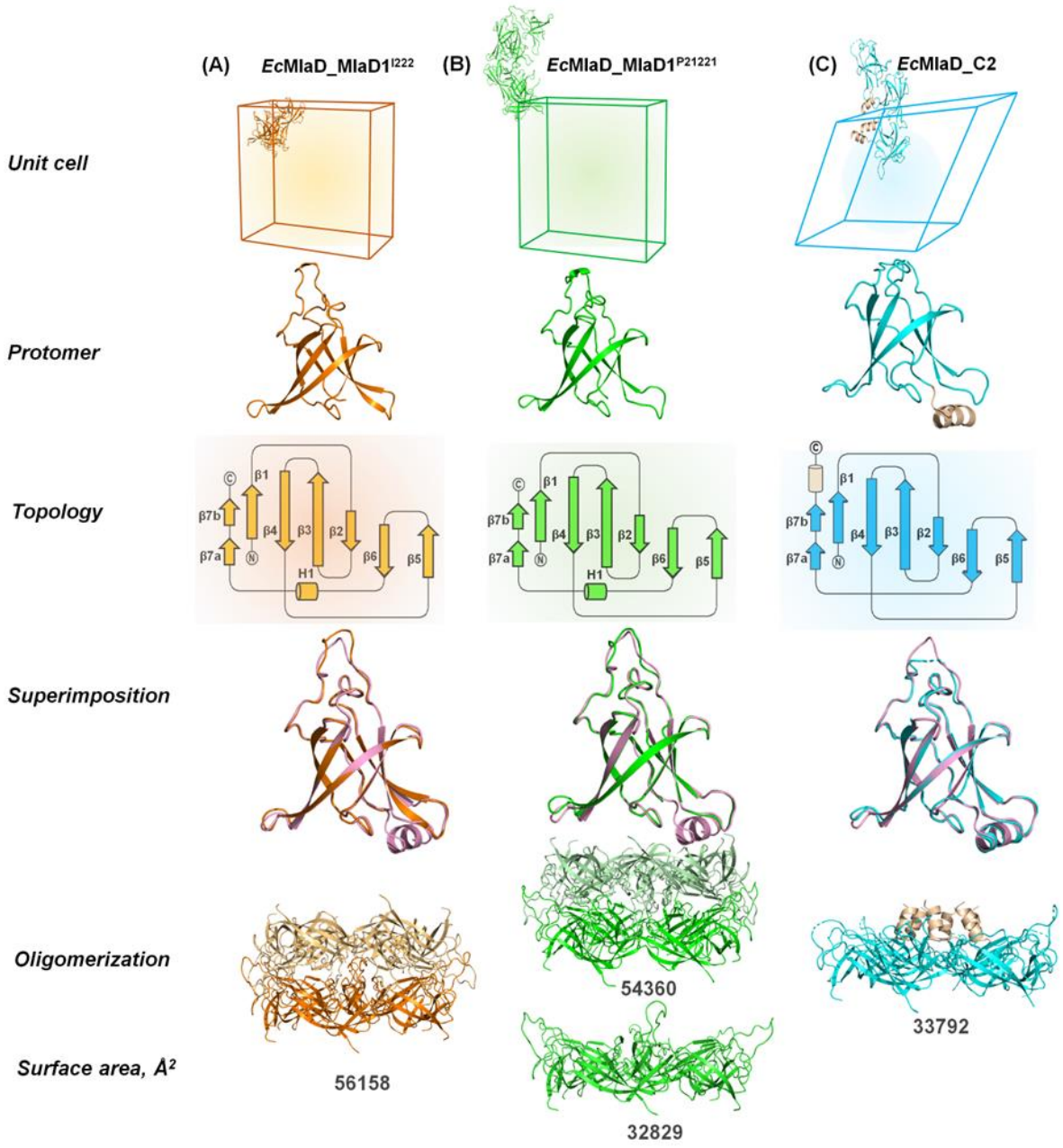


Figure C2. Comparison of the structures of EcMlaD crystallized in different space groups. *(A-C) Structural study of EcMlaD crystalized in the space groups I222 (EcMlaD_MlaD1^{I222}, orange, PDB id: 8HPZ), P2$_1$22$_1$ (EcMlaD_MlaD1^{P21221}, green, PDB id: 8HQ9) and previously reported structure in C2 (EcMlaD_C2, cyan, PDB id: 5UWA), respectively. The analysis includes the study of unit cells (first row), structures of the protomers (second row), topologies (third row), structural superimpositions with EcMlaD (pink) (fourth row), oligomeric organizations (fifth row) and their surface areas (sixth row).*

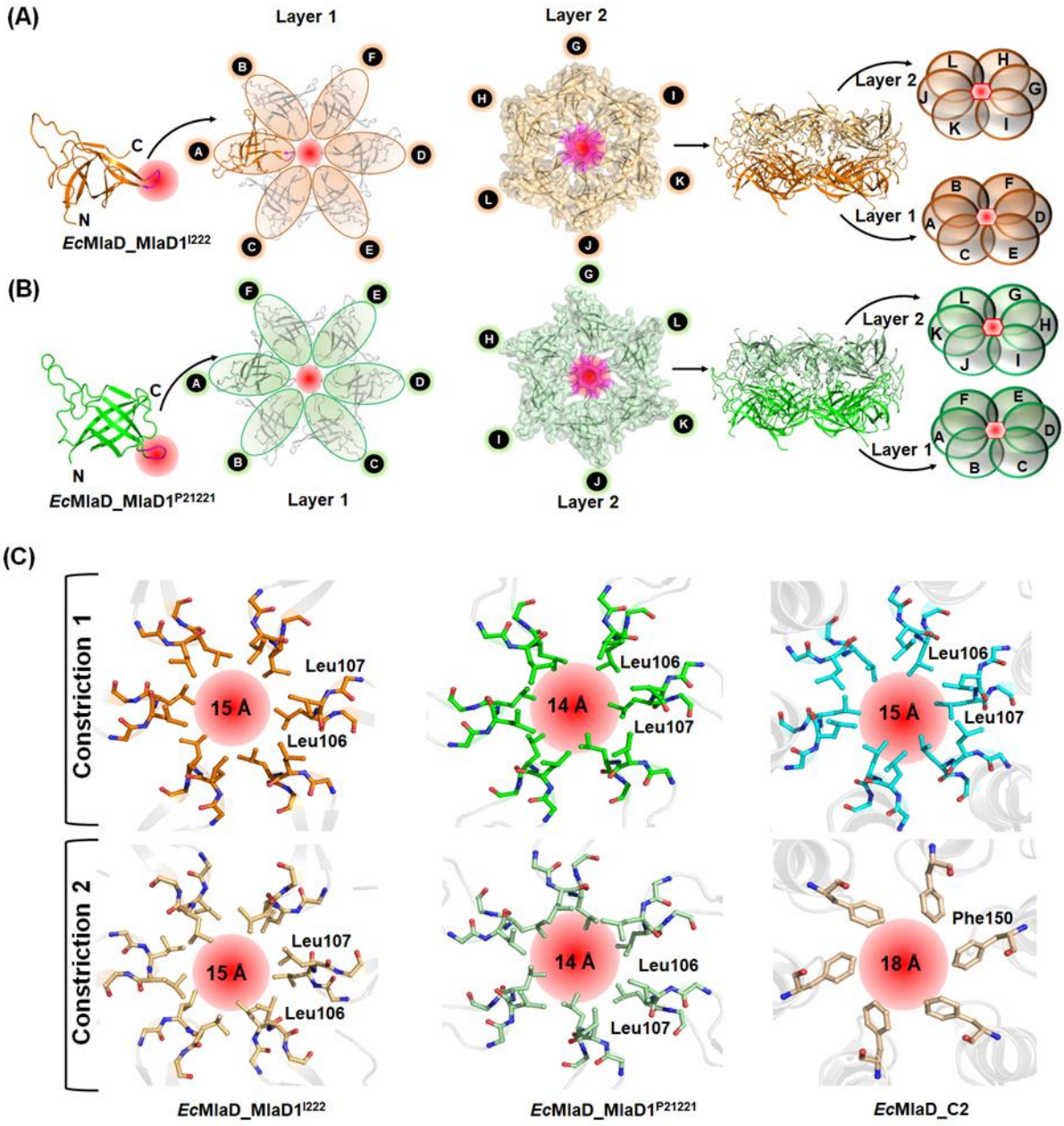


Figure C3. Analysis of the oligomeric arrangement of EcMlaD. *(A-B) Oligomerization of EcMlaD_MlaD1^{I222} (Orange) and EcMlaD_MlaD1^{P21221} (Green) protomers, respectively. (First) Orientation of EcMlaD_MlaD1$^{I222/P21221}$ protomers with respect to the PLP (highlighted in magenta). (Second) Homo-hexameric arrangement of EcMlaD_MlaD1$^{I222/P21221}$ protomers (Layer 1). (Third) Homo-hexameric arrangement of EcMlaD_MlaD1$^{I222/P21221}$ protomers (Layer 2). (Fourth) Dodecameric arrangement of EcMlaD_MlaD1$^{I222/P21221}$ protomers in two layers. (Fifth) Schematic representation of Layer 1 and Layer 2 of EcMlaD_MlaD1$^{I222/P21221}$ protomers. (C) (Left, center, right) Overview of the central channels of EcMlaD_MlaD1$^{I222/P21221/C2}$ hexamers. (Top, bottom) Diameters of constrictions 1 and 2 of EcMlaD_MlaD1$^{I222/P21221/C2}$ hexamers, respectively. The central channel is highlighted in red.*

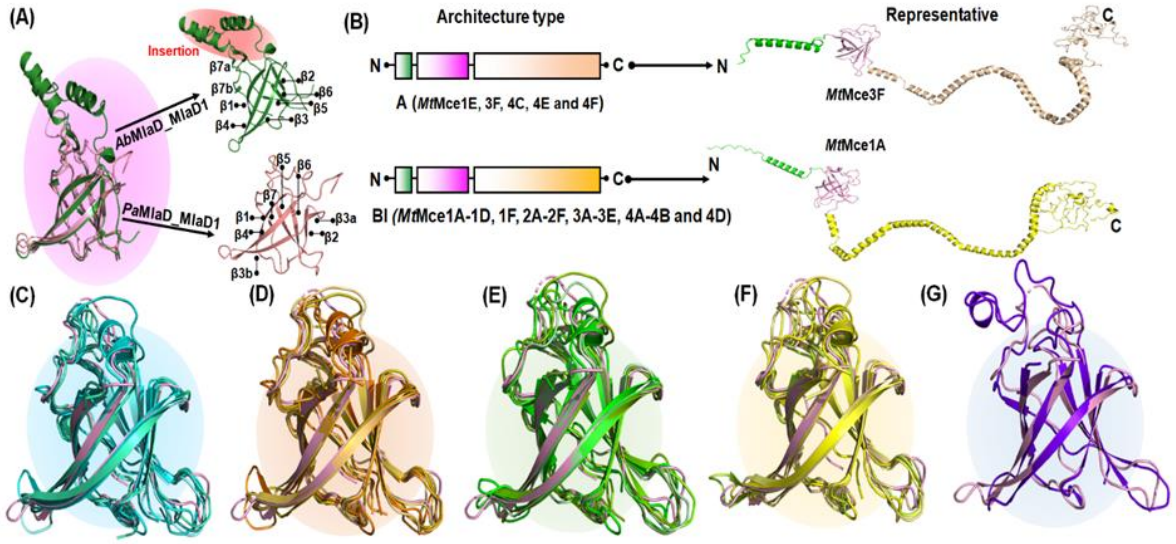


Figure C4. Comparison of the MlaD domains. *(A) (Left) Structural superimposition of EcMlaD_MlaD1^{P41212} (pink, PDB id: 8HQA) with AbMlaD_MlaD1 (Forest, PDB id: 6Z5U) and PaMlaD_MlaD1 (Salmon, PDB id: 7CHA). The overall structure of AbMlaD_MlaD1 (Right top) and PaMlaD_MlaD1 (Right bottom). The insertion between β4 and β5 strands of AbMlaD_MlaD1 is highlighted in red. (B) Schematic and cartoon representations of the domain architecture of MtMce proteins have been shown on left and right, respectively. The TM, MlaD, Mce_CUP4 and HD domains are shown in green, pink, wheat and yellow rectangles, respectively. (C-G) Structural superimposition of EcMlaD_MlaD1^{P41212} (pink) with MtMce1A-1F_MlaD1 (cyan, aquamarine, greencyan, teal, deepteal and lightteal, respectively) and MtMce2A-2F_MlaD1 (orange, tv_orange, brightorange, lightorange, yelloworange and olive, respectively), MtMce3A-3F_MlaD1 (green, tv_green, chartreuse, splitpea, smudge and palegreen, respectively), MtMce4A-4F_MlaD1 (yellow, tv_yellow, paleyellow, yelloworange, limon and sand, respectively) and AtTGD2_MlaD1 (purpleblue), respectively.*

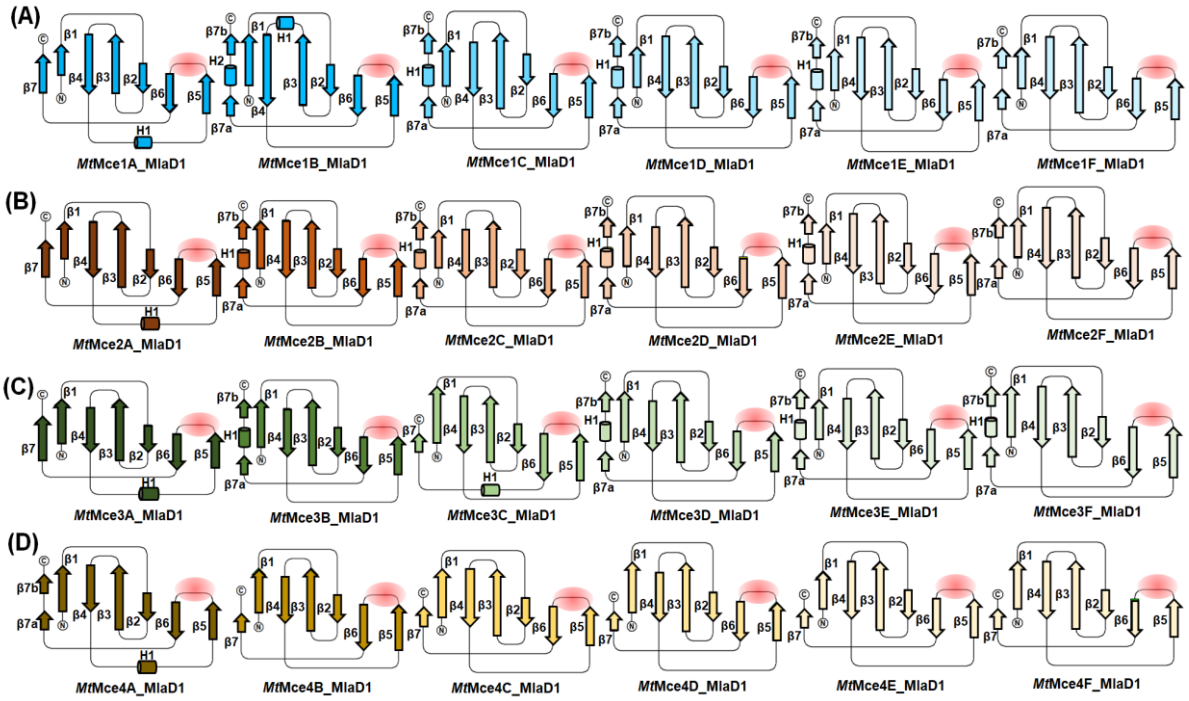


Figure C5. Comparison of the MlaD domain topologies from M. tuberculosis. *(A-D) Topology diagrams of MtMce1A-1F_MlaD1 (different shades of cyan), MtMce2A-2F_MlaD1 (different shades of orange), MtMce3A-3F_MlaD1 (different shades of green) and MtMce4A-4F_MlaD1 (different shades of yellow), respectively. The α-helices and β-strands are represented as cylinders and arrows, respectively. The PLP region is highlighted in red.*

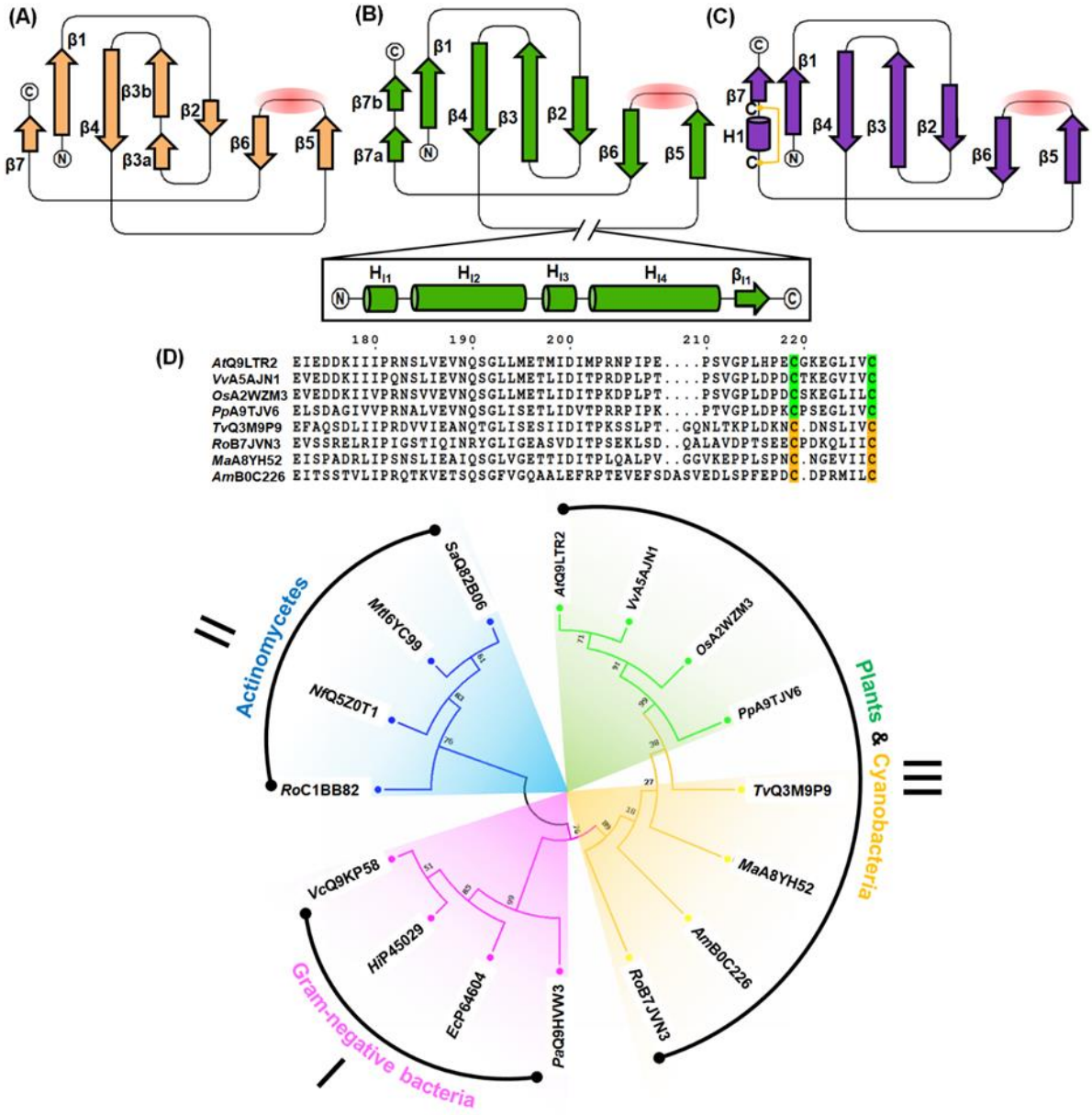


Figure C6. Structure and sequence comparison of MlaD domains. *(A-C) Topology diagrams of PaMlaD_MlaD1 (orange), AbMlaD_MlaD1 (green), and AtTGD2_MlaD1 (purple), respectively. The helices and β-strands are represented as cylinders and arrows, respectively. The PLP region is highlighted in red and the disulfide bond is in orange. The magnified view of the insertion between β4 and β5 strands in AbMlaD_MlaD1 is provided at the bottom of Figure C6B. (D) Sequence and evolutionary analysis of EcMlaD and its orthologs. (Top) MSA of protein sequences from plants and cyanobacteria. The conservation of cysteine residue in plants and cyanobacteria is highlighted in green and orange, respectively. (Bottom) Phylogenetic tree showing the evolutionary relationship of EcMlaD and its orthologs from Gram-negative bacteria, actinomycetes, plants and cyanobacteria. The three major clusters are labeled I, II and III. Numbers at each internal node of the phylogenetic tree signify the probability of its occurrence (in percentage) out of 1000 bootstrap replicates*

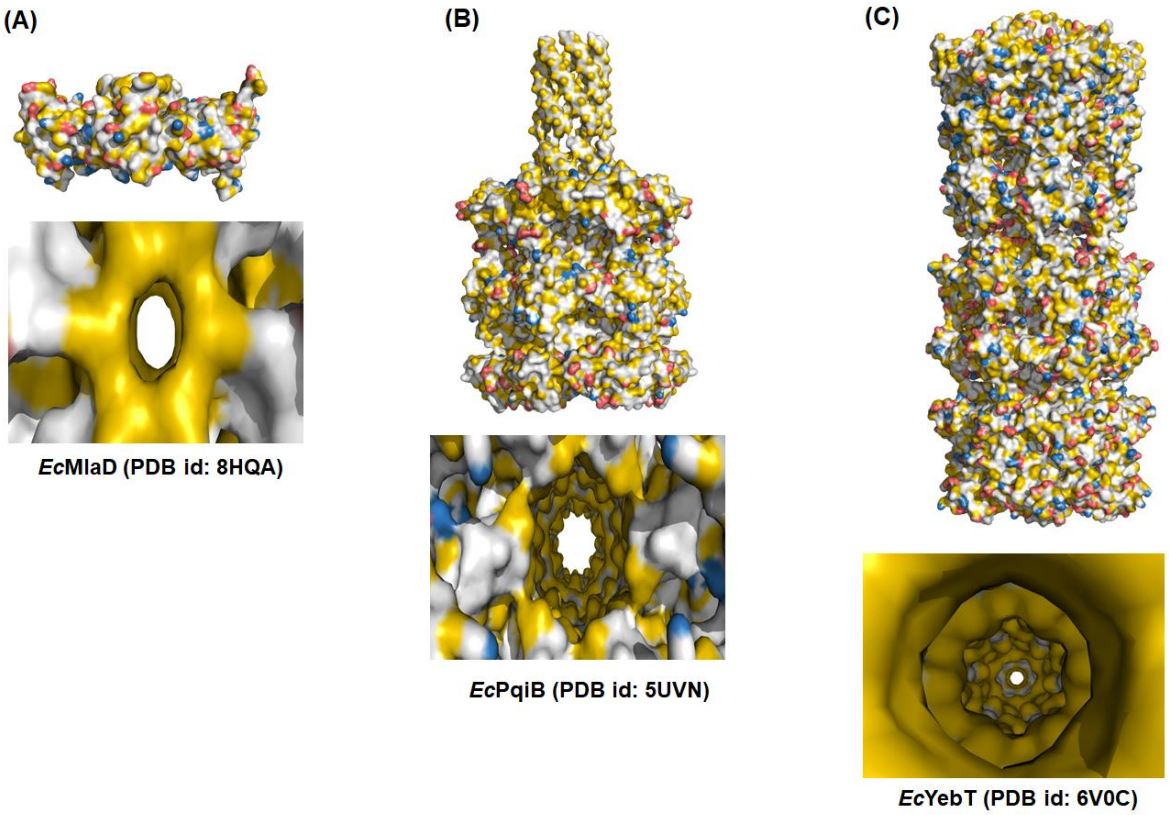


Figure C7. Hydrophobicity mapping of MlaD domain-containing proteins from E. coli. *(A–C) (top) The side view of the EcMlaD_P4$_1$2$_1$2 hexamer (PDB id: 8HQA), EcPqiB hexamer (PDB id: 5UVN), and EcYebT hexamer (PDB ids: 6V0C), respectively. (Bottom) The bottom view of the hydrophobic channels present in these proteins, respectively. The hydrophobicity mapping of the proteins was performed by using the YRB color scheme in the PyMOL program.*

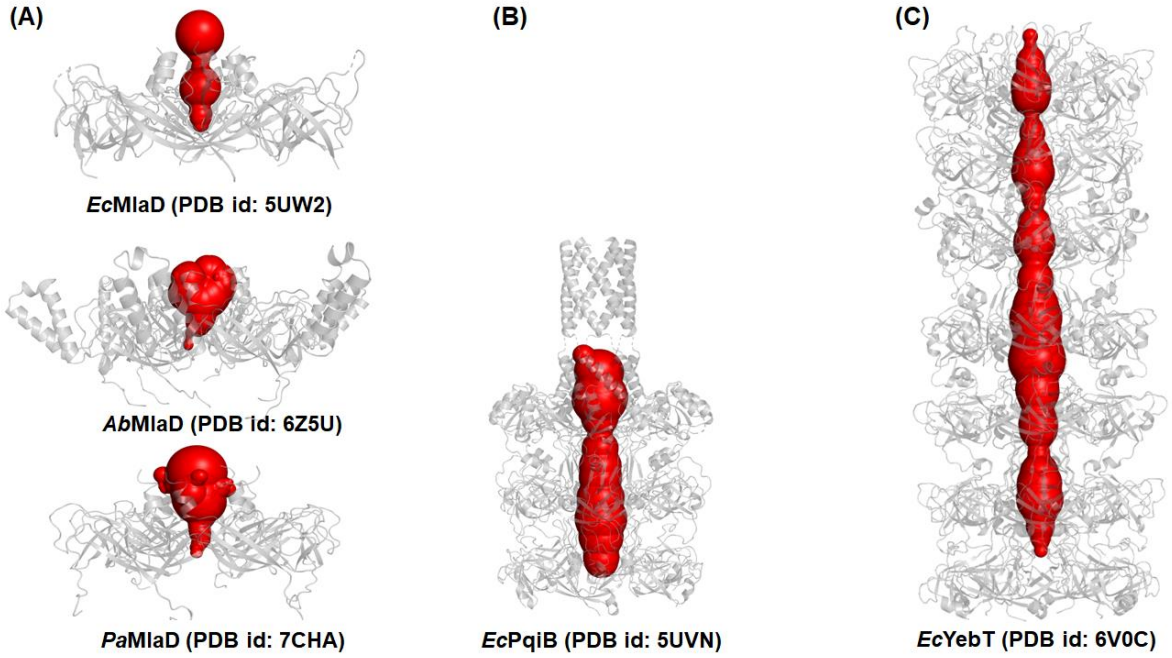


Figure C8. Mapping of the central channels. *The central channel in (A) EcMlaD (PDB id: 5UW2, top), AbMlaD (PDB id: 6Z5U, middle) and PaMlaD (PDB id: 7CHA, bottom), (B) EcPqiB (PDB id: 5UVN) and (C) EcYebT (PDB id: 6V0C), respectively. The central tunnels are depicted as red surfaces.*

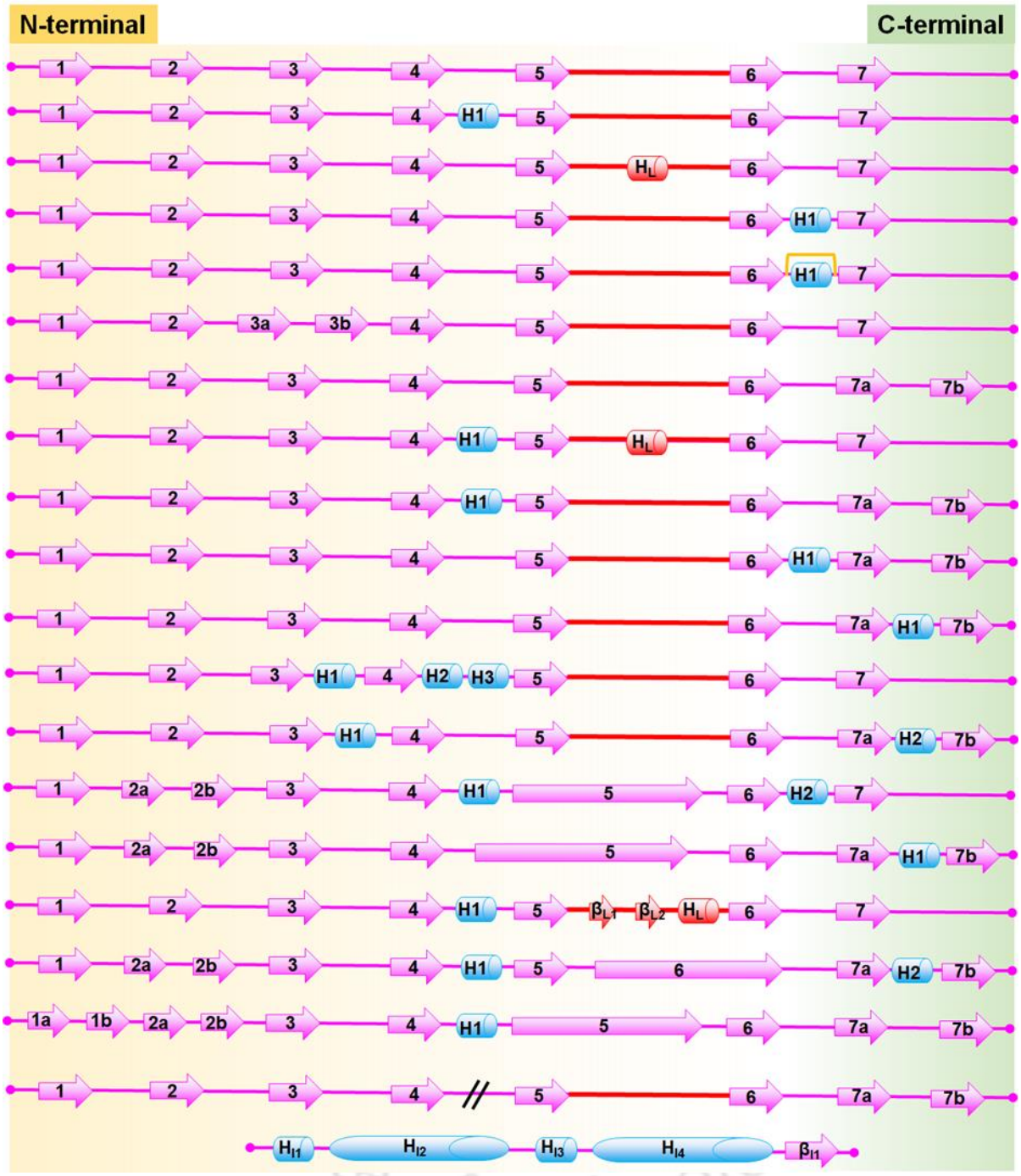


Figure C9. Overview of the secondary structure profiles of MlaD domains. *The β-strands and α-helices are represented as arrows (pink) and cylinders (cyan), respectively. The PLP region containing the β-strands and α-helices is shown in red. The insertion between β4 and β5 strands is represented as two slanted lines. The disulfide bond is highlighted in orange.*

Figure C10. MSA of the MlaD domains from E. coli*. The conservation of hydrophobic, glycine and serine/threonine residues are highlighted in yellow, green and cyan, respectively.*

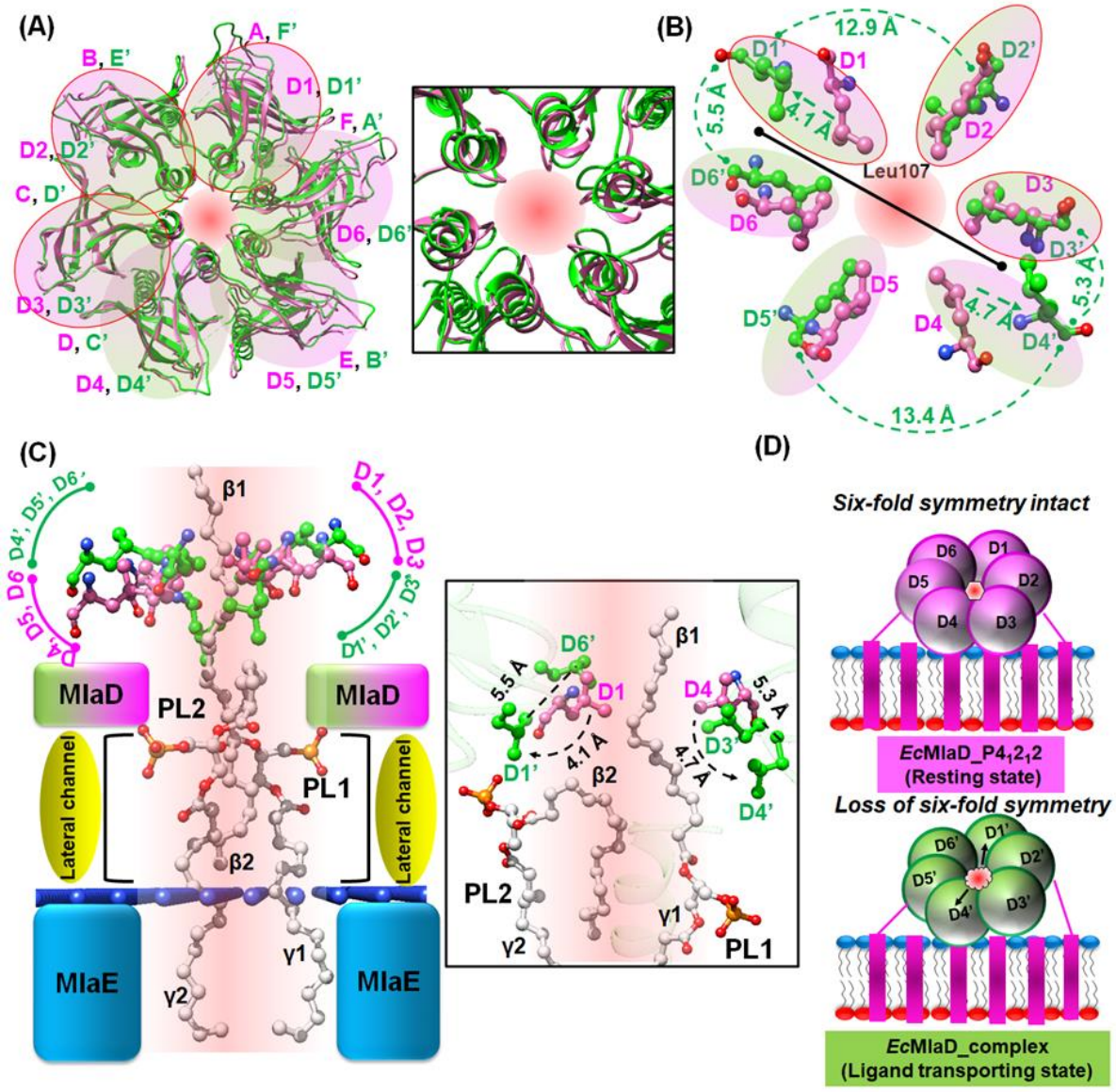


Figure C11. Mechanism of PL transport. *(A) Structural superimposition of EcMlaD_P4₁2₁2 and EcMlaD_complex and the magnified view of the central channels. (B) Structural superimposition of the residue Leu107 of the PLPs from EcMlaD_P4₁2₁2 (pink) and EcMlaD_complex (green). The outward movement of Leu107 is represented by green dotted arrows. The central line (black) shows the hexamers in two halves, while the curved green dotted lines represent the distance between the adjacent Leu107 residues. The residue Leu107 of the protomers demonstrating upward motion is encircled in red. (C) (Left) Orientation of PL1 (β1 & γ1 chains) and PL2 (β2 & γ2 chains) in the translocation pathway of EcMlaFEDB complex (PDB id: 6XBD). (Right) Magnified view of PL1 and PL2 in the translocation pathway. Curved-dotted arrows and straight-dotted lines indicate the movement of Leu107 in D1/D4 and the distance between Leu107 from D1'-D6' & D3'-D4', respectively. The translocation pathway is highlighted in red. (D) Schematic representation of EcMlaD hexamer during (top) resting state (pink) and ligand transporting state (green). In the former, the six-fold symmetry is intact, whereas, in the latter, the six-fold symmetry is lost. The outward movements of D1' and D4' are shown by arrows.*

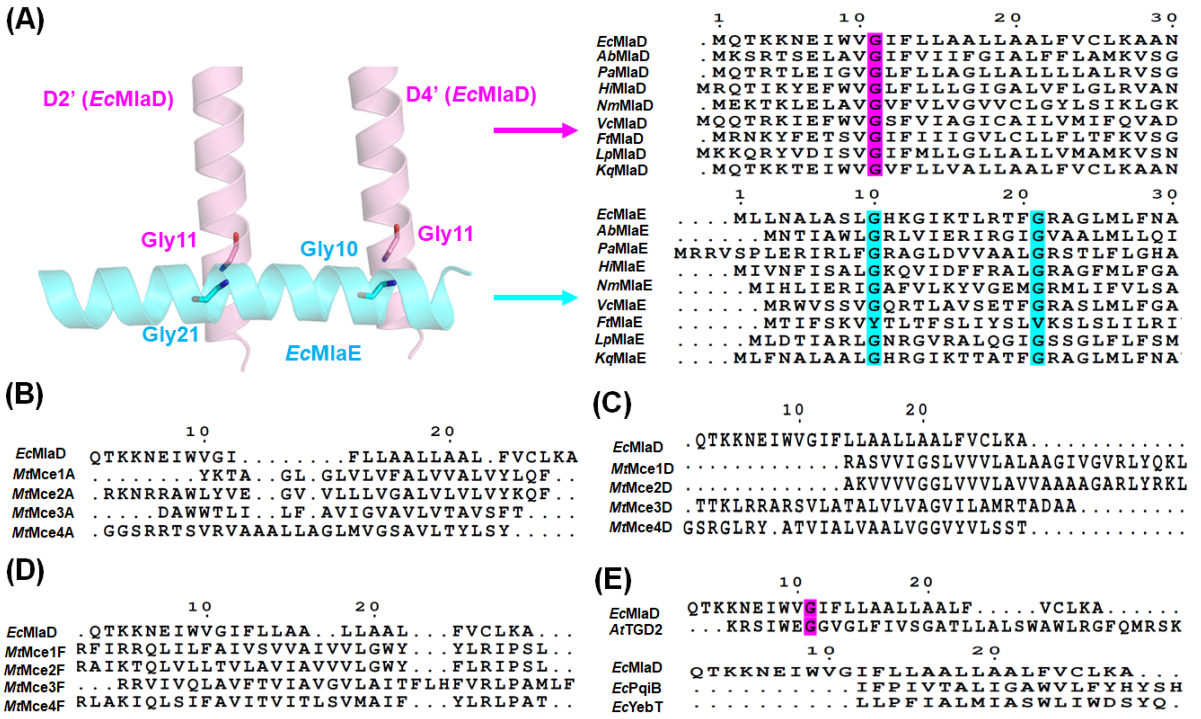


Figure C12. Sequence analysis of the TM of MlaD domain-containing proteins. *(A) (Left) Interaction between the TM helix of EcMlaD (pink) and IF1 helix of EcMlaE (cyan) (PDB id: 6XBD). (Right) MSA of (Top) TM helix of EcMlaD orthologs and (Bottom) IF1 helices of EcMlaE orthologs from Gram-negative bacteria. The conservation of Gly11 in EcMlaD orthologs and Gly10 & Gly21 in EcMlaE is highlighted in magenta and cyan, respectively. MSA of the TM helix from EcMlaD and (B) MtMce(1-4)A, (C) MtMce(1-4)D and (D) MtMce(1-4)F. (E) Sequence comparison of EcMlaD & AtTGD2 (top) and EcMlaD, EcPqiB and & EcYebT (bottom). The conservation of Gly11 is highlighted in magenta.*

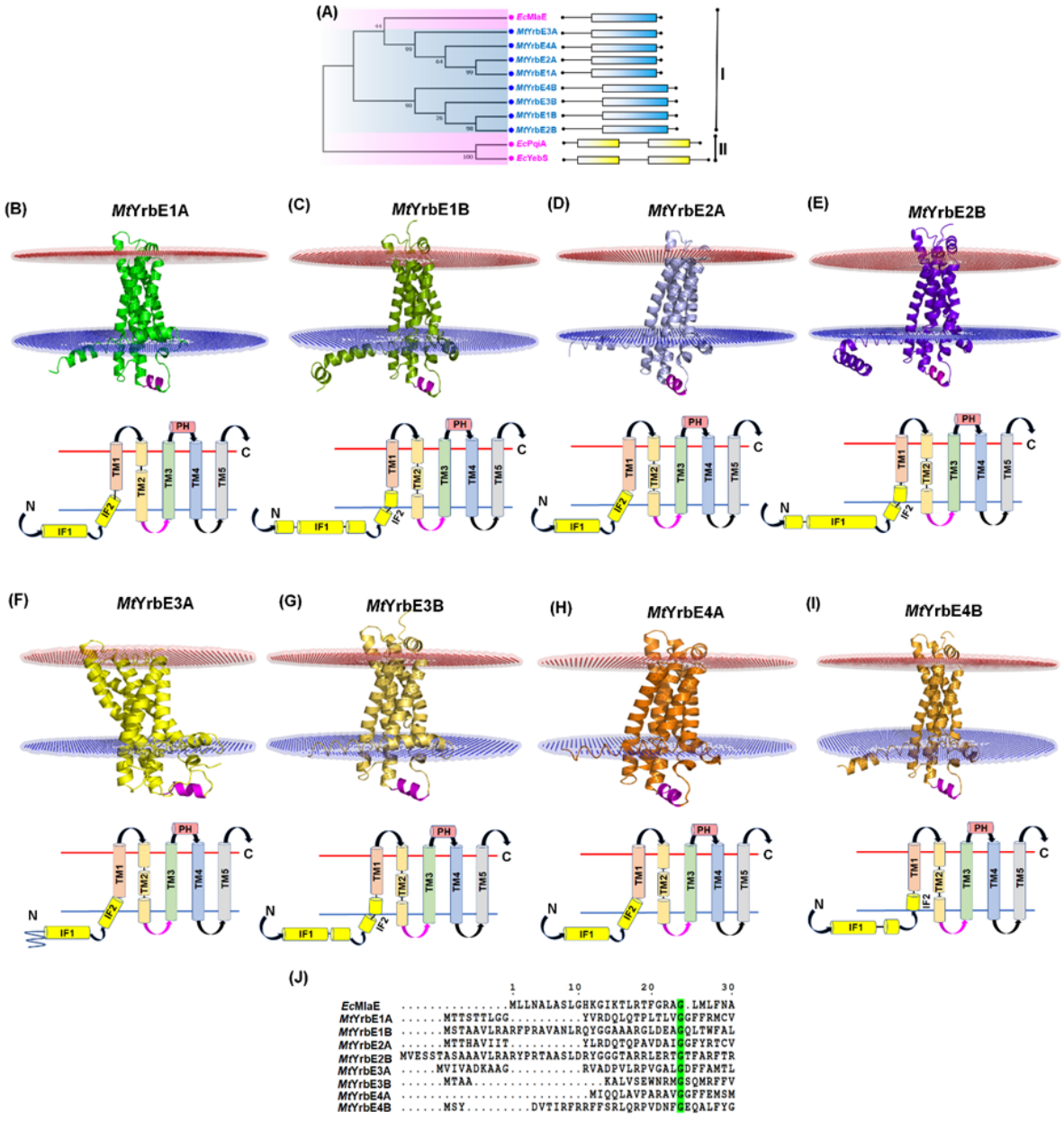


Figure C13. Evolutionary and structural analysis of TMDs. *(A) Phylogenetic tree showing the evolutionary relationship of EcMlaE with EcPqiA, EcYebS and MtYrbE1A-4B. Numbers at each internal node of the phylogenetic tree signify the probability of its occurrence (in percentage) out of 1000 bootstrap replicates. Besides, the domain arrangement of the proteins MlaE and PqiA are represented as blue and yellow rectangles, respectively. (B-I) (Top-bottom) Orientation and topology analyses of MtYrbE1A (green), 1B (splitpea), 2A (lightblue), 2B (purpleblue), 3A (yellow), 3B (yelloworange), 4A (orange) and 4B (brightorange), respectively. The coupling helices are in magenta. The outer and inner leaflets of IM are represented by red and blue planes, respectively. IF: interfacial helix, TM: transmembrane helix and PH: periplasmic helix. (J) MSA of EcMlaE with MtYrbE1A-4B showing the conservation of Gly24 in green.*

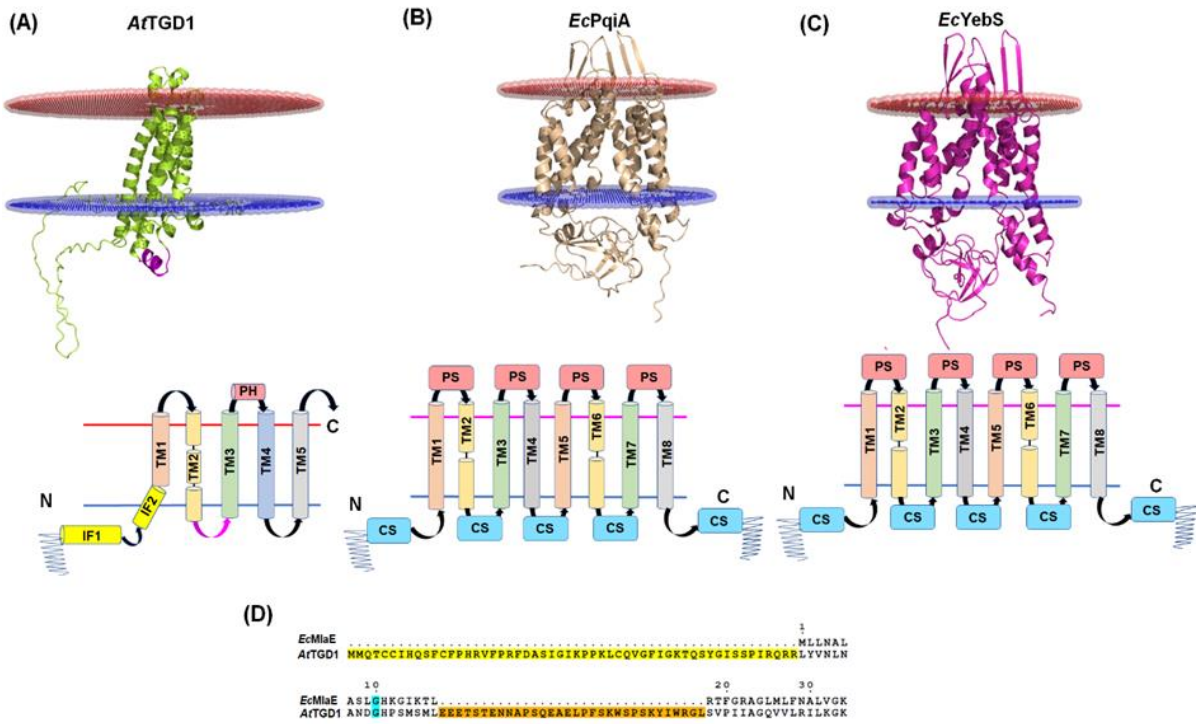


Figure C14. Structure and sequence analysis of TMDs. *(A-C) (Top-bottom) Orientation and topology analyses of AtTGD1 (limon) and EcPqiA (wheat) & EcYebS (lightmagenta), respectively. The coupling helices are in magenta. The outer and inner leaflets of IM are represented by red and blue planes, respectively. IF: interfacial helix, TM: transmembrane helix, PH: periplasmic helix, PS: periplasmic segment and CS: cytoplasmic segment. (D) Pairwise sequence alignment of EcMlaE and AtTGD1 showing the elongation (yellow) and insertion (orange) as well as the conservation of Gly10 in cyan.*

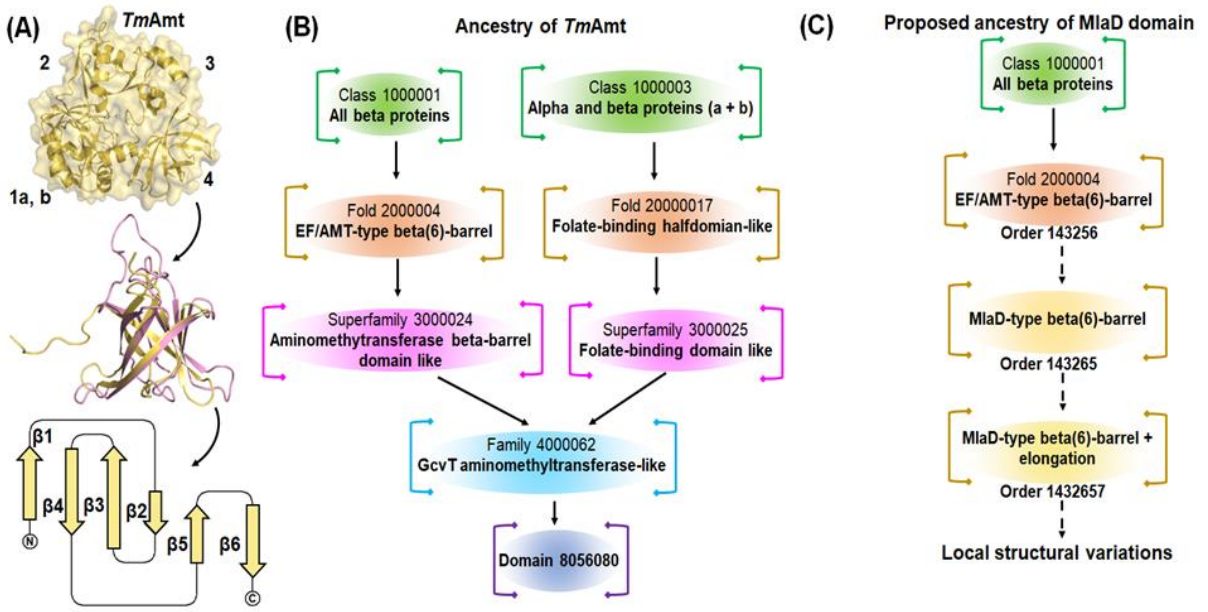


Figure C15. Ancestry analysis of the MlaD domain. *(A) (Top) Overall structure of TmAmt (PDB id: 1WOS) highlighting the four domains. (Middle) Structural superimposition of EcMlaD_MlaD1 and Domain 4 of TmAmt. (Bottom) Topology diagram of Domain 4 of TmAmt. (B) Ancestry of TmAmt as per SCOP database. (C) Proposed ancestry of the MlaD domain.*

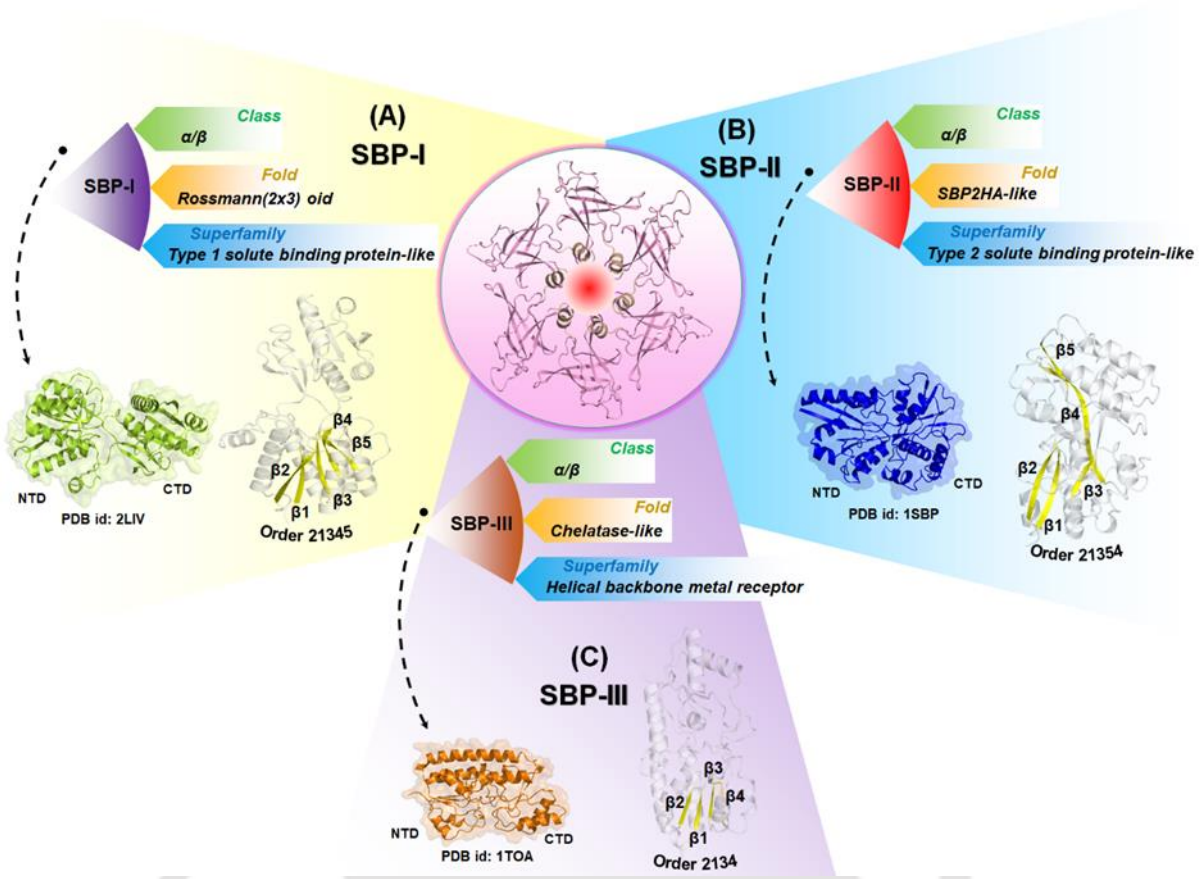


Figure C16. Classification of canonical SBPs. *(Centre) three-dimensional structure of EcMlaD (pink). (A) (Top) Classification of SBP-I, (bottom, left) the three-dimensional structure of Leu/Ile/Val-binding protein (PDB id: 2LIV, limon) and (bottom, right) the general organization of the β-strands. (B) (Top) Classification of SBP-II, (bottom, left) the three-dimensional structure of the sulfate-binding protein (PDB id: 1SBP, blue) and the general organization of the β-strands. (C) (Top) classification of SBP-III, (bottom, left) the three-dimensional structure zinc-binding protein (PDB id: 1TOA, orange) and (bottom, right) the general organization of the β-strands.*

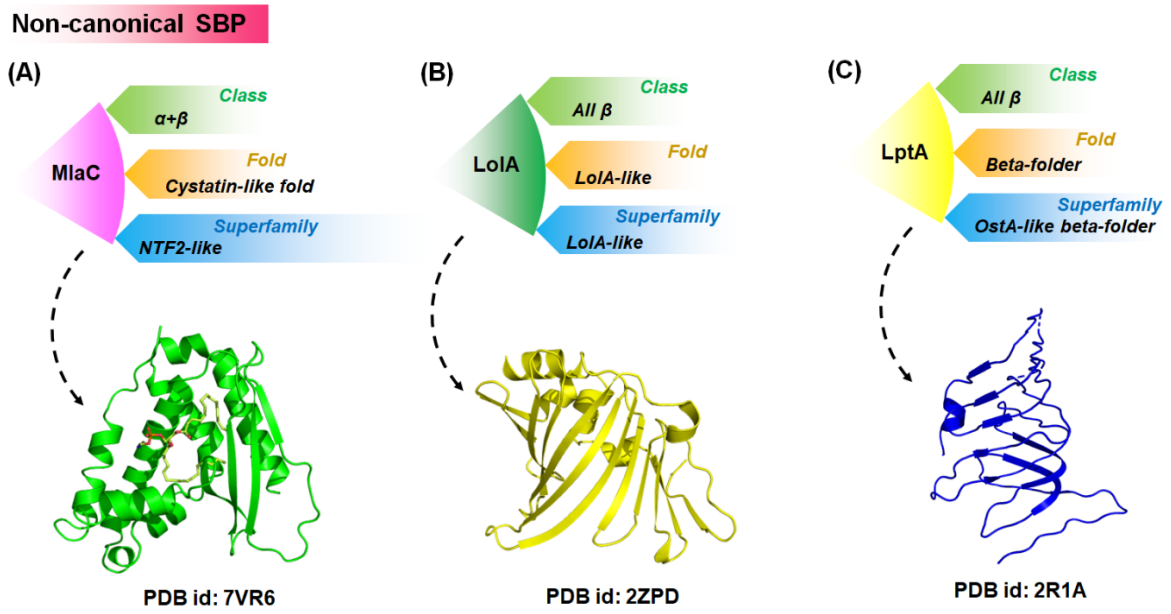


Figure C17. Classification of non-canonical SBPs. *SCOP database classification of (A-C) (Top) MlaC, LolA and LptA, respectively, and (bottom) their crystal structures (PDB ids: 7VR6 (green), 2ZPD (yellow), 2R1A (blue)).*

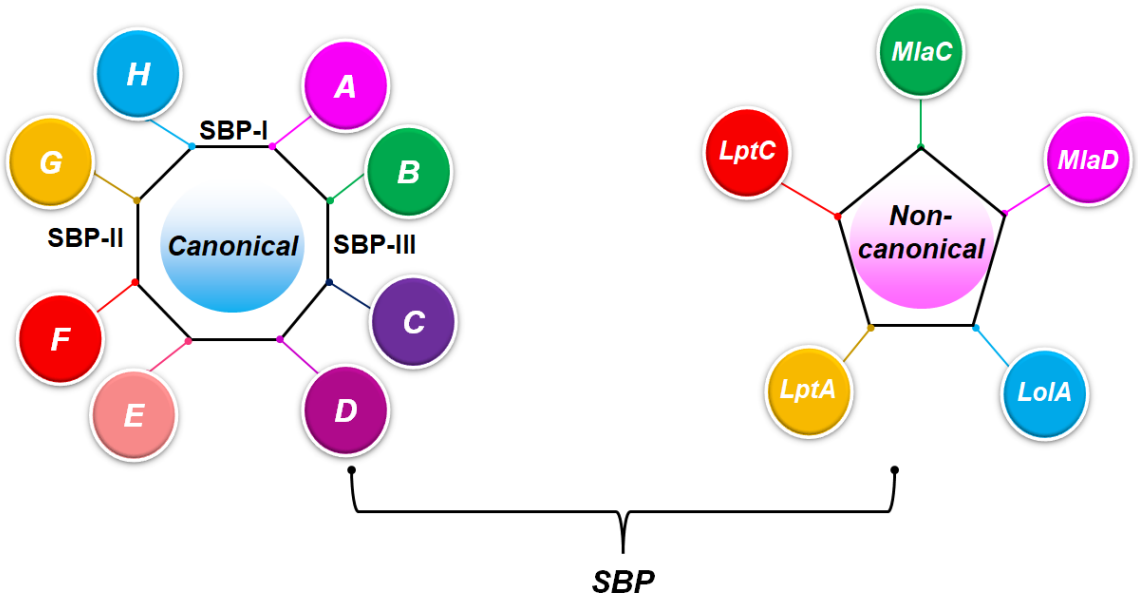


Figure C18. An updated classification of SBPs. *(Left) Canonical SBPs can be divided into eight clusters, A to H. These eight clusters belong to either one of the three SBP classes, I, II and III. (Right) Non-canonical SBPs consist of five different proteins, MlaC, MlaD, LptA, LptC and LolA. The proteins MlaD and LptC are membrane-associated.*

Appendix D

Supplementary data to Chapter 6

Supplementary tables

Table D1. List of the crystal structures analyzed in the study.

PDB ids	Description	Organism	Bound ligand(s)	Group	Annotation	State
6GKI	Structure of *E. coli* MlaC in Variously Loaded States	*Escherichia coli*	-	Group I	*Ec*MlaC(Apo)	Closed, unliganded
8DTE	Crystal Structure of a toluene tolerance periplasmic transport protein from *Neisseria gonorrhoeae*	*Neisseria gonorrhoeae*	-		*Ng*MlaC(Apo)	Open unliganded
7VR6	Crystal structure of MlaC from *Escherichia coli* in quasi-open state	*Escherichia coli*	PEF	Group II	*Ec*MlaC(Holo1)	Open liganded
5UWA	Structure of *E. coli* phospholipid binding protein MlaC		8ND		*Ec*MlaC(Holo2)	Open liganded
2QGU	Three-dimensional structure of the phospholipid-binding protein from *Ralstonia solanacearum* Q8XV73_RALSQ in complex with a phospholipid at the resolution 1.53 A. Northeast Structural Genomics Consortium target RsR89	*Ralstonia solanacearum*	PEF		*Rs*MlaC(Holo)	Open liganded
5UWB	Re-refined 4FCZ: lipid-bound crystal structure of toluene-tolerance protein from *Pseudomonas putida*	*Pseudomonas putida*	PEF, PEF	Group III	*Pp*MlaC(Holo)	Open liganded
6HSY	Two-phospholipid-bound crystal structure of the substrate-binding protein Ttg2D from *Pseudomonas aeruginosa*	*Pseudomonas aeruginosa*	GOT, H3T		*Pa*MlaC(Holo)	Open liganded

Table D2. Summary of the sequence and structure comparison of MlaC orthologues.

Pairwise structural comparison (RMSD in Å)							
PDB id	**7VR6**	**6GKI**	**5UWA**	**2QGU**	**5UWB**	**6HSY**	**8DTE**
7VR6	0	1. 4	0.7	1.7	3.2	3.7	2.0
6GKI	-	0	1.8	2.1	3.6	3.9	2.4
5UWA	-	-	0	1.9	3.6	3.9	1.9
2GQU	-	-	-	0	1.4	1.2	1.7
5UWB	-	-	-	-	0	1.0	2.8
6HSY	-	-	-	-	-	0	2.5
8DTE	-	-	-	-	-	-	0
Pairwise sequence comparison (identity/similarity (query coverage) in percentage)							
UniProt (PDB) id	**P0ADV7 (7VR6)**	**P0ADV7 (6GKI)**	**P0ADV7 (5UWA)**	**Q8XV73 (2QGU)**	**Q88P91 (5UWB)**	**Q9HVW4 (6HSY)**	**B4RRD5 (8DTE)**
P0ADV7 (7VR6)	100/100 (100)	100/100 (100)	100/100 (100)	29/51 (63)	22/46 (78)	22/42 (81)	30/56 (63)
P0ADV7 (6GKI)	-	100/100 (100)	100/100 (100)	29/51 (63)	22/46 (78)	22/42 (81)	30/56 (63)
P0ADV7 (5UWA)	-	-	100/100 (100)	29/51 (63)	22/46 (78)	22/42 (81)	30/56 (63)
Q8XV73 (2GQU)	-	-	-	100/100 (100)	24/50 (73)	26/48 (88)	32/50 (85)
Q88P91 (5UWB)	-	-	-	-	100/100 (100)	66/80 (95)	29/48 (56)
Q9HVW4 (6HSY)	-	-	-	-	-	100/100 (100)	NS
B4RRD5 (8DTE)	-	-	-	-	-	-	100/100 (100)

Table D3. Subdomain boundaries in the MlaC structures.

S. No.	Subdomain (superfamily)	PDB id	Subdomain boundary
1.	D1R1 (NTF2-like)	7VR6	25-69
		5UWA	24-69
		6GKI	23-69
		2QGU	29-73
		5UWB	23-68
		6HSY	23-68
		8DTE	23-63
2.	D2R1 (PBP)	7VR6	70-112
		5UWA	70-112
		6GKI	70-112
		2QGU	74-116
		5UWB	69-112
		6HSY	69-112
		8DTE	64-106
3.	D1R2 (NTF2-like)	7VR6	113-172
		5UWA	113-172
		6GKI	113-172
		2QGU	117-172
		5UWB	113-168
		6HSY	113-169
		8DTE	107-164
4.	D2R2 (PBP)	7VR6	173-209
		5UWA	173-208
		6GKI	173-207
		2QGU	173-207
		5UWB	169-209
		6HSY	170-212
		8DTE	165-194

Table D4. Overview of the molecular docking studies

Sl. No.	Ligand description	Ligand id	Formula	6GKI	7VR6	6HSY	
				Binding energy (kcal/mol)			
1	N-{(2S,3R,4E)-1-[(3,6-di-O-sulfo-beta-D-galactopyranosyl)oxy]-3-hydroxyoctadec-4-en-2-yl}dodecanamide	0SG	$C_{36}H_{69}NO_{14}S_2$	NB	-4.45	-6.21	-5.05
2	(1S)-2-HYDROXY-1-[(NONANOYLOXY)METHYL]ETHYL MYRISTATE	1EM	$C_{26}H_{50}O_5$		-6.02	-5.71	-5.61
3	(2S,3R,4E)-2-(hexadecanoylamino)-3-hydroxyoctadec-4-en-1-yl dihydrogen phosphate	1PX	$C_{34}H_{68}NO_6P$		-3.89	-6.24	-5.92
4	(2S,3R,4E)-2-(dodecanoylamino)-3-hydroxyoctadec-4-en-1-yl dihydrogen phosphate	1PZ	$C_{30}H_{60}NO_6P$		-6.13	-6.16	-5.13
5	(2S,3R,4Z)-3-hydroxy-2-[(9E)-octadec-9-enoylamino]octadec-4-en-1-yl dihydrogen phosphate	1Q0	$C_{36}H_{70}NO_6P$		-4.25	-5.95	-5.55
6	(2S,3R,4E)-3-hydroxy-2-(octanoylamino)octadec-4-en-1-yl dihydrogen phosphate	1T9	$C_{26}H_{52}NO_6P$		-6.8	-6.82	-5.28
7	(2R)-1-{[(R)-hydroxy{[(1R,2R,3R,4R,5S,6R)-2,3,5,6-tetrahydroxy-4-(phosphonooxy)cyclohexyl]oxy}phosphoryl]oxy}-3-(octadecanoyloxy)propan-2-yl (5Z,8Z,11Z,14Z)-icosa-5,8,11,14-tetraenoate	2Y5	$C_{47}H_{84}O_{16}P_2$		-3.94	-2.11	-2.28
8	1,2-DIACYL-SN-GLYCERO-3-PHOSPHOETHANOLAMINE	3PE	$C_{41}H_{82}NO_8P$		-3.87	-5.55	-3.05
9	1,2-DIACYL-GLYCEROL-3-SN-PHOSPHATE	3PH	$C_{39}H_{77}O_8P$		-5.65	-5.84	-5.26
10	(2R)-3-HYDROXYPROPANE-1,2-DIYL DIHEXADECANOATE	4AG	$C_{35}H_{68}O_5$		-5.78		-4.69
11	(1R,4S,6R)-6-({[2-(ACETYLAMINO)-2-DEOXY-ALPHA-D-GLUCOPYRANOSYL]OXY}METHYL)-4-HYDROXY-1-{[(15-METHYLHEXADECANOYL)OXY]METHYL}-4-OXIDO-7-OXO-3,5-DIOXA-8-AZA-4-PHOSPHAHEPTACOS-1-YL 15-METHYLHEXADECANOATE	5PL	$C_{67}H_{129}N_2O_{15}P$		NB		
12	N-((E,2S,3R)-1,3-DIHYDROXYOCTADEC-4-EN-2-YL)HEXANAMIDE	6CM	$C_{24}H_{47}NO_3$		-5.45	-5.96	-5.62

13	(4R,7S)-4-hydroxy-N,N,N-trimethyl-4,9-dioxo-7-[(pentanoyloxy)methyl]-3,5,8-trioxa-4lambda~5~-phosphatetradecan-1-aminium	6O8	$C_{19}H_{39}NO_8P$		-4.49	-4.85	-5.38
14	(2S)-2-(acetyloxy)-3-{[(R)-(2-aminoethoxy)(hydroxy)phosphoryl]oxy}propyl pentanoate	6O9	$C_{12}H_{24}NO_8P$		-4.21	NB	
15	(2S)-3-{[(S)-(2-aminoethoxy)(hydroxy)phosphoryl]oxy}-2-(hexanoyloxy)propyl hexanoate	6OE	$C_{17}H_{34}NO_8P$		-3.76	-4.02	-4.02
16	[(2~{R})-1-[2-azanylethoxy(oxidanyl)phosphoryl]oxy-3-hexadecanoyloxy-propan-2-yl] (~{Z})-octadec-9-enoate	6OU	$C_{39}H_{76}NO_8P$		-3.9	-4.09	-3.87
17	(1R)-2-(phosphonooxy)-1-[(tridecanoyloxy)methyl]ethyl pentadecanoate	6PH	$C_{31}H_{61}O_8P$		-6.44	-6.01	-5.47
18	[(2~{S})-3-[[(2~{R})-2,3-bis(oxidanyl)propoxy]-oxidanyl-phosphoryl]oxy-2-hexadec-9-enoyloxy-propyl] hexadecanoat	6V6	$C_{38}H_{73}O_{10}P$		-3.32	-5.33	-3.62
19	[(2R)-2-heptanoyloxy-3-phosphonooxy-propyl] nonanoate	7P9	$C_{19}H_{37}O_8P$		-4.83	-5.2	-4.86
20	(1R)-2-(dodecanoyloxy)-1-[(phosphonooxy)methyl]ethyl tetradecanoate	7PH	$C_{29}H_{57}O_8P$		-6.73	-5.97	-5.08
21	(2S)-3-(2-aminoethoxy)propane-1,2-diyl dihexadecanoate	8ND	$C_{37}H_{73}NO_5$		-6.91	-4.73	-4.4
22	(2R)-3-{[(S)-(2-aminoethoxy)(hydroxy)phosphoryl]oxy}-2-(tetradecanoyloxy)propyl octadecanoate	8PE	$C_{37}H_{74}NO_8P$		-5.95	-4.93	-3.48
23	(1R)-2-{[(S)-(2-aminoethoxy)(hydroxy)phosphoryl]oxy}-1-[(heptanoyloxy)methyl]ethyl octadecanoate	9PE	$C_{30}H_{60}NO_8P$		-4.5	-4.3	-3.82
24	2-azaniumylethyl (2R)-2,3-bis{[(9Z)-octadec-9-enoyl]oxy}propyl phosphate	9TL	$C_{41}H_{78}NO_8P$		-4.11	-5.41	-3.74
25	(2R)-3-(((2-aminoethoxy)(hydroxy)phosphoryl)oxy)-2-(palmitoyloxy)propyl (E)-octadec-9-enoate	9Y0	$C_{39}H_{76}NO_8P$		-2.93	-3.74	-4.03
26	N-((E,2S,3R)-1,3-DIHYDROXYOCTADEC-4-EN-2-YL)PALMITAMIDE	16C	$C_{34}H_{67}NO_3$		-6.66	-5.18	-5.61
27	N-((E,2S,3R)-1,3-DIHYDROXYOCTADEC-4-EN-2-YL)STEARAMIDE	18C	$C_{36}H_{71}NO_3$		-5.88	-5.22	-5.32
28	(2R)-3-(phosphonooxy)propane-1,2-diyl dihexanoate	44E	$C_{15}H_{29}O_8P$		-5.46	-6.28	-5.69

29	(2S)-3-{[(R)-{[(2R)-2,3-dihydroxypropyl]oxy}(hydroxy)phosphoryl]oxy}-2-(hexanoyloxy)propyl hexanoate	44G	$C_{18}H_{35}O_{10}P$		-3.65	-4.17	-3.22
30	(2R)-3-{[(S)-(2-aminoethoxy)(hydroxy)phosphoryl]oxy}-2-(tetradecanoyloxy)propyl tetradecanoate	46E	$C_{33}H_{66}NO_8P$		-5.77	-3.44	-4.85
31	(2S)-2-(BUTYRYLOXY)-3-HYDROXYPROPYL NONANOATE	B3H	$C_{16}H_{30}O_5$		-5.92	-4.44	-4.48
32	(1R)-2-{[(S)-hydroxy{[(1S,2R,3R,4S,5S,6R)-2,3,4,5,6-pentahydroxycyclohexyl]oxy}phosphoryl]oxy}-1-[(octadecanoyloxy)methyl]ethyl (9Z)-octadec-9-enoate	B7N	$C_{45}H_{85}O_{13}P$		-2.22	-2.97	-2.58
33	(2R)-1-(dodecanoyloxy)-3-hydroxypropan-2-yl (5E,8E,11E)-tetradeca-5,8,11-trienoate	BQ9	$C_{29}H_{50}O_5$		-6.03	-6.1	-6.25
34	Cardiolipin	CDL	$C_{81}H_{156}O_{17}P_2$		NB		
35	(15Z)-N-((1S,2R,3E)-2-HYDROXY-1-{[(3-O-SULFO-BETA-D-GALACTOPYRANOSYL)OXY]METHYL}HEPTADEC-3-ENYL)TETRACOS-15-ENAMIDE	CIS	$C_{48}H_{91}NO_{11}S$		-1.38	-4.36	-5.11
36	1-PALMITOYL-2-LINOLEOYL-SN-GLYCERO-3-PHOSPHOCHOLINE	CPL	$C_{42}H_{80}NO_8P$		-4.05	-4.05	-4.62
37	(19S,22R,25R)-22,25,26-trihydroxy-16,22-dioxo-17,21,23-trioxa-22lambda~5~-phosphahexacosan-19-yl (9E)-octadec-9-enoate	D3D	$C_{40}H_{77}O_{10}P$		-5.47	-4.59	-3.27
38	(2R)-1-(hexadecanoyloxy)-3-(phosphonooxy)propan-2-yl (9Z)-octadec-9-enoate	D21	$C_{37}H_{71}O_8P$		-6.22	-3.27	-5.75
39	(2S)-3-hydroxypropane-1,2-diyl didecanoate	DDR	$C_{23}H_{44}O_5$		-5.27	-4.96	-5.15
40	DIACYL GLYCEROL	DGA	$C_{39}H_{76}O_5$		-6	-5.17	-4.62
41	1-[GLYCEROLYLPHOSPHONYL]-2-[8-(2-HEXYL-CYCLOPROPYL)-OCTANAL-1-YL]-3-[HEXADECANAL-1-YL]-GLYCEROL	DGG	$C_{39}H_{75}O_{10}P$		-4.86	-5.06	-3.88
42	1,2-DILINOLEOYL-SN-GLYCERO-3-PHOSPHOCHOLINE	DLP	$C_{44}H_{80}NO_8P$		-4.91	-4.13	-5.26
43	1-CIS-9-OCTADECANOYL-2-CIS-9-HEXADECANOYL PHOSPHATIDYL GLYCEROL	DR9	$C_{40}H_{75}O_{10}P$		-3.14	-4.9	-4.63
44	(2R)-2-(hexadecanoyloxy)propyl nonadecanoate	E2V	$C_{38}H_{74}O_4$		-7.3	-5.88	-6.77

45	N-{(2S,3R,4E)-3-hydroxy-1-[(3-O-sulfo-beta-D-galactopyranosyl)oxy]octadec-4-en-2-yl}dodecanamide	EIS	$C_{36}H_{69}NO_{11}S$		-5.47	-4.92	-5.54
46	[(2S)-2-octadecanoyloxypropyl] octadecanoate	EL6	$C_{39}H_{76}O_4$		-5.76	-4.21	-4.9
47	L-ALPHA-PHOSPHATIDYL-BETA-OLEOYL-GAMMA-PALMITOYL-PHOSPHATIDYLETHANOLAMINE	EPH	$C_{39}H_{68}NO_8P$		-5.8	-5.43	-4.88
48	(2R)-1-(decanoyloxy)-3-(phosphonooxy)propan-2-yl octadecanoate	F57	$C_{31}H_{61}O_8P$		-6.23	-5.13	-5.23
49	(2S)-3-hydroxypropane-1,2-diyl ditetradecanoate	FAW	$C_{31}H_{60}O_5$		-6.94	-4.77	-4.95
50	(2R)-3-hydroxypropane-1,2-diyl dihexanoate	FGJ	$C_{15}H_{28}O_5$		-5.34	-4.35	-4.35
51	1,2-Dioctanoyl-SN-Glycero-3-Phosphoethanolamine	G8C	$C_{21}H_{42}NO_8P$		-3.94	-4.84	-5.35
52	N-{1-[(HEXOPYRANOSYLOXY)METHYL]-2-HYDROXYNONADECYL}TETRACOSANAMIDE	GM3	$C_{50}H_{99}NO_8$		-3.03	-1.22	NB
53	Phospholipid PG(16:0/cy17:0)	GOT	$C_{39}H_{75}O_{10}P$		-3.13	-4.14	-4.24
54	(1R)-2-{[(S)-(2-aminoethoxy)(hydroxy)phosphoryl]oxy}-1-[(pentadecanoyloxy)methyl]ethyl (12E)-hexadeca-9,12-dienoate	GP7	$C_{36}H_{68}NO_8P$		-3.85	-4.33	NB
55	Phospholipid PG(16:0/cy17:0)	H3T	$C_{39}H_{75}O_{10}P$		-2.78	-3.78	-3.87
56	[(2~{S})-1-octadecanoyloxy-3-[oxidanyl-[(1~{R},2~{R},3~{S},4~{S},5~{S},6~{S})-2,3,6-tris(oxidanyl)-4,5-diphosphonooxy-cyclohexyl]oxy-phosphoryl]oxy-propan-2-yl] icosa-5,8,11,14-tetraenoate	IEP	$C_{47}H_{85}O_{19}P_3$		NB	-3.16	-0.22
57	(2S)-3-HYDROXY-2-(NONANOYLOXY)PROPYL LAURATE	L2C	$C_{24}H_{46}O_5$		-4.73	-4.34	-4.01
58	(1S)-2-{[(S)-(2-aminoethoxy)(hydroxy)phosphoryl]oxy}-1-[(octadecanoyloxy)methyl]ethyl (9Z)-octadec-9-enoate	L9Q	$C_{41}H_{80}NO_8P$		-2.06	-4.05	-5.18
59	(2S)-1-hydroxy-3-(tetradecanoyloxy)propan-2-yl docosanoate	L44	$C_{39}H_{76}O_5$		-4.65	-4.23	-2.09
60	1,2-DIPALMITOYL-PHOSPHATIDYL-GLYCEROLE	LHG	$C_{38}H_{75}O_{10}P$		-3.27	-4.13	-4.38
61	(1R)-2-{[(R)-(2-AMINOETHOXY)(HYDROXY)PHOSPHORYL]OXY}-1-[(DODECANOYLOXY)METHYL]ETHYL (9Z)-OCTADEC-9-ENOATE	LOP	$C_{35}H_{68}NO_8P$		-4.43	-5.12	-4.61

62	2-(HEXADECANOYLOXY)-1-[(PHOSPHONOOXY)METHYL]ETHYL HEXADECANOATE	LPP	$C_{35}H_{69}O_8P$		-6.8	-4.25	-4.87
63	(2R)-2-(hexadecanoyloxy)-3-{[(10R)-10-methyloctadecanoyl]oxy}propyl phosphate	M7U	$C_{38}H_{75}O_8P$		-6.24	-6.08	-6.4
64	1,2-DIMYRISTOYL-RAC-GLYCERO-3-PHOSPHOCHOLINE	MC3	$C_{36}H_{72}NO_8P$		-4.69	-4.56	-4.2
65	(2R)-3-(phosphonooxy)propane-1,2-diyl (9Z,9'Z)bis-octadec-9-enoate	MX7	$C_{39}H_{73}O_8P$		-5.07	-5	-5.7
66	(2R)-3-(PHOSPHONOOXY)-2-(TETRADECANOYLOXY)PROPYL PALMITATE	MYY	$C_{33}H_{65}O_8P$		-5.43	-6.64	-5.38
67	(7R,17E)-4-HYDROXY-N,N,N,7-TETRAMETHYL-7-[(8E)-OCTADEC-8-ENOYLOXY]-10-OXO-3,5,9-TRIOXA-4-PHOSPHAHEPTACOS-17-EN-1-AMINIUM 4-OXIDE	OPC	$C_{45}H_{87}NO_8P$		-3.31	-4.26	-4.13
68	(2R)-3-{[(R)-{[(2S)-2,3-dihydroxypropyl]oxy}(hydroxy)phosphoryl]oxy}-2-[(6Z)-tridec-6-enoyloxy]propyl (9Z)-octadec-9-enoate	OZ2	$C_{37}H_{69}O_{10}P$		-3.64	-2.02	-3.92
69	O-[(R)-{[(2R)-2,3-bis(octadecanoyloxy)propyl]oxy}(hydroxy)phosphoryl]-L-serine	P5S	$C_{42}H_{82}NO_{10}P$		-2.66	-3.31	-3.91
70	(2S)-3-{[{[(2S)-2,3-DIHYDROXYPROPYL]OXY}(HYDROXY)PHOSPHORYL]OXY}-2-[(6E)-HEXADEC-6-ENOYLOXY]PROPYL (8E)-OCTADEC-8-ENOATE	P6L	$C_{40}H_{75}O_{10}P$		-1.96	-4.87	-2.1
71	(R)-2-(FORMYLOXY)-3-(PHOSPHONOOXY)PROPYL PENTANOATE	PA6	$C_9H_{17}O_8P$		NB		
72	1,2-DIOCTANOYL-SN-GLYCERO-3-PHOSPHATE	PA8	$C_{19}H_{36}O_8P$		-6.07	-6.19	-6.31
73	1,2-DIACYL-SN-GLYCERO-3-PHOSPHOCHOLINE	PC1	$C_{44}H_{88}NO_8P$		-4.86	-5.58	-5.5
74	1,2-DIACYL-SN-GLYCERO-3-PHOSHOCHOLINE	PCF	$C_{40}H_{80}NO_8P$		-3.6	-2.52	-4.36
75	1,2-DIOLEOYL-SN-GLYCERO-3-PHOSPHOCHOLINE	PCW	$C_{44}H_{85}NO_8P$		-4.53	-4.54	-6.34
76	(2R)-3-(phosphonooxy)propane-1,2-diyl diheptanoate	PD7	$C_{17}H_{33}O_8P$		-5.48	-5.87	-5.84
77	1,2-Dioleoyl-sn-glycero-3-phosphoethanolamine	PEE	$C_{41}H_{83}NO_8P$		-2.99	-5.31	-4.84
78	DI-PALMITOYL-3-SN-PHOSPHATIDYLETHANOLAMINE	PEF	$C_{37}H_{74}NO_8P$		-4	-4.77	-5.48

79	DI-STEAROYL-3-SN-PHOSPHATIDYLETHANOLAMINE	PEH	$C_{41}H_{82}NO_8P$		-3.06	-3.56	-4.71
80	(1S)-2-{[(2-AMINOETHOXY)(HYDROXY)PHOSPHORYL]OXY}-1-[(STEAROYLOXY)METHYL]ETHYL (5E,8E,11E,14E)-ICOSA-5,8,11,14-TETRAENOATE	PEK	$C_{43}H_{78}NO_8P$		-0.73	-4.73	-3.66
81	(1S)-2-{[(2-AMINOETHOXY)(HYDROXY)PHOSPHORYL]OXY}-1-[(PALMITOYLOXY)METHYL]ETHYL STEARATE	PEV	$C_{39}H_{78}NO_8P$		-6.2	-5.65	-4.2
82	1,2-DIDECANOYL-SN-GLYCERO-3-PHOSPHOETHANOLAMINE	PEX	$C_{25}H_{49}NO_8P$		-4.75	-3.56	-4.31
83	(1R)-2-{[{[(2R)-2,3-DIHYDROXYPROPYL]OXY}(HYDROXY)PHOSPHORYL]OXY}-1-[(PALMITOYLOXY)METHYL]ETHYL (9S,10S)-9,10-DIBROMOOCTADECANOATE	PGK	$C_{40}H_{77}Br_2O_{10}P$		-1.72	-4.32	-2.46
84	(1S)-2-{[{[(2R)-2,3-DIHYDROXYPROPYL]OXY}(HYDROXY)PHOSPHORYL]OXY}-1-[(PALMITOYLOXY)METHYL]ETHYL STEARATE	PGT	$C_{40}H_{79}O_{10}P$		-3.98	-3.77	-3.06
85	(1R)-2-{[{[(2S)-2,3-DIHYDROXYPROPYL]OXY}(HYDROXY)PHOSPHORYL]OXY}-1-[(PALMITOYLOXY)METHYL]ETHYL (11E)-OCTADEC-11-ENOATE	PGV	$C_{40}H_{77}O_{10}P$		-2.97	-0.72	-0.78
86	(1R)-2-{[(S)-{[(2S)-2,3-dihydroxypropyl]oxy}(hydroxy)phosphoryl]oxy}-1-[(hexadecanoyloxy)methyl]ethyl (9Z)-octadec-9-enoate	PGW	$C_{40}H_{77}O_{10}P$		-4.66	-4.02	-5.45
87	(2R)-2-(dodecanoyloxy)propyl (4E,6E,8E,10E,12E)-pentadeca-4,6,8,10,12-pentaenoate	PJZ	$C_{30}H_{48}O_4$		-6.54	-5.05	-5.07
88	(2S)-3-(hexadecanoyloxy)-2-[(9Z)-octadec-9-enoyloxy]propyl 2-(trimethylammonio)ethyl phosphate	POV	$C_{42}H_{82}NO_8P$		-3.4	-4.52	-4.39
89	(7R,17E,20E)-4-HYDROXY-N,N,N-TRIMETHYL-9-OXO-7-[(PALMITOYLOXY)METHYL]-3,5,8-TRIOXA-4-PHOSPHAHEXACOSA-17,20-DIEN-1-AMINIUM 4-OXIDE	PSC	$C_{42}H_{81}NO_8P$		-3.83	-5.12	-3.14
90	1,2-DICAPROYL-SN-PHOSPHATIDYL-L-SERINE	PSF	$C_{18}H_{34}NO_{10}P$		-3.87	-4.38	-3.77
91	PHOSPHATIDYLETHANOLAMINE	PTY	$C_{40}H_{80}NO_8P$		-2.68	-1.22	-4.32
92	1,2-DILAUROYL-SN-GLYCERO-3-PHOSPHATE	PX2	$C_{27}H_{52}O_8P$		-6.99	-5.38	-4.96

93	1,2-DIPALMITOYL-SN-GLYCERO-3-PHOSPHATE	PX6	$C_{35}H_{68}O_8P$		-1.57	-6.82	-5.74
94	1,2-DISTEAROYL-SN-GLYCERO-3-PHOSPHATE	PX8	$C_{39}H_{76}O_8P$		-5.56	-6.25	-6.34
95	(2S)-propane-1,2-diyl dihexadecanoate	PXS	$C_{35}H_{68}O_4$		-6.23	-5.35	-6.17
96	(2S)-1-(nonanoyloxy)-3-(phosphonooxy)propan-2-yl tetradecanoate	QNP	$C_{26}H_{51}O_8P$		-5.57	-7.34	-4.87
97	(7Z,19R,22R)-25-amino-22-hydroxy-16,22-dioxo-17,21,23-trioxa-22lambda~5~-phosphapentacos-7-en-19-yl (9Z)-octadec-9-enoate	RXY	$C_{39}H_{74}NO_8P$		-4.03	-4.21	-5.28
98	[(2~{S})-1-[2-azanylethoxy(oxidanyl)phosphoryl]oxy-3-octanoyloxy-propan-2-yl] octadecanoate	SBJ	$C_{31}H_{62}NO_8P$		-4.37	-1.27	-4.2
99	Phosphatidylinositol	T7X	$C_{47}H_{83}O_{13}P$		-1.81	-3.32	NB
100	1,2-DIMYRISTOYL-SN-GLYCERO-3-PHOSPHATE	XP4	$C_{31}H_{60}O_8P$		-4.57	-5.19	-5.46
101	(2S)-2-(hexadecanoyloxy)-3-hydroxypropyl (9Z)-octadec-9-enoate	YZY	$C_{37}H_{70}O_5$		-7.31	-4.51	-5.46
102	(2S)-1-(hexadecanoyloxy)-3-hydroxypropan-2-yl (11Z)-octadec-11-enoate	Z0P	$C_{37}H_{70}O_5$		-5.31	-4.78	-4.95
103	(2S)-3-hydroxypropane-1,2-diyl dihexadecanoate	Z41	$C_{35}H_{68}O_5$		-6.46	-5.13	-4.66
104	(2R)-3-{[(S)-(2-aminoethoxy)(hydroxy)phosphoryl]oxy}-2-(tetradec-5-enoyloxy)propyl (11Z)-octadec-11-enoate	ZPE	$C_{37}H_{70}NO_8P$		-3.84	-1.66	-5.08

NB stands for no binding.

Table D5. List of structures used for molecular docking and molecular dynamics study.

PDB ids	Description	Organism	Bound ligand(s)	Group	Annotation†	State
6GKI	Structure of *E. coli* MlaC in Variously Loaded States	*Escherichia coli*	-	Group I	*Ec*MlaC(Apo)	Closed, unliganded
8DTE	Crystal Structure of a toluene tolerance periplasmic transport protein from *Neisseria gonorrhoeae*	*Neisseria gonorrhoeae*			*Ng*MlaC(Apo)	Open unliganded
7VR6	Crystal structure of MlaC from *Escherichia coli* in quasi-open state	*Escherichia coli*	PEF	Group II	*Ec*MlaC(Holo)	Open liganded
					*Ec*MlaC(Apo*)	Open unliganded
6HSY	Two-phospholipid-bound crystal structure of the substrate-binding protein Ttg2D from *Pseudomonas aeruginosa*	*Pseudomonas aeruginosa*	GOT, H3T	Group III	*Pa*MlaC(Holo)	Open liganded
					*Pa*MlaC(Apo*)	Open unliganded

†The apo systems generated by removing the endogenously bound PL(s) are designated with asterisks.

Table D6. Summary of the RMSDs and convergence of the simulated systems.

Sl. No.	System	RMSD range (Å)	Convergence (ns)
1	*Ec*MlaC(Apo)	1.00-2.00	300
2	*Ng*MlaC(Apo)	2.50-3.50	400
3	*Ec*MlaC(Holo1)	1.50-2.25	500
4	*Ec*MlaC(Apo*)	2.50-3.00	500
5	*Pa*MlaC(Holo)	1.50-3.00	-
6	*Pa*MlaC(Apo*)	4.00-5.00	-

Table D7. Overview of the residues around the binding pocket.

Protein	Distance 1	Distance 2	Angle forming residues
*Ec*MlaC(Apo)	Ala108-Glu169	Tyr112-Glu169	Tyr112-Gly182-Glu169
*Ec*MlaC(Holo1)			
*Ng*MlaC(Apo)	Thr102-Glu161	Phe106-Glu161	Phe106-Gly174-Glu161
*Pa*MlaC(Holo)	Ala108-Asn166	Tyr112-Asn166	Tyr112-Ala179-Asn166

Supplementary figures

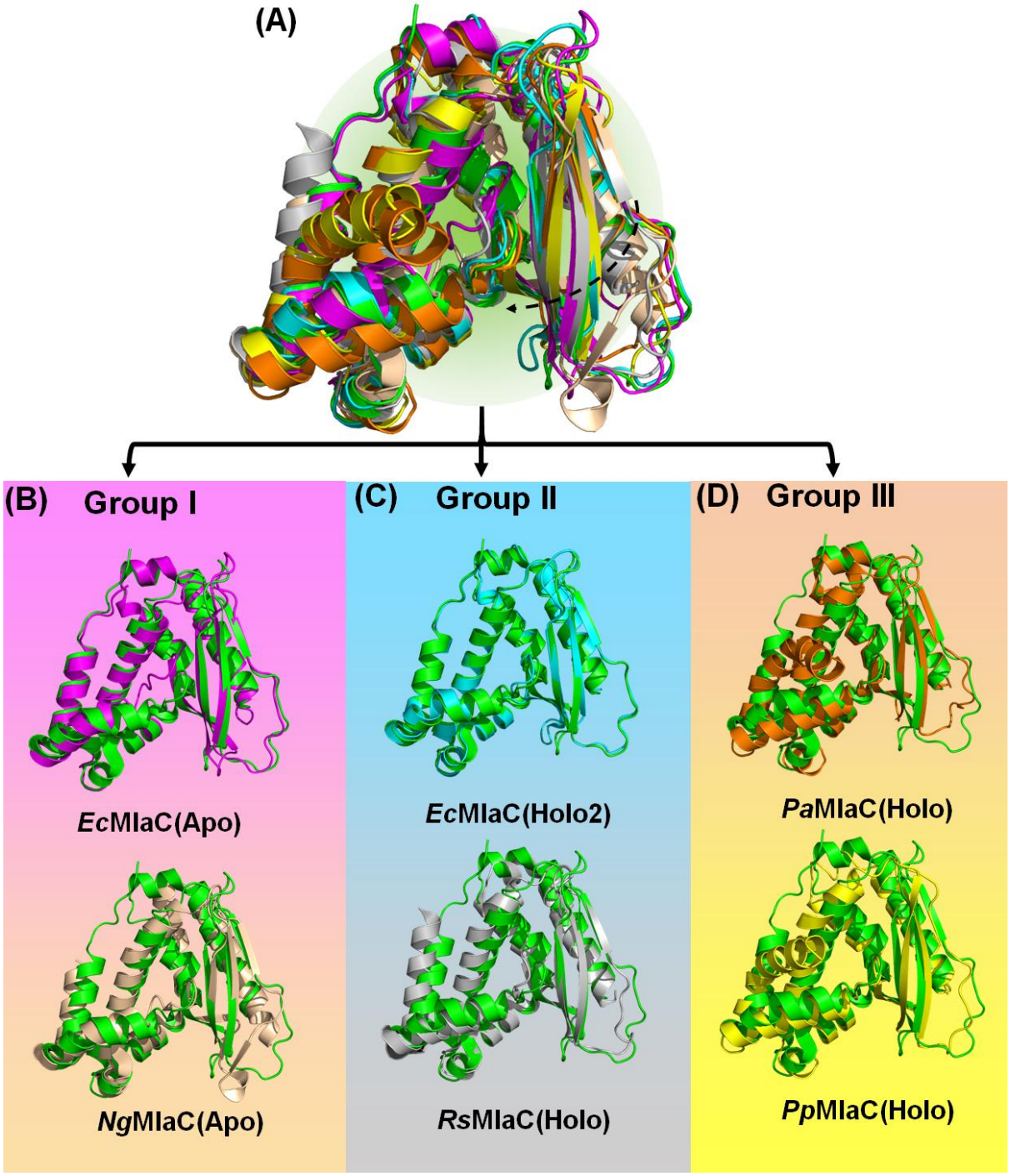


Figure D1. Structural comparison of EcMlaC orthologues. *(A) Structural superimposition of EcMlaC(Holo1)(Green) with EcMlaC(Apo)(Magenta), NgMlaC(Apo)(Wheat), EcMlaC(Holo2)(Cyan), RsMlaC(Holo)(Grey), PaMlaC(Holo)(Orange) and PpMlaC(Holo)(Yellow), respectively. The curved arrow denotes the change in curvature of β-sheets. (B-D) Structutal superimposition of EcMlaC(Holo1)(Green) with Group I, II and III members, respectively.*

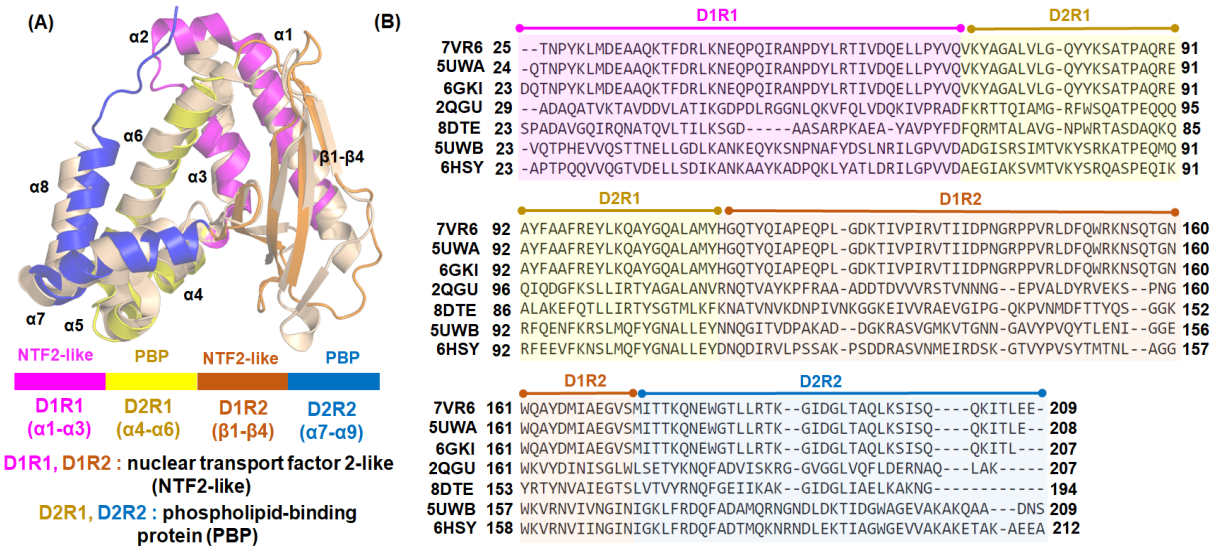


Figure D2. Domain analysis of EcMlaC orthologs. *(A) Structural superimposition of EcMlaC(Holo1)(Magenta-yellow-orange-blue) and NgMlaC(Apo)(Wheat). The subdomains D1R1 and D1R2 (NTF2-like family) are highlighted in magenta and orange. The subdomains D2R1 and D2R2 (PBP family) are highlighted in yellow and blue, respectively. (B) Structure-based sequence alignment of EcMlaC orthologues. The subdomains are coloured as per the previous colour scheme.*

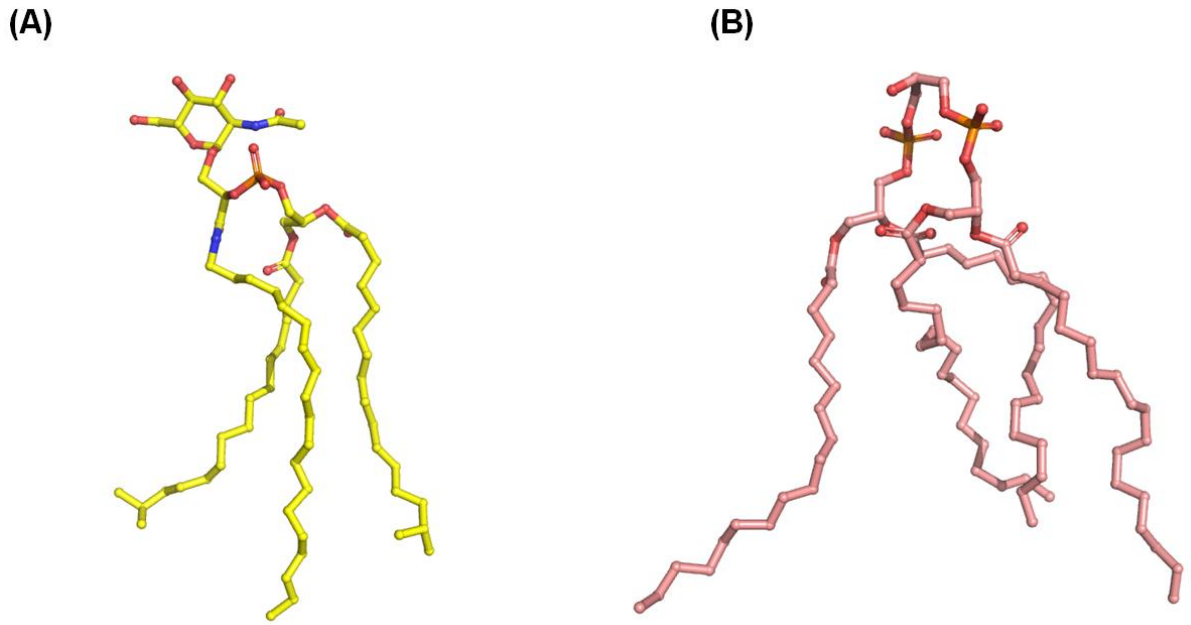


Figure D3. Overview of the PLs that did not bind to EcMlaC and PaMlaC. *(A-B) The overall structures of 5PL and CDL, respectively.*

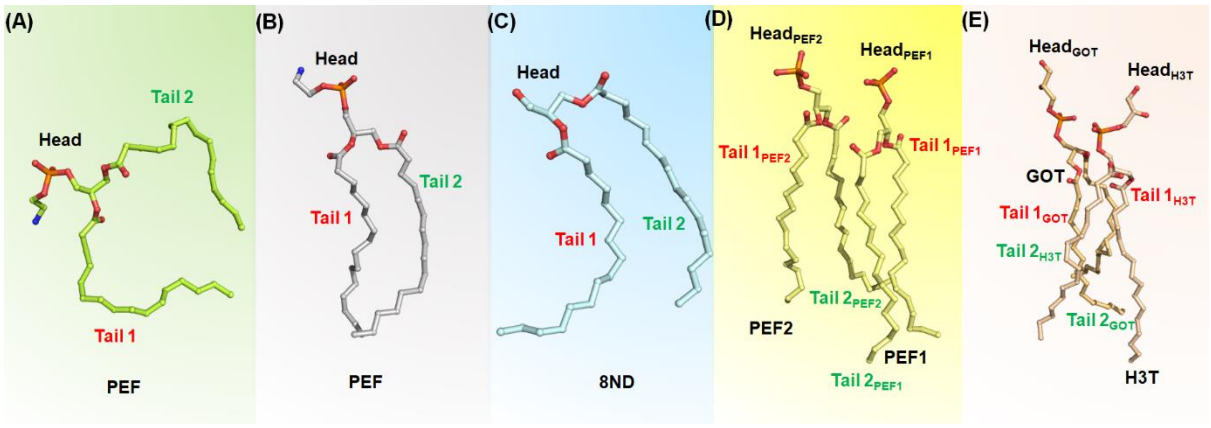


Figure D4. Conformational variation of the bound PLs in different crystal structures of EcMlaC orthologues. *(A-E) Conformation of PEF (limon, bound to EcMlaC(Holo1)), PEF (grey, bound to RsMlaC(Holo)), 8ND (palecyan, bound to EcMlaC(Holo2)), PEF$_{1/2}$ (lightyellow, bound to PpMlaC(Holo)) and GOT/H3T (lightorange and wheat, respectively, bound to PaMlaC(Holo), respectively.*

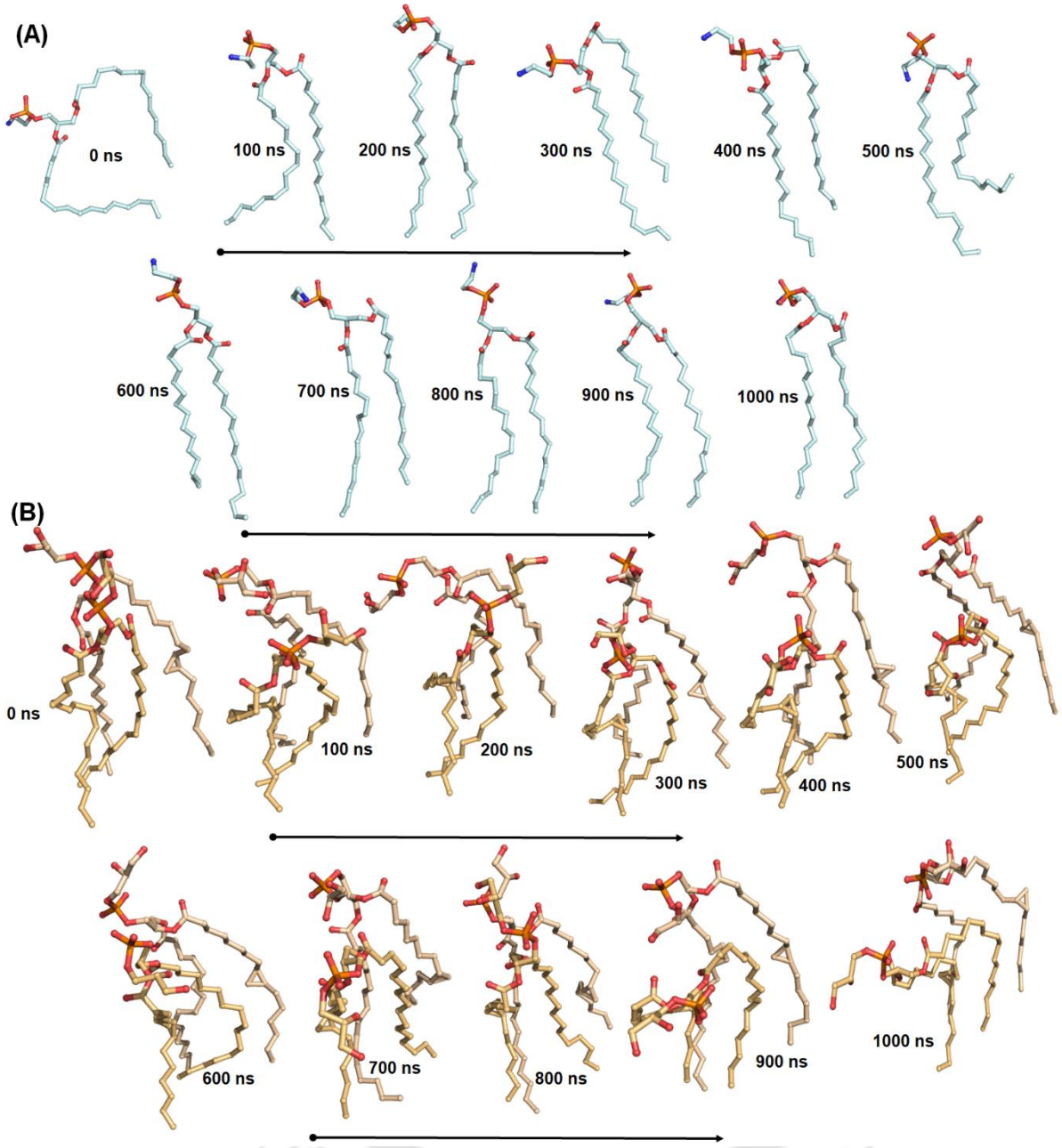


Figure D5. Change in the conformation of PLs during simulation. *(A-B) Conformations of PEF (palecyan, bound to EcMlaC(Holo1)) and GOT/H3T (lightorange and wheat, respectively, bound to PaMlaC(Holo)), respectively after every 100 ns till the completion of 1000 ns.*

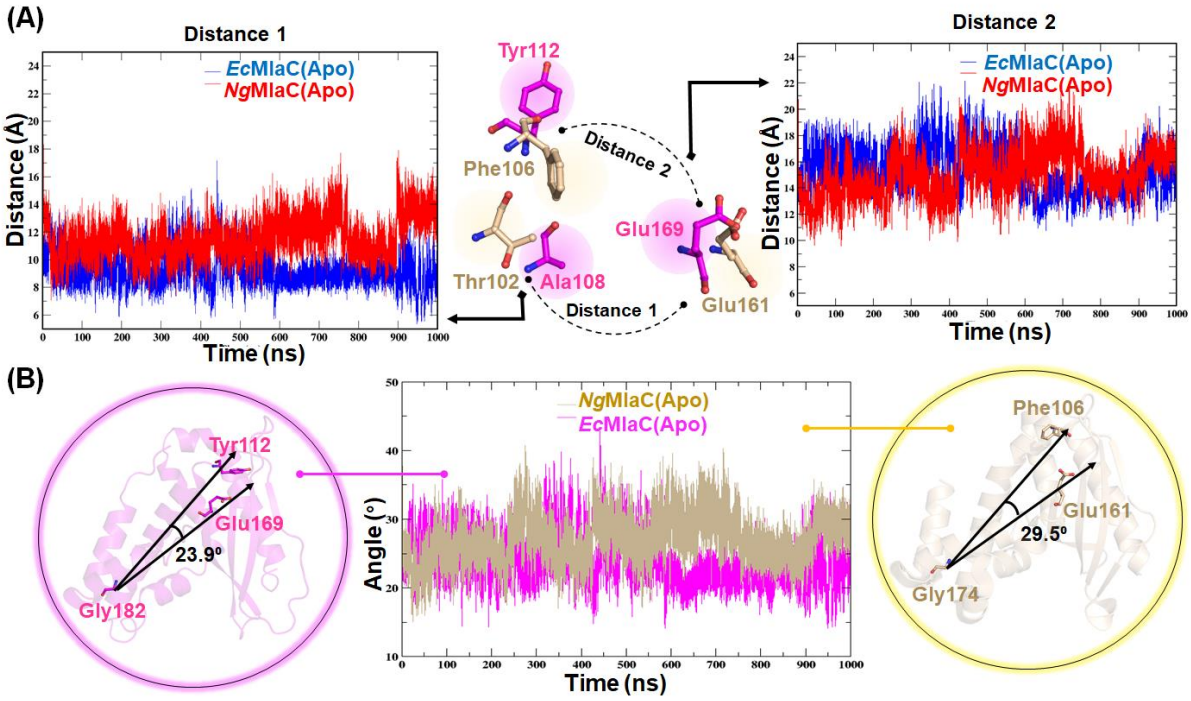


Figure D6. Analysis of the change in distance between residues around the binding pocket and angular motions. *(A) (Centre) Overview of Distance 1 and 2. (Left, right) Fluctuations of Distance 1 and 2 in EcMlaC(Apo*) and NgMlaC(Apo), respectively. (B) (Centre) Fluctuations of angular motions in EcMlaC(Apo*) and NgMlaC(Apo). (Left, right) Angles formed by the residues around the binding pocket in EcMlaC(Apo*) and NgMlaC(Apo), respectively.*

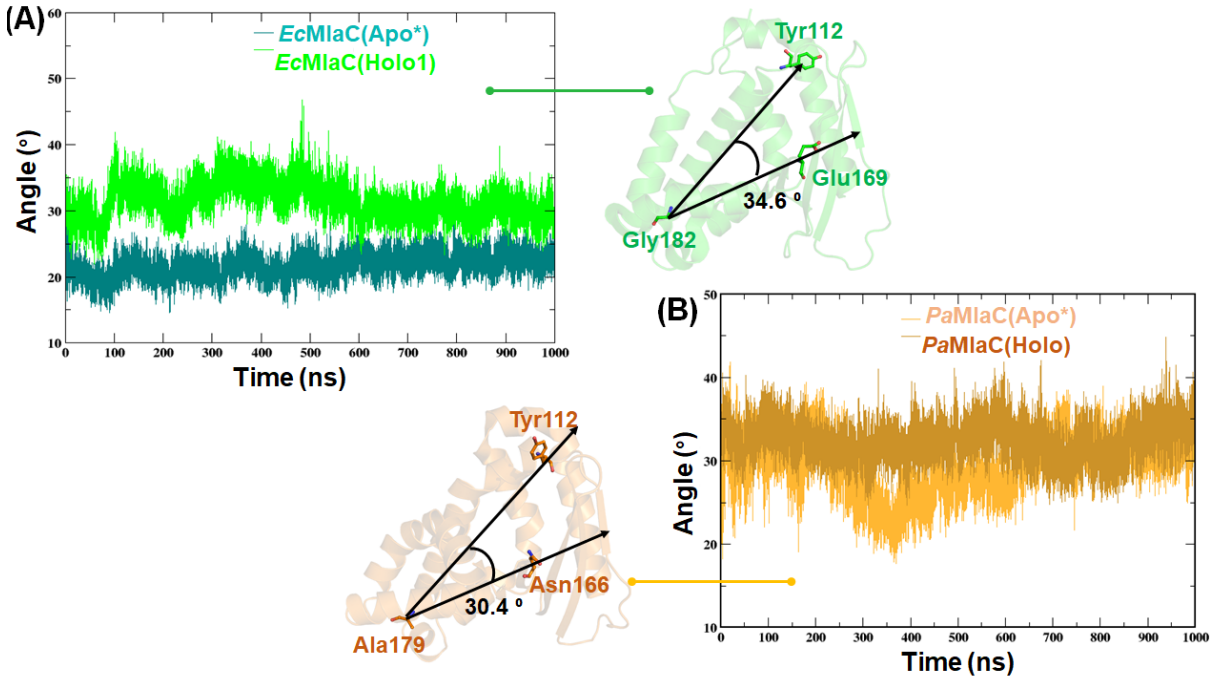


Figure D7. Analysis of the fluctuations of angular motions. *(A-B) (Left) Fluctuations of angular motion in EcMlaC(Apo*/Holo1) and PaMlaC(Apo*/Holo1), respectively. (Right) Angles formed by the residues around the binding pocket in EcMlaC(Holo1) and PaMlaC(Holo), respectively.*

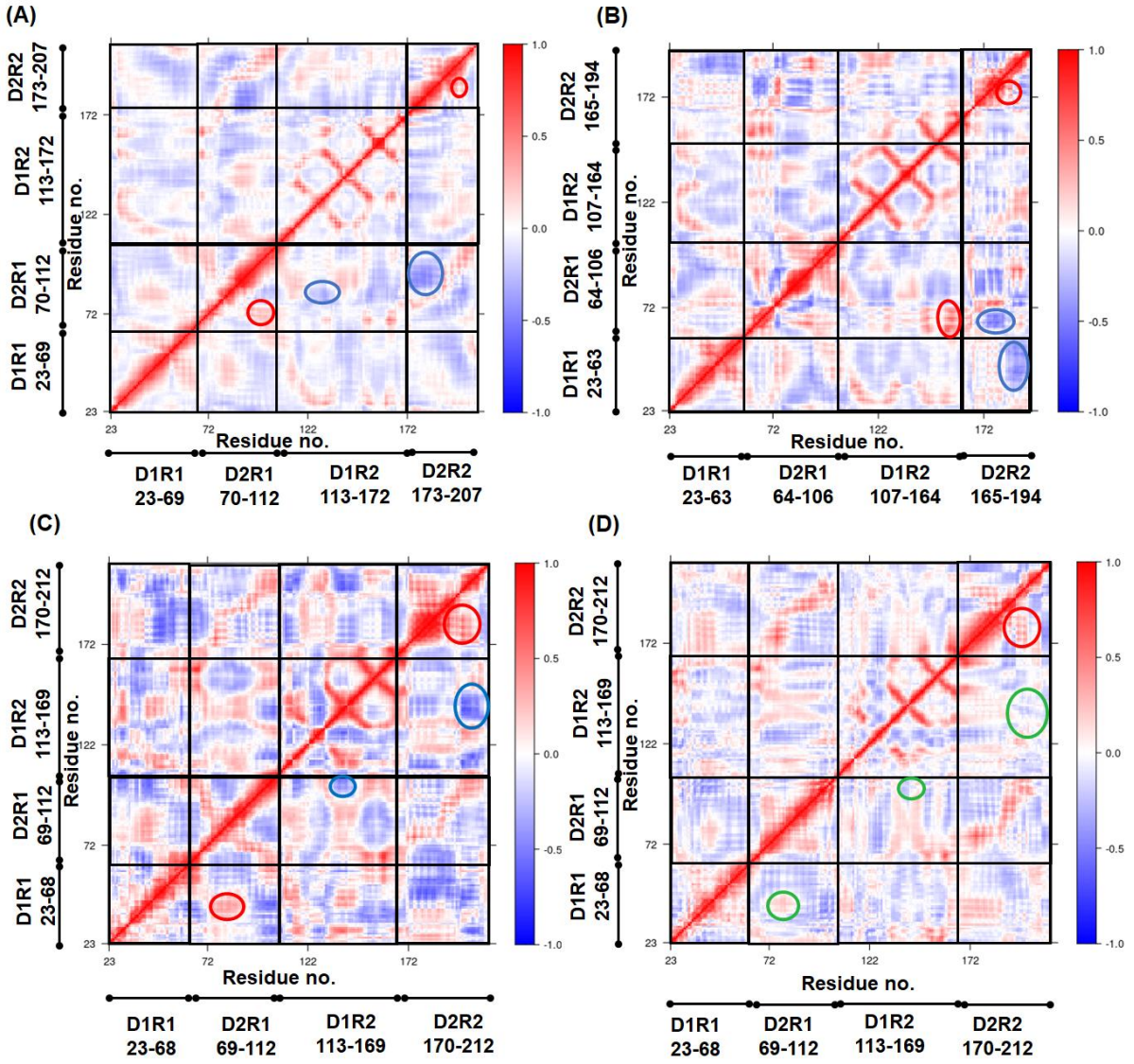


Figure D8. Correlation maps of EcMlaC systems. *(A-D) Correlation map of EcMlaC(Apo), NgMlaC(Apo), PaMlaC(Apo) and PaMlaC(Holo), respectively. The change in the colour bar from blue to red signifies the transition from anti-correlated to correlated motions. The regions encircled with blue and red circles represent anti-correlated and correlated motions of interest. The regions encircled in green represent the absence of any motion in the corresponding regions. The maps are divided into quadrants that correspond to the four subdomains (D1R1, D2R1, D1R2, and D2R2) as per the position of the residues. The boundaries of the subdomains are provided along the axes.*

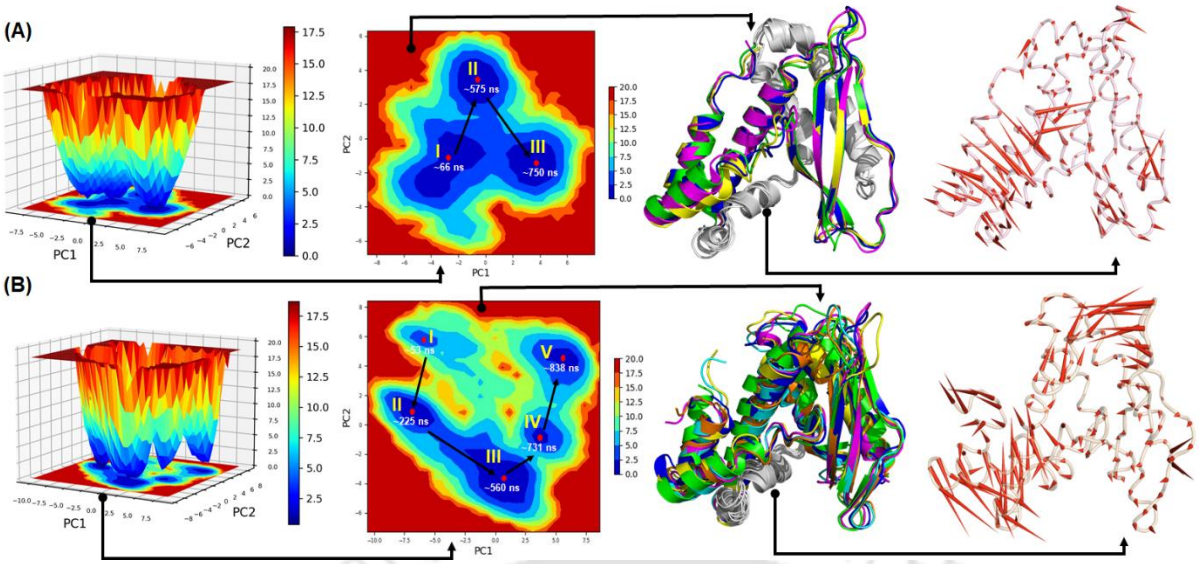


Figure D9. Free energy landscape (FEL) analysis and porcupine plots of EcMlaC systems. *(A-B) FEL plots and structural analysis of EcMlaC(Apo) and NgMlaC(Apo), respectively. (First) 3D FEL between PC1 (X-axis) and PC2 (Y-axis). (Second) 2D FEL between PC1 (X-axis) and PC2 (Y-axis). The low-energy minima states of the systems and the paths connecting these minima are marked in red dots (the time frames are mentioned below) in the 2D FEL plots. The colour bar in the first and second panels represents the Gibbs free energy (kcal mol^{-1}) ranging from the lowest energy (blue) to the highest (red) conformational states. (Third) Structural superimposition of minima extracted from FEL plots. For EcMlaC(Apo), conformations from three minima (I-III; yellow, blue, and magenta, respectively; green, reference structure) and NgMlaC(Apo) conformations from five minima (I-V; yellow, blue, magenta, orange, and cyan respectively; green, reference) were superimposed. The regions showing movements are coloured, while the rest of the regions are grey. (Fourth) Porcupine plots of the systems. The red arrows represent the direction of motion.*

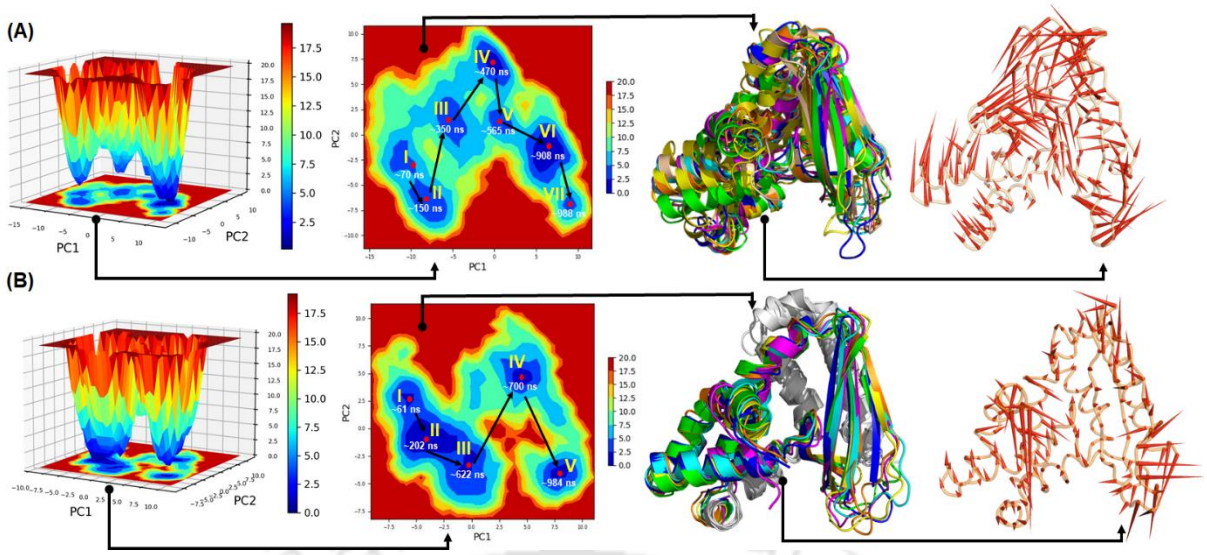


Figure D10. Free energy landscape (FEL) analysis and porcupine plots of EcMlaC systems. *(A-B) FEL plots and structural analysis of PaMlaC(Apo*) and PaMlaC(Holo), respectively. (First) 3D FEL between PC1 (X-axis) and PC2 (Y-axis). (Second) 2D FEL between PC1 (X-axis) and PC2 (Y-axis). The low-energy minima states of the systems and the paths connecting these minima are marked in red dots (the time frames are mentioned below) in the 2D FEL plots. The colour bar in the first and second panels represents the Gibbs free energy (kcal mol^{-1}) ranging from the lowest energy (blue) to the highest (red) conformational states. (Third) Structural superimposition of minima extracted from FEL plots. For PaMlaC(Apo*), conformations from seven minima (I-VII; yellow, blue, magenta, orange, cyan, wheat, and olive, respectively; green, reference structure) and PaMlaC(Holo) conformations from five minima (I-V; yellow, blue, magenta, orange, and cyan respectively; green, reference) were superimposed. The regions showing movements are coloured, while the rest of the regions are grey. (Fourth) Porcupine plots of the systems. The red arrows represent the direction of motion.*

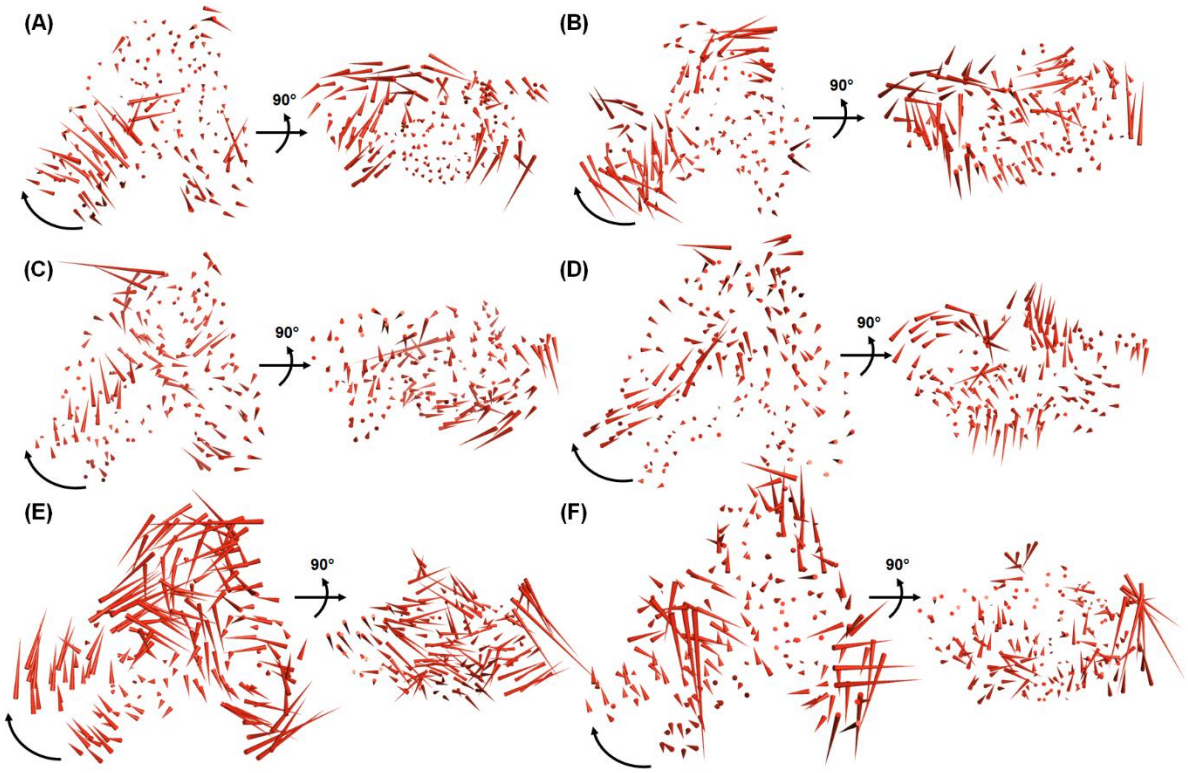


Figure D11. Protein motions of EcMlaC orthologues. *(A-F) Porcupine plots of EcMlaC(Apo), NgMlaC(Apo), EcMlaC(Apo*), EcMlaC(Holo1), PaMlaC(Apo*) and PaMlaC(Holo), respectively. The red arrows represent the direction of movement. The curved black arrow represents the twisted motion of α7 and α8.*

Bibliography

1. Abellón-Ruiz J, Kaptan SS, Baslé A, Claudi B, Bumann D, Kleinekathöfer U and van den Berg B (2017). Structural basis for maintenance of bacterial outer membrane lipid asymmetry. *Nat. Microbiol.* 2(12), 1616-1623.
2. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B and Lindahl E (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1, 19-25.
3. Adhikari R, Singh D, Chandravanshi M, Dutta A and Kanaujia SP (2017). UgpB, a periplasmic component of the UgpABCE ATP-binding cassette transporter, predominantly follows the Sec translocation pathway. *Meta Gene* 13, 129-139.
4. Al-Saad KA, Zabrouskov V, Siems WF, Knowles NR, Hannan RM and Hill Jr HH (2003). Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry of lipids: ionization and prompt fragmentation patterns. *Rapid Commun. Mass. Spectrom.* 17(1), 87-96.
5. Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ (1990). Basic local alignment search tool. *J. Mol. Biol.* 215(3), 403-410.
6. Andreeva A, Kulesha E, Gough J and Murzin AG (2020). The SCOP database in 2020: expanded classification of representative family and superfamily domains of known protein structures. *Nucleic Acids Res.* 48(D1), D376-D382.
7. Arruda S, Bomfim G, Knights R, Huima-Byron T and Riley LW (1993). Cloning of an *M. tuberculosis* DNA fragment associated with entry and survival inside cells. *Science* 261(5127), 1454-1457.
8. Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T and Ben-Tal N (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* 44(W1), W344-W350.
9. Asthana P, Singh D, Pedersen JS, Hynönen MJ, Sulu R, Murthy AV, Laitaoja M, Jänis J, Riley LW and Venkatesan R (2021). Structural insights into the substrate-binding proteins Mce1A and Mce4A from *Mycobacterium tuberculosis*. *IUCrJ* 8, 757-774.
10. Awai K, Xu C, Tamot B and Benning C (2006). A phosphatidic acid-binding protein of the chloroplast inner envelope membrane involved in lipid trafficking. *Proc. Natl. Acad. Sci. U. S. A*. 103(2006), 10817-10822.
11. Balamurugan B, Roshan MN, Shaahul Hameed B, Sumathi K, Senthilkumar R, Udayakumar A, Venkatesh Babu KH, Kalaivani M, Sowmiya G, Sivasankari P and Saravanan S (2007). PSAP: protein structure analysis package. *J. Appl. Crystallogr.* 40, 773-777.
12. Banerjee A, Moreno A, Pata J, Falson P and Prasad R (2021). ABCG: a new fold of ABC exporters and a whole new bag of riddles! *Adv. Protein Chem. Struct. Biol.* 123, 163-191.
13. Basanta B, Bick MJ, Bera AK, Norn C, Chow CM, Carter LP, Goreshnik I, Dimaio F and Baker D (2020). An enumerative algorithm for de novo design of proteins with diverse pocket structures. *Proc. Natl. Acad. Sci. U. S. A.* 117(36), 22135-22145.
14. Battye TG, Kontogiannis L, Johnson O, Powell HR and Leslie AG (2011). iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta. Cryst.* D67, 271-281.
15. Berman HM, Bhat TN, Bourne PE, Feng Z, Gilliland G, Weissig H and Westbrook J (2000). The Protein Data Bank and the challenge of structural genomics. *Nat. Struct. Mol. Biol.* 7, 957-959
16. Berntsson RP, Smits SH, Schmitt L, Slotboom DJ and Poolman B (2010). A structural classification of substrate-binding proteins. *FEBS Lett.* 584(12), 2606-2617.
17. Bertani B and Ruiz N (2018). Function and biogenesis of lipopolysaccharides. *EcoSal Plus*, 8(1).

18. Bilati U, Pasquarello C, Corthals GL, Hochstrasser DF, Allémann E and Doelker E (2005). Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry for quantitation and molecular stability assessment of insulin entrapped within PLGA nanoparticles. *J. Pharm. Sci.* 94(3), 688-694.
19. Bishop RE (2008). Structural biology of membrane-intrinsic β-barrel enzymes: Sentinels of the bacterial outer membrane. *Biochim. Biophys. Acta – Biomembr.* 1778(9), 1881-1896.
20. Bligh EG and Dyer WJ (1959). A rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiol.* 37(8), 911-917.
21. Blum M, Chang HY, Chuguransky S, Grego T, Kandasaamy S, Mitchell A, Nuka G, Paysan-Lafosse T, Qureshi M, Raj S, Richardson L, Salazar GA, Williams L, Bork P, Bridge A, Gough J, Haft DH, Letunic I, Marchler-Bauer A, Mi H, Natale DA, Necci M, Orengo CA, Pandurangan AP, Rivoire C, Sigrist CJA, Sillitoe I, Thanki N, Thomas PD, Tosatto SCE, Wu CH, Bateman A and Finn RD (2021). The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.* 49, D344–D354.
22. Bond CS (2003). TopDraw: a sketchpad for protein structure topology cartoons. *Bioinformatics* 19(2), 311-312.
23. Brünger AT (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* 355, 472-475.
24. Bussi G, Donadio D and Parrinello M (2007). Canonical sampling through velocity rescaling. *J Chem. Phys.* 126(1), 014101.
25. Campbell EA, Masuda S, Sun JL, Muzzin O, Olson CA, Wang S and Darst SA (2002). Crystal structure of the *Bacillus stearothermophilus* anti-σ factor SpoIIAB with the sporulation σ factor σF. *Cell* 108(6), 795-807.
26. Casali N and Riley LW (2007). A phylogenomic analysis of the Actinomycetales *mce* operons. *BMC Genom.* 8(1), 1-23.
27. Cevc G (1993). Phospholipids Handbook. Marcel Dekker Inc., New York, USA.
28. Chandravanshi M, Gogoi P and Kanaujia SP (2020a). Structural and thermodynamic correlation illuminates the selective transport mechanism of disaccharide α-glycosides through ABC transporter. *FEBS J.* 287(8), 1576-1597.
29. Chandravanshi M, Samanta R and Kanaujia SP (2020b). Conformational trapping of a β-glucosides-binding protein unveils the selective two-step ligand-binding mechanism of ABC importers. *J. Mol. Biol.* 432(20), 5711-5734.
30. Chandravanshi M, Tripathi SK and Kanaujia SP (2021). An updated classification and mechanistic insights into ligand binding of the substrate-binding proteins. *FEBS Letts.* 595(18), 2395-2409.
31. Chen J, Fruhauf A, Fan C, Ponce J, Ueberheide B, Bhabha G and Ekiert DC (2022). Structure of an endogenous mycobacterial MCE lipid transporter. *bioRxiv* https://doi.org/10.1101/2022.12.08.519548.
32. Chi X, Fan Q, Zhang Y, Liang K, Wan L, Zhou Q and Li Y (2020). Structural mechanism of phospholipids translocation by MlaFEDB complex. *Cell Res.* 30(12), 1127-1135.
33. Chng SS, Gronenberg LS and Kahne D (2010). Proteins required for lipopolysaccharide assembly in *Escherichia coli* form a trans-envelope complex. *Biochemistry* 49(22), 4565-4567.
34. Chong ZS, Woo WF and Chng SS (2015). Osmoporin OmpC forms a complex with MlaA to maintain outer membrane lipid asymmetry in *Escherichia coli*. *Mol. Microbiol.* 98(6), 1133-1146.
35. Chovancova E, Pavelka A, Benes P, Strnad O, Brezovsky J, Kozlikova B, Gora, A, Sustr V, Klvana M, Medek P and Biedermannova L (2012). CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. *PLoS Comput. Biol.* 8(10), e1002708.

36. Clifton LA, Skoda MW, Daulton EL, Hughes AV, Le Brun AP, Lakey JH and Holt SA (2013). Asymmetric phospholipid: lipopolysaccharide bilayers; a Gram-negative bacterial outer membrane mimic. *J. R. Soc. Interface* 10(89), 20130810.
37. Cole S, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry III CE, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail MA, Rajandream MA, Rogers J, Rutter S, Seeger K, Skelton J, Squares R, Squares S, Sulston JE, Taylor K, Whitehead S and Barrell BG (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393, 537-544.
38. Coudray N, Isom GL, MacRae MR, Saiduddin MN, Bhabha G and Ekiert DC (2020). Structure of bacterial phospholipid transporter MlaFEDB with substrate bound. *eLife* 9, e62518.
39. Crooks GE, Hon G, Chandonia JM and Brenner SE (2004). WebLogo: a sequence logo generator. *Genome Res.* 14(6), 1188-1190.
40. Dahl AC, Chavent M and Sansom MS (2012). Bendix: intuitive helix geometry analysis and abstraction. *Bioinformatics* 28(16), 2193-2194.
41. Dalebroux ZD, Edrozo MB, Pfuetzner RA, Ressl S, Kulasekara BR, Blanc MP and Miller SI (2015). Delivery of cardiolipins to the *Salmonella* outer membrane is necessary for survival within host tissues and virulence. *Cell Host Microbe*. 17(4), 441-451.
42. Dassa E (2011). Natural history of ABC systems: not only transporters. *Essays. Biochem.* 50, 19-42.
43. Dassa E and Bouige P (2001). The ABC of ABCs: a phylogenetic and functional classification of ABC systems in living organisms. *Res. Microbiol.* 152(3-4), 211-229.
44. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall III WB, Snoeyink J, Richardson JS and Richardson DC (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* 35, W375-W383.
45. Dawson NL, Lewis TE, Das S, Lees JG, Lee D, Ashford P, Orengo CA & Sillitoe I (2017). CATH: an expanded resource to predict protein function through structure and sequence. *Nucleic Acids Res.* 45, D289-D295.
46. Delcour AH (2009). Outer membrane permeability and antibiotic resistance. *Biochim. Biophys. Acta - Proteins Proteom.* 1794(5), 808-816.
47. Dever TE, Glynias MJ and Merrick WC (1987). GTP-binding domain: three consensus sequence elements with distinct spacing. *Proc. Natl. Acad. Sci. U. S. A.* 84(7), 1814-1818.
48. Dong H, Tang X, Zhang Z and Dong C (2017). Structural insight into lipopolysaccharide transport from the gram-negative bacterial inner membrane to the outer membrane. *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids.* 1862(11):1461-1467.
49. Dong H, Zhang Z, Tang X, Huang S, Li H, Peng B and Dong C (2016). Structural insights into cardiolipin transfer from the inner membrane to the outer membrane by PbgA in Gram-negative bacteria. *Sci. Rep.* 6(1), 30815.
50. Duong F, Eichler J, Price A, Leonard MR and Wickner W (1997). Biogenesis of the gram-negative bacterial envelope. *Cell* 91(5), 567-573.
51. Dutta A and Kanaujia SP (2022). MlaC belongs to a unique class of non-canonical substrate-binding proteins and follows a novel phospholipid-binding mechanism. *J. Struct. Biol.* 214, 107896.
52. Dutta A, Chandravanshi M and Kanaujia SP (2021) Conserved features of the MlaD domain aid the trafficking of hydrophobic molecules. *Proteins* 89, 1473-1488.

53. Ekiert DC, Bhabha G, Isom GL, Greenan G, Ovchinnikov S, Henderson IR, Cox JS and Vale RD (2017). Architectures of lipid transport systems for the bacterial outer membrane. *Cell* 169(2), 273-285.
54. Ekiert DC, Coudray N and Bhabha G (2022). Structure and mechanism of the bacterial lipid ABC transporter, MlaFEDB. *Curr. Opin. Struct. Biol.* 76, 102429.
55. El Rayes J, Rodríguez-Alonso R and Collet JF (2021). Lipoproteins in Gram-negative bacteria: New insights into their biogenesis, subcellular targeting and functional roles. *Curr. Opin. Microbiol.* 61, 25-34.
56. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE and Finn RD (2019). The Pfam protein families database in 2019. *Nucleic Acids Res.* 47(D1) D427-D432.
57. Emsley P, Lohkamp B, Scott WG and Cowtan K (2010). Features and development of Coot. *Acta. Cryst. D*66, 486-501.
58. Engel, K.M., Prabutzki, P., Leopold, J., Nimptsch, A., Lemmnitzer, K., Vos, D.N., Hopf, C. and Schiller, J (2022). A new update of MALDI-TOF mass spectrometry in lipid research. *Prog. Lipid. Res.* 86, 101145.
59. Ercan B, Low WY, Liu X and Chng SS (2018). Characterization of interactions and phospholipid transfer between substrate binding proteins of the OmpC-Mla system. *Biochemistry* 58(2), 114-119.
60. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H and Pedersen LG (1995). A smooth particle mesh Ewald method. *J. Chem. Phys.* 103(19), 8577-8593.
61. Evans P and McCoy A (2008). An introduction to molecular replacement. *Acta. Cryst. D*64(Pt 1), 1-10.
62. Evans PR and Murshudov GN (2013). How good are my data and what is the resolution? *Acta. Cryst.* D69, 1204-1214.
63. Eyal E, Lum G and Bahar I (2015). The anisotropic network model web server at 2015 (ANM 2.0). *Bioinformatics* 31(9), 1487-1489.
64. Fan J, Zhai Z, Yan C and Xu C (2015). *Arabidopsis* TRIGALACTOSYLDIACYLGLYCEROL5 interacts with TGD1, TGD2, and TGD4 to facilitate lipid transfer from the endoplasmic reticulum to plastids. *Plant Cell* 27(10), 2941-2955.
65. Fenn K, Wong CT and Darbari VC (2020). *Mycobacterium tuberculosis* uses *mce* proteins to interfere with host cell signaling. *Front. Mol. Biosci.* 6, 149.
66. Forslund SK, Kaduk M and Sonnhammer ELL (2019). Evolution of protein domain architectures. In: Anisimova M, eds. *Evolutionary Genomics, Methods in Molecular Biology*. New York: Humana Press, pp. 469-504.
67. Fuchs B, Süß R and Schiller J (2010). An update of MALDI-TOF mass spectrometry in lipid research. *Prog. Lipid. Res.* 49(4), 450-475.
68. Fujiwara K, Ebisawa S, Watanabe Y, Fujiwara H and Ikeguchi M (2015). The origin of β-strand bending in globular proteins. *BMC Struct. Biol.* 15(1), 1-2.
69. Fukami-Kobayashi K, Tateno Y and Nishikawa K (1999). Domain dislocation: a change of core structure in periplasmic binding proteins in their evolutionary history. *J. Mol. Biol.* 286(1), 279-290.
70. Garde S, Chodisetti PK and Reddy M (2021). Peptidoglycan: Structure, synthesis, and regulation. *EcoSal Plus* 9(2).
71. Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD and Bairoch A (2005). Protein identification and analysis tools on the ExPASyserver. In: Walker JM, ed. *The Proteomics Protocols Handbook* Humana Press, Totowa, NJ, 571-607.
72. Gasteiger J and Marsili M (1980). Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. *Tetrahedron* 36(22), 3219-3228.

73. Giacometti SI, MacRae MR, Dancel-Manning K, Bhabha G and Ekiert DC (2022). Lipid transport across bacterial membranes. *Annu. Rev. Cell Dev. Biol.* 38, 125-153.
74. Gioffré A, Infante E, Aguilar D, De la Paz Santangelo M, Klepp L, Amadio A, Meikle V, Etchechoury I, Romano IM, Cataldi A and Hernández RP (2005). Mutation in *mce* operons attenuates *Mycobacterium tuberculosis* virulence. *Microb. Infect.* 7(3), 325-334.
75. Gouet P, Robert X and Courcelle E (2003). ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res.* 31, 3320-3323.
76. Grabowicz M (2019). Lipoproteins and their trafficking to the outer membrane. *EcoSal Plus*, 8(2).
77. Grant BJ, Rodrigues AP, ElSawy KM, McCammon JA and Caves LS (2006). Bio3d: an R package for the comparative analysis of protein structures. *Bioinfor.* 22(21), 2695-6.
78. Grant BJ, Skjærven L and Yao XQ (2021). The Bio3D packages for structural bioinformatics. *Protein Sci.* 30(1), 20-30.
79. Hagemans D, van Belzen IA, Morán Luengo T and Rüdiger SG (2015) A script to highlight hydrophobicity and charge on protein surfaces. *Front. Mol. Biosci.* 2, 56.
80. Hassell AM, An G, Bledsoe RK, Bynum JM, Carter HL, Deng SJ, Gampe RT, Grisard TE, Madauss KP, Nolte RT and Rocque WJ (2007). Crystallization of protein-ligand complexes. *Acta. Cryst.* D63, 72-79.
81. Henderson JC, Zimmerman SM, Crofts AA, Boll JM, Kuhns LG, Herrera CM and Trent MS (2016). The power of asymmetry: architecture and assembly of the Gram-negative outer membrane lipid bilayer. *Annu. Rev. Microbiol.* 70, 255-278.
82. Hess B (2008). P-LINCS: A parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comp.* 4(1), 116-22.
83. Hilton KL, Manwani C, Boles JE, White LJ, Ozturk S, Garrett MD and Hiscock JR (2021). The phospholipid membrane compositions of bacterial cells, cancer cell lines and biological samples from cancer patients. *Chem. Sci.* 12(40), 13273-13282.
84. Hoh F, Pons JL, Gautier MF, De Lamotte F and Dumas C (2005). Structure of a liganded type 2 non-specific lipid-transfer protein from wheat and the molecular basis of lipid binding. *Acta. Cryst.* D61(4), 397-406.
85. Hollingsworth SA and Dror RO (2018). Molecular dynamics simulation for all. *Neuron* 99(6), 1129-1143.
86. Holm L (2020). Using Dali for Protein Structure Comparison. *Methods Mol. Biol.* 2112, 29-42.
87. Holm L (2022) Dali server: structural unification of protein families. *Nucleic Acids Res.* 50, W210-W215.
88. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A and Simmerling C (2006). Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* 65(3), 712-725.
89. Huang YM, Miao Y, Munguia J, Lin L, Nizet V and McCammon JA (2016). Molecular dynamic study of MlaC protein in Gram-negative bacteria: conformational flexibility, solvent effect and protein-phospholipid binding. *Protein Sci.* 25(8), 1430-1437.
90. Huang YM, Munguia J, Miao Y, Nizet V and McCammon JA (2019). Docking simulation and antibiotic discovery targeting the MlaC protein in Gram-negative bacteria. *Chem. Biol. Drug. Des.* 93(4), 647-652.
91. Hughes GW, Hall SC, Laxton CS, Sridhar P, Mahadi AH, Hatton C, Piggot TJ, Wotherspoon PJ, Leney AC, Ward DG, Jamshad M, Spana V, Cadby IT, Harding C, Isom GL, Bryant JA, Parr RJ, Yakub Y, Jeeves M, Huber D, Henderson IR, Clifton LA, Lovering AL and Knowles TJ (2019). Evidence for phospholipid export from the

bacterial inner membrane by the Mla ABC transport system. *Nat. Microbiol.* 4(10), 1692-1705.
92. Humphrey W, Dalke A and Schulten K (1996). VMD: visual molecular dynamics. *J. Mol. Graph.* 14(1), 33-38.
93. Hussain AS, Shanthi V, Sheik SS, Jeyakanthan J, Selvarani P and Sekar K (2002). PDB Goodies–a web-based GUI to manipulate the Protein Data Bank file. *Acta. Cryst.* D58, 1385-1386.
94. Isom GL, Coudray N, MacRae MR, McManus CT, Ekiert DC and Bhabha G (2020). LetB structure reveals a tunnel for lipid transport across the bacterial envelope. *Cell* 181(3), 653-664.
95. Isom GL, Davies NJ, Chong ZS, Bryant JA, Jamshad M, Sharif M, Cunningham AF, Knowles TJ, Chng SS, Cole JA and Henderson IR (2017). MCE domain proteins: conserved inner membrane lipid-binding proteins required for outer membrane homeostasis. *Sci. Rep.* 7(1), 8608.
96. Ito Y, Kanamaru K, Taniguchi N, Miyamoto S and Tokuda H (2006). A novel ligand bound ABC transporter, LolCDE, provides insights into the molecular mechanisms underlying membrane detachment of bacterial lipoproteins. *Mol. Microbiol.* 62(4), 1064-1075.
97. Joosten RP, Te Beek TA, Krieger E, Hekkelman ML, Hooft RW, Schneider R, Sander C and Vriend G (2010). A series of PDB related databases for everyday needs. *Nucleic Acids Res.* 39, D411-D419.
98. Joseph AP, Valadié H, Srinivasan N and de Brevern AG (2012). Local structural differences in homologous proteins: specificities in different SCOP classes. *PloS One* 7, e38805.
99. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A and Bridgland A (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583-589.
100. Kabsch W and Sander C (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22(12), 2577-2637.
101. Kamio Y and Nikaido H (1976). Outer membrane of *Salmonella typhimurium*: accessibility of phospholipid head groups to phospholipase c and cyanogen bromide activated dextran in the external medium. *Biochemistry*15(12), 2561-2570.
102. Kamischke C, Fan J, Bergeron J, Kulasekara HD, Dalebroux ZD, Burrell A, Kollman JM and Miller SI (2019). The *Acinetobacter baumannii* Mla system and glycerophospholipid transport to the outer membrane. *eLlife* 8, e40171.
103. Kanehisa M and Goto S (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27-30.
104. Kaplan E, Greene NP, Crow A and Koronakis V (2018). Insights into bacterial lipoprotein trafficking from a structure of LolA bound to the LolC periplasmic domain. *Proc. Natl. Acad. Sci. U. S. A.* 115(31), E7389-E7397.
105. Kapopoulou A, Jones LM and Cole ST (2011). The MycoBrowser portal: a comprehensive and manually annotated resource for mycobacterial genomes. *Tuberculosis* 91(1), 8-13.
106. Keseler IM, Mackie A, Santos-Zavaleta A, Billington R, Bonavides-Martínez C, Caspi R, Fulcher C, Gama-Castro S, Kothari A, Krummenacker M and Latendresse M (2017). The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12. *Nucleic Acids Res.* 45(D1), D543-50.
107. Khan S, Khan FI, Mohammad T, Khan P, Hasan GM, Lobb KA, Islam A, Ahmad F and Hassan MI (2018). Exploring molecular insights into the interaction mechanism of cholesterol derivatives with the Mce4A: A combined spectroscopic and molecular dynamic simulation studies. *Int. J. Biol. Macromol.* 111, 548-560.

108. Khazanov NA and Carlson HA (2013). Exploring the composition of protein-ligand binding sites on a large scale. *PLoS Comput. Biol.* 9(11), e1003321.
109. Khoo K and Norton RS (2011). Role of disulfide bonds in peptide and protein conformation. In: Hughes AB, eds. Amino Acids, Peptides and Proteins in Organic Chemistry: Analysis and Function of Amino Acids and Peptides. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA, pp. 395-417.
110. Kim KH, Aulakh S and Paetzel M (2012). The bacterial outer membrane β-barrel assembly machinery. *Protein Sci.* 21(6), 751-768.
111. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, Li Q, Shoemaker BA, Thiessen PA, Yu B, Zaslavsky L, Zhang J and Bolton EE (2023). PubChem 2023 update. *Nucleic Acids Res.* 51(D1), D1373-D1380.
112. Kleanthous C and Armitage JP (2015). The bacterial cell envelope. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 370(1679), 20150019.
113. Klepp LI, y Garcia JS and FabianaBigi (2022). Mycobacterial MCE proteins as transporters that control lipid homeostasis of the cell wall. *Tuberculosis* 132, 102162.
114. Koča J, Kříž Z and Carlsen PH (1994). Computer study of conformational flexibility of 20 common amino acids. *J. Mol. Struct. THEOCHEM* 306(2-3), 157-164.
115. Koebnik R, Locher KP and Van Gelder P (2000). Structure and function of bacterial outer membrane proteins: barrels in a nutshell. *Mol. Microbiol.* 37(2), 239-253.
116. Kolich LR, Chang YT, Coudray N, Giacometti SI, MacRae MR, Isom GL, Teran EM, Bhabha G and Ekiert DC (2020). Structure of MlaFB uncovers novel mechanisms of ABC transporter regulation. *eLife* 9, e60030.
117. Konovalova A, Kahne DE and Silhavy TJ (2017). Outer membrane biogenesis. *Annu. Rev. Microbiol.* 71, 539-556.
118. Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, Yueh C, Beglov D and Vajda S (2017). The ClusPro web server for protein–protein docking. *Nat. Protoc.* 12(2), 255-278.
119. Krissinel E and Henrick K (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta. Cryst.* D60(12), 2256-2268.
120. Krissinel E and Henrick K (2007). Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* 372(3), 774-797.
121. Krissinel E and Henrick K (2007). Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* 372, 774-797.
122. Krogh A, Larsson B, Von Heijne G and Sonnhammer EL (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305, 567-580.
123. Kumar S, Stecher G and Tamura K (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870-1874.
124. Lal R, Pandey G, Sharma P, Kumari K, Malhotra S, Pandey R, Raina V, Kohler HP, Holliger C, Jackson C and Oakeshott JG (2010). Biochemistry of microbial degradation of hexachlorocyclohexane and prospects for bioremediation. *Microbiol. Mol. Biol. Rev.* 74(1), 58-80.
125. Laskowski RA, Jabłońska J, Pravda L, Vařeková RS and Thornton JM (2018). PDBsum: Structural summaries of PDB entries. *Protein Sci.* 27(1), 129-134.
126. Laskowski RA, MacArthur MW, Moss DS and Thornton JM (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26(2), 283-291.
127. Lee YH, Deka RK, Norgard MV, Radolf JD and Hasemann CA (1999). *Treponema pallidum* TroA is a periplasmic zinc-binding protein with a helical backbone. *Nat. Struct. Biol.* 6(7), 628-633.

128. Lewinson O and Livnat-Levanon N (2017). Mechanism of action of ABC importers: conservation, divergence, and physiological adaptations. *J. Mol. Biol.* 429(5), 606-619.
129. Lewis VG, Ween MP and McDevitt CA (2012). The role of ATP-binding cassette transporters in bacterial pathogenicity. *Protoplasma* 249, 919-942.
130. Li H, Jamal J, Delker S, Plaza C, Ji H, Jing Q, Huang H, Kang S, Silverman RB and Poulos TL (2014). The mobility of a conserved tyrosine residue controls isoform-dependent enzyme-inhibitor interactions in nitric oxide synthases. *Biochemistry* 53(32), 5272-5279.
131. Linton KJ (2007). Structure and function of ABC transporters. *Physiology* 22(2), 122-130.
132. Liu C, Ma J, Wang J, Wang H and Zhang L (2020). Cryo-EM structure of a bacterial lipid transporter YebT. *J. Mol. Biol.* 432(4), 1008-1119.
133. Locher KP (2009). Structure and mechanism of ATP-binding cassette transporters. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 364(1514), 239-245.
134. Locher KP, Lee AT and Rees DC (2002). The *E. coli* BtuCD structure: a framework for ABC transporter architecture and mechanism. *Science* 296(5570), 1091-1098.
135. Lomize MA, Pogozheva ID, Joo H, Mosberg HI and Lomize AL (2012). OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res.* 40, D370-D376.
136. Low WY and Chng SS (2021). Current mechanistic understanding of intermembrane lipid trafficking important for maintenance of bacterial outer membrane lipid asymmetry. *Curr. Opin. Chem. Biol.* 65, 163-171.
137. Low WY, Thong S and Chng SS (2021). ATP disrupts lipid-binding equilibrium to drive retrograde transport critical for bacterial outer membrane asymmetry. *Proc. Natl. Acad. Sci. U.S.A.* 118(50), e2110055118.
138. Lu B and Benning C (2009). A 25-amino acid sequence of the *Arabidopsis* TGD2 protein is sufficient for specific binding of phosphatidic acid. *J. Biol. Chem.* 284(26), 17420-17427.
139. Madan Babu M and Sankaran K (2002). DOLOP-database of bacterial lipoproteins. *Bioinformatics* 18, 641-643.
140. Madni ZK, Kumar A, Kumar U, Jaiswal D and Salunke DM (2022). Dynamics of lipid displacement inside the hydrophobic cavity of a nonspecific lipid transfer protein from *Solanum melongena*. *J. Biomol. Struct. Dynamics* 4, 1-1.
141. Makise M, Mima S, Tsuchiya T and Mizushima T (2001). Molecular mechanism for functional interaction between DnaA protein and acidic phospholipids: identification of important amino acids. *J. Biol. Chem.* 276(10), 7450-7456.
142. Malinverni JC and Silhavy TJ (2009). An ABC transport system that maintains lipid asymmetry in the gram-negative outer membrane. *Proc. Natl. Acad. Sci. U.S.A.* 106(19), 8009-8014.
143. Mann D, Fan J, Somboon K, Farrell DP, Muenks A, Tzokov SB, DiMaio F, Khalid S, Miller SI and Bergeron JR (2021). Structure and lipid dynamics in the maintenance of lipid asymmetry inner membrane complex of *A. baumannii*. *Commun. Biol.* 4(1), 817.
144. Mao B, Pear MR, McCammon JA and Quiocho FA (1982). Hinge-bending in L-arabinose-binding protein. The “Venus’s-flytrap” model. *J. Biol. Chem.* 257(3), 1131-1133.
145. Maqbool A, Horler RS, Muller A, Wilkinson AJ, Wilson KS and Thomas GH (2015). The substrate-binding protein in bacterial ABC transporters: dissecting roles in the evolution of substrate specificity. *Biochem. Soc. Trans.* 43(5), 1011-1017.

146. Martorana AM, Motta S, Di Silvestre D, Falchi F, Deho G, Mauri P, Sperandeo P and Polissi A (2014). Dissecting *Escherichia coli* outer membrane biogenesis using differential proteomics. *PLoS One* 9(6), e100941.
147. May KL and Grabowicz M (2018). The bacterial outer membrane is an evolving antibiotic barrier. *Proc. Natl. Acad. Sci. U.S.A.* 115(36), 8852-8854.
148. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC and Read RJ (2007). Phaser crystallographic software. *J. Appl. Crystallogr.* 40, 658-674.
149. Mi W, Li Y, Yoon SH, Ernst RK, Walz T and Liao M (2017). Structural basis of MsbA-mediated lipopolysaccharide transport. *Nature* 549, 233-237.
150. Miller SI and Salama NR (2018). The gram-negative bacterial periplasm: Size matters. *PLoS Biol.* 16(1), e2004935.
151. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer EL, Tosatto SC, Paladin L, Raj S, Richardson LJ and Finn RD (2021). Pfam: The protein families database in 2021. *Nucleic Acids Res*. 49, D412-D419.
152. Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, Brown SD, Chang HY, El-Gebali S, Fraser MI, Gough J, Haft DR, Huang H, Letunic I, Lopez R, Luciani A, Madeira F, Marchler-Bauer A, Mi H, Natale DA, Necci M, Nuka G, Orengo C, Pandurangan AP, Paysan-Lafosse T, Pesseat S, Potter SC, Qureshi MA, Rawlings ND, Redaschi N, Richardson LJ, Rivoire C, Salazar GA, Sangrador-Vegas A, Sigrist CJA, Sillitoe I, Sutton GG, Thanki N, Thomas PD, Tosatto SCE, Yong SY and Finn RD (2019). InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.* 47(D1), D351-D360.
153. Mittal L, Srivastava M, Kumari A, Tonk RK, Awasthi A and Asthana S (2021). Interplay among structural stability, plasticity, and energetics determined by conformational attuning of flexible loops in PD-1. *J. Chem. Infor. Modeling* 61(1), 358-84.
154. Miyadai H, Tanaka-Masuda K, Matsuyama SI and Tokuda H (2004). Effects of lipoprotein overproduction on the induction of DegP (HtrA) involved in quality control in the *Escherichia coli* periplasm. *J. Biol. Chem.* 279(38), 39807-39813.
155. Morris GM, Huey R, Lindstrom W and Sanner MF, Belew RK, Goodsell DS and Olson AJ (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* 16, 2785-2791.
156. Mourez M, Hofnung M and Dassa E (1997). Subunit interactions in ABC transporters: a conserved sequence in hydrophobic membrane proteins of periplasmic permeases defines an important site of interaction with the ATPase subunits. *EMBO J.* 16(11), 3066-3077.
157. Moussatova A, Kandt C, O'Mara ML and Tieleman DP (2008). ATP-binding cassette transporters in *Escherichia coli*. *Biochim. Biophys. Acta - Biomembr.* 1778(9), 1757-1771.
158. Murahari P, Anishetty S and Pennathur G (2013). Understanding the lid movements of LolA in *Escherichia coli* using molecular dynamics simulation and *in silico* point mutation. *Comput. Biol. Chem.* 47, 71-80.
159. Nagata Y, Natsui S, Endo R, Ohtsubo Y, Ichikawa N, Ankai A, Oguchi A, Fukui S, Fujita N and Tsuda M (2011). Genomic organization and genomic structural rearrangements of *Sphingobium japonicum* UT26, an archetypal γ-hexachlorocyclohexane-degrading bacterium. *Enzyme Microb. Technol.* 49(6–7), 499-508.
160. Najafi SM, Harris DA and Yudkin MD (1996). The SpoIIAA protein of Bacillus subtilis has GTP-binding properties. *J. Bacteriol.* 178(22), 6632-6634.
161. Nakayama T and Zhang-Akiyama QM (2017). *pqiABC* and *yebST*, putative *mce* operons of *Escherichia coli*, encode transport pathways and contribute to membrane integrity. *J. Bacteriol.* 199(1), e00606-16.

162. Narita SI and Tokuda H (2017). Bacterial lipoproteins; biogenesis, sorting and quality control. *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids.* 1862(11), 1414-1423.
163. Nazarova EV, Montague CR, La T, Wilburn KM, Sukumar N, Lee W, Caldwell S, Russell DG and VanderVen BC (2017). Rv3723/LucA coordinates fatty acid and cholesterol uptake in *Mycobacterium tuberculosis*. *eLife* 6, e26969.
164. Neumann J, Rose-Sperling D and Hellmich UA (2017). Diverse relations between ABC transporters and lipids: An overview. *Biochim. Biophys. Acta - Biomembr.* 1859(4), 605-618.
165. Nikaido H (2003). Molecular basis of bacterial outer membrane permeability revisited. *Microbiol. Mol. Biol. Rev.* 67(4), 593-656.
166. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T and Hutchison GR (2011). Open Babel: An open chemical toolbox. *J. Cheminfor.* 3(1): 1-4.
167. Oguchi Y, Takeda K, Watanabe S, Yokota N, Miki K and Tokuda H (2008). Opening and closing of the hydrophobic cavity of LolA coupled to lipoprotein binding and release. *J. Biol. Chem.* 283(37), 25414-25420.
168. Okuda S and Tokuda H (2011). Lipoprotein sorting in bacteria. *Annu. Rev. Microbiol.* 65, 239-259.
169. Oldham ML, Davidson AL and Chen J (2008). Structural insights into ABC transporter mechanism. *Curr. Opin. Struct. Biol.* 18(6), 726-733.
170. Pandey AK and Sassetti CM (2008). Mycobacterial persistence requires the utilization of host cholesterol. *Proc. Natl. Acad. Sci. U. S. A.* 105(11), 4376-4380.
171. Pandey S, Modak A, Phale PS and Bhaumik P (2016). High resolution structures of periplasmic glucose-binding protein of *Pseudomonas putida* CSV86 reveal structural basis of its substrate specificity. *J. Biol. Chem.* 291(15), 7844–7857.
172. Parrinello M and Rahman A (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *J. App. Phys.* 52(12): 7182-90.
173. Pei J, Kim BH and Grishin NV (2008). PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.* 36(7), 2295-2300.
174. Perkowski EF, Miller BK, McCann JR, Sullivan JT, Malik S, Allen IC, Godfrey V, Hayden JD and Braunstein M (2016). An orphaned Mce-associated membrane protein of *Mycobacterium tuberculosis* is a virulence factor that stabilizes Mce transporters. *Mol. Microbiol.* 100(1), 90-107.
175. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC and Ferrin TE (2004). UCSF Chimera-a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25(13), 1605-1612.
176. Pie J, Kim BH and Grishin NV (2008). PROMALS3D: a tool for multiple sequence and structure alignment. *Nucleic Acids Res.* 36(7), 2295-2300.
177. Ponder JW and Case DA (2003). Force fields for protein simulations. *Adv. Protein Chem.* 66, 27-85.
178. Pontiggia F, Pachov DV, Clarkson MW, Villali J, Hagan MF, Pande VS and Kern D (2015). Free energy landscape of activation in a signalling protein at atomic resolution. *Nat. Comm.* 6(1), 7284.
179. Powers MJ and Trent MS (2018). Phospholipid retention in the absence of asymmetry strengthens the outer membrane permeability barrier to last-resort antibiotics. *Proc. Natl. Acad. Sci. U.S.A.* 115(36), E8518-E8527.
180. Powers MJ, Simpson BW and Trent MS (2020). The Mla pathway in *Acinetobacter baumannii* has no demonstrable role in anterograde lipid transport. *eLife* 9, e56571.
181. Price DJ and Brooks CL (2004). A modified TIP3P water potential for simulation with Ewald summation. *J. Chem. Phys.* 121, 10096-10103.
182. Raetz CR and Dowhan W (1990). Biosynthesis and function of phospholipids in *Escherichia coli*. *J. Biol. Chem.* 265(3), 1235–1238.

183. Rees DC, Johnson E and Lewinson O (2009). ABC transporters: the power to change. *Nat. Rev. Mol. Cell Biol.* 10(3), 218-227.
184. Reynolds CH, Tounge BA and Bembenek SD (2008). Ligand binding efficiency: trends, physical basis, and implications. *J. Med. Chem.* 51(8), 2432-2438.
185. Rice AJ, Park A and Pinkett HW (2014). Diversity in ABC transporters: type I, II and III importers. *Crit. Rev. Biochem. Mol. Biol.* 49(5), 426-437.
186. Roier S, Zingl FG, Cakar F, Durakovic S, Kohl P, Eichmann TO, Klug L, Gadermaier B, Weinzerl K, Prassl R and Lass A (2016). A novel mechanism for the biogenesis of outer membrane vesicles in Gram-negative bacteria. *Nat. Commun.* 7(1), 10515.
187. Rollauer SE, Sooreshjani MA, Noinaj N and Buchanan SK (2015). Outer membrane protein biogenesis in Gram-negative bacteria. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 370(1679), 20150023.
188. RStudio Team (2015). RStudio: Integrated Development for R. Boston, MA (Computer Software v0.98.1074: RStudio, Inc.
189. Rujoi M, Estrada R and Yappert MC (2004). In situ MALDI-TOF MS regional analysis of neutral phospholipids in lens tissue. *Anal. Chem.* 76(6), 1657-1663.
190. Sandhya S, Rani SS, Pankaj B, Govind MK, Offmann B, Srinivasan N and Sowdhamini R (2009). Length variations amongst protein domain superfamilies and consequences on structure and function. *PLoS One* 4, e4981.
191. Scheepers GH, Lycklama a Nijeholt JA and Poolman B (2016). An updated structural classification of substrate-binding proteins. *FEBS Letts.* 590(23), 4393-4401.
192. Schiller J, Arnhold J, Benard S, Müller M, Reichl S and Arnold K (1999). Lipid analysis by matrix-assisted laser desorption and ionization mass spectrometry: A methodological approach. *Anal. Biochem.* 267(1), 46-56.
193. Schultz KM, Fischer MA, Noey EL and Klug CS (2018). Disruption of the *E. coli* LptC dimerization interface and characterization of lipopolysaccharide and LptA binding to monomeric LptC. *Protein Sci.* 27, 1407-1417.
194. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B and Ideker T (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13(11), 2498-2504.
195. Sharma V, Siedenburg G, Birke J, Mobeen F, Jendrossek D and Prakash T (2018). Metabolic and taxonomic insights into the Gram-negative natural rubber degrading bacterium *Steroidobacter cummioxidans* sp. nov., strain 35Y. *PLoS One* 13, e0197448.
196. Shaw Stewart PD, Kolek SA, Briggs RA, Chayen NE and Baldock PF (2011). Random microseeding: a theoretical and practical exploration of seed stability and seeding techniques for successful protein crystallization. *Cryst. Growth Des.* 11(8), 3432-3441.
197. Shrivastava R and Chng SS (2019). Lipid trafficking across the Gram-negative cell envelope. *J. Biol. Chem.* 294(39), 14175-14184.
198. Shrivastava R, Jiang XE and Chng SS (2017). Outer membrane lipid homeostasis via retrograde phospholipid transport in *Escherichia coli*. *Mol. Microbiol.* 106(3), 395-408.
199. Shukolyukov SA (2009). Aggregation of frog rhodopsin to oligomers and their dissociation to monomer: application of BN-and SDS-PAGE. *Biochemistry (Moscow)* 74, 599-604.
200. Sievers F and Higgins DG (2014). Clustal Omega, Accurate Alignment of Very Large Numbers of Sequences. In *Multiple Sequence Alignment Methods. Methods in Molecular Biology (Methods and Protocols)* (Russell D, eds), pp. 105-116. Humana Press, Totowa, NJ.

201. Silhavy TJ, Kahne D and Walker S (2010). The bacterial cell envelope. *Cold Spring Harb. Perspect. Biol.* 2(5), a000414.
202. Smith RD, Engdahl AL, Dunbar JB Jr and Carlson HA (2012). Biophysical limits of protein–ligand binding. *J. Chem. Inf. Model* 52(8), 2098-2106.
203. Sohlenkamp C and Geiger O (2016). Bacterial membrane lipids: diversity in structures and pathways. *FEMS Microbiol. Rev.* 40(1), 133-159.
204. Sousa da Silva AW and Vranken WF (2012). ACPYPE-Antechamber python parser interface. *BMC Res. Notes* 5(1), 1-8.
205. Stock JB, Rauch B and Roseman S (1977). Periplasmic space in *Salmonella typhimurium* and *Escherichia coli*. *J. Biol. Chem.* 252(21), 7850-7861.
206. Sumathi K, Ananthalakshmi P, Roshan MM and Sekar K (2006). 3dSS: 3D structural superposition. *Nucleic Acids Res.* 34(suppl_2), W128-W132.
207. Surovtsev NV and Dzuba SA (2014). Flexibility of phospholipids with saturated and unsaturated chains studied by Raman scattering: The effect of cholesterol on dynamical and phase transitions. *J. Chem. Phys.* 140(23), 235103.
208. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, Jensen LJ and Mering CV (2017). The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* 45, D362-D368.
209. Takeda K, Miyatake H, Yokota N, Matsuyama SI, Tokuda H and Miki K (2003). Crystal structures of bacterial lipoprotein localization factors, LolA and LolB. *EMBO J.* 22(13), 3199-3209.
210. Tan BK, Bogdanov M, Zhao J, Dowhan W, Raetz CR and Guan Z (2012). Discovery of a cardiolipin synthase utilizing phosphatidylethanolamine and phosphatidylglycerol as substrates. *Proc. Natl. Acad. Sci. U.S.A.* 109(41), 16504-16509.
211. Tanaka KJ, Song S, Mason K and Pinkett HW (2018). Selective substrate uptake: the role of ATP-binding cassette (ABC) importers in pathogenesis. *Biochim. Biophys. Acta Biomembr.* 1860, 868-877.
212. Tang X, Chang S, Qiao W, Luo Q, Chen Y, Jia Z, Coleman J, Zhang K, Wang T, Zhang Z, Zhang C, Zhu X, Wei X, Dong C, Zhang X and Dong H (2021). Structural insights into outer membrane asymmetry maintenance in Gram-negative bacteria by MlaFEDB. *Nat. Struct. Mol. Biol.* 28(1), 81-91.
213. Taniguchi N and Tokuda H (2008). Molecular events involved in a single cycle of ligand transfer from an ATP binding cassette transporter, Lol CDE, to a molecular chaperone, LolA. *J. Biol. Chem.* 283, 8538-8544.
214. Taylor GL (2010). Introduction to phasing. *Acta. Cryst.* D66, 325-338.
215. ter Beek J, Guskov A and Slotboom DJ (2014). Structural diversity of ABC transporters. *J. Gen. Physiol.* 143(4), 419-435.
216. Terasaka K, Blakeslee JJ, Titapiwatanakun B, Peer WA, Bandyopadhyay A, Makam SN, Lee OR, Richards EL, Murphy AS, Sato F and Yazaki K (2005). PGP4, an ATP binding cassette P-glycoprotein, catalyzes auxin transport in *Arabidopsis thaliana* roots. *Plant Cell* 17(11), 2922-2939.
217. The UniProt Consortium (2023). UniProt: the universal protein knowledgebase in 2023. *Nucleic. Acids Res.* 51, D523-D531.
218. Thomas C, Aller SG, Beis K, Carpenter EP, Chang G, Chen L, Dassa E, Dean M, Duong Van Hoa F, Ekiert D, Ford R, Gaudet R, Xin Gong, Holland BI, Huang Y, Kahne DK, Kato H, Koronakis V, Koth CM, Lee Y, Lewinson O, Lill R, Martinoia E, Murakami S, Pinkett HW, Poolman B, Rosenbaum D, Sarkadi B, Schmitt L, Schneider E, Shi Y, Shyng SL, Slotboom DJ, Tajkhorshid E, Tieleman DP, Ueda K, Váradi A, Wen PC, Yan N, Zhang P, Zheng H, Zimmer J and Tampé R (2020).

Structural and functional diversity calls for a new classification of ABC transporters. *FEBS Letts.* 594(23), 3767-3775.
219. Thong S, Ercan B, Torta F, Fong ZY, Wong HY, Wenk MR and Chng SS (2016). Defining key roles for auxiliary proteins in an ABC transporter that maintains bacterial outer membrane lipid asymmetry. *eLife* 5, e19042.
220. Tian W, Chen C, Lei X, Zhao J and Liang J (2018). CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Res.* 46(W1), W363-W367.
221. Tiwari A and Singh S (2022). Computational approaches in drug designing. In: Singh DB and Pathak RK, ed. *Bioinformatics: Methods and Application* Academic Press, 207-217.
222. Tokuda H and Matsuyama SI (2004). Sorting of lipoproteins to the outer membrane in *E. coli*. *Biochim. Biophys. Acta - Mol. Cell Res.* 1693(1), 5-13.
223. Tremblay JM, George HM and Yarbrough LR (1996). Limited proteolysis of rat phosphatidylinositol transfer protein by trypsin cleaves the C terminus, enhances binding to lipid vesicles, and reduces phospholipid transfer activity. *J. Biol. Chem.* 271(35), 21075-21080.
224. Vagin AA, Steiner RA, Lebedev AA, Potterton L, McNicholas S, Long F and Murshudov GN (2004). REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta. Cryst.* D60, 2184-2195.
225. van der Heide T and Poolman B (2002). ABC transporters: one, two or four extracytoplasmic substrate-binding sites? *EMBO Rep.* 3, 938-943.
226. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G, Laydon A and Žídek A (2022). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* 50, D439-D444.
227. Vázquez-Laslop N, Lee H, Hu R and Neyfakh AA (2001). Molecular sieve mechanism of selective release of cytoplasmic proteins by osmotically shocked *Escherichia coli*. *J. Bacteriol.* 183(8), 2399-404.
228. Viljoen A , Foster SJ, Fantner GE, Hobbs JK and Dufrêne YF (2020). Scratching the surface: bacterial cell envelopes at the nanoscale. *MBio* 11(1), e03020-19.
229. Villar HO and Kauvar LM (1994). Amino acid preferences at protein binding sites. *FEBS Lett.* 349(1), 125-130.
230. Vollmer W, Blanot D and De Pedro MA (2008). Peptidoglycan structure and architecture. *FEMS Microbiol. Rev.* 32(2), 149-167.
231. Wilburn KM, Fieweger RA and VanderVen BC (2018). Cholesterol and fatty acids grease the wheels of *Mycobacterium tuberculosis* pathogenesis. *Pathog. Dis.* 76(2), fty021.
232. Wilkens S (2015). Structure and mechanism of ABC transporters. *F1000 Prime Rep.* 7, 14.
233. Wilman HR, Shi J and Deane CM (2014). Helix kinks are equally prevalent in soluble and membrane proteins. *Proteins* 82(9), 1960-1970.
234. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A and McNicholas SJ (2011). Overview of the CCP4 suite and current developments. *Acta. Cryst.* D67, 235-242.
235. Wong LH, Čopič A and Levine TP (2017). Advances on the transfer of lipids by lipid transfer proteins. *Trends Biochem. Sci.* 42(7), 516-530.
236. Wong LH, Gatta AT and Levine TP (2019). Lipid transfer proteins: the lipid commute via shuttles, bridges and tubes. *Nat. Rev. Mol. Cell Biol.* 20(2), 85-101.
237. Woods AS, Ugarov M, Egan T, Koomen J, Gillig KJ, Fuhrer K, Gonin M and Schultz JA (2004). Lipid/peptide/nucleotide separation with MALDI-ion mobility-TOF MS. *Anal. Chem.* 76(8), 2187-2195.

238. Yero D, Díaz-Lobo M, Costenaro L, Conchillo-Solé O, Mayo A, Ferrer-Navarro M, Vilaseca M, Gibert I and Daura X (2021). The *Pseudomonas aeruginosa* substrate-binding protein Ttg2D functions as a general glycerophospholipid transporter across the periplasm. *Comm. Biol.* 4(1), 1-6.
239. Yoder MD, Thomas LM, Tremblay JM, Oliver RL, Yarbrough LR and Helmkamp GM (2001). Structure of a multifunctional protein: mammalian phosphatidylinositol transfer protein complexed with phosphatidylcholine. *J. Biol. Chem.* 276(12), 9246-9252.
240. Zhou C, Shi H, Zhang M, Zhou L, Xiao L, Feng S, Im W, Zhou M, Zhang X and Huang Y (2021). Structural insight into phospholipid transport by the MlaFEBD complex from *P. aeruginosa*. *J. Mol. Biol.* 433(13), 166986.

List of publications, conferences and workshops

Publications from thesis

1. **Dutta, A.**, Chandravanshi, M. and Kanaujia, S.P. (2021). Conserved features of the MlaD domain aid the trafficking of hydrophobic molecules. ***Proteins: Structure, Function, and Bioinformatics***, 89(11), 1473-1488.
2. **Dutta, A.** and Kanaujia, S.P. (2022). MlaC belongs to a unique class of non-canonical substrate-binding proteins and follows a novel phospholipid-binding mechanism. ***Journal of Structural Biology***, 214, 107896.
3. **Dutta, A.** and Kanaujia, S.P. The structural features of MlaD illuminate its unique ligand-transporting mechanism and ancestry **(*Under communication)***.
4. **Dutta, A.** and Kanaujia, S.P. *In silico* characterization of Mla proteins **(*Under preparation)***.
5. **Dutta, A.**, Patel S. and Kanaujia, S.P. Molecular dynamics simulation studies of MlaC **(*Under preparation)***.
6. **Dutta, A.**, and Kanaujia, S.P. Review on the Mla system **(*Under preparation)***.

Publications from other work

1. Adhikari, R., Singh, D., Chandravanshi, M., **Dutta, A.** and Kanaujia, S.P. (2017). UgpB, a periplasmic component of the UgpABCE ATP-binding cassette transporter, predominantly follows the sec translocation pathway. ***Meta Gene***, 13, 129-139.
2. Sinha, A.K., **Dutta, A.**, Chandravanshi, M. and Kanaujia, S.P. (2019). An insight into bacterial phospholipase C classification and their translocation through Tat and Sec pathways: A data mining study. ***Meta Gene***, 20, 100547.

Conferences and workshops attended

1. **Angshu Dutta** and Shankar Prasad Kanaujia. MlaD is a non-canonical substrate-binding protein that follows a novel phospholipid-binding mechanism. Research and Industrial Conclave 2023 (RIC 2023). May 14-16, 2023, IIT Guwahati, Guwahati, Assam, India. ***[Best Oral Presentation Award]***.
2. **Angshu Dutta** and Shankar Prasad Kanaujia. Exploring the structural and mechanistic intricacies of non-canonical substrate-binding proteins. Research and Industrial Conclave 2023 (RIC 2023). May 14-16, 2023, IIT Guwahati, Guwahati, Assam, India. ***[Three Minutes Thesis Presentation]***.
3. **Angshu Dutta** and Shankar Prasad Kanaujia. MlaC belongs to a unique class of non-canonical substrate-binding proteins and follows a novel phospholipid-binding mechanism. National Seminar on Crystallography (NSC-49). November 28-30, 2022, University of Jammu, Jammu, India. **[*Professor Kailasam Venkatesan Award for Best Oral Presentation Award*]**.
4. **Angshu Dutta** and Shankar Prasad Kanaujia. Unveiling the mechanistic aspects of a novel substrate-binding protein in Gram-negative bacteria. National Seminar on Crystallography (NSC-48). November 25-27, 2021, IIT Roorkee, Uttarakhand, India. **[*Best Oral Presentation Award*]**.
5. **Angshu Dutta** and Shankar Prasad Kanaujia. Exploring the structural aspects of a novel phospholipid-binding protein. PDB 50th Anniversary Symposium in Asia. November 24, 2021, PDBj and Institute for Protein Research, Osaka University, Japan. **[*Best Poster Award*]**.
6. **Angshu Dutta** and Shankar Prasad Kanaujia. Deciphering the structural aspects of a novel substrate-binding protein. Bringing Molecular Structure to Life: 50 Years of the PDB. October 20-22, 2021, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany. **[*Poster Presentation; Recipient of EMBL Corporate Partnership Programme Registration Fee Waiver*]**.

7. **Angshu Dutta** and Shankar Prasad Kanaujia. MlaC represents a unique class of solute-binding proteins. 25th Congress of the International Union of Crystallography (IUCr)-2021. August 14-22, 2021, Prague, Czech Republic. **[*Poster Presentation*]**.
8. **Angshu Dutta**. Indo-Italian Elettra Beamline Workshop: Application in Crystallography and Drug Discovery. Jointly organized by All India Institute of Medical Sciences, Jawaharlal Nehru University, National Institute of Immunology, The International Centre for Genetic Engineering, CSIR Institute of Genomics and Integrative Biology & Biotechnology and Regional Centre of Biotechnology. November 11-12, 2019, All India Institute of Medical Sciences, New Delhi, India.
9. **Angshu Dutta** and Shankar Prasad Kanaujia. Functional annotation, classification and assignment of translocation pathway of phospholipases C. Research Conclave 2018. March 08-11, 2018, IIT Guwahati, India. **[*Springer award* for *Best Poster Presentation*]**.
10. **Angshu Dutta**. Hands-on training workshop on crystallization in lipid bilayer. Organized by TPPLabtech at ICGEB. May 29-30, 2018, New Delhi, India.
11. **Angshu Dutta**. International workshop on ‘’X-ray Absorption Spectroscopy for the
12. Crystallographer’’. Organized by 24th Congress and General Assembly of International Union of Crystallography (IUCr 2017). August 21, 2017, Hyderabad, India.
13. **Angshu Dutta** and Shankar Prasad Kanaujia. UgpB protein dominantly follows Sec translocation pathway. 24th Congress and General Assembly of International Union of Crystallography (IUCr 2017). August 21-28, 2017, Hyderabad, India. **[*Poster Presentation; Recipient of bursary from Department of Science and Technology, Government of India*]**.
14. **Angshu Dutta** and Shankar Prasad Kanaujia. Deciphering the structural aspects of an antimicrobial peptide importer in Gram-negative bacteria for developing drugs. National Seminar on Crystallography (NSC-45). July 9-12, 2017, IIT (BHU) Varanasi, India. **[*Poster Presentation].***